DISEÑO ESTADISTICO DE EXPERIMENTOS

1 9 8 2

DIRECTORIO DE PROFESORES

1. DR. OCTAVIO A. RASCON CHAVEZ
   SUBDIRECTOR
   INSTITUTO DE INGENIERIA, UNAM
   CIUDAD UNIVERSITARIA,
   MEXICO 20, D.F.
   TEL: 548. 54. 79.

2. M. EN I. AUGUSTO VILLARREAL ARANDA
   GERENTE DE OPERACIONES
   GRUPO V E A
   ASIA No. 31
   MEXICO 21, D.F.
   TEL: 554.45. 31 y 554.41.31

3. M. EN I. BERNARDO FRONTANA DE LA CRUZ
   INVESTIGADOR
   INSTITUTO DE INGENIERIA, UNAM
   CIUDAD UNIVERSITARIA
   MEXICO 20, D.F.
   TEL: 550.52. 15 ext. 3600 y 3669

pmc.

DISEÑO ESTADISTICO DE EXPERIMENTOS

DIA	HORA	TEMAS	PROFESORES
19, 21, 23 y 26 de abril	18 a 21 cada día	1. INTRODUCCION	Dr. Octavio A. Rascón Chávez
		1.1 Unidad de Experimentación. Factores. Variable aleatoria	M. en I. Augusto Villarreal Aranda
		1.2 Etapas en el diseño de un experimento	
		1.3 Procesamiento elemental de datos	
		2. DISEÑO Y ANALISIS DE EXPERIMENTOS CON UNA SOLA VARIABLE	M. en I. Augusto Villarreal Aranda
		2.1 Distribución de probabilidades. Distribuciones binominal, multinomial y normal	
		2.2 Estadísticas. Estimación puntual de parámetros	
		2.3 Distribuciones muestrales t, χ^2 y F	
		2.4 Estimación de parámetros por intervalos de confianza	
		2.5 Pruebas de Hipótesis	
		3. DISEÑO Y ANALISIS DE EXPERIMENTOS CON DOS VARIABLES	M. en I. Augusto Villarreal Aranda
		3.1 Comparación de valores medios: variables independientes y variables dependientes	
		3.2 Comparación de medidas de dispersión (de variancias)	
28 y 30 de abril y 3 de mayo	18 a 21 cada día	3.3 Prueba de independencia de dos variables nominales Tablas de contingencia	Dr. Octavio A. Rascón Chávez
		K TRATAMIENTOS	
		4.1 Comparación de valores medios. Análisis de variancia: niveles fijos y niveles aleatorios Comparaciones múltiples: métodos de Tukey, Dunnett y Duncan	
		4.2 Comparación de variancias	
		5. DISEÑO Y ANALISIS DE EXPERIMENTOS CON BLOQUES ALEATORIZADOS	
		5.1 Análisis de variancia	
		5.2 Estimación de efectos y residuos	
7 de mayo	18 a 21	6. DISEÑO Y ANALISIS DE EXPERIMENTOS CON CLASIFICACION DE DOS FACTORES	Ing. Bernardo Frontana C.
		6.1 Factores no cruzados: niveles fijos y niveles aleatorios	
		6.2 Factores cruzados: niveles fijos y niveles aleatorios	
12, 14 y 17 de mayo	18 a 21 cada día	7. ANALISIS DE EXPERIMENTO DE CUADROS LATINOS	Dr. Octavio A. Rascón Chávez
		8. ANALISIS DE EXPERIMENTO DE CUADROS GRECO-LATINOS	
		9. ANALISIS DE EXPERIMENTOS DE BLOQUES ALEATORIZADOS INCOMPLETOS Y DE CUADROS DE YUDEN	

DIA	HORA	TEMAS	PROFESORES
19, 21 y 24 de mayo.	18 a 21 cada día	10. ANALISIS DE EXPERIMENTOS FACTORIALES 2^K	Ing. Bernardo Frontana C.
		11. ANALISIS DE VARIANCIA EN REGRESION LINEAL	
		11.1 Regresión lineal. Cálculo de la recta de regresión. Prueba de independencia de las dos variables. Error de la predicción	
		11.2 Variancia explicada e inexplicada	
		11.3 Análisis de covariancia	

DIVISION DE EDUCACION CONTINUA
FACULTAD DE INGENIERIA U.N.A.M.

DISEÑO ESTADISTICO DE EXPERIMENTOS

FUNDAMENTOS DEL DISEÑO Y ANALISIS ESTADISTICO DE EXPERIMENTOS

DR. OCTAVIO A. RASCON CHAVEZ

ABRIL, 1982

de Minería Calle de Tacuba 5 primer piso Deleg. Cuauhtemoc 06000 México, D.F. Tel.: 521-40-20 Apdo. Postal M-2285

INDICE

1. Tener un método eficiente para diseñar un experimento que conduzca a resultados que permitan obtener las respuestas a las preguntas que se plantean, y que sean afectados lo menos posible por alguna fuente de error.

2. Contar con algún método para analizar los resultados y sacar conclusiones.

De estos factores el más importante es el primero, ya que si el experimento no se diseña adecuadamente no se podrá obtener la información necesaria para extraer las conclusiones deseadas, aun cuando se cuenta con métodos de análisis sofisticados.

Dificultades confrontadas por los investigadores

Las dificultades usuales que tiene que vencer un investigador son:

a. Error experimental
b. Confusión de correlación con causalidad
c. Complejidad de los efectos estudiados

a. Error experimental. Toda variación en los resultados ocasionada por factores disturbantes, conocidos o no, se llama error experimental.

La confusión que ocasiona el error experimental se puede reducir grandemente mediante un diseño adecuado del experimento y mediante el uso de métodos estadísticos de análisis.

b. Confusión de correlación con causalidad. Es necesario saber discernir cuándo una correlación aparente entre dos parámetros es casual o causal; en el primer caso ésta aparecerá por casualidad; en el segundo, se tendrá cuando en realidad la variación de un parámetro se puede explicar por la variación del otro, es decir, que un cambio en uno causa un cambio en el otro.

c. Complejidad de los efectos estudiados. No siempre es fácil detectar si un parámetro influye en los resultados experimentales, y si sí, en qué rangos de valores lo hace, y de qué manera interactúa con otros parámetros para influir junto con ellos (efectos cruzados).

Por ejemplo, los parámetros vino y café pueden influir en el tiempo de reacción de un individuo ante cierto estímulo; los efectos pueden ser de manera individual (debido sólo al café o sólo al vino) o combinada (debido a ambos a la vez). Asimismo, los efectos pueden cambiar en función del número de tazas de café o de copas de vino.

Es muy importante que en cualquier investigación experimental:

a) se definan claramente los objetivos que se persiguen

b) se asegure de que todas las partes interesadas estén de acuerdo con ellos

c) se defina el criterio bajo el cual se probará si se cumplieron los objetivos, es decir, se seleccionan el diseño experimental que se considere adecuado y el método estadístico de prueba; y

d) se tengan acuerdos preliminares con las partes interesadas sobre las acciones a tomar en caso de que no se cumplan los objetivos.

En lo que sigue se entederá por espécimen o unidad experimental a la persona, animal u objeto sobre el cual se hace la medición de la propiedad o característica bajo estudio.

Por su parte, se entenderá por tratamiento a un nivel o valor de un factor o a una combinación de niveles de factores.

Por ejemplo, al comparar el rendimiento (en km/lt) que se tiene con cuatro aditivos para gasolina y dos marcas diferentes de automóvil:

- se tendrán dos factores, aditivo y marca, el primero con cuatro niveles y el segundo con dos
- cada tratamiento será una de las combinaciones aditivo-marca
- las unidades experimentales serán los vehículos a los cuales se les "apliquen" los tratamientos
- el rendimiento es la característica o variable en estudio
- los resultados de cada medición (km/lt) serán los datos u observaciones
- el conjunto de datos para cada tratamiento conforma la

muestra correspondiente.

En este ejemplo cada muestra debe ser representativa de la respectiva población; las poblaciones son las colecciones de resultados (rendimientos) que se tendrían si todo el aditivo disponible de cada tipo se usara en todos los automóviles de ambas marcas; obviamente sería no sólo antieconómico sino improcedente el usar todo el volumen fabricado de cada aditivo para hacer la comparación de rendimientos, puesto que no quedaría nada para usarse con el fin previsto (en este caso, escoger el mejor aditivo para los vehículos de una empresa), y la verificación teóricamente no terminaría nunca ya que las fábricas de aditivos y vehículos pueden producir continua e indefinidamente (se trataría de poblaciones teóricamente infinitas).

2. TABLAS DE CONTINGENCIA

CON FRECUENCIA SE DESEA DETERMINAR SI LA CLASIFICACION DE UNA MUESTRA EN TERMINOS DE 2 O MAS CRITERIOS ES TAL QUE PERMITA INFERIR SI ESOS CRITERIOS SON INDEPENDIENTES ENTRE SI.

POR EJEMPLO, UNA MUESTRA DE PERSONAS QUE HAN FALLECIDO SE PUEDE CLASIFICAR DE LA SIGUIENTE MANERA:

	MUERTE POR CANCER DEL PULMON	MUERTE POR OTRAS CAUSAS
FUMADORES	348	3152
NO FUMADORES	82	1418

EN UN CASO COMO ESTE SE PRETENDERIA PROBAR LA HIPOTESIS DE QUE FUMAR Y MORIR POR CANCER DEL PULMON SON CARACTERISTICAS INDEPENDIENTES.

CUANDO LOS DATOS SE CATEGORIZAN DE ESTA MANERA, SE DICE QUE SE FORMA UNA TABLA DE CONTINGENCIA.

SEA UNA MUESTRA DE TAMAÑO n Y QUE EL EXPERIMENTO SE HA DISEÑADO PARA CLASIFICARLA EN DOS CATEGORIAS, UNA CON r NIVELES Y LA OTRA CON c NIVELES.

SEA X_{ij} EL NUMERO (LA FRECUENCIA) DE ELEMENTOS DE LA MUESTRA QUE QUEDAN EN LA CELDA (i,j).

LA TABLA DE CONTINGENCIA SERÍA

CLASI- FICACION 1	CLASIFICACION 2 1	2	3	. . . .	c	TOTAL
1	x_{11}	x_{12}	x_{13}		x_{1c}	$x_{1.}$
2	x_{21}	x_{22}	x_{23}		x_{2c}	$x_{2.}$
3	x_{31}	x_{32}	x_{33}		x_{3c}	$x_{3.}$
.						
.						
.						
r	x_{r1}	x_{r2}	x_{r3}		x_{rc}	$x_{r.}$
TOTAL	$x_{.1}$	$x_{.2}$	$x_{.3}$		$x_{.c}$	n

LOS TOTALES POR RENGLON SE DENOTAN CON $x_{i.}$, ES DECIR

$$x_{i.} = \sum_{j=1}^{c} x_{ij}$$

LOS TOTALES POR COLUMNA SE DENOTAN

$$x_{.j} = \sum_{i=1}^{r} x_{ij}$$

LA SUMA DE LOS TOTALES POR COLUMNA O POR RENGLON DEBE SER EL TAMAÑO DE LA MUESTRA, ES DECIR

$$\sum_{i=1}^{r} x_{i.} = \sum_{j=1}^{c} x_{.j} = n = \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}$$

EL PROBLEMA DE VERIFICAR SI LAS CATEGORIAS SON INDEPENDIENTES EQUIVALE AL DE VERIFICAR SI LA PROBABILIDAD DE QUE EL ESPÉCIMEN CUMPLA CON ALGÚN NIVEL DE LA CATEGORIA 1 DEPENDE DE EN QUÉ NIVEL DE LA CATEGORIA 2 SE ENCUENTRA. ASÍ, EN EL

EJEMPLO ANTERIOR SE TRATARIA DE VERIFICAR SI LA MUERTE POR CANCER PULMONAR DEPENDE O NO DE SI LA PERSONA ES O NO FUMADORA.

EN INFERENCIA ESTADISTICA SE DEMUESTRA QUE LA ESTADISTICA

$$V = \sum_{i=1}^{k} \frac{(\chi_i - nP_i)^2}{nP_i} \qquad (1)$$

TIENDE A UNA VARIABLE ALEATORIA CON DISTRIBUCION DE PROBABILIDADES χ^2 CON k-r-1 GRADOS DE LIBERTAD CONFORME CRECE n, EN DONDE n ES EL TAMAÑO DE LA MUESTRA, χ_i ES LA FRECUENCIA CON QUE SE OBSERVO EL EVENTO i Y P_i ES LA PROBABILIDAD DE OBSERVARLO EN UNA REALIZACION DEL EXPERIMENTO.

EN NUESTRO CASO, SI P_{ij} ES LA PROBABILIDAD DE QUE UN RESULTADO TENGA EL VALOR i DE LA CARACTERISTICA 1 Y EL VALOR j DE LA 2, Y SI LOS DOS METODOS DE CLASIFICACION SON REALMENTE INDEPENDIENTES, ENTONCES.

$$P_{ij} = \omega_i S_j, \ i = 1, 2, \ldots r; \ j = 1, 2, \ldots, c$$

DONDE ω_i ES LA PROBABILIDAD DE QUE EL ELEMENTO OBSERVADO CAIGA EN EL I-ESIMO NIVEL DE LA CLASIFICACION 1, Y S_j ES LA PROBABILIDAD DE QUE CAIGA EN EL J-ESIMO NIVEL DE LA CLASIFICACION 2.

POR OTRA PARTE, LOS ESTIMADORES DE MAXIMA VEROSIMILITUD DE ω_i Y S_i SON

$$\hat{\omega}_i = \frac{\chi_{i.}}{n} \quad , \quad \hat{S}_j = \frac{\chi_{.j}}{n}$$

POR LO TANTO, CON LA EC.(1) SE OBTIENE QUE

$$V = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(x_{ij} - n\hat{\omega}_i \hat{S}_j)^2}{n \hat{\omega}_i \hat{S}_j} \qquad (2)$$

TIENE DISTRIBUCION χ^2 CON $(r-1)(c-1)$ GRADOS DE LIBERTAD PARA n GRANDE. ESTE NUMERO DE GRADOS DE LIBERTAD SE JUSTIFICA DE LA SIGUIENTE MANERA: SE TIENEN $K = rc$ CLASES Y PARA ESTIMAR LAS P_{ij} SE REQUIERE ESTIMAR $r-1$ VALORES DE ω Y $c-1$ VALORES DE S, ES DECIR, SE ESTIMAN $(r-1) + (c-1)$ PARAMETROS; POR LO TANTO LOS GRADOS DE LIBERTAD SON

$$rc-(r-1)-(c-1)-1 = (r-1)(c-1)$$

EJEMPLO

CON EL FIN DE VERIFICAR SI EL FUMAR Y EL MORIR POR CANCER PULMONAR SON INDEPENDIENTES, SE DISEÑO UN EXPERIMENTO ESTADISTICO QUE CONSISTIO EN SACAR UNA MUESTRA ALEATORIA DE 5000 EXPEDIENTES CLINICOS DE PERSONAS FALLECIDAS EN UNA CADENA DE HOSPITALES, Y CLASIFICARLA EN UNA TABLA DE CONTINGENCIA. EL RESULTADO FUE EL SIGUIENTE:

	MUERTE POR CANCER PULMONAR	MUERTE POR OTRAS CAUSAS	TOTAL $X_{i.}$	$\hat{\omega}_i$
FUMADORES	348	3152	3500	0.7
NO FUMADORES	82	1418	1500	0.3
TOTAL : $X_{.j}$	430	4570	5000	1.0
$\hat{S}_j$	0.086	0.914	1.000	

PARA REALIZAR LA PRUEBA DE INDEPENDENCIA SE UTILIZA LA EC (2), Y SE DETERMINA EL VALOR CRITICO DE χ^2 QUE CORRESPONDA A UN NIVEL DE CONFIANZA PRESTABLECIDO, $1-\alpha$, USANDO $(2-1)\times(2-1) = 1$ GRADOS DE LIBERTAD, YA QUE $r = c = 2$.

$$\hat{\omega}_1 = \frac{3500}{5000} = 0.7, \quad \hat{\omega}_2 = \frac{1500}{5000} = 0.3$$

$$\hat{S}_1 = \frac{430}{5000} = 0.086, \quad \hat{S}_2 = \frac{4570}{5000} = 0.914$$

$$V = \frac{[348-5000\ (0.7)(0.086)]^2}{5000\ (0.7)(0.086)} + \frac{[3152-5000\ (0.7)(0.914)]^2}{5000\ (0.7)(0.914)} + \frac{[82-5000\ (0.3)(0.086)]^2}{5000\ (0.086)(0.3)} + \frac{[1418-5000\ (0.3)(0.914)]^2}{5000\ (0.3)(0.914)}$$

$$V = \frac{2209}{301.00} + \frac{2209}{3199.00} + \frac{2209}{129.00} + \frac{2209}{1371.00} =$$

$$= 7.34 + 0.69 + 17.12 + 1.61 = 26.76$$

SI $1-\alpha = 0.99$, ENTONCES

$$(\chi_c^2)_{0.99,1} = 6.63 < 26.76$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE INDEPENDENCIA.

<u>EJEMPLO</u>

UN CLUB DE PESCA DEPORTIVA ESTA INTERESADO EN SABER SI SE PESCA CADA TIPO DE PESCADO CON LA MISMA FRECUENCIA EN LOS MESES DE JUNIO A SEPTIEMBRE. PARA ELLO SE DISEÑO UN EXPERIMENTO CONSISTENTE EN REGISTRAR LA PESCA MENSUAL EN UNO DE LOS BARCOS DE LOS TRES TIPOS DE PECES DE LA ZONA: ABADEJO, PEZ AZUL Y COLA AMARILLA.

LA TABLA DE CONTINGENCIA QUE SE FORMULO FUE LA SIGUIENTE:

	ABADEJOS	PECES AZULES	COLAS AMARILLAS	TOTAL	$\hat{\omega}$
JUNIO	315	1347	620	2282	0.2611
JULIO	270	1250	514	2034	0.2327
AGOSTO	295	1480	710	2485	0.2843
SEPTIEMBRE	246	1200	494	1940	0.2219
TOTAL	1126	5277	2338	8741	1.0000
$\hat{S}$	0.1288	0.6037	0.2675	1.0000	

PARA LA PRUEBA DE INDEPENDENCIA CON CONFIABILIDAD $1-\alpha = 0.95$ Y $(4-1)(3-1) = 6$ GRADOS DE LIBERTAD, SE TIENE QUE $(\chi_c^2)_{0.95,6} = 12.6$.

EL VALOR DE LA ESTADISTICA V SE OBTIENE EN LA EC (2)

$$V = \sum_{i=1}^{4} \sum_{j=1}^{3} \frac{(x_{ij} - 8741\ \hat{\omega}_i\ \hat{S}_j)^2}{8741\ \hat{\omega}_i\ \hat{S}_j}$$

CON $\hat{\omega}_1 = \frac{2782}{8741} = 0.2611$, $\hat{\omega}_2 = \frac{2034}{8741} = 0.2327$

$$\hat{\omega}_3 = \frac{2485}{8741} = 0.2843, \quad \hat{\omega}_4 = \frac{1940}{8741} = 0.2219$$

$$\hat{S}_1 = \frac{1126}{8741} = 0.1288, \quad \hat{S}_2 = \frac{5277}{8741} = 0.6037$$

$$\hat{S}_3 = \frac{2338}{8741} = 0.2675$$

$$8741\ \hat{\omega}_1\ \hat{S}_1 = 8741\ (0.2611)\ (0.1288) = 293.957$$

$$8741\ \hat{\omega}_1\ \hat{S}_2 = 8741\ (0.2611)\ (0.6037) = 1377.809$$

$$8741\ \hat{\omega}_1\ \hat{S}_3 = 8741\ (0.2611)\ (0.2675) = 610.509$$

$$8741\ \hat{\omega}_2\ \hat{S}_1 = 8741\ (0.2327)\ (0.1288) = 261.983$$

$$8741\ \hat{\omega}_2\ \hat{S}_2 = 8741\ (0.2327)\ (0.6037) = 1227.944$$

$$8741\ \hat{\omega}_2\ \hat{S}_3 = 8741\ (0.2327)\ (0.2675) = 544.103$$

$$8741\ \hat{\omega}_3\ \hat{S}_1 = 8741\ (0.2843)\ (0.1288) = 320.077$$

$$8741\ \hat{\omega}_3\ \hat{S}_2 = 8741\ (0.2843)\ (0.6037) = 1500.235$$

$$8741\ \hat{\omega}_3\ \hat{S}_3 = 8741\ (0.2843)\ (0.2675) = 664.755$$

$$8741\ \hat{\omega}_4\ \hat{S}_1 = 8741\ (0.2219)\ (0.1288) = 249.824$$

$$8741\ \hat{\omega}_4\ \hat{S}_2 = 8741\ (0.2219)\ (0.6037) = 1170.953$$

$$8741\ \hat{\omega}_4\ \hat{S}_3 = 8741\ (0.2219)\ (0.2675) = 518.850$$

$$v = \frac{(315-293.957)^2}{293.957} + \frac{(1347-1377.809)^2}{1377.809} + \frac{(620-610.509)^2}{610.509} +$$

$$\frac{(270-261.983)^2}{261.983} + \frac{(1250-1227.944)^2}{1227.944} + \frac{(514-544.103)^2}{544.103} +$$

$$\frac{(295-320.077)^2}{320.077} + \frac{(1480-1500.235)^2}{1500.235} + \frac{(710-664.755)^2}{664.755} +$$

$$\frac{(246-249.824)^2}{249.824} + \frac{(1200-1170.953)^2}{1170.953} + \frac{(494-518.850)^2}{518.850}$$

$$v = 1.506+0.689+0.148+0.245+0.396+1.665+1.965+0.273+3.079$$

$$0.059+0.721+1.190$$

$$v = 11.936 < 12.6$$

POR LO TANTO SE ACEPTA LA HIPOTESIS DE QUE LA CANTIDAD DE PECES ES INDEPENDIENTE DEL MES EN EL PERIODO DE JUNIO A SEPTIEMBRE, CON UN 95% DE NIVEL DE CONFIANZA.

EJEMPLO

CON EL FIN DE VERIFICAR SI LAS VARIABLES REGION GEOGRAFICA, PARTIDO DE AFILIACION Y SEXO SON INDEPENDIENTES, SE DISEÑO UN EXPERIMENTO ESTADISTICO QUE CONSISTIO EN SACAR UNA MUESTRA ALEATORIA DE 1500 PERSONAS Y CLASIFICAR A CADA UNA DE ACUERDO CON ESAS VARIABLES; CON ESTO SE OBTUVO LA SIGUIENTE TABLA DE CONTINGENCIA:

	ESTE		OESTE			
PARTIDO	MASCULINO	FEMENINO	MASCULINO	FEMENINO	TOTAL	$\hat{w}_i$
DEMOCRATA	183	217	223	227	850	0.5667
REPUBLICANO	196	154	137	113	600	0.4000
OTRO	12	8	14	16	50	0.0333
	391 +	379	374 +	356	1500	
TOTAL	=770		=730			
$\hat{s}_j$	770/1500 = 0.5133		730/1500 = 0.4867			

$391 + 374 = 765$, $379 + 356 = 735$, $\hat{r}_1 = 765/1500=0.51$

$\hat{r}_2 = 735/1500=0.49$, $r = 3$, $c = 2$, $m = 2$

LA ESTADISTICA $V = \sum_{i=1}^{3} \sum_{j=1}^{2} \sum_{k=1}^{2} \frac{(x_{ijk} - 1500\hat{w}_i\hat{s}_j\hat{r}_k)^2}{1500\hat{w}_i\hat{s}_j\hat{r}_k} = 30.88$

TIENE DISTRIBUCION χ^2 CON $rcm-(r+c+m)+2=7$ GRADOS DE LIBERTAD.

SI $1-\alpha = 95\%$, ENTONCES

$$\chi^2_{0.95,7} = 14.1 < 30.88$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE QUE LAS TRES VARIABLES SON INDEPENDIENTES.

FORMULA CORTA PARA TABLAS DE CONTINGENCIA DE 2 x 2

SI DENOTAMOS A LAS FRECUENCIAS DE LA TABLA CON a, b, c Y d, O SEA $x_{11} = a$, $x_{12} = b$, $x_{21} = c$ Y $x_{22} = d$, SE PUEDE DEMOSTRAR QUE EL VALOR DE LA ESTADISTICA V SE CALCULA CON LA FORMULA

$$V = \frac{(ad - bc)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

EJEMPLO

CON EL FIN DE VERIFICAR SI DOS FABRICANTES DE TELEVISORES DE COLOR TIENEN IGUAL NIVEL DE CALIDAD SE DISEÑO UN EXPERIMENTO CONSISTENTE EN PREGUNTAR A 412 COMPRADORES DE LAS MISMAS SI SE REQUIRIO DE SERVICIO DE GARANTIA EN LOS DOS PRIMEROS AÑOS DE FUNCIONAMIENTO, CON LO CUAL SE INTEGRO LA SIGUIENTE TABLA DE CONTINGENCIA:

FABRICA	REQUIRIO SERVICIO	NO REQUIRIO SERVICIO	TOTAL
A	111 = a	152 = b	273
B	85 = c	54 = d	139
TOTAL	196	216	412

$$V = \frac{[(111)(54) - (152)(85)]^2\, 412}{(273)(139(196)(216)} = 15.51$$

$$\chi^2_{0.95,1} = 3.84 < 15.51$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE IGUALDAD DE CALIDAD, A UN 95% DE NIVEL DE CONFIANZA.

CORRECCION DE YATES

CON EL FIN DE MEJORAR LA APROXIMACION DE LA DISTRIBUCION χ^2 COMO DENSIDAD DE PROBABILIDADES DE LA ESTADISTICA V, CUANDO SE TIENEN POCAS CELDAS EN LA TABLA DE CONTINGENCIA, SE HA PROPUESTO INTRODUCIR UNA CORRECCION A LAS DIFERENCIAS DE LAS FRECUENCIAS OBSERVADAS MENOS LAS ESPERADAS, CONSISTENTE EN SUSTRAERLE 0.5 AL VALOR ABSOLUTO DE CADA DIFERENCIA, ES DECIR,

$$V = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(|x_{ij} - n\hat{w}_i\hat{s}_j| - 0.5)^2}{n\hat{w}_i\hat{s}_j}$$

CON ESTA CORRECCION LA FORMULA CORTA PARA TABLAS DE 2 x 2 QUEDA EN LA FORMA

$$V = \frac{(|ad - bc| - 0.5n)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$

EN EL EJEMPLO INMEDIATO ANTERIOR, AL APLICAR ESTA CORRECCION SE OBTIENE:

$$V = \frac{[|(111)(54)-(162)(85)|-0.5(412)]^2\,412}{(273)(139)(196)(216)} = \frac{(7570)^2 412}{(273)(139)(196)(216)} = 14.69$$

EJEMPLO

CON EL FIN DE VERIFICAR SI EL GRADO DE MEJORIA EN EL FUNCIONAMIENTO DE UN TIPO DE PROTESIS ES INDEPENDIENTE DEL HOSPITAL DONDE SE COLOCA, SE DISEÑO UN EXPERIMENTO CONSISTENTE EN FORMULAR UNA TABLA DE CONTINGENCIA; PARA ELLO SE OBTUVO UNA MUESTRA ALEATORIA DE PACIENTES DE 5 HOSPITALES CON ESTE TIPO DE PROTESIS, Y A CADA UNO SE LE CALIFICO COMO: FUNCIONAMIENTO NORMAL, PARCIAL O NULO. LOS RESULTADOS FUERON

	HOSPITAL				
FUNCIONAMIENTO	A	B	C	D	E
NULO	13	5	8	21	43
PARCIAL	18	10	36	56	29
NORMAL	16	16	35	51	10

a. PROBAR LA HIPOTESIS DE INDEPENDENCIA

b. ¿SON LOS RESULTADOS DE LOS HOSPITALES A, B, C Y D INDEPENDIENTES DEL FUNCIONAMIENTO?

c. ¿SI SE JUNTAN LOS RESULTADOS DE LOS HOSPITALES A, B, C Y D, ¿RESULTAN INDEPENDIENTES DE LOS DEL HOSPITAL E?

USAR 95% DE NIVEL DE CONFIANZA.

SOLUCION

a)

FUNCIONAMIENTO	HOSPITAL					TOTALES	$\hat{w}_i$
	A	B	C	D	E		
NULO	13	5	8	21	43	90	0.245
PARCIAL	18	10	36	56	29	149	0.406
NORMAL	16	16	35	51	10	128	0.349
TOTALES	47	31	79	128	82	367	
$\hat{S}_j$	0.128	0.0845	0.215	0.349	0.223		

$$v = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(x_{ij} - n\,\hat{w}_i\,\hat{S}_j)^2}{n\,\hat{w}_i\,\hat{S}_j}$$

$$v = \frac{(13 - (367)(0.245)(0.128))^2}{367 \times 0.245 \times 0.128} + \frac{(5 - (367)(0.245)(0.0845))^2}{367 \times 0.245 \times 0.0845} +$$

$$\frac{(8 - (367)(0.245)(0.215))^2}{367 \times 0.245 \times 0.215} + \frac{(21 - (367)(0.245)(0.349))^2}{367 \times 0.245 \times 0.349} +$$

$$\frac{(43 - (367)(0.245)(0.223))^2}{367 \times 0.245 \times 0.223} + \frac{(18 - (367)(0.406)(0.128))^2}{367 \times 0.406 \times 0.128} +$$

$$\frac{(10 - (367)(0.406)(0.0845))^2}{367 \times 0.406 \times 0.0845} + \frac{(36 - (367)(0.406)(0.215))^2}{367 \times 0.406 \times 0.215} +$$

$$\frac{(56 - (367)(0.406)(0.349))^2}{367 \times 0.406 \times 0.349} + \frac{(29 - (367)(0.406)(0.223))^2}{367 \times 0.406 \times 0.223} +$$

$$\frac{(16 - (367)(0.349)(0.128))^2}{367 \times 0.340 \times 0.128} + \frac{(16 - (367)(0.349)(0.0845))^2}{367 \times 0.349 \times 0.0845} +$$

$$\frac{(35 - (367)(0.349)(0.215))^2}{367 \times 0.349 \times 0.215} + \frac{(51 - (367)(0.349)(0.349))^2}{367 \times 0.349 \times 0.349}$$

$$\frac{(10 - (367)(0.349)(0.223))^2}{367 \times 0.349 \times 0.223}$$

$$v = \frac{2.2227232}{11.50912} + \frac{6.7486558}{7.5978175} + \frac{128.40799}{19.331725} + \frac{107.75135}{31.380335} +$$

$$\frac{526.65454}{20.051045} + \frac{1.1497329}{19.0722566} + \frac{6.745659}{12.590669} + \frac{15.717815}{32.03543} +$$

$$\frac{15.986419}{52.001698} + \frac{17.8713}{33.227446} + \frac{0.1557281}{16.394624} + \frac{26.801189}{10.823014} +$$

$$\frac{55.683757}{27.537845} + \frac{39.677817}{44.700967} + \frac{344.56674}{28.562509}$$

$$v = 0.1931271 + 0.8882361 + 6.6423452 + 3.4337233 +$$

$$26.26569 + 0.060283 + 0.5330587 + 0.4906385 +$$

$$0.3074211 + 0.5378475 + 0.0094987 + 2.4763149 +$$

$$2.0220811 + 0.8876277 + 12.063602 = 56.811495$$

$v = 56.81$

GRADOS DE LIBERTAD: $\nu = (r-1)(c-1) = (3-1)(5-1) = 2 \times 4 = 8$

DE LAS TABLAS DE LA DISTRIBUCION χ^2, PARA 95% DE NIVEL DE CONFIANZA Y 8 GRADOS DE LIBERTAD, SE TIENE:

$\chi^2_c = \chi^2_{0.95,8} = 15.5$

$v = 56.81 > \chi^2_c = 15.5$

POR TANTO SE RECHAZA LA HIPOTESIS DE INDEPENDENCIA A UN 95% DE NIVEL DE CONFIANZA, O SEA QUE SI HAY RELACION ENTRE EL FUNCIONAMIENTO DE LA PROTESIS Y EL HOSPITAL.

b)

FUNCIONAMIENTO	HOSPITAL				TOTALES	$\hat{r}_i$
	A	B	C	D	$X_{i.}$	
NULO	13	5	8	21	47	0.165
PARCIAL	18	10	36	56	120	0.421
NORMAL	16	16	35	51	118	0.414
TOTALES: $X_{.j}$	47	31	79	128	285	
$\hat{s}_j$	0.165	0.109	0.277	0.449		1.000

$$v = \frac{(13 - (285)(0.165)(0.165))^2}{285 \times 0.165 \times 0.165} + \frac{(5 - (285)(0.165)(0.109))^2}{285 \times 0.165 \times 0.109} +$$

$$\frac{(8 - (285)(0.165)(0.277))^2}{285 \times 0.165 \times 0.277} + \frac{(21 - (285)(0.165)(0.449))^2}{285 \times 0.165 \times 0.449} +$$

$$\frac{(18 - (285)(0.421)(0.165))^2}{285 \times 0.421 \times 0.165} + \frac{(10 - (285)(0.421)(0.109))^2}{285 \times 0.421 \times 0.109} +$$

$$\frac{(36 - (285)(0.421)(0.277))^2}{285 \times 0.421 \times 0.277} + \frac{(56 - (285)(0.421)(0.449))^2}{285 \times 0.421 \times 0.449} +$$

$$\frac{(16 - (285)(0.414)(0.165))^2}{285 \times 0.414 \times 0.165} + \frac{(16 - (285)(0.414)(0.109))^2}{285 \times 0.414 \times 0.109} +$$

$$\frac{(35 - (285)(0.414)(0.277))^2}{285 \times 0.414 \times 0.277} + \frac{(51 - (285)(0.414)(0.449))^2}{285 \times 0.414 \times 0.449} +$$

$$v = \frac{27.466771}{7.759125} + \frac{0.0158068}{5.125725} + \frac{25.259922}{13.025925} + \frac{0.0130474}{21.114225} +$$

$$\frac{3.2310961}{19.797525} + \frac{9.4763311}{13.078365} + \frac{7.6405529}{33.235845} + \frac{4.5230018}{53.873265} +$$

$$\frac{12.029452}{19.46835} + \frac{9.853886}{12.86091} + \frac{5.3674232}{32.68323} + \frac{3.9105458}{52.97751} +$$

$$v = 3.5399315 + 0.0030838 + 1.9392037 + 0.0006179 +$$

$$0.1632071 + 0.7245807 + 0.2298889 + 0.0839563 +$$

$$0.6178979 + 0.7661889 + 0.1642256 + 0.0738152$$

$$v = 8.3065975 \approx 8.31 .$$

GRADOS DE LIBERTAD: $\nu = (r-1)(c-1) = (3-1)(4-1) = 2 \times 3 = 6$

DE LAS TABLAS, PARA 95% DE CONFIANZA Y 6 GRADOS DE LIBERTAD, SE TIENE:

$$\chi_c^2 = \chi_{0.95,6}^2 = 12.6$$

$$v = 8.31 < \chi_c^2 = 12.6$$

POR LO TANTO SE ACEPTA LA HIPOTESIS. SE CONCLUYE QUE LAS VARIABLES SON INDEPENDIENTES, O SEA EL FUNCIONAMIENTO DE LAS PROTESIS ES INDEPENDIENTE DEL HOSPITAL.

c)

FUNCIONAMIENTO	HOSPITAL		TOTALES $X_{i.}$	$\hat{w}_i$
	(A+B+C+D)	E		
NULO	47	43	90	0.245
PARCIAL	120	29	149	0.406
NORMAL	118	10	128	0.349
TOTALES: $X_{.j}$	285	82	367	
$\hat{S}_j$	0.777	0.223		1.000

$$v = \frac{(47 - (367)(0.245)(0.777))^2}{367 \times 0.245 \times 0.777} + \frac{(43 - (367)(0.245)(0.223))^2}{367 \times 0.245 \times 0.223} +$$

$$\frac{(120 - (367)(0.406)(0.777))^2}{367 \times 0.406 \times 0.777} + \frac{(29 - (367)(0.406)(0.223))^2}{367 \times 0.406 \times 0.223} +$$

$$\frac{(118 - (367)(0.349)(0.777))^2}{367 \times 0.349 \times 0.777} + \frac{(10 - (367)(0.349)(0.223))^2}{367 \times 0.349 \times 0.223} +$$

$$v = \frac{522.76044}{69.863955} + \frac{526.65454}{20.051045} + \frac{17.854394}{115.77455} + \frac{17.8713}{33.227446} +$$

$$\frac{341.49225}{99.520491} + \frac{344.56674}{28.562509}$$

$v = 7.4825486 + 26.26569 + 0.1542169 + 0.5378475 +$

$3.4313763 + 12.063602 = 49.935281 \approx 49.94$

GRADOS DE LIBERTAD: $v = (r-1)(c-1) = (3-1)(2-1) = 2 \times 1 = 2$

DE LAS TABLAS, PARA UN 95% DE NIVEL DE CONFIANZA Y 2 GRADOS DE LIBERTAD, SE TIENE:

$\chi^2_c = \chi^2_{0.95,2} = 5.99$

$v = 49.94 > \chi^2_c = 5.99$

POR LO TANTO SE RECHAZA LA HIPOTESIS. SE CONCLUYE QUE CON 95% DE CONFIANZA LOS RESULTADOS DE LOS HOSPITALES A + B + C + D (JUNTOS) Y LOS DE E NO SON INDEPENDIENTES DEL FUNCIONAMIENTO DE LAS PROTESIS. EN GENERAL, SE PUEDE DECIR QUE LOS RESULTADOS DEL HOSPITAL E SON LOS QUE DAN LA DEPENDENCIA DE ESTE EXPERIMENTO.

3. EXPERIMENTOS PARA COMPARAR DOS TRATAMIENTOS

CUANDO INTERESA VERIFICAR SI DOS PROCEDIMIENTOS DISTINTOS PARA LOGRAR UN MISMO OBJETIVO CONDUCEN A RESULTADOS IGUALES, SE DISEÑA UN EXPERIMENTO QUE CONSISTE EN OBTENER UNA MUESTRA ALEATORIA DE LOS RESULTADOS LOGRADOS CON CADA TRATAMIENTO, Y COMPARAR ENTRE SI LAS MEDIAS Y VARIANCIAS CORRESPONDIENTES.

CUANDO LAS OBSERVACIONES SON INDEPENDIENTES, ESTO SE LOGRA MEDIANTE LAS PRUEBAS DE HIPOTESIS DE IGUALDAD DE MEDIAS Y DE VARIANCIAS.

CUANDO NO LO SON, LA COMPARACION SE HACE EN TERMINOS DE LAS DIFERENCIAS INDIVIDUALES DE CADA PAREJA DE RESULTADOS.

AL DISEÑAR EL EXPERIMENTO SE DEBEN CONSIDERAR DOS ALTERNATIVAS:

a. ASIGNAR AL AZAR A CADA ESPECIMEN EL TRATAMIENTO QUE LE SERA APLICADO; A ESTE PROCESO SE LE LLAMA DE ALEATORIZACION.

EJEMPLO

POR EJEMPLO, SI SE TRATARA DE VERIFICAR SI UN FERTILIZANTE ES MAS EFICIENTE QUE OTRO, UNA VEZ DEFINIDOS LOS LOTES PARA SIEMBRA NOMINALMENTE IGUALES, HABRIA QUE ASIGNAR AL AZAR CADA LOTE A CADA FERTILIZANTE. SUPONGAMOS QUE SE DISPONE DE 11 LOTES Y QUE 5 SE TRATARAN CON EL FERTILIZANTE A Y 6 CON EL B. EL EXPERIMENTO ALEATORIZADO SERIA

LOTE	1	2	3	4	5	6	7	8	9	10	11
FERTILIZANTE	A	A	B	B	A	B	B	B	A	A	B
COSECHA DE TOMATE	29.9	11.4	26.6	23.7	25.3	28.5	14.2	17.9	16.5	21.1	24.3

COSECHA CON FERTILIZANTE A	COSECHA CON FERTILIZANTE B
29.9	26.6
11.4	23.7
25.3	28.5
16.5	14.2
21.1	17.9
104.2	24.3
	135.2

$\overline{Y}_A = \frac{104.2}{5} = 20.84, \quad \overline{Y}_B = \frac{135.2}{6} = 22.53$

$\overline{Y}_B - \overline{Y}_A = 22.53 - 20.84 = 1.69$

LAS VARIANCIAS INSESGADAS VALEN

$S_A^2 = 52.50, \quad S_B^2 = 29.51$

PRUEBA DE HIPOTESIS PARA LA VARIANCIA:

$H_o: \sigma_A^2 = \sigma_B^2 \ ; \quad H_1: \sigma_A^2 > \sigma_B^2; \ 1-\alpha = 0.99$

$F = \frac{S_A^2}{S_B^2} = \frac{52.50}{29.51} = 1.78, \quad F_{0.01,4,5} = 11.4 > 1.78$

POR LO QUE SE ACEPTA H_o CON UN NIVEL DE CONFIANZA DEL 99%.

PRUEBA DE HIPOTESIS PARA LAS MEDIAS:

$H_o: \mu_A = \mu_B, \quad H_1: \mu_A \neq \mu_B, \quad 1-\alpha = 99\%$

$$T = \frac{\bar{y}_B - \bar{y}_A}{\varepsilon\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \,, \qquad \varepsilon = \sqrt{\frac{\nu_A S_A^2 + \nu_B S_B^2}{\nu_A + \nu_B}}$$ (CON VARIANCIAS INSESGADAS)

$$\nu_A = n_A - 1 = 4, \qquad \nu_B = n_B - 1 = 6 - 1 = 5$$

$$\varepsilon = \sqrt{\frac{4x52.50 + 5x29.51}{4 + 5}} = \sqrt{39.73} = 6.30$$

$$t = \frac{1.69}{6.30\sqrt{\frac{1}{5} + \frac{1}{6}}} = 0.44 < t_{0.01,9} = 3.25$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE IGUALDAD DE MEDIAS, O SEA, QUE CON 99% DE PROBABILIDAD EL RENDIMIENTO DE LAS TIERRAS CON AMBOS FERTILIZANTES ES EL MISMO.

b. APLICAR CADA TRATAMIENTO A GRUPOS O BLOQUES DE ESPECIMENES, EN ESTE CASO EN PAREJAS, QUE PERMITAN REDUCIR LA VARIANCIA O DISPERSION ALEATORIA DE LOS RESULTADOS, INVOLUCRANDO, A LA VEZ, UN PROCESO DE ALEATORIZACION EN LA ASIGNACION DE LOS BLOQUES; A ESTE PROCESO SE LA LLAMA DE AGRUPAMIENTO EN BLOQUES.

EJEMPLO

EN EL EJEMPLO INMEDIATO ANTERIOR LA INCERTIDUMBRE EN LOS RESULTADOS POR LOS EFECTOS ALEATORIOS INVOLUCRADOS SE PUEDE REDUCIR SI EN VEZ DE SORTEARSE LOS LOTES PARA CADA FERTILIZANTE, CADA

LOTE SE DIVIDE EN DOS PARTES IGUALES Y SE SORTEA QUE MITAD SE TRATARA CON CADA UNO DE ELLOS. CON ESTO LOS RESULTADOS QUEDAN AGRUPADOS POR PAREJAS (y_A, y_B), UNA PARA CADA LOTE, TENIENDOSE QUE y_B Y y_A NO SON INDEPENDIENTES. CON ESTO SE TIENE UN EXPERIMENTO CON AGRUPAMIENTO POR BLOQUES.

SOPONGAMOS QUE LAS PAREJAS DE DATOS QUEDARON DE LA SIGUIENTE MANERA PARA 5 LOTES:

y_A	y_B	$y_B - y_A = d$	d^2
29.9	26.6	-3.3	10.89
11.4	23.7	12.3	151.29
25.3	28.5	3.2	10.24
16.5	14.2	-2.3	5.29
21.1	17.9	-3.2	10.24
		6.7	187.95

$H_0 : \mu_d = 0; \ H_1 : \mu_d \neq 0; \ 1-\alpha = 0.99$

$\bar{d} = 6.7/5 = 1.34, \ \bar{d}^2 = 1.80$

$\overline{d^2} = 187.95/5 = 37.59$

$S_d^2 = 37.59 - 1.80 = 35.79$

$S_d = 5.98, \ t = \frac{\bar{d}}{S_d}\sqrt{n-1} = 0.448$

$t_{0.005,4} = 4.60 > 0.448;$ POR LO TANTO SE ACEPTA H_o.

EN ESTE CASO SE MANEJA LA ESTADISTICA $\bar{d}$ CON DISTRIBUCION t DE STUDENT.

EJEMPLO

CON EL FIN DE VERIFICAR SI DOS MATERIALES PARA FABRICAR SUELA DE ZAPATO SE DISEÑO UN EXPERIMENTO CON AGRUPAMIENTO POR BLOQUES Y ALEATORIZACION. EL AGRUPAMIENTO SE HIZO AL USAR EL ZAPATO DEL PIE IZQUIERDO CON UN MATERIAL Y EL DEL DERECHO CON EL OTRO; LA ALEATORIZACION SE HIZO AL ASIGNAR AL AZAR CUAL MATERIAL ESTARIA EN EL IZQUIERDO Y CUAL EN EL DERECHO, PARA CADA

NIÑO QUE USARIA LOS ZAPATOS DE PRUEBA.

LAS DURACIONES DE LOS ZAPATOS, EN MESES, FUERON:

NIÑO	MATERIAL A	MATERIAL B	DIFERENCIA = d	d^2
1	13.2 (I)	14.0 (D)	0.8	0.64
2	8.2 (I)	8.8 (D)	0.6	0.36
3	10.9 (D)	11.2 (I)	0.3	0.09
4	14.3 (I)	14.2 (D)	-0.1	0.01
5	10.7 (D)	11.8 (I)	1.1	1.21
6	6.6 (I)	6.4 (D)	-0.2	0.04
7	9.5 (I)	9.8 (D)	0.3	0.09
8	10.8 (I)	11.3 (D)	0.5	0.25
9	8.8 (D)	9.3 (I)	0.5	0.25
10	13.3 (I)	13.6 (D)	0.3	0.09
			4.1	3.03

$H_0: \mu_{\bar{d}} = 0; \quad H_1: \mu_d \neq 0; \quad 1-\alpha = 0.99$

$\bar{d} = 4.1/10 = 0.41, \quad S_d^2 = \overline{d^2} - \bar{d}^2 = \frac{3.03}{10} - 0.41^2 = 0.1349$

$S_d = 0.367, \quad t = \frac{0.41}{0.367}\sqrt{9} = 3.35 > t_{0.005,9} = 3.25$

POR LO TANTO SE RECHAZA LA HIPOTESIS DE IGUALDAD DE DURACION DE LAS SUELAS HECHAS CON AMBOS MATERIALES, CON UN 99% DE NIVEL DE CONFIANZA.

RESUMEN

1. LOS EXPERIMENTOS DEBEN SER COMPARABLES Y REPRODUCIBLES. CUANDO SE COMPARAN TRATAMIENTOS DEBE PROCURARSE QUE LOS

EXPERIMENTOS PARA CADA UN CORRAN EN PARALELO.

2. DEBE HABER REPLICAS DE CADA TRATAMIENTO. LAS VARIACIONES ENTRE LOS RESULTADOS DEBE PERMITIR ESTIMAR LOS "ERRORES" DEBIDOS AL AZAR

3. SIEMPRE QUE SEA POSIBLE SE DEBEN AGRUPAR LOS RESULTADOS EN BLOQUES PARA REDUCIR EL ERROR, AL HOMOGENIZAR LOS RESULTADOS DE CADA REPLICA.

4. EXPERIMENTOS PARA COMPARAR k TRATAMIENTOS

CON FRECUENCIA ES NECESARIO VERIFICAR SI MAS DE DOS "TRATAMIENTOS" CONDUCEN A RESULTADOS CON VALORES MEDIOS IGUALES. PARA HACER ESTO SE DISEÑA UN EXPERIMENTO EN EL QUE LOS ESPECIMENES (EL MATERIAL EXPERIMENTAL) SE ASIGNAN AL AZAR A CADA TRATAMIENTO.

SI LAS MEDIAS POBLACIONALES DE LOS TRATAMIENTOS SON η_A, η_B, η_C, ETC., INTERESA PROBAR LA HIPOTESIS NULA DE QUE $\eta_A = \eta_B = \eta_C \ldots$, EN CONTRA DE LA HIPOTESIS ALTERNATIVA DE QUE NO TODAS LA MEDIAS SON IGUALES. ESTA PRUEBA SE REALIZA MEDIANTE LA TECNICA ESTADISTICA CONOCIDA COMO ANALISIS DE VARIANCIA.

SUPONGAMOS, POR EJEMPLO, QUE SE TRATA DE VERIFICAR SI CUATRO MEDICINAS DIFERENTES CONDUCEN A TIEMPOS IGUALES DE COAGULACION DE LA SANGRE DE LOS PACIENTES. PARA ESTO SE OBTIENE UNA MUESTRA ALEATORIA DE 24 INDIVIDUOS, A LOS CUALES SE LES ASIGNAN AL AZAR LAS CUATRO MEDICINAS, SE LES APLICA EL TRATAMIENTO DURANTE EL TIEMPO PRESTABLECIDO Y SE LES SACA UNA MUESTRA DE SANGRE A CADA UNO PARA MEDIR LOS TIEMPOS INDIVIDUALES DE COAGULACION. EL NUMERO DE ESPECIMENES (INDIVIDUOS) NO NECESITA SER EL MISMO PARA CADA TRATAMIENTO.

SUPONGAMOS AHORA QUE LAS MUESTRAS DE TIEMPOS DE COAGULACION ASOCIADOS A CADA UNO DE LOS CUATRO TRATAMIENTOS SON LOS PRESENTADOS EN LA SIGUIENTE TABLA:

	MEDICINA			
	A	B	C	D
	62 seg	63 seg	68 seg	56 seg
	60	67	66	62
	63	71	71	60
	59	64	67	61
		65	68	63
		66	68	64
				63
				59
PROMEDIOS	61	66	68	61
PROMEDIO GLOBAL: 64 seg				

EL ANALISIS DE VARIANCIA, EN ESTE CASO, SERVIRIA PARA DISCRIMINAR SI LA VARIABILIDAD DE LOS RESULTADOS QUE SE TIENEN ENTRE LOS DIVERSOS TRATAMIENTOS ES IGUAL A LA QUE SE TIENE DENTRO DE CADA TRATAMIENTO Y, POR LO TANTO, PODER AFIRMAR QUE ESTA SE DEBE AL AZAR Y NO A DIFERENCIAS REALES ENTRE LOS RESULTADOS DE LOS TRATAMIENTOS.

EL ANALISIS DE VARIANCIA PARTE DE LA CONSIDERACION DE QUE CADA RESULTADO EXPERIMENTAL ES CONSECUENCIA DE LOS EFECTOS DEBIDOS A FACTORES O VARIABLES ALEATORIAS QUE SE SUJETAN A CONTROL, Y DE OTRAS QUE NO SE CONTROLAN; A ESTAS ULTIMAS SE LES LLAMA VARIABLES RESIDUALES, Y A SUS EFECTOS SE LES DENOMINA

EFECTOS RESIDUALES O ERROR. A MAYOR NUMERO DE VARIABLES BAJO CONTROL, CORRESPONDE UN MENOR EFECTO RESIDUAL.

BAJO ESTA PREMISA, EL ANALISIS DE VARIANCIA SE FUNDAMENTA EN LAS SIGUIENTES HIPOTESIS:

1. EL VALOR MEDIO DE CADA VARIABLE RESIDUAL ES CERO.
2. LAS VARIABLES RESIDUALES SON INDEPENDIENTES.
3. LAS VARIABLES RESIDUALES TIENEN IGUAL VARIANCIA.
4. LAS VARIABLES RESIDUALES TIENEN DISTRIBUCION NORMAL.

DE ESTAS HIPOTESIS LA QUE REQUIERE MAYOR ANALISIS, EN CUANTO A SU VERIFICACION, ES LA NUMERO 3. SI ESTA HIPOTESIS NO SE CUMPLE, SE RECOMIENDA OBTENER MUESTRAS IGUALES PARA CADA TRATAMIENTO, YA QUE EN ESE CASO EL EFECTO DE LA DIFERENCIA DE VARIANCIAS NO ES IMPORTANTE.

FACTORES. EN TERMINOS GENERALES, LLAMAREMOS FACTORES A LAS CUALIDADES O PROPIEDADES DE ACUERDO A LAS CUALES SE HACE LA CLASIFICACION DE LOS DATOS. POR EJEMPLO, SI UN PRODUCTO SE ELABORA CON DIFERENTES TIPOS DE MAQUINAS Y VARIOS OPERARIOS DURANTE LOS DIVERSOS DIAS, ENTONCES SE PUEDEN CONSIDERAR EN EL ANALISIS AL MENOS TRES FACTORES: MAQUINA, OPERARIO Y DIA. CADA UNO DE ESTOS FACTORES TENDRA SUS PROPIOS NIVELES; POR EJEMPLO, HABRA LAS MAQUINAS A, B Y C (3 NIVELES), LOS OPERARIOS JUAN Y JORGE (2 NIVELES) Y LOS DIAS DE LUNES A VIERNES (5 NIVELES).

CLASIFICACION EN UNA DIRECCION

SE TIENE UNA CLASIFICACION EN UNA DIRECCION CUANDO SE COMPARAN LOS RESULTADOS EN TERMINOS DE LOS DIVERSOS NIVELES QUE TIENE UN SOLO FACTOR. EN EL CASO DEL EJEMPLO DE LAS MEDICINAS DESCRITO ANTERIORMENTE, EL FACTOR UNICO ES MEDICINA Y TIENE CUATRO NIVELES; SE TRATA DE COMPARAR LOS RESULTADOS DE LA VARIABLE "TIEMPOS DE COAGULACION" QUE SE OBTIENEN CON CADA UNO DE LOS NIVELES, TRATAMIENTOS O GRUPOS.

LA FORMULACION DEL MODELO PUEDE TENER DOS VARIEDADES:

1. LOS NIVELES O TRATAMIENTOS SON FIJOS Y SE TOMAN TODOS EN EL EXPERIMENTO. A ESTE MODELO SE LE DENOMINA DE NIVELES FIJOS, PARAMETRICO O MODELO I.

2. LOS NIVELES QUE SE INCLUYEN EN EL EXPERIMENTO SON SOLO ALGUNOS DE LOS POSIBLES, Y SE SELECCIONAN AL AZAR; EN ESTE CASO EL FACTOR ES EN SI UNA VARIABLE ALEATORIA. A ESTE MODELO SE LE DENOMINA DE NIVELES ALEATORIOS O MODELO II.

SEA X_{ti} EL i-ESIMO RESULTADO DE APLICAR EL TRATAMIENTO t, $t = 1, 2, \ldots, k$; $i = 1, 2, \ldots, n_t$.

CADA RESULTADO ESTARA COMPUESTO DE UN TERMINO QUE REPRESENTA EL EFECTO DEL TRATAMIENTO RESPECTIVO, Y OTRO TERMINO QUE ES EL EFECTO RESIDUAL O ERROR.

SI DENOTAMOS CON Z_{ti} A DICHO EFECTO RESIDUAL, LAS HIPOTESIS

1 A 4 ANTERIORES SERIAN EN ESTE CASO:

1) $E(Z_{ti}) = 0$ PARA TODO t E i

2) LAS Z_{ti} SON MUTUAMENTE INDEPENDIENTES

3) $\sigma^2(Z_{ti}) = \sigma^2$ PARA TODO t E i

4) LAS Z_{ti} TIENEN DISTRIBUCION NORMAL

EN EL CASO DE QUE SE TUVIERAN FACTORES FIJOS, EL MODELO I CONSISTE EN DESCOMPONER CADA OBSERVACION EN DOS TERMINOS: UNO DEBIDO AL TRATAMIENTO, ξ_t, Y EL OTRO DEBIDO AL AZAR O RESIDUAL, Z_{ti}, ES DECIR

$$X_{ti} = \xi_t + Z_{ti}$$

POR CONVENIENCIA, REPRESENTEMOS A ξ_t EN LA FORMA

$$\xi_t = \xi + \gamma_t$$

DONDE ξ ES EL EFECTO MEDIO DE TODOS LOS TRATAMIENTOS Y γ_t ES LA DESVIACION RESPECTO A ξ QUE TIENE EL TRATAMIENTO t. AL HACER ESTO TENDREMOS k+1 TERMINOS, ξ, γ_1, $\gamma_2, \ldots, \gamma_k$, PARA REPRESENTAR A LOS k PARAMETROS, ξ_1, $\xi_2, \ldots \xi_k$, POR LO QUE SE LE DEBE IMPONER ALGUNA CONDICION A LAS γ_t; DICHA CONDICION SERA QUE

$$\sum_{t=1}^{k} n_t \gamma_t = 0 \qquad \text{(A)}$$

LO CUAL SIGNIFICA QUE LA MEDIA ξ ES UN PROMEDIO PESADO DE LAS ξ_t, ES DECIR

$$\xi = \sum_{t=1}^{k} n_t \xi_t / N; \quad N = \sum_{t=1}^{k} n_t$$

EN EL CASO DE LOS NIVELES ALEATORIOS EL MODELO II SERIA

$$X_{ti} = \xi + U_t + Z_{ti}$$

EN DONDE LAS U_t SON VARIABLES ALEATORIAS MUTUAMENTE INDEPENDIENTES CON MEDIA CERO Y VARIANCIA σ_U^2, CON DISTRIBUCION NORMAL E INDEPENDIENTES DE LA Z_{ti}.

MODELO PARAMETRICO

SI LOS NIVELES O TRATAMIENTOS SON FIJOS, EL MODELO I O PARAMETRICO SERA

$$X_{ti} = \xi + \gamma_t + Z_{ti} \tag{1}$$

EL PROMEDIO ARITMETICO DE LOS DATOS DE CADA GRUPO O TRATAMIENTO SERA

$$\bar{X}_{t.} = \sum_{i=1}^{n_t} X_{ti} / n_t, \quad t = 1, 2, \ldots, k \tag{2}$$

SUSTITUYENDO LA EC (1) EN LA EC (2):

$$\bar{X}_{t.} = \xi + \gamma_t + \bar{Z}_{t.} \tag{3}$$

DONDE

$$\bar{Z}_{t.} = \sum_{i=1}^{n_t} Z_{ti}/n_t \qquad (4)$$

LA MEDIA GLOBAL DE LAS OBSERVACIONES ES

$$\bar{X}_{..} = \sum_{t=1}^{k} n_t \bar{X}_{t.}/N \qquad (5)$$

SUSTITUYENDO LA EC (3) EN LA (5) Y CONSIDERANDO LA CONDICION (A):

$$\bar{X}_{..} = \xi + \bar{Z}_{..} \qquad (6)$$

DONDE $\bar{Z}_{..} = \sum_{t=1}^{k} n_t \bar{Z}_{t.}/N$, PUESTO QUE

$$\sum_{t=1}^{k} n_t \gamma_t = 0$$

EL PROBLEMA QUE NOS OCUPA ES EL PROBAR LA HIPOTESIS DE QUE $\gamma_1 = \gamma_2 = \ldots = \gamma_k = 0$; ES NATURAL, POR LO TANTO, QUE CALCULEMOS LAS DIFERENCIAS DEL PROMEDIO DE CADA TRATAMIENTO MENOS EL PROMEDIO GLOBAL

$$\bar{X}_{t.} - \bar{X}_{..} = \xi + \gamma_t + \bar{Z}_{t.} - \xi - \bar{Z}_{..} = \gamma_t + \bar{Z}_{t.} - \bar{Z}_{..} \qquad (7)$$

CUYA ESPERANZA ES PRECISAMENTE γ_t.

PARA CADA GRUPO, LA VARIANCIA DE LAS OBSERVACIONES SE OBTIENE EN TERMINOS DE LAS DIFERENCIAS

$$X_{ti} - \bar{X}_{t.} = \xi + \gamma_t + Z_{ti} - (\xi + \gamma_t + \bar{Z}_{t.}) = Z_{ti} - \bar{Z}_{t.} \qquad (8)$$

AHORA, SUMANDO Y RESTANDO $\bar{X}_{t.}$ a $X_{ti} - \bar{X}_{..}$ OBTENEMOS

$$X_{ti} - \bar{X}_{..} = (\bar{X}_{t.} - \bar{X}_{..}) + (X_{ti} - \bar{X}_{t.}) \qquad (9)$$

LA SUMA TOTAL DE LOS CUADRADOS DE ESTAS DIFERENCIAS PARA TODA LA MUESTRA SERA

$$\sum_{t=1}^{k}\sum_{i=1}^{n_t}(X_{ti} - \bar{X}_{..})^2 = \sum_{t=1}^{k}\sum_{i=1}^{n_t}(\bar{X}_{t.} - \bar{X}_{..})^2 + \sum_{t=1}^{k}\sum_{i=1}^{n_t}(X_{ti} - \bar{X}_{t.})^2$$

$$+ \sum_{t=1}^{k}\sum_{i=1}^{n_t} 2(\bar{X}_{t.} - \bar{X}_{..})(X_{ti} - \bar{X}_{t.})$$

$$= \sum_{t=1}^{k} n_t(\bar{X}_{t.} - \bar{X}_{..})^2 + \sum_{t=1}^{k}\sum_{i=1}^{n_t}(X_{ti} - \bar{X}_{t.})^2 \qquad (10)$$

YA QUE LA SUMATORIA DEL DOBLE PRODUCTO VALE CERO PORQUE

$$\sum_{i=1}^{n_t}(X_{ti} - \bar{X}_{t.}) = 0.$$

DE ESTA MANERA SE TIENE QUE:

[SUMA TOTAL DE CUADRADOS] = [SUMA DE CUADRADOS ENTRE GRUPOS] + [SUMA DE CUADRADOS DENTRO DE GRUPOS]

TOMANDO EN CUENTA LA EC (7), <u>LA ESPERANZA DE LA SUMA DE LOS CUADRADOS ENTRE GRUPOS ES</u>

$$E\{\sum_{t=1}^{k} n_t(\gamma_t + \bar{Z}_{t.} - \bar{Z}_{..})^2\} = \sum_{t=1}^{k} n_t\gamma_t^2 + E\{\sum_{t=1}^{k} n_t(\bar{Z}_{t.} - \bar{Z}_{..})^2\}$$

(11)

YA QUE

$$E\{\sum_{t=1}^{k} 2n_t\gamma_t (\bar{Z}_{t.} - \bar{Z}_{..})\} = 0$$

EN VIRTUD DE LA HIPOTESIS 1 DE QUE $E(Z_{ti} = 0)$.

ADEMAS

$$E\{\sum_{t=1}^{k} n_t(\bar{Z}_{t.} - \bar{Z}_{..})^2\} = E\{\sum_{t=1}^{k} n_t \bar{Z}_{t.}^2 + \sum_{t=1}^{k} n_t\bar{Z}_{..}^2 -$$

$$- 2\sum_{t=1}^{k} n_t \bar{Z}_{t.}\bar{Z}_{..}\} = E\{\sum_{t=1}^{k} n_t\bar{Z}_{t.}^2 + N\bar{Z}_{..}^2 - 2\bar{Z}_{..}(N\bar{Z}_{..})\}$$

PUESTO QUE $\sum_{t=1}^{k} n_t \bar{Z}_{t.} = N\bar{Z}_{..}$

$$E\{\sum_{t=1}^{k} n_t (\bar{Z}_{t.} - \bar{Z}_{..})^2\} = E\{\sum_{t=1}^{k} n_t \bar{Z}_{t.}^2 - N\bar{Z}_{..}^2\} =$$

$$= \sum_{t=1}^{k} n_t E(\bar{Z}_{t.}^2) - NE(\bar{Z}_{..}^2)$$

$$= \sum_{t=1}^{k} n_t \frac{\sigma^2}{n_t} - N\frac{\sigma^2}{N} = k\sigma^2 - \sigma^2 = (k-1)\sigma^2$$

(12)

PUESTO QUE $E(\bar{Z}_{t.}) = E(\bar{Z}_{..}) = 0$, DEBIDO A QUE $E(Z_{ti}) = 0$.

AQUI σ^2 ES LA VARIANCIA DEL ERROR O RESIDUAL.

SUSTITUYENDO LA EC (12) EN LA EC (11) LA SUMA DE LOS

CUADRADOS ENTRE GRUPOS QUEDA EN LA FORMA

$$\sum_{t=1}^{k} n_t(\bar{X}_{t.} - \bar{X}_{..})^2 = \sum_{t=1}^{k} n_t \gamma_t^2 + (k-1)\sigma^2 \qquad (13)$$

POR SU PARTE, LA ESPERANZA DE LA SUMA DE CUADRADOS DENTRO DE GRUPOS, TOMANDO EN CUENTA LA EC 8, ES

$$E\{\sum_{t=1}^{k}\sum_{i=1}^{n_t}(X_{ti} - \bar{X}_{t.})^2\} = E\{\sum_{t=1}^{k}\sum_{i=1}^{n_t}(Z_{ti} - \bar{Z}_{t.})^2\} =$$

$$= \sum_{t=1}^{k} E\{\sum_{i=1}^{n_t}(Z_{ti} - \bar{Z}_{t.})^2\} = \sum_{t=1}^{k}(n_t - 1)\sigma^2 = (N-k)\sigma^2 \qquad (14)$$

DIVIDIENDO LAS SUMAS DE LOS CUADRADOS DADAS EN LAS ECS (13) y (14), ENTRE LOS GRADOS DE LIBERTAD (k-1) y (N-k), RESPECTIVAMENTE, RESULTAN LOS VALORES MEDIOS CUADRATICOS ENTRE Y DENTRO DE LOS GRUPOS DADOS POR LOS TERMINOS

$\dfrac{\sum_{t=1}^{k} n_t \gamma_t^2}{k-1} + \sigma^2$ y σ^2, RESPECTIVAMENTE.

BAJO LA HIPOTESIS NULA DE QUE TODAS LAS γ_t SON CERO, EL VALOR MEDIO CUADRATICO ENTRE GRUPOS VALE σ^2, YA QUE EN TAL CASO $\sum_{t=1}^{k} n_t \gamma_t^2/(k-1) = 0$.

DE ESTA MANERA, BAJO LA HIPOTESIS NULA, LOS VALORES MEDIOS CUADRATICOS DENTRO Y ENTRE GRUPOS SON ESTIMADORES INSESGADOS DE LA VARIANCIA DEL ERROR O RESIDUAL, σ^2.

POR LO TANTO, PARA PROBAR LA HIPOTESIS DE QUE $\gamma_1 = \gamma_2 = \ldots = \gamma_k = 0$,

BASTA PROBAR LA HIPOTESIS DE QUE LOS VALORES MEDIOS CUADRATICOS DENTRO Y ENTRE GRUPOS SON IGUALES, LO CUAL SE PUEDE HACER MEDIANTE UNA PRUEBA F, AL TOMAR EN CUENTA LA HIPOTESIS 4, DE QUE LOS ERRORES TIENEN DISTRIBUCION NORMAL. LA ESTADISTICA F ES, ENTONCES

$$F = \frac{\text{VALOR MEDIO CUADRATICO ENTRE GRUPOS}}{\text{VALOR MEDIO CUADRATICO DENTRO DE GRUPOS}} = \frac{MSB}{MSW}$$

QUE TIENE DISTRIBUCION F CON k-1 Y N-k GRADOS DE LIBERTAD.

EJEMPLO

EN EL EJEMPLO DE LAS MEDICINAS DESCRITO ANTERIORMENTE, SE TIENE UN CASO DE UN SOLO FACTOR CON 4 NIVELES. LA SUMA DE CUADRADOS ENTRE GRUPOS ES

$$\sum_{t=1}^{k} n_t(\bar{X}_{t.} - \bar{X}_{..})^2 = 4(61-64)^2 + 6(66-64)^2 + 6(68-64)^2 + $$
$$+ 8(61-64)^2 = 228$$

LA SUMA DE CUADRADOS DENTRO DE GRUPOS ES

$$[(62-61)^2 + (60-61)^2 + (63-61)^2 + (59-61)^2] + [(63-66)^2 +$$
$$(67-66)^2 + (71-66)^2 + (64-66)^2 + (65-66)^2 + (66-66)^2] +$$
$$[(68-68)^2 + (66-68)^2 + (71-68)^2 + (67-68)^2 + (68-68)^2 + (68-68)^2]+$$
$$+[(56-61)^2 + (62-61)^2 + (60-61)^2 + (61-61)^2 + (63-61)^2 + (64-61)^2 +$$
$$+(63-61)^2 + (59-61)^2] = 10 + 40 + 14 + 48 = 112$$

EN TAL CASO: H_0: $E(MSB) = E(MSW)$; H_1: $E(MSB) > E(MSW)$; $1-\alpha = 0.05$

$$F = \frac{228/3}{112/20} = \frac{76}{5.6} = 13.6 > F_{0.95,\ 3,20} = 3.10$$

POR LO QUE SE RECHAZA LA HIPOTESIS NULA DE IGUALDAD DE TIEMPOS DE COAGULACION PARA LAS CUATRO MEDICINAS, A UN 95% DE NIVEL DE CONFIANZA.

FORMULAS SIMPLIFICADAS PARA EL ANALISIS DE VARIANCIA EN UNA DIRECCION

SUMA TOTAL DE CUADRADOS = SST = $\sum_{ti}\sum (X_{ti} - \bar{X}_{..})^2$

$$SST = \sum_{ti}\sum X_{ti}^2 - N\bar{X}_{..}^2$$

SUMA DE CUADRADOS ENTRE TRATAMIENTOS = SSB = $\sum_t (\bar{X}_{t.} - \bar{X}_{..}^2)$

$$SSB = \sum_t \frac{(\sum_i X_{ti})^2}{n_t} - \frac{(\sum_{ti}\sum X_{ti})^2}{N} = \sum_t \frac{(\sum_i X_{ti})^2}{n_t} - N\bar{X}_{..}^2$$

SUMA DE CUADRADOS DENTRO DE GRUPOS SSW = SST - SSB

$$SSW = \sum_{ti}\sum X_{ti}^2 - \sum \frac{(\sum_i X_{ti})^2}{n_t} = SST - SSB$$

EL RESUMEN DEL ANALISIS DE VARIANCIA SE PUEDE HACER EN LA SIGUIENTE TABLA

FUENTES DE VARIABILIDAD	SS	g. DE l.	MS	F
TRATAMIENTOS (ENTRE GRUPOS)	SSB	k-1	$\frac{SSB}{k-1}$ = MSB	$\frac{MSB}{MSW}$
ERROR (DENTRO DE GRUPOS)	SSW	N-k	$\frac{SSW}{N-k}$ = MSW	
TOTALES	SST	N-1		

EJEMPLO

CON EL FIN DE VERIFICAR SI CIERTO TIPO DE LESION CEREBRAL AFECTA LA CAPACIDAD DE APRENDIZAJE, SI ESTA APARECE EN EL LADO IZQUIERDO, DERECHO O EN AMBOS, SE DISEÑO UN EXPERIMENTO CONSISTENTE EN OCASIONAR DICHO TIPO DE LESION A UNAS MUESTRAS ALEATORIAS DE RATAS Y TOMAR COMO COMPARACION A OTRO GRUPO DE RATAS SIN DICHO TIPO DE LESION (GRUPO DE CONTROL, I).

LOS INTENTOS DE APRENDIZAJE DE CIERTA RUTINA SE PRESENTAN EN LA SIGUIENTE TABLA.

	GRUPOS				
	I	II	III	IV	
	20	24	20	27	
	18	22	22	35	
	26	25	30	18	
	19	25	27	24	
	26	20	22	28	
	24	21	24	32	
	26	34	28	16	
		18	21	18	
		32	23	25	
		23	25		
		22	18		
			30		
			32		
$\sum_{i=1}^{n_t} X_{ti} =$ TOTALES	159	266	322	223	970, $\overline{X}_{..} = 970/40 = 24.25$
n_t	7	11	13	9	$N = 40$
$\overline{X}_{t.}$	22.71	24.18	24.77	24.78	
τ_t	-1.54	-0.07	0.52	0.53	

$\sum_{t}\sum_{i} X_{ti}^2 = 20^2 + 18^2 + \ldots + 18^2 + 25^2 = 24,424;\ N\overline{X}_{..}^2 = 40(24.25)^2 = 23,522.5$

SST = 24, 424 - 23, 522.5 = 901.5

$$\sum_t \frac{(\sum_i X_{ti})^2}{n_t} = \frac{159^2}{7} + \frac{266^2}{11} + \frac{322^2}{13} + \frac{223^2}{9} = 23,545.1$$

SSB = 23, 545.1 - 23, 522.5 = 22.6

SSW = SST - SSB = 901.5 - 22.6 = 878.9

FUENTE DE VARIABILIDAD	SS	g. DE l.	MS	F
ENTRE GRUPOS	SSB = 22.6	3	$\frac{22.6}{3} = 7.5$	$\frac{7.5}{24.4}$
DENTRO DE GRUPOS	SSW = 878.9	36	$\frac{878.9}{36} = 24.4$	
TOTALES	SST = 901.5	39		

$$F = \frac{7.5}{24.4} < F_{0.95,3,36} = 2.8$$

POR LO TANTO SE ACEPTA LA HIPOTESIS DE QUE EL NUMERO MEDIO DE INTENTOS PARA APRENDER CIERTA RUTINA ES IGUAL EN LOS CUATRO TRATAMIENTOS O GRUPOS, CON UN 95% DE NIVEL DE CONFIANZA.

<u>ESTIMACION DE LOS EFECTOS</u>

SI SE OBTIENE LA ESPERANZA DE LA DIFERENCIA $\overline{X}_{t.} - \overline{X}_{..}$, DE LA EC (7) SE OBTIENE

$$E(\overline{X}_{t.} - \overline{X}_{..}) = E(\gamma_t)$$

O SEA QUE $\overline{X}_{t.} - \overline{X}_{..}$ ES UN ESTIMADOR PUNTUAL INSESGADO DE LA MAGNITUD DE LOS EFECTOS.

EJEMPLO

LOS SIGUIENTES DATOS SE OBTUVIERON DE UN EXPERIMENTO COMPLETAMENTE ALEATORIZADO PARA COMPARAR LAS PROPIEDADES REFLECTIVAS DE CUATRO TIPOS DE PINTURA. LOS RESULTADOS FUERON OBTENIDOS MEDIANTE UN INSTRUMENTO OPTICO SIENDO LOS SIGUIENTES:

	PINTURA #1	PINTURA #2	PINTURA #3	PINTURA #4	TOTALES
	195	45	230	110	
	150	40	115	55	
	205	195	235	120	
	120	65	225	50	
	160.	145		80	
		195			
n_t	5	6	4	5	N = 20
$\bar{X}_{t.}$	166	114.167	201.25	83.0	$\bar{X}_{..}$ =136.75
TOTALES: $\Sigma X_{t.}$	830	685	805	415	2735

a) ELABORAR LA TABLA DE ANALISIS DE VARIANCIA (MEDIANTE LOS PROCEDIMIENTOS ORIGINALES Y SIMPLIFICADOS).

$$\bar{X}_{1.} = \frac{195+150+205+120+160}{5} = 166; \quad \bar{X}_{2.} = \frac{45+40+195+65+145+195}{6} = 114.167$$

$$\bar{X}_{3.} = \frac{230+115+235+225}{4} = 201.25; \quad \bar{X}_{4.} = \frac{110+55+120+50+80}{5} = 83.0$$

$$\bar{X}_{..} = \frac{166 \times 5 + 114.167 \times 6 + 201.25 \times 4 + 5 \times 83}{20} = 136.75$$

SUMA DE CUADRADOS ENTRE GRUPOS:

$$SSB = \sum_{t=1}^{k} n_t (\bar{X}_{t.} - \bar{X}_{..})^2 = 5(166 - 136.75)^2 + 6(114.167 - 136.75)^2$$

$$+ 4(201.25 - 136.75)^2 + 5(83 - 136.75)^2$$

$$= 5 \times 855.56 + 6 \times 509.99 + 4 \times 4160.25 + 5 \times 2889.06$$

$$= 4277.81 + 3059.95 + 16641 + 14445.31 = \underline{38424.08}$$

SUMA DE CUADRADOS DENTRO DE GRUPOS:

$$SSW = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (X_{ti} - \bar{X}_{t.})^2$$

$$= [(195 - 166)^2 + (150 - 166)^2 + (205 - 166)^2 + (120 - 166)^2 + (160 - 166)^2]$$

$$+ [(45 - 114.167)^2 + (40 - 114.167)^2 + (195 - 114.167)^2 + (65 - 114.167)^2$$

$$+ (145 - 114.167)^2 + (195 - 114.167)^2] + [(230 - 201.25)^2 +$$

$$+ (115 - 201.25)^2 + (235 - 201.25)^2 + (225 - 201.25)^2] + [(110 - 83)^2$$

$$+ (55 - 83)^2 + (120 - 83)^2 + (50 - 83)^2 + (80 - 83)^2]$$

$$= 4770 + 26720.833 + 9968.75 + 3980$$

$$SSW = \underline{45439.583}$$

CON ESTOS DATOS PODEMOS FORMULAR LA TABLA DE ANALISIS DE VARIANCIA COMO SIGUE:

FUENTES DE VARIABILIDAD	SS	GRADOS DE LIBERTAD	MS	$\hat{F}$
TIPOS DE PINTURA (ENTRE GRUPOS)	SSB = 38424.08	#TIPOS DE PINT-1 = K - 1 = 4 - 1 = 3	$MSB = SSB/(k-1) = \frac{38424.08}{3} = 12808.03$	$\hat{F} = \frac{MSB}{MSW} = 4.51$
ERRAR (DENTRO DE GRUPOS)	SSW = 45439.583	#ELEM. DE LA M . -#TIPOS DE PINT. N - K = 20-4 = 16	$MSW = SSW/(N-k) = \frac{45439.583}{16} = 2839.974$	
TOTALES	SST = SSB + SSW = 83863.66			

EL VALOR DE F TEORICO PARA UN NIVEL DE CONFIANZA DEL 95% ES:

$$F_{0.95,3,16} = 3.24$$

CONCLUSION:

DADO QUE $F_{0.95,3,16} < \hat{F}$ (3.24 < 4.51), ENTONCES $\hat{F}$ CAE EN LA REGION DEL RECHAZO, POR LO CUAL NOSE ACEPTA LA HIPOTESIS DE QUE LOS VALORES MEDIOS DE LAS REFLECTANCIAS DE LOS 4 TIPOS DE PINTURA ES IGUAL EN TODOS LOS TIPOS DE PINTURA, CON UN NIVEL DE CONFIANZA DEL 95%.

CALCULO DE LAS SUMAS DE CUADRADOS POR EL METODO SIMPLIFICADO

$$SSB = \sum_{t=1}^{k} \frac{(\sum_i X_{ti})^2}{n_t} - N\bar{X}_{..}^2 = \frac{(195 + 150 + 205 + 120 + 160)^2}{5} +$$

$$\frac{(45 + 40 + 195 + 65 + 145 + 195)^2}{6} + \frac{(230 + 115 + 235 + 225)^2}{4} +$$

$$+ \frac{(110 + 55 + 120 + 50 + 80)^2}{5} - 20 \times 136.75^2$$

$$SSB = \frac{830^2}{5} + \frac{685^2}{6} + \frac{805^2}{4} + \frac{415^2}{5} - 20 \times 136.75^2$$

$$= 412435.42 - 374011.25 = \underline{38424.17}$$

$$SST = \sum_t \sum_i X_{ti}^2 - N\bar{X}_{..}^2 -$$

$$= 195^2 + 150^2 + 205^2 + \ldots + 45^2 + 40^2 + \ldots +$$

$$+ 230^2 + 115^2 + \ldots + 50^2 + 80^2$$

$$- 20 \times 136.75^2$$

$$SST = 457875 - 374011.25 = 83863.75$$

$$SST = SSB + SSW \Rightarrow \quad SSW = SST - SSB = 83863.75 - 38424.17$$

$$SSW = 45439.58$$

MEDIDA DE ASOCIACION ENTRE EL FACTOR Y LA VARIABLE

UNA MEDIDA DESCRIPTIVA DEL GRADO DE ASOCIACION O CORRELACION QUE EXISTE ENTRE LA VARIABLE DEPENDIENTE Y EL FACTOR (O VARIABLE INDEPENDIENTE), ES

$$\eta^2 = \frac{SSB}{SST} = \frac{SST - SSW}{SST} \tag{20}$$

QUE CORRESPONDE A LA PROPORCION DE LA SUMA DE CUADRADOS QUE SE EXPLICA POR LA RELACION ENTRE AMBAS VARIABLES.

SE OBSERVA QUE η^2 VALE UNO CUANDO TODA LA VARIACION SE EXPLICA POR LA RELACION, ES DECIR, QUE SE TIENE UNA RELACION PERFECTA, Y VALE CERO CUANDO SSB = 0, O SEA, CUANDO NO HAY NINGUNA RELACION.

MODELO DE NIVELES ALEATORIOS

ES EL ANALISIS DE VARIANCIA CON UN SOLO FACTOR EN EL QUE LOS NIVELES DEL MISMO NO CUBREN TODOS LOS VALORES POSIBLES DE ESTE, SINO SOLO ALGUNOS DE ELLOS, CADA OBSERVACION QUEDA EN LA FORMA

$$X_{ti} = \xi + U_t + Z_{ti} \tag{21}$$

EN ESTE CASO, LA ESPERANZA DE LA SUMA DE CUADRADOS DENTRO DE GRUPOS ES IGUAL QUE EN EL MODELO FACTORIAL, ES DECIR, $(N - k)\ \sigma^2$.

POR SU PARTE LA SUMA DE CUADRADOS ENTRE GRUPOS ES:

$$\sum_t n_t \, (\bar{X}_{t.} - \bar{X}_{..})^2 = \sum_k n_t \left[(U_t - \bar{U}) + (\bar{Z}_{t.} - \bar{Z}_{..}) \right]^2 \tag{22}$$

DONDE $\bar{U} = \sum_t n_t \, U_t / N$

AL ELEVAR AL CUADRADO EL BINOMIO DE LA EC (22) Y OBTENER LA ESPERANZA CORRESPONDIENTE APARECERA EL TERMINO

$$E\left[\sum_t 2 n_t (U_t - \bar{U})(\bar{Z}_{t.} - \bar{Z}_{..})\right] = 0$$

DEBIDO A QUE SE CONSIDERO LA HIPOTESIS DE QUE LAS U Y LAS Z SON VARIABLES ALEATORIAS INDEPENDIENTES.

LOS OTROS DOS TERMINOS SON:

$$E\left[\sum_t n_t (\bar{Z}_{t.} - \bar{Z}_{..})^2\right] = (k - 1)\,\sigma^2 \tag{23}$$

Y

$$E\left[\sum_t n_t (U_t - \bar{U})^2\right] = E\left[\sum_t n_t \, U_t^2 - N\bar{U}^2\right] \tag{24}$$

PUESTO QUE $E(U_t) = 0$ y $VAR(U_t) = \sigma_u^2$

SE TIENE QUE $E(U_t^2) = \sigma_u^2$ y $E(\bar{U}^2) = \sigma_u^2 \sum_t (n_t/N)^2$

(25)

POR LO TANTO, LA ESPERANZA DE LA SUMA DE CUADRADOS ENTRE GRUPOS ES

$$E(SSB) = (k-1)\sigma^2 + \sigma_u^2 \left(\sum_t n_t - N\sum_t (n_t/N)^2\right) = (k-1)\sigma^2 + \sigma_u^2 \left(N - \frac{1}{N}\sum_t n_t^2\right) \tag{26}$$

DIVIDIENDO ENTRE LOS GRADOS DE LIBERTAD RESPECTIVOS SE OBTIENEN

$$E(MSW) = \sigma^2 \qquad (27)$$

$$E(MSB) = \sigma^2 + \sigma_u^2 \{N - \frac{1}{N} \sum_t n_t^2\}/(k-1) \qquad (28)$$

PUESTO QUE EL COEFICIENTE DE σ_u^2 ES POSITIVO, UNA DIFERENCIA EXCESIVA DE MSB SOBRE MSW PUEDE DEBERSE A QUE σ_u^2 NO ES CERO, ESTO ES, A UNA VARIACION REAL ENTRE LOS GRUPOS O TRATAMIENTOS.

BAJO LA HIPOTESIS NULA DE QUE $\sigma_u^2=0$,TANTO MSB COMO MSW SON ESTIMADORES INSESGADOS DE σ^2, POR LO QUE LA PRUEBA DE HIPOTESIS SE REALIZA CON LA ESTADISTICA F

$$F = MSB/MSW \qquad (29)$$

CON DISTRIBUCION F CON k-1 Y N-k GRADOS DE LIBERTAD.

EN EL CASO PARTICULAR DE QUE TODAS LAS MUESTRAS DE LOS TRATAMIENTOS SEAN DE IGUAL TAMAÑO, ES DECIR, SI $n_t = n$, ENTONCES LA EC (28) SE REDUCE A

$$MSB = \sigma^2 + n\sigma_u^2 \qquad (30)$$

UNA ESTIMACION PUNTUAL DE σ_u^2 SE PUEDE OBTENER SI A LA EC (28) SE LE RESTA, MIEMBRO A MIEMBRO, LA EC (27) Y DEL RESULTADO SE DESPEJA A σ_u^2; EN TAL CASO

$$\hat{\sigma}_u^2 = \frac{(MSB - MSW)(k - 1)}{N - \frac{1}{N} \sum_t n_t^2} \qquad (31)$$

EN EL CASO EN QUE TODAS LAS n_t SEAN IGUALES, LA ESTIMACION DE $\hat{\sigma}_u^2$, EMPLEANDO LAS ECS (30) y (28), SERA

$$\hat{\sigma}_u^2 = (MSB - MSW)/n \qquad (32)$$

EJEMPLO

SE TIENE UN PROBLEMA DE APLICACION DE UN TEST PSICOLOGICO EN EL QUE SE TRATA DE VERIFICAR SI SE OBTIENEN LOS MISMOS RESULTADOS AL SER APLICADO POR DIFERENTES PERSONAS. PARA ESTO SE DISEÑO UN EXPERIMENTO CONSISTENTE EN SELECCIONAR AL AZAR A 5 PERSONAS, QUIENES APLICARON EL TEST A 8 SUJETOS ASIGNADOS AL AZAR A CADA UNA. LAS CALIFICACIONES QUE OBTUVIERON SE PRESENTAN EN LA SIGUIENTE TABLA

	EXPERIMENTADOR				
	1	2	3	4	5
	5.8	6.0	6.3	6.4	5.7
	5.1	6.1	5.5	6.4	5.9
	5.7	6.6	5.7	6.5	6.5
	5.9	6.5	6.0	6.1	6.3
	5.6	5.9	6.1	6.6	6.2
	5.4	5.9	6.2	5.9	6.4
	5.3	6.4	5.8	6.7	6.0
	5.2	6.3	5.6	6.0	6.3
Total	44.0	49.7	47.2	50.6	49.3

$\sum_t \sum_i X_{ti} = 240.8$

EN LA SIGUIENTE TABLA SE RESUMEN LOS RESULTADOS DEL ANALISIS DE VARIANCIA REALIZADO CON ESTOS DATOS:

Fuente	SS	g. de l.	MS	E(MS)	F
Entre experimentadores	3.47	4	0.868	$8\hat{\sigma}_u^2+\hat{\sigma}^2$	10.72
Dentro de experimentadores	2.85	35	0.081	$\hat{\sigma}^2$	
Total	6.32	39			

$$F_{0.99,4,35} = 4.12 < 10.72$$

POR LO TANTO SE RECHAZA LA HIPOTESIS DE IGUALDAD DE RESULTADOS A UN 99% DE NIVEL DE CONFIANZA.

LA ESTIMACION DE LA VARIANCIA ENTRE EXPERIMENTADORES VALE, DE ACUERDO CON LA EC (32):

$$\hat{\sigma}_u^2 = \frac{0.868 - 0.081}{8} = 0.098$$

LA ESTIMACION DE LA VARIANCIA TOTAL ES

$$\hat{\sigma}_x^2 = \hat{\sigma}_u^2 + \hat{\sigma}^2 = 0.098 + 0.081 = 0.179$$

LA ESTIMACION DE LA PROPORCION DE LA VARIANCIA EXPLICADA POR LAS DIFERENCIAS ENTRE LOS EXPERIMENTADORES RESULTA SER

$$\hat{\sigma}_u^2 / \hat{\sigma}_x^2 = 0.098/0.179 = 0.55$$

5. COMPARACIONES MULTIPLES

COMPARACION DE DOS MEDIAS

CON LOS RESULTADOS DEL ANALISIS DE VARIANCIA SE PUEDEN DETERMINAR INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE DOS MEDIAS CUALESQUIERA DE LA SIGUIENTE MANERA.

SEAN LAS MEDIAS DE LOS GRUPOS i Y j, $\bar{X}_{i.}$ Y $\bar{X}_{j.}$; LA DIFERENCIA $\bar{X}_{i.} - \bar{X}_{j.}$ ES UNA ESTADISTICA CON VARIANCIA $\sigma^2 (1/n_i - 1/n_j)$, EN DONDE σ^2 SE ESTIMA CON MSW = S^2. EN TAL CASO, EL INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE LAS MEDIAS PRESELECCIONADAS ES

$$\bar{X}_{i.} - \bar{X}_{j.} \pm (t_{\nu,\ \alpha/2})\ S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \qquad (33)$$

DONDE $\nu = \nu_R$ ES EL NUMERO DE GRADOS DE LIBERTAD ASOCIADO CON S^2, Y α ES EL NIVEL DE SIGNIFICANCIA DEL INTERVALO (ν_R=N-k).

EJEMPLO

PARA EL PROBLEMA ANALIZADO ANTERIORMENTE DE LOS TIEMPOS DE COAGULACION DE LA SANGRE ASOCIADOS A DIFERENTES MEDICINAS, CALCULEMOS EL INTERVALO DE CONFIANZA DE LA DIFERENCIA DE LAS MEDIAS DE LOS TRATAMIENTOS A Y B.

PARA ESTE PROBLEMA SE OBTUVO:

$$\bar{X}_{A.} = 61,\ \bar{X}_{B.} = 66,\ n_A=4,\quad n_B=6.$$

$S^2 = 5.6,\ \nu = 20.$

POR LO TANTO $\bar{X}_{B.} - \bar{X}_{A.} = 66 - 61 = 5$ Y $(t_{20,0.025}) = 2.09$

EL INTERVALO DE CONFIANZA RESULTA SER

$$5 \pm 2.09\sqrt{5.6}\sqrt{\frac{1}{4}+\frac{1}{6}} = 5 \pm 3.2 = (1.8, 8.2)$$

COMPARACION DE PARES DE MEDIAS

SI SE DESEA COMPARAR LAS DIFERENCIAS DE LAS MEDIAS DE k TRATAMIENTOS, SE TENDRAN $k(k-1)/2$ PAREJAS DIFERENTES DE COMPARACIONES POR HACER. EN CASO DE QUE SE TENGAN MUESTRAS DE IGUAL TAMAÑO PARA CADA TRATAMIENTO, LA SIGUIENTE FORMULA DEBIDA A TUKEY PARA CALCULAR LOS INTERVALOS DE CONFIANZA ES EXACTA; EN CASO CONTRARIO SERA SOLO APROXIMADA:

$$\bar{X}_{i.} - \bar{X}_{j.} \pm \frac{q_{k,\nu,\alpha/2}S}{\sqrt{2}}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}} \qquad (34)$$

DONDE $q_{k,\ \nu,\ \alpha/2}$ ES EL RANGO STUDENTIZADO PARA k MEDIAS Y ν GRADOS DE LIBERTAD. LOS VALORES DEL RANGO STUDENTIZADO SE HAN TABULADO EN ALGUNAS PUBLICACIONES, TALES COMO: PEARSON, E.S. Y HARTLEY, H.O., "BIOMETRIKA TABLES FOR STATISTICIANS", TABLA 29, VOL. 1, 3a. ED., 1966, CAMBRIDGE UNIV. PRESS.

EJEMPLO

SUPONGAMOS QUE EN UN EXPERIMENTO CON 7 TRATAMIENTOS SE OB

Table 11 Percentage points of the studentized range

Error df	α	*r* = number of treatment means: 2	3	4	5	6	7	8	9	10	11
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
	.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
	.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.30	5.35	5.49	5.61
	.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13
12	.05	3.08	3.77	4.20	4.52	4.75	4.95	5.12	5.27	5.39	5.51
	.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31
	.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
	.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46
17	.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19
24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	3.92	5.01
	.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02
30	.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92
	.01	3.89	4.45	4.60	5.05	5.24	5.40	5.54	5.65	5.76	5.85
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82
	.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73
	.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53
120	.05	2.80	3.36	3.65	3.92	4.10	4.24	4.36	4.47	4.56	4.64
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37
∞	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23

This table is abridged from Table 29 in *Biometrika Tables for Statisticians*, Vol. I, 2d ed. New York: Cambridge, 1958. Edited by E.S. Pearson and H.O. Hartley. Reproduced with the kind permission of the editors and the trustees of *Biometrika*.

Table 11 *(continued)*

t = number of treatment means										Error
12	13	14	15	16	17	18	19	20	α	df
7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21	.05	5
10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93	.01	
6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59	.05	6
9.48	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54	.01	
6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17	.05	7
8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65	.01	
6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87	.05	8
8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03	.01	
5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64	.05	9
7.78	7.91	8.03	8.13	8.23	8.33	8.41	8.49	8.57	.01	
5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47	.05	10
7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23	.01	
5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33	.05	11
7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95	.01	
5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21	.05	12
7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73	.01	
5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11	.05	13
6.90	7.01	7.10	7.19	7.27	7.35	7.42	7.48	7.55	.01	
5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03	.05	14
6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.33	7.39	.01	
5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96	.05	15
6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26	.01	
5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90	.05	16
6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15	.01	
5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84	.05	17
6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7.05	.01	
5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79	.05	18
6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97	.01	
5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75	.05	19
6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89	.01	
5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71	.05	20
6.28	6.37	6.45	6.52	6.59	6.65	6.71	6.77	6.82	.01	
5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59	.05	24
6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61	.01	
5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47	.05	30
5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41	.01	
4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36	.05	40
5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21	.01	
4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24	.05	60
5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01	.01	
4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13	.05	120
5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83	.01	
4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01	.05	∞
5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65	.01	

TUVIERON LOS SIGUIENTES RESULTADOS DEL ANALISIS DE VARIANCIA CORRESPONDIENTE:

	TRATAMIENTOS						
	A	B	C	D	E	F	G
$\bar{X}_{i.}$	63	62	67	65	65	70	60

MSW = S^2 = 9.0, $k = 7$, $n_i = n = 4$, $\nu = 28-7 = 21$. PARA

$\alpha = 0.05$, SE OBTIENE DE LA TABLA DE LOS RANGOS STUDENTIZADOS $(q_{7,21,0.025})/\sqrt{2} = 3.26$.

CON ESTOS VALORES LOS MARGENES QUE DEFINEN LOS LIMITES DE CONFIANZA PARA LA DIFERENCIA DE LAS MEDIAS DE CUALQUIER PAR DE TRATAMIENTOS SON:

$$\pm \frac{q_{k,\nu,\alpha/2}}{\sqrt{2}} S \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = \pm\ 3.26 \times 3\sqrt{\left(\frac{1}{4} + \frac{1}{4}\right)} = \pm\ 6.91$$

DE ESTA MANERA CUALQUIER DIFERENCIA DE PROMEDIOS QUE EN VALOR ABSOLUTO EXCEDA DE 6.91 PUEDE CONSIDERARSE ESTADISTICAMENTE SIGNIFICATIVA AL 95% DE NIVEL DE CONFIANZA. TODAS LAS DIFERENCIAS POSIBLES (7 x 6/2 = 21) SE ENCUENTRAN EN LA SIGUIENTE TABLA, Y SE HAN ENMARCADO LAS QUE RESULTARON SIGNIFICATIVAS.

TRATAMIENTO	A	B	C	D	E	F	G
PROMEDIO	63	62	67	65	65	70	60
$\bar{X}_{i.} - \bar{X}_{j.}$	*	1	-4	-2	-2	$\boxed{-7}$	3
		*	-5	-3	-3	$\boxed{-8}$	2
			*	2	2	3	$\boxed{7}$
				*	0	-5	5
					*	-5	5
						*	$\boxed{10}$

EN ESTA TABLA SE OBSERVA QUE LAS PAREJAS CUYAS MEDIAS TUVIERON DIFERENCIAS SIGNIFICATIVAS SON: A Y F, B Y F, C Y G, Y F Y G.

METODO DE DUNNETT PARA COMPARACION DE VARIOS TRATAMIENTOS CON UNO ESTANDAR

SI SE DESEA COMPARAR LAS MEDIAS DE VARIOS TRATAMIENTOS CON LA DE UN TRATAMIENTO ESTANDAR, A, SE TIENEN QUE HACER k-1 COMPARACIONES POR PARES. LOS INTERVALOS DE CONFIANZA DE LAS DIFERENCIAS RESULTAN SER

$$\bar{X}_{A.} - \bar{X}_{i.} \pm (t_{k,\nu,\alpha/2})\; s\sqrt{\frac{1}{n_A} + \frac{1}{n_i}} \qquad (35)$$

EN DONDE $t_{k,\nu,\alpha/2}$ ES LA t DE DUNNETT*.

SI EN EL EJEMPLO ANTERIOR EL TRATAMIENTO A ES EL ESTANDAR, ENTONCES $t_{7,21,0.025} = 2.80$, Y EL MARGEN DE LOS INTERVALOS RESULTA SER

$$\pm\, 2.80 \times 3 \sqrt{\frac{1}{4} + \frac{1}{4}} = \pm\, 5.94$$

POR LO TANTO, CUALQUIER DIFERENCIA DE PROMEDIOS QUE SEA SUPERIOR A 5.94 RESULTA SIGNIFICATIVAMENTE DIFERENTE CON $\alpha = 0.05$.

---

* DUNNETT, C.W., "NEW TABLES FOR MULTIPLE COMPARISONS WITH A CONTROL", BIOMETRICS, VOL 20, P 482.

TRATAMIENTO	A (CONTROL)	B	C	D	E	F	G
PROMEDIO	63	62	67	65	65	70	60
$\bar{X}_A - \bar{X}_{i.}$	*	1	-4	-2	-2	$\boxed{-7}$	3

COMO RECOMENDACION, CUANDO SE USE UN TRATAMIENTO O GRUPO DE CONTROL, SE DEBE PROCURAR QUE EL TAMAÑO DE LA MUESTRA DE ESTE SEA $\sqrt{k}$ VECES MAYOR QUE EL DE LOS DEMAS.

...nd a control for a joint confidence coefficient of $P = .95$ and $P = .99$

Error df	P	p = number of treatment means, excluding control								
		1	2	3	4	5	6	7	8	9
5	.95	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
	.99	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	.95	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
	.99	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	.95	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
	.99	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	.95	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
	.99	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	.95	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
	.99	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10	.95	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
	.99	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	.95	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
	.99	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	.95	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
	.99	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	.95	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
	.99	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	.95	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
	.99	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	.95	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
	.99	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	.95	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
	.99	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	.95	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
	.99	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	.95	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
	.99	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	.95	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
	.99	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	.95	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
	.99	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	.95	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
	.99	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	.95	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
	.99	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	.95	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
	.99	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	.95	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
	.99	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	.95	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
	.99	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
∞	.95	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42
	.99	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

Source: This table is reproduced from "A multiple comparison procedure for comparing several treatments with a control," *J. Am. Stat. Assn.*, 50: 1096–1121 (1955), with permission of the author, C. W. Dunnett, and the editor.

Table A.9b Table of t for two-sided comparisons between p treatment means and a control for a joint confidence coefficient of $P = .95$ and $P = .99$

Error df	P	p = number of treatment means, excluding control								
		1	2	3	4	5	6	7	8	9
5	.95	2.57	3.03	3.39	3.66	3.88	4.06	4.22	4.36	4.49
	.99	4.03	4.63	5.09	5.44	5.73	5.97	6.18	6.36	6.53
6	.95	2.45	2.86	3.18	3.41	3.60	3.75	3.88	4.00	4.11
	.99	3.71	4.22	4.60	4.88	5.11	5.30	5.47	5.61	5.74
7	.95	2.36	2.75	3.04	3.24	3.41	3.54	3.66	3.76	3.86
	.99	3.50	3.95	4.28	4.52	4.71	4.87	5.01	5.13	5.24
8	.95	2.31	2.67	2.94	3.13	3.28	3.40	3.51	3.60	3.68
	.99	3.36	3.77	4.06	4.27	4.44	4.58	4.70	4.81	4.90
9	.95	2.26	2.61	2.86	3.04	3.18	3.29	3.39	3.48	3.55
	.99	3.25	3.63	3.90	4.09	4.24	4.37	4.48	4.57	4.65
10	.95	2.23	2.57	2.81	2.97	3.11	3.21	3.31	3.39	3.46
	.99	3.17	3.53	3.78	3.95	4.10	4.21	4.31	4.40	4.47
11	.95	2.20	2.53	2.76	2.92	3.05	3.15	3.24	3.31	3.38
	.99	3.11	3.45	3.68	3.85	3.98	4.09	4.18	4.26	4.33
12	.95	2.18	2.50	2.72	2.88	3.00	3.10	3.18	3.25	3.32
	.99	3.05	3.39	3.61	3.76	3.89	3.99	4.08	4.15	4.22
13	.95	2.16	2.48	2.69	2.84	2.96	3.06	3.14	3.21	3.27
	.99	3.01	3.33	3.54	3.69	3.81	3.91	3.99	4.06	4.13
14	.95	2.14	2.46	2.67	2.81	2.93	3.02	3.10	3.17	3.23
	.99	2.98	3.29	3.49	3.64	3.75	3.84	3.92	3.99	4.05
15	.95	2.13	2.44	2.64	2.79	2.90	2.99	3.07	3.13	3.19
	.99	2.95	3.25	3.45	3.59	3.70	3.79	3.86	3.93	3.99
16	.95	2.12	2.42	2.63	2.77	2.88	2.96	3.04	3.10	3.16
	.99	2.92	3.22	3.41	3.55	3.65	3.74	3.82	3.88	3.93
17	.95	2.11	2.41	2.61	2.75	2.85	2.94	3.01	3.08	3.13
	.99	2.90	3.19	3.38	3.51	3.62	3.70	3.77	3.83	3.89
18	.95	2.10	2.40	2.59	2.73	2.84	2.92	2.99	3.05	3.11
	.99	2.88	3.17	3.35	3.48	3.58	3.67	3.74	3.80	3.85
19	.95	2.09	2.39	2.58	2.72	2.82	2.90	2.97	3.04	3.09
	.99	2.86	3.15	3.33	3.46	3.55	3.64	3.70	3.76	3.81
20	.95	2.09	2.38	2.57	2.70	2.81	2.89	2.96	3.02	3.07
	.99	2.85	3.13	3.31	3.43	3.53	3.61	3.67	3.73	3.78
24	.95	2.06	2.35	2.53	2.66	2.76	2.84	2.91	2.96	3.01
	.99	2.80	3.07	3.24	3.36	3.45	3.52	3.58	3.64	3.69
30	.95	2.04	2.32	2.50	2.62	2.72	2.79	2.86	2.91	2.96
	.99	2.75	3.01	3.17	3.28	3.37	3.44	3.50	3.55	3.59
40	.95	2.02	2.29	2.47	2.58	2.67	2.75	2.81	2.86	2.90
	.99	2.70	2.95	3.10	3.21	3.29	3.36	3.41	3.46	3.50
60	.95	2.00	2.27	2.43	2.55	2.64	2.71	2.77	2.81	2.85
	.99	2.66	2.90	3.04	3.14	3.22	3.28	3.33	3.38	3.42
120	.95	1.98	2.24	2.40	2.51	2.59	2.66	2.71	2.76	2.80
	.99	2.62	2.84	2.98	3.08	3.15	3.21	3.25	3.30	3.33
∞	.95	1.96	2.21	2.37	2.47	2.55	2.62	2.67	2.71	2.75
	.99	2.58	2.79	2.92	3.01	3.08	3.14	3.18	3.22	3.25

Source: This table is reproduced from "A multiple comparison procedure for comparing several treatments with a control," *J. Am. Stat. Assn.*, 50: 1096–1121 (1955), with permission of the author, C. W. Dunnett, and the editor.

EJEMPLO

PARA EL PROBLEMA DE LOS TIEMPOS DE COAGULACION CON DIVERSAS MEDICINAS , TRATADO EN CLASE, HACER, CONSIDERANDO UNICAMENTE LOS DATOS DE LOS TRATAMIENTOS A Y B, LA SIGUIENTE PRUEBA DE HIPOTESIS PARA LAS MEDIAS: $\mu_A - \mu_B = 0$; $\mu_A \neq \mu_B$. ESTIMAR, EN LAS MISMAS CONDICIONES, EL INTERVALO DE CONFIANZA DE LA DIFERENCIA DE MEDIAS.

DATOS:

	TRATAMIENTOS			
	A	B	C	D
	62	63	68	56
	60	67	66	62
	63	71	71	60
	59	64	67	61
	244	65	68	63
		66	68	64
		396		63
				59
PROMEDIOS	61	66	68	61
VARIANZAS INSESGADAS	2.5	6.67	2.33	6

PROMEDIO GLOBAL = 64 seg.

a) INTERVALO DE CONFIANZA

SI SE CONSIDERAN SOLAMENTE LOS DATOS DE LOS TRATAMIENTOS A Y B, LA ECUACION PARA OBTENER EL INTERVALO DE CONFIANZA SERA:

$$\bar{X}_i - \bar{X}_j \pm t_{\nu,\alpha/2} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

DONDE s^2 ES LA VARIANCIA OBTENIDA DEL ANALISIS DE VARIANCIA DE LOS TRATAMIENTOS A Y B, EXCLUSIVAMENTE.

UTILIZANDO LAS FORMULAS SIMPLIFICADAS PARA EL ANALISIS DE VARIANCIA :

$$SST = \sum_{t}\sum_{i} X_{ti}^{2} - N\bar{X}_{..}$$

CON $N = 10$, $\bar{X}_{..} = (61)0.4 + 66(0.6) = 64$, Y $N\bar{X}_{..}^{2} = 10(64)^{2} = 40{,}960$

POR OTRA PARTE:

$$\sum_{t}\sum_{i} X_{ti}^{2} = 62^2+60^2+63^2+59^2+63^2+67^2+71^2+64^2+65^2+66^2 = 41{,}070$$

ENTONCES, SUSTITUYENDO:

$$SST = \sum_{t}\sum_{i} X_{ti}^{2} - N\bar{X}_{..} = 41{,}070 - 40{,}960 = 110$$

LA SUMA DE CUADRADOS ENTRE TRATAMIENTOS VALE:

$$SSB = \sum_{t} \frac{(\sum_{i} X_{ti})^{2}}{n_t} - N\bar{X}_{..}^{2}$$

HACIENDO OPERACIONES

$$\sum_{i} X_{Ai} = 62+60+63+59 = 244$$

$$\sum_{i} X_{Bi} = 63+67+71+64+65+66 = 396$$

POR TANTO

$$\sum_{t} \frac{(\sum_{i} X_{ti})^{2}}{n_t} = \frac{244^2}{4} + \frac{396^2}{6} = 41{,}020$$

SUSTITUYENDO EN LA ECUACION:

$$SSB = 41,020 - 40,960 = 60$$

LA SUMA DE CUADRADOS DENTRO DE GRUPOS VALE:

$$SSW = SST - SSB = 110 - 60 = 50$$

CON $N - k = 10 - 2 = 8$ GRADOS DE LIBERTAD

POR LO TANTO: $MSW = 50/8 = 6.25 = S^2$

PARA UN NIVEL DE CONFIANZA $\alpha = 0.05$ Y 8 GRADOS DE LIBERTAD, $t_{8,0.025} = 2.31$

SUSTITUYENDO LOS VALORES OBTENIDOS EN LA ECUACION PARA EL INTERVALO DE CONFIANZA SE OBTIENE:

$$66 - 61 \pm 2.31\sqrt{6.25}\sqrt{\frac{1}{4} + \frac{1}{6}} = 5 \pm 3.73$$

CONSIDERANDO TODOS LOS DATOS DEL ANALISIS DE VARIANCIA SE OBTUVO 5 ± 3.2

b) <u>PRUEBA DE HIPOTESIS (CON EL ANALISIS DE VARIANCIA DE A Y B)</u>

CONTINUANDO EL ANALISIS DE VARIANCIA PARA LOS TRATAMIENTOS A Y B:

$$MSB = \frac{SSB}{k-1} = \frac{60}{2-1} = 60$$

$$MSW = \frac{SSW}{N-k} = \frac{50}{8} = 6.25$$

ENTONCES LA ESTADISTICA F VALDRA:

$$F = \frac{MSB}{MSW} = \frac{60}{6.75} = 9.6$$

EN TABLAS, CON UN NIVEL DE CONFIANZA DE 95%, $F_{.05,1.8} = 5.32$

COMO 9.6 > 5.32 SE RECHAZA LA HIPOTESIS DE IGUALDAD DE MEDIAS.

c) PRUEBAS DE HIPOTESIS

SE TRATA DE PROBAR LA HIPOTESIS NULA H_0: $\mu_1 = \mu_2$ CONTRA LAS HIPOTESIS ALTERNATIVAS:

a) $\mu_1 \neq \mu_2$ b) $\mu_1 > \mu_2$ c) $\mu_1 < \mu_2$, PARA LOS TRATAMIENTOS A Y B.

c.1) SE PROBARA: HIPOTESIS NULA H_0: $\mu_1 = \mu_2$
HIPOTESIS ALTERNATIVA H_1: $\mu_1 \neq \mu_2$

EN PRIMER TERMINO DEBEMOS ASEGURARNOS QUE LAS VARIANCIAS CUMPLAN LA HIPOTESIS DE IGUALDAD DE VARIANCIAS:

H_0: $S_A^2 = S_B^2$; H_1: $S_A^2 \neq S_B^2$

DE LA TABLA DE DATOS OBTENEMOS:

$S_A^2 = 10/3 = 3.33$ CON 4-1 = 3 GRADOS DE LIBERTAD
$S_B^2 = 40/5 = 8$ CON 6-1 = 5 GRADOS DE LIBERTAD

ENTONCES LA ESTADISTICA F VALE:

$$F = \frac{8}{3.33} = 2.40$$

SE TRATA DE UNA PRUEBA DE DOS COLAS, POR LO QUE EL VALOR EMPIRICO CALCULADO F_0, DEBE CUMPLIR, RESPECTO AL TEORICO F_1:

$$F_0 < F_{\alpha/2,\nu_1,\nu_2}$$

O

$$F_0 > F_{1-\frac{\alpha}{2},\nu_1,\nu_2}$$

EN ESTE CASO:

$$F_{\frac{\alpha}{2},\nu_1,\nu_2} = F_{0.025,3,5} = 7.76$$

$$F_{1-\frac{\alpha}{2},\nu_1,\nu_2} = F_{1-0.025,3,5} = F_{0.975,3,5} = \frac{1}{F_{0.025,5,3}} = \frac{1}{14.88} = 0.0672$$

ENTONCES, COMPARANDO CON EL RESULTADO EMPIRICO: $0.0672 < 2.40 < 7.76$

POR LO TANTO ESTAMOS EN LA REGION DE ACEPTACION Y SE ADMITE QUE, CON UN NIVEL DE CONFIANZA DEL 95%, $S_A^2 = S_B^2$.

CON LO ANTERIOR, PODEMOS PROCEDER A EFECTUAR LA PRUEBA DE IGUALDAD DE MEDIAS:

$H_0: \mu_1 = \mu_2;\ H_1: \mu_1 \neq \mu_2$

PARA EFECTUARLA, SE CALCULARA LA ESTADISTICA T COMO:

$$T = \frac{\overline{Y}_B - \overline{Y}_A}{\varepsilon\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}}$$

DONDE

$$\varepsilon = \sqrt{\frac{\nu_A S_A^2 + \nu_B S_B^2}{\nu_A + \nu_B}}$$

DE LA TABLA DE DATOS:

$$S_A^2 = 3.33,\ \overline{Y}_A = 61,\ n_A = 4,\ \nu_A = 4-1 = 3$$

$$S_B^2 = 8,\ \overline{Y}_B = 66,\ n_B = 6,\ \nu_B = 6-1 = 5$$

SUSTITUYENDO

$$\varepsilon = \sqrt{\frac{3(3.33)+(5(8)}{5+3}} = \sqrt{\frac{10 + 40}{8}} = \sqrt{6.25} = 2.5$$

$$t = \frac{66-61}{2.5\sqrt{\frac{1}{4} + \frac{1}{6}}} = 3.1$$

COMO SE TRATA DE UNA PRUEBA DE DOS COLAS, LA ESTADISTICA TEORICA SERA:

$$t_{1-\frac{\alpha}{2},\nu_A+\nu_B} = t_{0.975,8} = 2.31$$

COMO 3.1 > 2.31 SE RECHAZA LA HIPOTESIS DE IGUALDAD DE MEDIAS, PARA UN NIVEL DE CONFIANZA DE 95%, EN CONTRA DE LA HIPOTESIS $\mu_1 \neq \mu_2$.

POR OTRA PARTE

$$T^2 = 3.1^2 = 9.6 = F$$

POR LO QUE SE VERIFICA QUE SI K=2, SE OBTIENE EL MISMO RESULTADO SI SE HACE LA PRUEBA CON LA DISTRIBUCION t O CON EL ANALISIS DE VARIANCIA.

c.2) PARA ESTE CASO SE PROBARA LA HIPOTESIS H_0: $\mu_1 = \mu_2$

CONTRA H_1: $\mu_1 > \mu_2$

SE HABIA CALCULADO EL VALOR EMPIRICO $T_0 = 3.1$

COMO SE TRATA DE UNA PRUEBA DE UNA COLA, ENTONCES:

$$t_{\alpha,\ \nu_1 + \nu_2} = t_{0.05,8} = 1.86 < 3.1$$

POR LO QUE <u>SE RECHAZA</u> LA HIPOTESIS DE IGUALDAD DE MEDIAS, CONTRA LA HIPOTESIS ALTERNATIVA $\mu_1 > \mu_2$

c.3) SE PROBARA FINALMENTE H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 < \mu_2$

PARA LA PRUEBA DE MEDIAS SE HABIA OBTENIDO $T_0 = 3.1$ Y $t_{0.05,8} = 1.86$; COMO $3.1 > 1.86$, <u>SE ACEPTA</u> LA HIPOTESIS DE IGUALDAD DE MEDIAS, CONTRA LA ALTERNATIVA $\mu_1 < \mu_2$

6. PRUEBA DE IGUALDAD DE VARIANCIAS

PARA APLICAR EL METODO DE ANALISIS DE VARIANCIA SE TIENE QUE CUMPLIR CON LA CONDICION DE QUE LAS VARIANCIAS DE LA PARTE ALEATORIA, Z_{ti}, DEL MODELO SEAN IGUALES, ES DECIR, QUE $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$.

PARA HACER ESTO, SI $k = 2$, SE PUEDE UTILIZAR LA PRUEBA USUAL DE IGUALDAD DE DOS VARIANCIAS. SI $k > 2$ SE PUEDE USAR CUALQUIERA DE LOS DOS METODOS SIGUIENTES, LOS CUALES SON APLICABLES SI SE TIENE QUE TODAS LAS MUESTRAS, n_i, SON DE IGUAL TAMAÑO:

a. PRUEBA DE COCHRAN, QUE USA COMO CRITERIO AL COCIENTE $SSW_{máx} / \sum_{t=1}^{k} SSW_t$

b. PRUEBA QUE USA COMO CRITERIO AL COCIENTE $SSW_{máx}/SSW_{mín}$

EN LA PUBLICACION BIOMETRIKA TABLES, REFERIDA ANTERIORMENTE, SE TIENEN TABLAS DE LOS VALORES CRITICOS DE LAS ESTADISTICAS ANTERIORES, SEMEJANTES A LA TABLA QUE SE PRESENTA A CONTINUACION PARA $\alpha = 0.01$.

Table 18 Percentage points of $F_{max} = s^2_{max}/s^2_{min}$

Upper 5% points

df_2 \ k	2	3	4	5	6	7	8	9	10	11	12
2	39.0	87.5	142	202	266	333	403	475	550	626	704
3	15.4	27.8	39.2	50.7	62.0	72.9	83.5	93.9	104	114	124
4	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.1	44.6	48.0	51.4
5	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
6	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
7	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
8	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
9	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
10	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
12	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
15	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
20	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
30	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
60	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Upper 1% points

df_2 \ k	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
3	47.5	85	120	151	184	21(6)	24(9)	28(1)	31(0)	33(7)	36(1)
4	23.2	37	49	59	69	79	89	97	106	113	120
5	14.9	22	28	33	38	42	46	50	54	57	60
6	11.1	15.5	19.1	22	25	27	30	32	34	36	37
7	8.89	12.1	14.5	16.5	18.4	20	22	23	24	26	27
8	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21
9	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
30	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
60	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
∞	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

s^2_{max} is the largest and s^2_{min} the smallest in a set of k independent mean squares, each based on $df_2 = n - 1$ degrees of freedom.

Values in the column $k = 2$ and in the rows $df_2 = 2$ and ∞ are exact. Elsewhere the third digit may be in error by a few units for the 5% points and several units for the 1% points. The third digit figures in brackets for $df_2 = 3$ are the most uncertain.

VALORES CRITICOS PARA LA PRUEBA DE
IGUALDAD DE VARIANCIAS (α = 0.01)

	ν	NUMERO DE VARIANCIAS 4	5	6	7	8	9	10
PRUEBA DE COCHRAN	2	0.864	0.788	0.722	0.664	0.615	0.573	0.536
	3	0.781	0.696	0.626	0.568	0.521	0.481	0.447
	4	0.721	0.633	0.564	0.508	0.463	0.425	0.393
	6	0.641	0.553	0.487	0.435	0.393	0.359	0.331
	8	0.590	0.504	0.440	0.391	0.352	0.321	0.294
	10	0.554	0.470	0.408	0.362	0.325	0.295	0.270
PRUEBA DE $\frac{SSW_{máx}}{SSW_{mín}}$	2	729	1036	1362	1705	2063	2432	2813
	3	120	151	184	216	249	281	310
	4	49	59	69	79	89	97	106
	6	19.1	22	25	27	30	32	34
	8	11.7	13.2	14.5	15.8	16.9	17.9	18.9
	10	8.6	9.6	10.4	11.1	11.8	12.4	12.9

EJEMPLO

SE DESEA PROBAR LA HIPOTESIS DE QUE LAS VARIANCIAS DE LOS DATOS DE SIETE DIFERENTES NIVELES DE UN FACTOR SON IGUALES; LAS MUESTRAS FUERON DE NUEVE ELEMENTOS CADA UNA. LOS VALORES DE SSW FUERON: 6.24, 5.16, 6.34, 8.26, 5.93, y 5.74 y 5.86. SE TOMARA α = 0.01.

PRUEBA DE COCHRAN: $SSW_{máx}/\sum_{t=1}^{9} SSW_t = 8.26/45.53 = 0.190$

VALOR CRITICO = 0.391 > 0.190

POR LO QUE SE ACEPTA LA HIPOTESIS DE

IGUALDAD DE LAS VARIANCIAS

PRUEBA b) $SSW_{máx}/SSW_{mín} = 8.26/5.16 = 1.60$

VALOR CRITICO = 15.8 > 1.60

POR LO QUE SE ACEPTA LA HIPOTESIS BAJO

PRUEBA.

7. DISEÑO DE EXPERIMENTOS CON BLOQUES ALEATORIZADOS

CONSIDEREMOS, COMO EJEMPLO, QUE INTERESA EL PROCESO DE MANUFACTURA DE PENICILINA, PARA LO CUAL HAY CUATRO METODOS O "TRATAMIENTOS". ADEMAS, SUPONGAMOS QUE UNA DE LAS MATERIAS PRIMAS PROVIENE DE CINCO FUENTES DIFERENTES, A LAS QUE LLAMAREMOS BLOQUES. EL PRINCIPAL INTERES ESTA EN VERIFICAR SI LOS CUATRO TRATAMIENTOS DAN RESULTADOS ESTADISTICAMENTE DIFERENTES; EL INTERES SECUNDARIO ES VERIFICAR SI LAS FUENTES DE MATERIAS PRIMAS INFLUYEN EN LOS RESULTADOS.

EN ESTE CASO SE ALEATORIZA UNA MUESTRA ASIGNANDOLE A CADA TRATAMIENTO UNA DE LAS MATERIAS PRIMAS, QUEDANDO UNA TABLA DE RESULTADOS COMO LA SIGUIENTE:

BLOQUE (MATERIA PRIMA)	TRATAMIENTO				PROMEDIO DE LOS BLOQUES
	A	B	C	D	
1	89	88	97	94	92
2	84	77	92	79	83
3	81	87	87	85	85
4	87	92	89	84	88
5	79	81	80	88	82
PROMEDIO DE LOS TRATAMIENTOS	84	85	89	86	

PROMEDIO GLOBAL = $\bar{X}_{..}$ = 86

EL DISEÑO DE UN EXPERIMENTO MEDIANTE BLOQUES ALEATORIZADOS TIENE LAS SIGUIENTES VENTAJAS:

1. SE PUEDEN ELIMINAR LAS VARIACIONES DE LOS BLOQUES AL HACER LA COMPARACION DE LOS TRATAMIENTOS
2. SE PUEDE ESTUDIAR EL EFECTO DE LOS BLOQUES EN LOS RESULTADOS EXPERIMENTALES, CUANDO ESTOS SON PREVISIBLES

EL MODELO MATEMATICO QUE EMPLEAREMOS PARA ESTUDIAR ESTE EXPERIMENTO ES EL DE ADITIVIDAD DE EFECTOS

$$X_{ti} = \eta + \beta_i + \tau_t + \epsilon_{ti} \qquad (1)$$

DONDE X_{ti} ES LA OBSERVACION CORRESPONDIENTE AL TRATAMIENTO t Y AL BLOQUE i, η ES LA MEDIA GLOBAL, β_i ES EL EFECTO DEL BLOQUE i, τ_t ES EL EFECTO DEL TRATAMIENTO t, Y ϵ_{ti} ES EL ERROR.

DE ACUERDO CON ESTE MODELO LAS OBSERVACIONES SE PUEDEN DESCOMPONER EN LA SIGUIENTE FORMA:

$$X_{ti} = \bar{X}_{..} + (\bar{X}_{.i} - \bar{X}_{..}) + (\bar{X}_{t.} - \bar{X}_{..}) + (X_{ti} - \bar{X}_{.i} - \bar{X}_{t.} + \bar{X}_{..}) \qquad (2)$$

AL ULTIMO DE LOS TERMINOS DE ESTA ECUACION SE LE LLAMA EL RESIDUO, POR SER LO QUE RESULTA AL QUITARLE A LA MEDIA GLOBAL LOS EFECTOS DE LOS BLOQUES $(\bar{X}_{.i} - \bar{X}_{..})$ Y EL DE LOS TRATAMIENTOS $(\bar{X}_{t.} - \bar{X}_{..})$.

EL CASO GENERAL DE UN DISEÑO CON BLOQUES ALEATORIZADOS QUEDA EN UNA TABLA COMO LA SIGUIENTE:

	TRATAMIENTOS					PROMEDIOS
BLOQUES	1	2	3	...	k	$\bar{X}_{.i}$
1	X_{11}	X_{21}	X_{31}	...	X_{k1}	$\bar{X}_{.1}$
2	X_{12}	X_{22}	X_{32}	...	X_{k2}	$\bar{X}_{.2}$
3	.	.	.		.	.
.	.					.
.	.					.
.	.					
n	X_{1n}	X_{2n}	X_{3n}	...	X_{kn}	$\bar{X}_{.n}$
PROMEDIOS						
$\bar{X}_{t.}$	$\bar{X}_{1.}$	$\bar{X}_{2.}$	$\bar{X}_{3.}$	...	$\bar{X}_{k.}$	$\bar{X}_{..}$ = PROMEDIO GLOBAL

SE PUEDE DEMOSTRAR QUE LA SUMA DE CUADRADOS DE LAS OBSERVACIONES SE PUEDE DESCOMPONER EN LA FORMA:

$$SS = SS\bar{X}_{..} + SSb + SSt + SSr \qquad (3)$$

EN DONDE $SS\bar{X}_{..}$ = SUMA DE CUADRADOS DE LA MEDIA GLOBAL = $nk\bar{X}^2_{..}$,

Y TIENE 1 GRADO DE LIBERTAD

SSb = SUMA DE CUADRADOS ENTRE BLOQUES =

$= k \sum_{i=1}^{n} (\bar{X}_{.i} - \bar{X}_{..})^2$, Y TIENE $n-1$ GRADOS DE LIBERTAD

SSt = SUMA DE CUADRADOS ENTRE TRATAMIENTOS =

$$= n \sum_{t=1}^{k} (\bar{X}_{t.} - \bar{X}_{..})^2,$$ Y TIENE k-1 GRADOS DE LIBERTAD

SSr = SUMA DE CUADRADOS RESIDUAL

$$= \sum_{i=1}^{n} \sum_{t=1}^{k} (X_{ti} - \bar{X}_{.i} - \bar{X}_{t.} - \bar{X}_{..})^2 = SS - SSb - SSt$$

Y TIENE (n-1)(k-1) GRADOS DE LIBERTAD

LA MEJOR ESTIMACION DEL RESULTADO X_{ti} ES

$$\hat{X}_{ti} = \bar{X}_{..} + (\bar{X}_{.i} - \bar{X}_{..}) + (\bar{X}_{t.} - \bar{X}_{..}) \qquad (4)$$

LOS RESIDUOS SERAN $\epsilon_{ti} = X_{ti} - \hat{X}_{ti}$

EJEMPLO

EN EL EJEMPLO DE LA PENICILINA TRATADO ANTERIORMENTE SE TENDRA:

$SS = 89^2 + 84^2 + 81^2 + 87^2 + 79^2 + 88^2 + \ldots + 84^2 + 88^2 = 148,480$

$SS\bar{X}_{..} = 5 \times 4 \times 86^2 = 147,920$

$SSb = 4[(92-86)^2 + (83-86)^2 + (85-86)^2 + (88-86)^2 + (82-86)^2] =$

$= 4 \times 66 = 264$

$SSt = 5[(84-86)^2 + (85-86)^2 + (89-86)^2 + (86-86)^2] = 5 \times 14 = 70$

$SSr = 148,480 - 147,920 - 264 - 70 = 226$

ESTOS CALCULOS Y LAS ESTIMACIONES $\hat{X}_{ti}$ SE PUEDEN FACILITAR MEDIANTE LA SIGUIENTE TABULACION, EN LA CUAL SE ENCUENTRAN ANOTADAS TAMBIEN LAS ESTIMACIONES $\hat{X}_{ti}$.

RESIDUOS Y ESTIMACIONES							
BLOQUES	A	B	C	D	$\bar{X}_{.i}$	$\bar{X}_{.i}-\bar{X}_{..}$	$(\bar{X}_{.i}-\bar{X}_{..})^2$
1	-1	-3	2	2	92	92-86=6	36
	(90)	(91)	(95)	(92)			
2	3	-5	6	-4	83	83-86=-3	9
	(81)	(82)	(86)	(83)			
3	-2	3	-1	0	85	85-86=-1	1
	(83)	(84)	(88)	(85)			
4	1	5	-2	-4	88	88-86=2	4
	(86)	(87)	(91)	(88)			
5	-1	0	-5	6	82	82-86=-4	16
	(80)	(81)	(85)	(82)			66
	84	85	89	86			
$\bar{X}_{t.}-\bar{X}_{..}$	84-86=-2	85-86=-1	89-86=3	86-86=0			
$\Sigma(\bar{X}_{t.}-\bar{X}_{..})^2$	4 +	1 +	9 +	0	= 14		

RESIDUOS: $\epsilon_{ti} = X_{ti} - \hat{X}_{ti} = X_{ti} - \bar{X}_{.i} - \bar{X}_{t.} + \bar{X}_{..}$

$\epsilon_{11} = 89 - 92 - 84 + 86 = -1$

$\epsilon_{12} = 84 - 83 - 84 + 86 = 3$

$\epsilon_{13} = 81 - 85 - 84 + 86 = -2$

ETC.

ESTIMACIONES: $\hat{X}_{ti} = \bar{X}_{..} + (\bar{X}_{.i} - \bar{X}_{..}) + (\bar{X}_{t.} - \bar{X}_{..})$

$\hat{X}_{11} = 86 + 6 + (-2) = 90$

$\hat{X}_{21} = 86 + 6 + (-1) = 91$

$\hat{X}_{31} = 86 + 6 + 3 = 95$

$\hat{X}_{41} = 86 + 6 + 0 = 92$

$\hat{X}_{12} = 86 + (-3) + (-2) = 81$, ETC.

ESTOS VALORES ESTIMADOS ESTAN ANOTADOS EN LA TABLA ANTERIOR ENTRE LOS PARENTESIS.

BAJO LA HIPOTESIS DE QUE LOS RESIDUOS O ERRORES ε_{ti} SON VARIABLES ALEATORIAS CON DISTRIBUCION NORMAL DE MEDIA CERO Y VARIANCIA σ^2, LAS ESPERANZAS DE LOS VALORES MEDIOS CUADRATICOS

$$
\begin{aligned}
MSb &= SSb/(n-1) \\
MSt &= SSt/(k-1) \qquad (5) \\
MSr &= SSr/(n-1)(k-1)
\end{aligned}
$$

SON

$$
\begin{aligned}
E\{MSb\} &= \sigma^2 + k \sum_i \beta_i^2/(n-1) \\
E\{MSt\} &= \sigma^2 + n \sum_t \tau_t^2/(k-1) \qquad (6) \\
E\{MSr\} &= \sigma^2
\end{aligned}
$$

BAJO LA HIPOTESIS NULA DE QUE TODAS LAS τ_t SON NULAS, ES DECIR, QUE NO HAY EFECTOS DEBIDOS A LOS TRATAMIENTOS, LA ESTADISTICA

$$F_t = MSt/MSr \qquad (7)$$

TIENE LA DISTRIBUCION F CON (k-1) Y (n-1)(k-1) GRADOS DE LIBERTAD.

DE IGUAL MANERA, BAJO LA HIPOTESIS NULA DE QUE TODAS LAS β_i SON NULAS, ES DECIR, QUE NO HAY EFECTOS DEBIDOS A LOS

BLOQUES, LA ESTADISTICA

$$F_b = MSb/MSr \tag{8}$$

TIENE DISTRIBUCION F CON (n-1) Y (n-1)(k-1) GRADOS DE LIBERTAD.

PARA EL EJEMPLO DE LA PENICILINA SI $\alpha = 0.05$, SE TIENEN

$MSb = 264/4 = 66$

$MSt = 70/3 = 23.3$

$MSr = 226/(4x3) = 18.8$

$F_{4,12,0.05} = 3.26$

$F_{3,12,0.05} = 3.49$

$F_b = 66/18.8 = 3.51 > 3.26$: SE RECHAZA LA HIPOTESIS DE QUE NO HAY EFECTOS DE BLOQUES

$F_t = 23.3/18.8 = 1.24 < 3.49$: SE ACEPTA LA HIPOTESIS DE QUE NO HAY EFECTOS DE TRATAMIENTOS

EJEMPLO

SE HIZO UN EXPERIMENTO ALEATORIZADO Y SE ENCONTRARON LOS SIGUIENTES RESULTADOS:

BLOQUES	TRATAMIENTOS		
	A	B	C
1	6.5	7.4	7.4
2	6.8	7.3	6.9
3	6.4	7.2	8.0
4	6.7	6.9	6.5

PROBAR SI LOS EFECTOS DE LOS BLOQUES Y TRATAMIENTOS SON NULOS

SOLUCION

LOS CALCULOS SE RESUMEN EN LA SIGUIENTE TABLA:

BLOQUES	TRATAMIENTOS A	B	C	$\bar{X}_{.i}$	$\bar{X}_{.i} - \bar{X}_{..}$	$(\bar{X}_{.i} - \bar{X}_{..})^2$
1	6.5 42.25 6.56 -0.06	7.4 54.76 7.7 -0.3	7.4 54.76 7.7 -0.3	7.1	0.1	0.01
2	6.8 46.24 6.76 -0.04	7.3 53.29 7.5 -0.2	6.9 47.61 7.1 -0.2	7	0	0
3	6.4 40.96 6.56 -0.16	7.2 51.84 7.6 -0.4	8.0 6 4 8.4 -0.4	7.2	0.2	0.04
4	6.7 44.89 7.04 -0.34	6.9 47.61 6.8 0.1	6.5 42.25 6.4 0.1	6.7	-0.3	0.09
						$\Sigma = 0.14$
$\bar{X}_{t.}$	6.6	7.2	7.2			
$\bar{X}_{t.} - \bar{X}_{..}$	-0.04	0.2	0.2			
$(\bar{X}_{t.} - \bar{X}_{..})^2$	0.16	0.04	0.04	$\Sigma = 0.24$		

DONDE EN CADA CELDA SE INDICA:

$$\boxed{\begin{array}{c} X_{ti} \\ (X_{ti})^2 \\ \hat{X}_{ti} \\ \hat{\varepsilon}_{ti} \end{array}}$$

SUMANDO TODOS LOS VALORES DE $(X_{ti})^2$ INDICADOS EN LAS CELDAS OBTENEMOS:

$$SS = \sum_{i}\sum_{t} (X_{ti})^2 = 6.5^2 + 7.4^2 + \dots + 6.9^2 + 6.5^2 = 590.46 \qquad \underline{SS = 590.46}$$

LAS SUMAS DE CUADRADOS DE LAS DESVIACIONES SON:

$$SS\bar{X}_{..} = nk(\bar{X}_{..})^2 = 4(3)(7)^2 = 588 \qquad \underline{SS\bar{X}_{..} = 588}$$

$$SSb = K \sum_{i=1}^{n} (\bar{X}_{.i} - \bar{X}_{..})^2 = 3(0.14) = 0.42 \qquad \underline{SSb = 0.42}$$

$$SSt = n \sum_{t=1}^{k} (\bar{X}_{t.} - \bar{X}_{..})^2 = 4(0.24) = 0.96 \qquad \underline{SSt = 0.96}$$

$$SSr = SS - SSb - SSt - SS\bar{X}_{..} = 590.46 - 588 - 0.42 - 0.96 = 1.08$$

$$\underline{SSr = 1.08}$$

SSb TIENE n - 1 = 4 - 1 = 3 GRADOS DE LIBERTAD

SSt TIENE k - 1 = 3 - 1 = 2 GRADOS DE LIBERTAD

SSr TIENE (n-1)(k-1) = 2 x 3 = 6 GRADOS DE LIBERTAD

LOS VALORES MEDIOS CUADRATICOS ENTRE BLOQUES, TRATAMIENTOS Y ERRORES, SON:

$$MSb = \frac{SSb}{n-1} = \frac{0.42}{3} = 0.14 \qquad \underline{MSb = 0.14}$$

$$MSt = \frac{SSt}{k-1} = \frac{0.96}{2} = 0.48 \qquad \underline{MSt = 0.48}$$

$$MSr = \frac{SSr}{(k-1)(n-1)} = \frac{1.08}{6} = 0.18 \qquad \underline{MSr = 0.18}$$

ENTONCES SE PUEDEN HACER LAS PRUEBAS DE HIPOTESIS PEDIDAS UTILIZANDO LA ESTADISTICA F.

- TRATAMIENTOS

$$F_t = \frac{MSt}{MSr} = \frac{0.48}{0.18} = \underline{2.667}$$

CON UN NIVEL DE CONFIANZA DEL 95%; EN TABLAS:

$$F_{0.05,2,6} = \underline{5.14}$$

COMO $2.667 < 5.14$

* SE ACEPTA LA HIPOTESIS DE QUE LOS EFECTOS DE LOS TRATAMIENTOS SON NULOS.

BLOQUES

$$F_b = \frac{MSb}{MSr} = \frac{0.14}{0.18} = \underline{0.778}$$

LA F, EN TABLAS:

$$F_{0.05,3,6} = \underline{4.76}$$

COMO $0.778 < 4.76$ SE ACEPTA LA HIPOTESIS DE QUE LOS EFECTOS DE LOS BLOQUES SON NULOS

EN LA TABLA SE INCLUYEN TAMBIEN, EN CADA CELDA, LOS VALORES DE LA MEJOR ESTIMACION Y DEL ERROR; QUE FUERON CALCULADOS CON LAS ECUACIONES:

MEJOR ESTIMACION: $\hat{X}_{ti} = \bar{X}_{..} + (\bar{X}_{.i} - \bar{X}_{..}) + (\bar{X}_{t.} - \bar{X}_{..})$

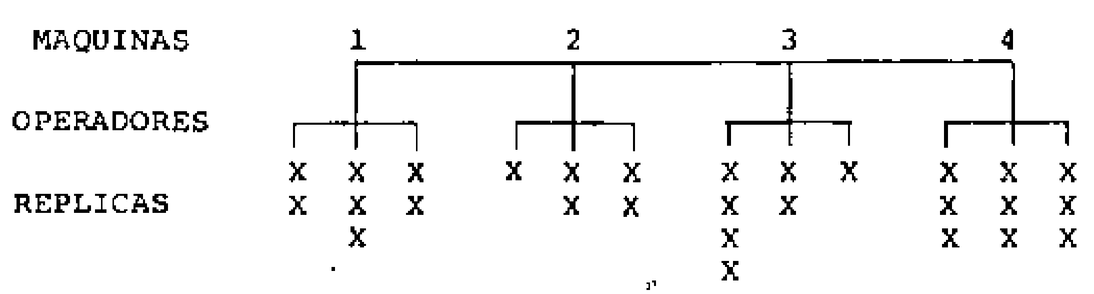
ERROR $\hat{\varepsilon}_{ti} = X_{ti} - \hat{X}_{ti}$

8. EXPERIMENTOS CON CLASIFICACION DE DOS FACTORES

EN OCASIONES NO INTERESA RELACIONAR AL FACTOR PRINCIPAL CON TODOS LOS NIVELES DEL FACTOR SECUNDARIO, POR LO CUAL A CADA NIVEL DEL FACTOR PRINCIPAL SE LE ASOCIA UN DIFERENTE CONJUNTO DE NIVELES DEL SECUNDARIO. EN TAL CASO SE TIENE UN EXPERIMENTO DE DOS FACTORES NO CRUZADO.

EN CAMBIO, CUANDO CADA NIVEL DEL FACTOR PRINCIPAL SE COMBINA CON TODOS LOS NIVELES DEL SECUNDARIO SE DICE QUE EL EXPERIMENTO ES DE DOS FACTORES CRUZADOS.

POR EJEMPLO, SI SE TIENE QUE EN UNA FABRICA SE DISPONE DE CUATRO MAQUINAS Y SE QUIERE ESTIMAR SU RENDIMIENTO, SE PUEDE DISEÑAR UN EXPERIMENTO EN EL QUE A CADA UNA SE LE ASIGNEN AL AZAR TRES OPERADORES. SI NO SE IDENTIFICA ALGUNA CARACTERISTICA DE LOS OPERADORES QUE SEÑALE LA CONVENIENCIA DE DISTINGUIRLOS EN TERMINOS DE ELLA, EL EXPERIMENTO CONSISTIRA EN REGISTRAR LOS RENDIMIENTOS INDIVIDUALES DE CADA PERSONA EN CADA VEZ QUE LA OPERE; ESTE EXPERIMENTO DE FACTORES NO CRUZADOS SE ILUSTRA EN LA SIGUIENTE FIGURA:



SI POR EL CONTRARIO, SE SABE QUE LOS OPERADORES TIENEN DI

FERENTE EXPERIENCIA EN EL USO DE MAQUINAS IGUALES A LAS DEL ESTUDIO, SERA NECESARIO DISEÑAR UN EXPERIMENTO CLASI FICANDOLOS EN TERMINOS DEL NIVEL DE EXPERIENCIA. SUPONGAMOS QUE ESTOS NIVELES SON 2, 4 Y 6 AÑOS DE EXPERIENCIA Y QUE A CADA MAQUINA SE LE ASIGNEN AL AZAR. EL EXPERIMENTO RESULTANTE SERA DE DOS FACTORES CRUZADOS, EL CUAL SE PUEDE REPRESENTAR EN UNA TABLA COMO LA SIGUIENTE:

EXPERIENCIA	MAQUINAS			
	I	II	III	IV
2 AÑOS	X X	X	X X X X	X X X
4 AÑOS	X X X	X X	X X	X X X
6 AÑOS	X X	X X	X	X X X

EN ESTE EJEMPLO EL NUMERO DE REPLICAS ES DIFERENTE PARA CADA NIVEL DE COMBINACION MAQUINA-EXPERIENCIA. EL ANALISIS DE ESTOS EXPERIMENTOS SE SIMPLIFICA GRANDEMENTE SI PARA CADA CELDA SE OBTIENE IGUAL NUMERO DE REPLICAS, n

EXPERIMENTO CON DOS FACTORES NO CRUZADOS O JERARQUIZADO

MODELO PARAMETRICO (I)

EL MODELO PARAMETRICO PARA ANALIZAR ESTE TIPO DE EXPERIMENTOS ES

$$X_{tij} = \xi + \gamma_t + \delta_{ti} + Z_{tij} \qquad (1)$$

DONDE $j = 1, 2, \ldots, n_{ti}$, ES EL NUMERO DE OBSERVACIONES (REPLICAS) DEL t-ESIMO GRUPO

PRINCIPAL Y DEL i-ESIMO GRUPO SECUNDARIO (SUBGRUPO)

$i = 1, 2, \ldots, m_t$, ES EL NUMERO DE SUBGRUPOS EN EL t-ESIMO NIVEL PRINCIPAL

$t = 1, 2, \ldots, k$, ES EL NUMERO DE GRUPOS EN EL FACTOR PRINCIPAL

ξ MEDIA GLOBAL

γ_t ES EL EFECTO MEDIO DEL TRATAMIENTO t

δ_{ti} ES EL EFECTO MEDIO DEL i-ESIMO SUBGRUPO EN EL t-ESIMO GRUPO PRINCIPAL

z_{tij} ES EL RESIDUO O ERROR ALEATORIO CON VARIANCIA σ^2 Y MEDIA CERO

AL IGUAL QUE EN EL MODELO DE CLASIFICACION EN UNA DIRECCION, A ESTOS EFECTOS SE LES IMPONEN LAS SIGUIENTES CONDICIONES:

$$\sum_{t=1}^{k} N_{t.}\gamma_t = 0, \quad \sum_{i=1}^{m_t} n_{ti}\delta_{ti} = 0, \text{ PARA TODA } t \quad (N_{t.} = \sum_{i=1}^{m_t} n_{ti})$$

LOS PROMEDIOS ARITMETICOS QUE RESULTAN DE ESTE MODELO SON:

PARA LOS SUBGRUPOS: $\bar{x}_{ti.} = \xi + \gamma_t + \delta_{ti} + \bar{z}_{ti.}$ (2)

PARA LOS GRUPOS PRINCIPALES: $\bar{x}_{t..} = \xi + \gamma_t + \bar{z}_{t..}$ (3)

PARA LA MEDIA GLOBAL: $\bar{x}_{...} = \xi + \bar{z}_{...}$ (4)

AL DEDUCIR ESTAS DOS ULTIMAS ECUACIONES SE HACE USO DE LAS DOS CONDICIONES ANTERIORES IMPUESTAS A γ_t Y δ_{ti}.

A PARTIR DE LAS ECS (2), (3) Y (4) SE OBTIENEN LAS SIGUIENTES DESVIACIONES:

$$\bar{X}_{t..} - \bar{X}_{...} = \gamma_t + \bar{Z}_{t..} - \bar{Z}_{...}; \quad E(\bar{X}_{t..} - \bar{X}_{...}) = \gamma_t \qquad (5)$$

$$\bar{X}_{ti.} - \bar{X}_{t..} = \delta_{ti} + \bar{Z}_{ti.} - \bar{Z}_{t..}; \quad E(\bar{X}_{ti.} - \bar{X}_{t..}) = \delta_{ti} \qquad (6)$$

$$X_{tij} - \bar{X}_{ti.} = Z_{tij} - \bar{Z}_{ti.}$$

POR LO ANTERIOR γ_t Y δ_{ti} SE PUEDEN ESTIMAR MEDIANTE LAS ESTADISTICAS:

$$\hat{\gamma}_t = \bar{X}_{t..} - \bar{X}_{...} \qquad \text{Y} \qquad (8)$$

$$\hat{\delta}_{ti} = \bar{X}_{ti.} - \bar{X}_{t..} \qquad \text{RESPECTIVAMENTE} \qquad (9)$$

LAS ESTADISTICAS PARA ANALIZAR LA INFORMACION DE UN EXPERIMENTO DE ESTE TIPO SE DEDUCEN DE LA SIGUIENTE PARTICION DE LA SUMA DE CUADRADOS:

$$\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{...})^2 = SSP + SSPW + SSR = \sum_t N_{t.} (\bar{X}_{t..} - \bar{X}_{...})^2 +$$

$$+ \sum_t \sum_i n_{ti} (\bar{X}_{ti.} - \bar{X}_{t..})^2 +$$

$$+ \sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2 \qquad (10)$$

LOS TERMINOS DEL MIEMBRO DERECHO SE DENOMINAN: SUMA DE CUADRADOS ENTRE GRUPOS PRINCIPALES, SUMA DE CUADRADOS ENTRE SUBGRUPOS DENTRO DE LOS GRUPOS PRINCIPALES, Y SUMA DE CUADRADOS RESIDUAL, RESPECTIVAMENTE.

LAS ESPERANZAS RESPECTIVAS SON:

ENTRE GRUPOS PRINCIPALES: $E(SSP) = (k-1)\sigma^2 + \sum_t N_{t.} \gamma_t^2$

ENTRE SUBGRUPOS DENTRO DE LOS GRUPOS PRINCIPALES:

$E(SSPW) = (\sum_t m_t - k)\sigma^2 + \sum_t \sum_i n_{ti} \delta_{ti}^2$

RESIDUAL: $E(SSR) = (N_{..} - \sum_t m_t)\sigma^2$; $(N_{..} = \sum_t \sum_i n_{ti})$

AL DIVIDIR ENTRE LOS NUMEROS CORRESPONDIENTES DE GRADOS DE LIBERTAD: $k-1$, $\sum_t m_t-k$ Y $N_{..} - \sum_t m_t$, SE OBTIENEN LOS RESPECTIVOS VALORES MEDIOS CUADRATICOS, A SABER

ENTRE GRUPOS PRINCIPALES: $E(MSP)=\sigma^2+\frac{1}{k-1} \sum_t N_{t.}\gamma_t^2$ (11)

ENTRE SUBGRUPOS DENTRO DE LOS GRUPOS PRINCIPALES:

$$E(MSPW) = \sigma^2 + \frac{1}{\sum_t m_t-k} \sum_t \sum_i n_{ti} \delta_{ti}^2 \quad (12)$$

RESIDUAL: $E(MSR) = \sigma^2$ (13)

COMPARANDO MSP CON MSR SE PUEDE PROBAR LA HIPOTESIS DE QUE $\gamma_1 = \gamma_2 = \ldots = \gamma_t = 0$ PARA TODA t. COMPARANDO MSPW CON MSR SE PUEDE PROBAR LA HIPOTESIS DE QUE $\delta_{ti} = 0$ PARA TODA t e i. AMBAS COMPARACIONES SE HACEN MEDIANTE LA ESTADISTICA F:

$$F_p = MSP/MSR \quad (14)$$

CON $k-1$ Y $N_{..}-\sum_t m_t$ GRADOS DE LIBERTAD.

$$F_{PW} = MSPW/MSR \tag{15}$$

CON $\sum_t m_t - k$ Y $N_{..} - \sum_t m_t$ GRADOS DE LIBERTAD.

SI SE DESEA PROBAR LA HIPOTESIS DE QUE $\delta_{ti} = 0$ PARA $i = 1, 2, \ldots, m_t$, Y CADA t POR SEPARADO, SE USA LA VARIANCIA

$$S_t^2 = \frac{1}{m_t - 1} \sum_i n_{ti} \; (\bar{X}_{ti.} - \bar{X}_{t..})^2 \tag{16}$$

QUE TIENE COMO ESPERANZA A

$$\sigma^2 + (m_t - 1)^{-1} \sum_i n_{ti} \; \delta_{ti}^2 \tag{17}$$

POR LO QUE SE PUEDE COMPARAR, PARA CADA t, CON MSR MEDIANTE LA ESTADISTICA

$$F_t = S_t^2/MSR \tag{18}$$

CON $m_t - 1$ y $N_{..} - \sum_t m_t$ GRADOS DE LIBERTAD

TODO ESTO SE PUEDE RESUMIR EN UNA TABLA DE ANALISIS DE VARIANCIA.

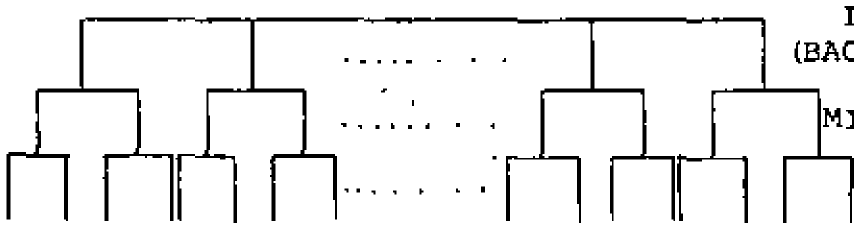
EN CASO DE QUE TODOS LOS SUBGRUPOS TENGAN IGUAL NUMERO DE OBSERVACIONES $n_{ti} = n$, Y DE QUE TODOS LOS GRUPOS PRINCIPALES TENGAN IGUAL NUMERO DE SUBGRUPOS $m_t = m$, DOS DE LOS GRADOS DE LIBERTAD SE PUEDEN ESCRIBIR DE LA SIGUIENTE MANERA

$$N_{..} - \sum_{t=1}^{k} m_t = kmn - km = km\ (n-1) \tag{19}$$

$$\sum_{t=1}^{k} m_t - k = km - k = k(m-1) \tag{20}$$

EJEMPLO

EN EL PROCESO DE FABRICACION DE UN COLORANTE INTERVIENE COMO VARIABLE IMPORTANTE EL CONTENIDO DE HUMEDAD DEL PRODUCTO. SE QUIERE VERIFICAR SI EL METODO DE PRUEBA PARA MEDIR LA HUMEDAD INTRODUCE UNA VARIACION APRECIABLE EN LOS RESULTADOS QUE SE REPORTAN. PARA ESTO SE DISEÑO UN EXPERIMENTO NO CURZADO EN QUE EL FACTOR PRINCIPAL ES EL LOTE Y EL SECUNDARIO ES PARTE DEL LOTE; SE DISPUSO DE k=15 LOTES, CON DOS MITADES CADA UNO (m_t=m=2) Y SE HICIERON n_{ti}=n=2 DETERMINACIONES DE HUMEDAD DE CADA MUESTRA. HACER EL ANALISIS DE VARIANCIA DE ESTE EXPERIMENTO.



BACHADA	MUESTRA	HUMEDAD
1	1	40, 39
	2	30, 30
2	3	26, 28
	4	25, 26
3	5	29, 28
	6	14, 15
4	7	30, 31

BACHADA	MUESTRA	HUMEDAD
	8	24, 24
5	9	19, 20
	10	17, 17
6	11	33, 32
	12	26, 24
7	13	23, 24
	14	32, 33
8	15	34, 34
	16	29, 29
9	17	27, 27
	18	31, 31
10	19	13, 16
	20	27, 24
11	21	25, 23
	22	25, 27
12	23	29, 29
	24	31, 32
13	25	19, 20
	26	29, 30
14	27	23, 24
	28	25, 25
15	29	39, 37
	30	26, 28

t	i	X_{ti1}	X_{ti2}	$\overline{X}_{ti.}$	$\overline{X}_{t..}$
1	1	40	39	39.5	34.75
	2	30	30	30	
2	1	26	28	27	26.25
	2	25	26	25.5	
3	1	28	29	28.5	21.5
	2	14	15	14.5	
4	1	30	31	30.5	27.25
	2	24	24	24	
5	1	19	20	19.5	18.25
	2	17	17	17.0	
6	1	33	32	32.5	28.75
	2	26	24	25.0	
7	1	23	24	23.5	28
	2	32	33	32.5	
8	1	34	34	34.0	31.5
	2	29	29	29.0	
9	1	27	27	27	29
	2	31	31	31	
10	1	13	16	14.5	20
	2	27	24	25.5	
11	1	25	23	24	25
	2	25	27	26	
12	1	29	29	29	30.25
	2	31	32	31.5	
13	1	19	20	19.5	24.5
	2	29	30	29.5	
14	1	23	24	23,5	24.25
	2	25	25	25	
15	1	39	37	38.0	32.5
	2	26	28	27	

Donde
$$\overline{X}_{ti.} = \frac{\sum_{j=1}^{2} X_{tij}}{n_{ti}} = \frac{\sum_{j=1}^{2} X_{tij}}{2}$$

$$\overline{X}_{t..} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} X_{tij}}{N_{t.}} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} X_{tij}}{4}$$

$$N_{..} = \sum_{t=1}^{15}\sum_{i=1}^{2} n_{ti} = \sum_{t=1}^{15}\sum_{i=1}^{2} 2 = 15x2x2=60$$

$$\sum_{t=1}^{15}\sum_{i=1}^{2}\sum_{j=1}^{2} X_{tij} = 40+39+30+30+26+28+\ldots+25+25+39+37+26+28 = 1607$$

$$\overline{X}_{...} = \frac{1607}{60} = 26.783$$

$$N_{t.} = \sum_{i=1}^{2} n_{ti} = 4$$

$$SSP = \sum_{t=1}^{15} N_{t.}(\bar{X}_{t..} - \bar{X}_{...})^2 = 4\sum_{t=1}^{15}(\bar{X}_{t..} - \bar{X}_{...})^2$$

$$= 4\left[(34.75-26.783)^2 + (26.25-26.783)^2 + (21.5-26.783)^2 + \ldots + (32.5-26.783)^2\right]$$

$$= 4(63.47 + 0.28 + 27.91 + \ldots + 32.684) = 1211.0$$

$$SSPW = \sum_t \sum_i n_{ti}(\bar{X}_{ti.} - \bar{X}_{t..})^2 = 2\sum_{t=1}^{15}\sum_{i=1}^{2}(\bar{X}_{ti.} - \bar{X}_{t..})^2$$

$$= 2\left[(39.5-34.75)^2 + (30-34.75)^2 + (27-26.25)^2 + (25.5-26.25)^2 + \ldots + (27-32.5)^2\right]$$

$$= 2\left[22.56 + 22.56 + \ldots\right]$$

$$= 2 \times 424.88 = 869.7$$

$$\sum_{t=1}^{15}\sum_{i=1}^{2}\sum_{j=1}^{2} X_{tij}^2 = 40^2 + 39^2 + 30^2 + 30^2 + \ldots + 39^2 + 37^2 + 26^2 + 28^2$$

$$= 45149$$

$$kmn\bar{X}_{...}^2 = 15 \times 2 \times 2 \times 26.783^2 = 43040.82$$

$$SST = \sum_{t=1}^{15}\sum_{i=1}^{2}\sum_{j=1}^{2} X_{tij}^2 - kmn\bar{X}_{...}^2 = 45149 - 43040.8 = 2108.2$$

$$SSR = SST - SSP - SSPW = 2108.2 - 1211.0 - 869.7 = 27.5$$

G. de L.: $k - 1 = 15 - 1 = 14$

$$\sum_{t=1}^{15} m_t - k = 15 \times 2 - 15 = 15$$

$$N_{..} - \sum_{t=1}^{15} m_t = 60 - 30 = 30$$

LA TABLA DEL ANALISIS DE VARIANCIA DE ESTE EXPERIMENTO ES:

FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MS	F
ENTRE BACHADAS	1211.0	14	86.6	92.2
ENTRE MITADES DE LOS GRUPOS PRINCIPALES	869.7	15	58.0	64.4
RESIDUO (ENTRE PRUEBAS)	27.5	30	0.9	
TOTAL	2108.2	59		

$$F_{0.01,14,30} = 2.75 < 92.2$$

$$F_{0.01,15,30} = 2.70 < 64.4$$

POR LO QUE SE RECHAZAN LAS HIPOTESIS DE QUE NO HAY EFECTOS DE BACHADAS Y DE MITADES A UN 99% DE NIVEL DE CONFIANZA. ADEMAS, AL COMPARAR LOS VALORES MEDIOS CUADRATICOS SE CONFIRMA QUE NO ES EL METODO DE PRUEBA, SINO LAS BACHADAS Y LAS MITADES LAS QUE INTRODUCEN LA MAYOR VARIABILIDAD DE LOS RESULTADOS, PUESTO QUE EL MS DEL RESIDUO ES MUY PEQUEÑO EN COMPARACION CON LOS OTROS DOS.

MODELO CON DOS FACTORES NO CRUZADOS. MODELO CON FACTORES ALEATORIOS (II) -

SI TANTO EL FACTOR PRINCIPAL COMO EL SECUNDARIO SON VARIABLES ALEATORIAS Y EN EL EXPERIMENTO SOLO SE INCLUYEN ALGUNOS NIVELES (O VALORES) DE LAS MISMAS, ENTONCES EL MODELO ES DE FACTORES ALEATORIOS O MODELO II. EN ESTE CASO EL MODELO MATEMATICO PARA REPRESENTAR A CADA OBSERVACION, X_{tij}, ES:

$$X_{tij} = \xi + U_t + V_{ti} + Z_{tij} \qquad (21)$$

DONDE U_t ES UNA VARIABLE ALEATORIA QUE REPRESENTA EL EFECTO MEDIO DEL FACTOR PRINCIPAL, V_{ti} ES OTRA VARIABLE ALEATORIA QUE REPRESENTA AL EFECTO MEDIO DEL I-ESIMO SUBGRUPO EN EL T-ESIMO GRUPO PRINCIPAL. ξ Y Z_{tij} TIENEN EL MISMO SIGNIFICADO QUE EN EL SUBCAPITULO ANTERIOR. SE SUPONE QUE U_t, V_{ti} Y Z_{tij} SON INDEPENDIENTES ENTRE SI, CON DISTRIBUCION NORMAL Y QUE $E(U_t) = 0$, $E(V_{ti}) = 0$ Y $E(Z_{tij} = 0)$; PARA LAS VARIANCIAS USAREMOS LOS SIGUIENTES SIMBOLOS:

$$\text{Var } (U_t) = \sigma_u^2; \quad \text{Var } (V_{ti}) = \sigma_v^2$$

CON ESTE MODELO SE TIENE QUE:

$$\bar{X}_{t..} - \bar{X}_{...} = U_t - \bar{U} + \bar{V}_{t.} - \bar{V}_{..} + \bar{Z}_{t..} - \bar{Z}_{...} \qquad (22)$$

$$\bar{X}_{ti.} - \bar{X}_{t..} = V_{ti} - \bar{V}_{t.} + \bar{Z}_{ti.} - \bar{Z}_{t..} \qquad (23)$$

$$X_{tij} - \bar{X}_{ti.} = Z_{tij} - \bar{Z}_{ti.} \qquad (24)$$

DONDE

$$\bar{U} = \sum_{t=1}^{k} N_{t.} U_{t}/N_{..}\,, \quad \bar{V}_{t.} = \sum_{i=1}^{m_t} n_{ti} V_{ti}/N_{t.}\,, \quad \bar{V}_{..} = \sum_{t=1}^{k} N_t \bar{V}_{t.} \qquad (25)$$

EN ESTE CASO LA DESCOMPOSICION DE CUADRADOS CONDUCE A LOS SI-GUIENTES VALORES MEDIOS CUADRATICOS:

$$E\left\{\frac{\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2}{N_{..} - \sum_t m_t}\right\} = E(MSR) = \sigma^2 \qquad (26)$$

$$E\left\{\frac{\sum_t N_{t.} (\bar{X}_{t..} - \bar{X}_{...})^2}{k-1}\right\} = E(MSP) = \sigma^2 + \frac{\sum_t (N_{t.}^{-1} - N_{..}^{-1}) \sum_i n_{ti}^2}{k-1} \sigma_v^2 + \frac{N_{..} - \sum_t N_{t.}^2/N_{..}}{k-1} \sigma_u^2 \qquad (27)$$

$$E\left\{\frac{\sum_t \sum_i n_{ti} (\bar{X}_{ti.} - \bar{X}_{t..})^2}{\sum_t m_t - k}\right\} = E(MSPW) = \sigma^2 + \frac{N_{..} - \sum_t N_{t.}^{-1} \sum_i n_{ti}^2}{\sum_t m_t - k} \sigma_v^2 \qquad (28)$$

EN ESTAS ECUACIONES SE OBSERVA QUE SI $\sigma_v^2 = 0$, ENTONCES $E(MSR) = E(MSPW)$, POR LO QUE PARA PROBAR LA HIPOTESIS DE QUE $\sigma_v^2 = 0$ BASTA FORMULAR LA ESTADISTICA

$$F_{PW} = MSPW/MSR \qquad (29)$$

QUE TIENE DISTRIBUCION F CON $(\Sigma\, m_t - k)$ Y $(N_{..} - \sum_t m_t)$ GRADOS DE LIBERTAD.

POR SU PARTE LA HIPOTESIS DE QUE $\sigma_u^2 = 0$ NO SE PUEDE PROBAR COM-

PARANDO MSP CON MSR, YA QUE EN E(MSP) INTERVIENEN TANTO σ_u^2 COMO σ_v^2. EN EL CASO PARTICULAR DE QUE $n_{ti}=n$ PARA TODO t E i, ENTONCES $N_{t.}=m_t n$ Y:

$$E(MSP) = \sigma^2 + n^2\sigma_v^2 + \frac{(\sum_t m_t)^2 - \sum_t m_t^2}{(k-1)\sum_t m_t} n\sigma_u^2 \qquad (30)$$

$$E(MSPW) = \sigma^2 + n\sigma_v^2 \qquad (31)$$

POR LO QUE LA HIPOTESIS DE QUE $\sigma_u^2=0$ SE PUEDE PROBAR COMPARANDO MSP CON MSPW MEDIANTE LA ESTADISTICA

$$F_p = MSP/MSPW \qquad (32)$$

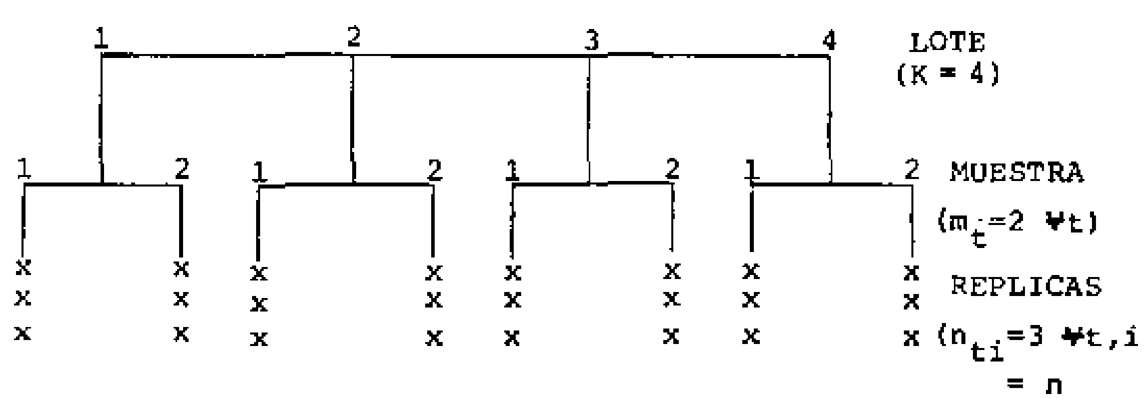
QUE TIENE DISTRIBUCION F CON (k-1) y $(\sum_t m_t - k)$ GRADOS DE LIBERTAD.

EJEMPLO

SE MUESTREARON CUATRO LOTES DE HULE CRUDO. DE CADA LOTE SE TOMARON DOS MUESTRAS. TRES PRUEBAS INDEPENDIENTES DE ESPECIMENES SE PREPARARON Y ANALIZARON PARA CADA UNO. ABAJO SE MUESTRAN LOS DATOS QUE DAN EL MODULO DE ELASTICIDAD OBTENIDO EN PORCENTAJE. CONSIDERE QUE SE APLICA EL MODELO DE VARIANCIA DE UNA COMPONENTE, CONSTRUYA LA TABLA ANOVA (ANALISIS DE VARIANCIA). USANDO LA TABLA OBTENGA ESTIMACIONES DE LA VARIANCIA DE CADA COMPONENTE.

LOTE O BACHADA	MODULO DE ELASTICIDAD (%)			
	1	2	3	4
MUESTRA 1	560	600	600	680
	580	640	610	700
	600	620	640	730
MUESTRA 2	660	580	580	720
	610	630	660	770
	600	670	620	740

SOLUCION



SE TRATA DE UN EXPERIMENTO CON DOS FACTORES NO CRUZADOS.

LAS ECUACIONES A EMPLEAR SON

$$SST = \sum_{tij}\sum\sum (X_{tij} - \overline{X}_{...})^2 = \sum_{tij}\sum\sum X_{tij}^2 - kmn\,\overline{X}_{...}^2$$

$$SSP = \sum_{t} N_{t.}(\overline{X}_{t..} - \overline{X}_{...})^2$$

$$SSPW = \sum_{ti}\sum n_{ti}(\overline{X}_{ti.} - \overline{X}_{t..})^2$$

$$SSR = \sum_{tij}\sum\sum (X_{tij} - \overline{X}_{ti})^2 = SST - SSP - SSPW$$

APLICANDO LAS ECUACIONES TENEMOS

$k = 4$

$m = 2$

$n = 3$

LOTE O BACHADA	MUESTRA	MODULO DE ELASTICIDAD x_{tij}	x^2_{tij}	$\bar{x}_{ti.}$	$\bar{x}_{t..}$	$(\bar{x}_{ti.}-\bar{x}_{t..})^2$	$(x_{t..}-\bar{x}_{...})^2$
1	1	560	313600	580.0	601.6667	469.444	1599.99
		580	336400				
		600	360000				
	2	660	435600	623.333		469.444	
		610	372100				
		600	360000				
2	1	600	360000	620.0	623.333	11.1111	336.11
		640	409600				
		620	384400				
	2	580	336400	626.667		11.1111	
		630	396900				
		670	448900				
3	1	600	360000	616.667	618.333	2.7778	560.11
		610	372100				
		640	409600				
	2	580	336400	620.0		2.7778	
		660	435600				
		620	384400				
4	1	680	462400	703.333	723.333	400.0001	6615.11
		700	490000				
		730	532900				
	2	720	518400	743.333		400.0001	
		770	592900				
		740	547600				
TOTALES		15,400	9956200			1766.667	9111.33

$\overline{X}_{...} = 15400/24 = 641.6667$

$kmn = 4 \times 2 \times 3 = 24$

$SST = \sum_{tij}\sum\sum X^2_{tij} - kmn\,\overline{X}^2_{...}$

$\sum_{tij}\sum\sum X^2_{tij} = 560^2+580^2+600^2+660^2+610^2+600^2+600^2+640^2+620^2+580^2+630^2+670^2+$

$600^2+610^2+640^2+580^2+660^2+620^2+680^2+700^2+730^2+720^2+770^2+740^2 =$

$2,177,700+2,336,200+2,298,100+3,144,200 = 9,956,200$

$SST = 9,956,200 - 24(641.667)^2 = 74,533.33$

$SSP = 6(9111.33) = 54,667.98$

$SSPW = 3(1766.667) = 5,300.00$

$SSR = SST - SSP - SSPW = 74,533.33 - 54,667.98 - 5,300.00 = 14,565.35$

$MSP = \frac{SSP}{k-1} = \frac{54,667.98}{4-1} = 18,222.66$

$MSPW = \frac{SSPW}{k(m-1)} = \frac{5300.00}{4(2-1)} = 1,325.00$

$MSR = \frac{SSR}{km(n-1)} = \frac{14,565.35}{4x2(3-1)} = 910.33$

DE TABLAS PARA UN 99% DE

NIVEL DE CONFIANZA

$F_{PW} = \frac{MSPW}{MSR} = \frac{1,325.00}{910.33} = 1.46 \quad < \quad F_{0.01,4,16} = 4.77$

$F_{P} = \frac{MSP}{MSPW}\ \frac{18,222.66}{1,325.00} = 13.75 \quad < \quad F_{0.01,3,4} = 16.69$

POR LO TANTO, PARA LOS SUBGRUPOS SE ACEPTA LA HIPOTESIS, O SEA NO HAY DIFERENCIA SIGNIFICATIVA ENTRE LAS MUESTRAS A UN NIVEL DE CONFIANZA DE 99%. PARA LOS LOTES SE ACEPTA LA HIPOTESIS, O SEA NO HAY DIFERENCIA SIGNIFICATIVA ENTRE LOTES A UN NIVEL DE CONFIANZA DE 99%.

ANOVA

FUENTE DE VARIACION	SS	GRADOS DE LIBERTAD	MS	F (CALC)	F (DE TABLAS) (α = 0.01)
ENTRE LOTES O BACHADAS	SSP=54,667.98	k - 1 3	MSP=18,222.66	F_P=13.75<	16.69
ENTRE PARTES DE LAS BACHADAS	SSPW=5,300.00	k(m-1) 4	MSPW=1,325.00	F_{PW}=1.46<	4.77
RESIDUAL (ENTRE PRUEBAS)	SSR=14,565.35	km(n-1) 16	MSR=910.33		
TOTAL	74,533.33	23			

$F_{3,4,0.99} = 16.69$, $F_{4,16,0.99} = 4.77$

ESTIMACIONES DE LAS VARIANCIAS DE CADA COMPONENTE

PUESTO QUE $E(MSR) = \sigma^2$, SE TIENE

$\hat{\sigma}^2 = MSR = 910.33$

DE LA EC 31, $\sigma_V^2 = (E(MSPW) - \sigma^2)/n$, POR LO QUE

$\hat{\sigma}_V^2 = (MSPW - MSR)/n = (1325.00 - 910.33)/3 = 138.22$

RESTANDO LA EC. 31 A LA EC. 30:

$$E(MSP) - E(MSPW) = \frac{\frac{(\Sigma m_t)^2 - \Sigma m_t^2}{t}}{(k-1)\Sigma m_t} n\sigma_u^2 = sn\sigma_u^2$$

POR LO QUE

$$\sigma_u^2 = \{E(MSP) - E(MSPW)\}/sn$$

Y

$$\hat{\sigma}_u^2 = (MSP - MSPW)/sn$$

EN NUESTRO PROBLEMA

$$S = \frac{8^2 - 4 \times 4}{3 \times 8} = 2$$

$$\hat{\sigma}_u^2 = (18,222.66 - 1325.00)/6 = 2816.28$$

9. EXPERIMENTO CON DOS FACTORES CRUZADOS. MODELO PARAMETRICO

EL MODELO PARA REPRESENTAR LA j-ESIMA OBSERVACION, X_{tij}, CORRESPONDIENTE AL NIVEL t DEL PRIMER FACTOR Y AL NIVEL i DEL SEGUNDO FACTOR ES

$$X_{tij} = \xi + \rho_t + \kappa_i + (\rho\kappa)_{ti} + Z_{tij} \qquad (1)$$

DONDE ρ_t Y κ_i SON EL EFECTO DEL t-ESIMO NIVEL (RENGLON) DEL PRIMER FACTOR Y DEL i-ESIMO NIVEL (COLUMNA) DEL SEGUNDO FACTOR, RESPECTIVAMENTE, $(\rho\kappa)_{ti}$ ES EL EFECTO DE INTERACCION DE LOS DOS FACTORES EN SUS NIVELES t E i, Y Z_{tij} ES EL RESIDUO, ERROR O EFECTO NO EXPLICABLE POR LOS FACTORES; LAS Z_{tij} SON VARIABLES ALEATORIAS INDEPENDIENTES CON DISTRIBUCION NORMAL DE MEDIA CERO E IDENTICA VARIANCIA, σ^2.

SI $t = 1, 2, \ldots, r$, E $i = 1, 2, \ldots, c$, SE DICE QUE SE TIENE UN EXPERIMENTO CRUZADO rxc; SE DICE QUE ESTE ES ORTOGONAL SI TIENE IGUAL NUMERO DE DATOS EN CADA CELDA (t, i), Y SI TODOS ESTOS SON RESULTADO DE OBSERVACIONES INDEPENDIENTES DE UNA POBLACION CON DISTRIBUCION NORMAL.

PUESTO QUE EL TOTAL DE PARAMETROS INVOLUCRADOS EN LA EC (1) PARA PRESENTAR A rc VALORES ESPERADOS ES $1 + r + c + rc$, ES NECESARIO IMPONER OTRAS $r + c + 1$ CONDICIONES; ELLAS SON:

$$\sum_{t=1}^{r} \rho_t = 0 \qquad (2)$$

$$\sum_{i=1}^{c} \kappa_i = 0 \qquad (3)$$

$$\sum_{t=1}^{r} (\rho\kappa)_{ti} = 0 \quad \text{PARA TODA } i \tag{4}$$

$$\sum_{i=1}^{c} (\rho\kappa)_{ti} = 0 \quad \text{PARA TODA } t \tag{5}$$

EN DONDE HAY $r + c + 2$ CONDICIONES, PERO UNA DE LAS DE LA EC (5) ES REDUNDANTE ($\sum_{i=1}^{c} (\rho\kappa)_{ri}$), YA QUE QUEDA OBLIGADA EN TERMINOS DE LAS $r + c - 1$ CONDICIONES RESTANTES IMPUESTAS POR LAS ECS (4) y (5).

DE ACUERDO CON ESTE MODELO SE OBTIENEN LOS SIGUIENTES PROMEDIOS:

PROMEDIO POR RENGLONES: $\bar{X}_{t..} = \xi + \rho_t + \bar{Z}_{t..}$ (6)

PROMEDIO POR COLUMNAS: $\bar{X}_{.i.} = \xi + \kappa_i + \bar{Z}_{.i.}$ (7)

PROMEDIO POR CELDAS: $\bar{X}_{ti.} + \xi + \rho_t + \kappa_i + (\rho\kappa)_{ti} + \bar{Z}_{ti}$ (8)

PROMEDIO GLOBAL: $\bar{X}_{...} = \xi + \bar{Z}_{...}$ (9)

LOS EFECTOS DE CADA PARAMETRO SE PUEDEN SEPARAR MEDIANTE LAS ESTADISTICAS, QUE SE OBTIENEN CON LAS ECUACIONES (6) A (9):

$$\bar{X}_{t..} - \bar{X}_{...} = \rho_t + \bar{Z}_{t..} - \bar{Z}_{...};\ E(\bar{X}_{t..} - \bar{X}_{...}) = \rho_t \tag{10}$$

$$\bar{X}_{.i.} - \bar{X}_{...} = \kappa_i + \bar{Z}_{.i.} - \bar{Z}_{...};\ E(\bar{X}_{.i.} - \bar{X}_{...}) = \kappa_i \tag{11}$$

$$\bar{X}_{ti.} - \bar{X}_{t..} - \bar{X}_{.i.} + \bar{X}_{...} = (\rho\kappa)_{ti} + \bar{Z}_{ti.} - \bar{Z}_{t..} - \bar{Z}_{.i.} + \bar{Z}_{...}$$

$$E(\bar{X}_{ti.} - \bar{X}_{t..} - \bar{X}_{.i.} + \bar{X}_{...}) = (\rho\kappa)_{ti} \tag{12}$$

$$X_{tij} - \bar{X}_{ti.} = Z_{tij} - \bar{Z}_{ti.};\ E(X_{tij} - \bar{X}_{ti.}) = 0 \tag{13}$$

PARA ANALIZAR LAS FUENTES DE VARIABILIDAD DE LOS DATOS, LA SUMA DE CUADRADOS SE PUEDE DIVIDIR, EN UNA PRIMERA ETAPA, EN LA SUMA DE CUADRADOS ENTRE CELDAS Y DENTRO DE LAS CELDAS:

$$\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{...})^2 = \underbrace{\sum_t \sum_i n_{ti} (\bar{X}_{ti.} - \bar{X}_{...})^2}_{\text{ENTRE CELDAS}} + \underbrace{\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2}_{\text{DENTRO DE LAS CELDAS}} \qquad (14)$$

UTILIZANDO LA EC (13) SE DEMUESTRA QUE

$$E\{\text{SUMA DE CUADRADOS DENTRO DE LAS CELDAS}\} = E\{\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2\} =$$

$$= (N_{..} - rc)\sigma^2 \qquad (15)$$

POR LO QUE EL NUMERO DE GRADOS DE LIBERTAD DE LA SUMA DE CUADRADOS DENTRO DE LAS CELDAS ES $(N_{..} - rc)$ Y, POR LO TANTO, LA ESTADISTICA VALOR MEDIO CUADRATICO DENTRO DE LAS CELDAS O RESIDUAL:

$$MSR = \sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2 / (N_{..} - rc) \qquad (16)$$

ES UN ESTIMADOR INSESGADO DE σ^2.

LA SUMA DE CUADRADOS ENTRE CELDAS SE PUEDE DIVIDIR EN TRES PARTES SOLO SI LAS n_{ti} SON IGUALES PARA TODA CELDA $(n_{ti}=n)$, O SI SE SATISFACEN CIERTAS CONDICIONES DE PROPORCIONALIDAD*; AQUI SOLO TRATAREMOS EL PRIMERO DE ESTOS CASOS, EN EL QUE SE OBTIENE:

$$n \sum_t \sum_i (\bar{X}_{ti.} - \bar{X}_{...})^2 = \underbrace{nc \sum_t (\bar{X}_{t..} - \bar{X}_{...})^2}_{\text{ENTRE RENGLONES} = SSBR} +$$

*BANCROFT, T. A., "TOPICS IN INTERMEDIATE STATISTICAL METHODS", IOWA UNIVERSITY PRESS, 1968.

$$+ \underbrace{nr\sum_{i}(\bar{X}_{.i.} - \bar{X}_{...})^2}_{\text{ENTRE COLUMNAS = SSBC}} + \underbrace{n\sum_{t}\sum_{i}(\bar{X}_{ti.} - \bar{X}_{t..} - \bar{X}_{.i.} + \bar{X}_{...})^2}_{\text{INTERACCION = SSI}} \qquad (17)$$

LAS ESPERANZAS DE LOS TERMINOS DEL MIEMBRO DERECHO DE LA EC (17) SON:

ENTRE RENGLONES: $$E(SSBR) = (r-1)\sigma^2 + nc\sum_{t}\rho_t^2 \qquad (18)$$

ENTRE COLUMNAS: $$E(SSBC) = (c-1)\sigma^2 + nr\sum_{i}\kappa_i^2 \qquad (19)$$

INTERACCION: $$E(SSI) = (r-1)(c-1)\sigma^2 + n\sum_{t}\sum_{i}(\rho\kappa)_{ti}^2 \qquad (20)$$

POR LO QUE LOS GRADOS DE LIBERTAD RESPECTIVOS SON $(r-1)$, $(c-1)$ Y $(r-1)(c-1)$; EN ESTAS CONDICIONES LOS VALORES MEDIOS CUADRATICOS SON:

ENTRE RENGLONES: $$E(MSBR) = \sigma^2 + (r-1)^{-1}\, nc\sum_{t}\rho_t^2 \qquad (21)$$

ENTRE COLUMNAS: $$E(MSBC) = \sigma^2 + (c-1)^{-1}\, nr\sum_{i}\kappa_i^2 \qquad (22)$$

INTERACCION: $$E(MSI) = \sigma^2 + (r-1)^{-1}(c-1)^{-1}\, n\sum_{t}\sum_{i}(\rho\kappa)_{ti}^2 \qquad (23)$$

POR LO ANTERIOR, LA PRUEBA DE HIPOTESIS DE QUE $\rho_1 = \rho_2 = \ldots = \rho_r = 0$ SE PUEDE HACER PROBANDO LA HIPOTESIS DE IGUALDAD DE VARIANCIAS ENTRE RENGLONES Y RESIDUAL, PARA LO CUAL SE UTILIZA LA ESTADISTICA

$$F = MSBR/MSR \qquad (24)$$

QUE TIENE DISTRIBUCION F CON $(r-1)$ Y $(N_{..}-rc) = rc(n-1)$ GRADOS DE LIBERTAD.

ASIMISMO, LA PRUEBA DE HIPOTESIS DE QUE $\kappa_1 = \kappa_2 = \dots \kappa_c = 0$ SE PUEDE HACER PROBANDO LA HIPOTESIS DE IGUALDAD DE VARIANCIAS ENTRE COLUMNAS Y RESIDUAL, CON LA ESTADISTICA

$$F = MSBC/MSR \tag{25}$$

QUE TIENE DISTRIBUCION F CON $(c-1)$ Y $rc(n-1)$ GRADOS DE LIBERTAD.

LA PRUEBA DE HIPOTESIS DE INTERACCION NULA, O SEA, DE QUE $(\rho\kappa)_{ti} = 0$ PARA TODA t E i SE PRUEBA CON

$$F = MSI/MSR \tag{26}$$

QUE TIENE DISTRIBUCION F CON $(r-1)(c-1)$ Y $rc(n-1)$ GRADOS DE LIBERTAD.

EN EL CASO PARTICULAR DE UNA OBSERVACION POR CELDA $(n=1)$, NO SE REQUIERE EL TERCER INDICE (j) Y EL MODELO ES

$$X_{ti} = \xi + \rho_t + \kappa_i + (\rho\kappa)_{ti} + Z_{ti} \tag{27}$$

EN ESTAS CONDICIONES NO SE OBTIENE NINGUNA SUMA DE CUADRADOS RESIDUAL Y NO ES POSIBLE ESTIMAR A σ^2 DE MANERA SEPARADA DE ρ_t, κ_i Y $(\rho\kappa)_{ti}$ Y, EN CONSECUENCIA, NO SE PUEDEN HACER LAS COMPARACIONES DE VARIANCIAS DADAS POR LAS ECS (24), (25) Y (26). PARA SALVAR ESTE OBSTACULO EL MODELO DE LA EC (27) SE REDUCE A

$$X_{ti} = \zeta + \rho_t + \kappa_i + Z_{ti} \qquad (28)$$

EL CUAL IMPLICA QUE $(\rho\kappa)_{ti} = 0$ PARA TODA t E i, ES DECIR, QUE NO HAY INTERACCION ENTRE LOS PARAMETROS; EN ESTE CASO LA ESTADISTICA

$$\sum_t \sum_i (X_{ti} - \bar{X}_{t.} - \bar{X}_{.i} + \bar{X}_{..})^2/(r-1)(c-1) \qquad (29)$$

ES EL VALOR MEDIO CUADRATICO RESIDUAL, MSR.

EL EXPERIMENTO DE BLOQUES ALEATORIZADOS VISTO ANTERIORMENTE ES, COMO PUEDE VERSE, EL CASO PARTICULAR DE UN EXPERIMENTO DE DOS FACTORES CRUZADOS CON $n=1$.

FORMULAS SIMPLIFICADAS PARA LAS SUMAS DE CUADRADOS (SS)

TOTAL: $SST = \sum_t \sum_i \sum_j X_{tij}^2 - nrc\bar{X}_{...}^2 \qquad (30)$

ENTRE RENGLONES: $SSBR = nc\sum_t \bar{X}_{t..}^2 - nrc\bar{X}_{...}^2 \qquad (31)$

ENTRE COLUMNAS: $SSBC = nr\sum_i \bar{X}_{.i.}^2 - nrc\bar{X}_{...}^2 \qquad (32)$

DENTRO CELDAS (RESIDUAL): $SSR = \sum_t \sum_i \sum_j X_{tij}^2 - n\sum_t \sum_i \bar{X}_{ti.}^2 \qquad (33)$

INTERACCION: $SSI = SST - SSBR - SSBC - SSR \qquad (34)$

SI $n = 1$, $SSR = SST - SSBR - SSBC$.

LA TABLA DEL ANALISIS DE VARIANCIA EN DOS DIRECCIONES CON FACTORES CRUZADOS QUEDA EN LA FORMA:

FUENTE DE VARIACION	G. DE L.	SS	MS	F
ENTRE RENGLONES	r-1	SSBR	MSBR	MSBR/MSR
ENTRE COLUMNAS	c-1	SSBC	MSBC	MSBC/MSR
INTERACCION	(r-1)(c-1)	SSI	MSI	MSI/MSR
RESIDUAL (DENTRO DE LAS CELDAS)	rc(n-1)	SSR	MSR	
TOTAL	rcn-1	SST		

EJEMPLO

EN UN EXPERIMENTO PARA DETERMINAR EL COEFICIENTE DE EXPANSION DE ALGUNAS ALEACIONES DE TITANIO, FABRICADAS CON DOS PROCEDIMIENTOS DIFERENTES, SE ELABORARON 16 ESPECIMENES A LOS CUALES SE LES MIDIO EL COEFICIENTE DE EXPANSION TERMICA. SE DESEA SABER SI LAS <u>ALEACIONES</u> Y <u>PROCEDIMIENTOS</u> INFLUYEN EN DICHO COEFICIENTE.

LOS RESULTADOS EXPERIMENTALES SE PRESENTAN EN LA TABLA SIGUIENTE, EN LA CUAL CADA CELDA TIENE LAS SIGUIENTES ANOTACIONES:

X_{ti1}

X_{ti2} $\bar{X}_{ti.}$ $\bar{X}^2_{ti.}$

COEFICIENTES DE EXPANSION						
PROCEDIMIENTOS	ALEACIONES					
	A	B	C	D	$\bar{X}_{t..}$	$\bar{X}^2_{t..}$
1	4.78 4.28 4.53 20.5209	3.84 5.28 4.56 20.7936	5.82 5.77 5.795 33.5820	4.57 5.44 5.005 25.0500	4.9725	24.7258
2	4.46 4.79 4.625 21.3906	4.73 3.36 4.045 16.3620	4.76 3.31 4.035 16.2812	4.30 3.86 4.08 16.6464	4.1963	17.6085
TOTALES	18.31	17.21	19.66	18.17		42.3343
$\bar{X}_{.i.}$	4.5775	4.3025	4.915	4.5425		
$\bar{X}^2_{.i.}$	20.9535	18.5115	24.1572	20.6343		

EL MODELO AQUI ES

$$X_{tij} = \xi + \rho_t + \kappa_i + (\rho\kappa)_{ti} + Z_{tij}$$

$$t = 1,2; \quad i = 1,2,3,4; \quad j = 1,2$$

POR LO TANTO: $r = 2,\ c = 4,\ n = 2,\ \bar{X}_{...} = 73.35/16 = 4.5844$

$$\bar{X}^2_{...} = 21.0164,\ nrc\ \bar{X}^2_{...} = 336.2639$$

$SSBR = 2 \times 4\ (24.7258 + 17.6085) - 336.2639$

$= 338.6741 - 336.2639 = 2.4102$

SSBC = 2 x 2 (20.9535 + 18.5115 + 24.1572 + 20.6343)-

- 336.2639

SSBC = 337.0260 -336.2639 = 0.7621

TABLA DE CUADRADOS

X^2_{tij}

	A	B	C	D
1	22.8484 18.3184	14.7456 27.8784	33.8724 33.2929	20.8849 29.5936
2	19.8916 22.9441	22.3729 11.2896	22.6576 10.9561	18.4900 14.8996
TOTAL	84.0025	76.2865	100.7790	83.8681

$\sum_{tij}\sum\sum X^2_{tij} = 344.9361$

SSR = 344.9361 - 2(20.5209 + 20.7936 + 33.5820 + 25.0500 +

+ 21.3906 + 16.3620 + 16.2812 + 16.6464) = 344.9361 -

- 341.2534 = 3.6827

SST = 344.9361 - 336.2639 = 8.6722

SSI = 8.6722 - 0.7621 - 2.4102 - 3.6827 = 1.8172

LA TABLA DEL ANALISIS DE VARIANCIA RESULTA SER:

FUENTE DE VARIACION	SS	G. DE L.	MS	F
ENTRE RENGLONES (PROCEDIMIENTOS)	2.4102	1	2.4102	5.236
ENTRE COLUMNAS (ALEACIONES)	0.7621	3	0.2541	0.552
INTERACCION	1.8172	3	0.6057	1.316
RESIDUAL	3.6827	8	0.4603	
TOTAL	8.6722			

$F_{0.95,1,8} = 5.32 > 5.236$ SE ACEPTA $H_0: \rho_t = 0 \ \forall t$

$F_{0.95,3,8} = 4.07 > 0.552$ SE ACEPTA $H_0: \kappa_i = 0 \ \forall i$

$F_{0.95,3,8} = 4.07 > 1.316$ SE ACEPTA $H_0: (\rho\kappa)_{ti} = 0 \ \forall t,i$

EJEMPLO

PARA DETERMINAR EL EFECTO DE CUATRO DIFERENTES PESTICIDAS EN LA PRODUCCION DE TRES TIPOS DE FRUTA CITRICA, SE DISEÑO UN EXPERIMENTO DE DOS FACTORES CRUZADOS EN EL QUE SE ASIGNARON AL AZAR DOS ARBOLES FRUTALES DE CADA TIPO PARA SER FUMIGADOS POR CADA PESTICIDA. LAS PRODUCCIONES DE FRUTA EN KG/ARBOL SE MUESTRAN EN LA TABLA SIGUIENTE:

VARIEDAD DE FRUTA	PESTICIDA 1	2	3	4
1	49 39	50 55	43 38	53 48
2	55 41	67 58	53 42	85 73
3	66 68	85 92	69 62	85 99

REALIZAR EL ANALISIS DE VARIANCIA Y HACER ESTIMACIONES PUNTUALES DE LOS EFECTOS, DE LAS INTERACCIONES Y DE σ^2.

LAS HIPOTESIS A PROBAR SON:

LOS EFECTOS DE LAS FRUTAS SON NULOS: H_o: $\rho_1 = \rho_2 = \rho_3 = 0$

H_1: NO TODOS LOS EFECTOS DE LAS FRUTAS SON IGUALES A CERO

LA CUAL PUEDE HACERCE PROBANDO LA HIPOTESIS DE IGUALDAD DE VARIANCIAS ENTRE VARIEDADES DE FRUTAS (RENGLONES) Y RESIDUAL, MEDIANTE LA ESTADISTICA:

$$F_R = MSBR/MSR \quad \text{VERSUS} \quad F_{0.01,2,12} = F_{CR}$$

b) LA PRUEBA DE LA HIPOTESIS DE QUE LOS EFECTOS ENTRE LOS PESTICIDAS SON NULOS: H_0: $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = 0$

CONTRA LA HIPOTESIS DE QUE LOS EFECTOS NO SON TODOS NULOS,

LA CUAL PUEDE HACERSE PROBANDO LA HIPOTESIS DE IGUALDAD DE VARIANCIAS ENTRE PESTICIDAS Y RESIDUAL:

$$F_c = MSBC/MSR \quad \text{VERSUS} \quad F_{0.01,3,12}$$

c) FINALMENTE LA PRUEBA DE HIPOTESIS DE INTERACCION NULA H_0: $(\rho K)_{ti} = 0 \forall t, \forall i$, CONTRA LA HIPOTESIS H_1 DE QUE NO TODAS LAS INTERACCIONES SON NULAS, PUEDE HACERSE PROBANDO LA HIPOTESIS DE IGUALDAD DE VARIANCIA ENTRE LAS INTERACCIONES Y LA RESIDUAL, CON LA ESTADISTICA:

$$F_I = MSI/MSR \quad \text{VERSUS} \quad F_{0.01,6,12}$$

DESARROLLEMOS LA TABLA DE ANALISIS DE VARIANCIA EN 2 DIRECCIONES CON FACTORES CRUZADOS.

VARIEDAD DE FRUTA	PESTICIDA 1		2		3		4		TOTALES	$\bar{X}_{t..}$	$\bar{X}^2_{t..}$
		44		52.5		40.5		50.5			
1	49		50		43		53		375	46.88	2197.27
	39	1936	55	2756.25	38	1640.25	48	2550.25			
		48		62.5		47.5		79			
2	55		67		53		85		474	59.25	3510.56
	41	2304	58	3906.25	42	2256.25	73	6241			
		67		88.5		65.5		92			
3	66		85		69		85		626	78.25	6123.06
	68	4489	92	7832.25	62	4290.25	99	8464			
TOTALES	318		407		307		443		$N_{..}$ = 1475		11830.89
$\bar{X}_{.i.}$	53		67.83		51.17		73.83			$\bar{X}_{...}$ = 61.46	
$\bar{X}^2_{.i.}$	2809		4601.36		2618.03		5451.36		15479.75		$\bar{X}^2_{...}$ = 3777.23

	$\bar{X}_{ti.}$
X_{ti1}	
X_{ti2}	$\bar{X}^2_{ti.}$

$$\sum_t \sum_i \bar{X}^2_{ti} = 48{,}667.75$$

TOTAL: $SST = \sum_t \sum_i \sum_j X^2_{tij} - nrc\bar{X}^2_{...} = 49^2+39^2+55^2+41^2+\ldots+73^2+85^2+55^2-2x3x4x$

$= 97839 - 90,655.92$

$= 7183.04$

ENTRE RENGLONES:

$SSBR = nc\sum_t \bar{X}^2_{t..} - nrc\bar{X}^2_{...} = 2 \times 4 \times 11830.89 - 90,655.92$

$= 94647.12 - 90,655.92$

$= 3991.20$

ENTRE COLUMNAS:

$SSBC = nr\sum_i \bar{X}^2_{.i.} - nrc\bar{X}^2_{...} = 2 \times 3 \times 15479.75 - 90,655.92$

$= 92878.50 - 90655.92$

$= 2222.58$

ENTRE CELDAS:

$SSR = \sum_t \sum_i \sum_j X^2_{tij} - n\sum_t \sum_i \bar{X}^2_{ti.} = 97839 - 2 \times 48667.75$

$= 507.5$

INTERACCION:

$SSI = SST - SSBR - SSBC - SSR$

$= 7183.04 - 3991.20 - 2222.58 - 507.5$

$= 461.76$

PUDIENDO CON LO ANTERIOR COMPLETAR LA TABLA DE ANALISIS DE VARIANZA:

FUENTE DE VARIACION	G. DE L	SS	MS	F_E	F_C
ENTRE VARIEDADES DE FRUTA	r-1=3-1=2	SSBR = 3991.20	SSBR/(r-1)= 1995.60	MSBR/MSR= = 47.19 >	$F_{CR}=F_{0.01,2,12}$ = 6.93
ENTRE PESTICIDA	c-1=4-1=3	SSBC = 2222.58	SSBC/(c-1)= 740.86	MSBC/MSR= = 17.52 >	$F_{CC}=F_{0.01,3,12}$ = 5.95
INTERACCION	(r-1)(c-1) = 6	SSI = 461.76	SSI/(r-1)(c-1)= 76.96	MSI/MSR = = 1.82 <	$F_{CI}=F_{0.01,6,12}$ = 4.82
RESIDUAL (DENTRO DE CELDAS)	rc(n-1)=12	SSR = 507.5	SSR/rc(n-1) 42.29		
TOTAL	rcn-1=23	SST = 7183.04			

COMO PUEDE OBSERVARSE EN LAS F_E (F.ESTIMADA) Y LAS F_C (F CRITICAS) SE TENDRAN LAS SIGUIENTES CONCLUSIONES DEL ANALISIS DE VARIANCIA (VER LAS 2 ULTIMAS COLUMNAS)

1. DADO QUE $F_{ER} > F_{CR} \Rightarrow$ SE RECHAZA LA HIPOTESIS H_0 $\therefore$ SI HAY EFECTO ENTRE VARIEDADES DE FRUTAS

2. DADO QUE $F_{EC} > F_{CC} \Rightarrow$ SE RECHAZA LA HIPOTESIS H_0 $\therefore$ SI HAY EFECTO ENTRE LOS DIFERENTES TIPOS DE PESTICIDAS.

3. DADO QUE $F_{EI} < F_{CI} \Rightarrow$ SE APLICA LA HIPOTESIS H_0 $\therefore$ NO HAY EFECTO DE INTERACCION

CALCULO DE LOS ESTIMADORES DE LOS EFECTOS:

EFECTO DE LA VARIEDAD DE FRUTAS

DADO QUE $E\{\bar{X}_{t..} - \bar{X}_{...}\} = \rho_t \Rightarrow \hat{\rho}_t = \bar{X}_{t..} - \bar{X}_{...}$

$\hat{\rho}_1 = 46.88 - 61.46 = -14.58$

$\hat{\rho}_2 = 59.25 - 61.46 = -2.21$

$\hat{\rho}_3 = 78.25 - 61.46 = 16.79$

EFECTOS DE LA VARIEDAD DE PESTICIDAS

DADO QUE $E\{\bar{X}_{.i.} - \bar{X}_{...}\} = k_i \Rightarrow \hat{k}_i = \bar{X}_{.i.} - \bar{X}_{...}$

$\hat{k}_1 = 53 - 61.46 = -8.46$

$\hat{k}_2 = 67.83 - 61.46 = 6.37$

$\hat{k}_3 = 51.17 - 61.46 = -10.29$

$\hat{k}_4 = 73.83 - 61.46 = 12.37$

ESTIMACIONES PUNTUALES DE LAS INTERACCIONES:

DADO QUE $E\{\bar{X}_{ti.} - \bar{X}_{t..} - \bar{X}_{.i.} + \bar{X}_{...}\} = (\rho k)_{ti} \Rightarrow$

TENDREMOS:

$\widehat{(\rho k)}_{ti} = \bar{X}_{ti.} - \bar{X}_{t..} - \bar{X}_{.i.} + \bar{X}_{...}$

$(\hat{\rho k})_{1,1} = 44 - 46.88 - 53 + 61.46 = 5.58$

$(\hat{\rho k})_{1,2} = 52.5 - 46.88 - 67.83 + 61.46 = -0.75$

$(\hat{\rho k})_{1,3} = 40.5 - 46.88 - 51.17 + 61.46 = 3.91$

$(\hat{\rho k})_{1,4} = 50.5 - 46.88 - 73.83 + 61.46 = -8.75$

$(\hat{\rho k})_{2,1} = 48 - 59.25 - 53 + 61.46 = -2.79$

$(\hat{\rho k})_{2,2} = 62.5 - 59.25 - 67.83 + 61.46 = -3.12$

$(\hat{\rho k})_{2,3} = 47.5 - 59.25 - 51.17 + 61.46 = -1.46$

$(\hat{\rho k})_{2,4} = 79 - 59.25 - 73.83 + 61.46 = 7.38$

$(\hat{\rho k})_{3,1} = 67 - 78.25 - 53 + 61.46 = -2.79$

$(\hat{\rho k})_{3,2} = 88.5 - 78.25 - 67.83 + 61.46 = 3.88$

$(\hat{\rho k})_{3,3} = 67.5 - 78.25 - 51.17 + 61.46 = -2.46$

$(\hat{\rho k})_{3,4} = 92 - 78.25 - 73.83 + 61.46 = 1.38$

FINALMENTE, DADO QUE EL VALOR DE MSW (O MSR) <u>ES UN ESTIMADOR INSESGADO DE $\sigma^2 \Rightarrow \hat{\sigma}^2 = 42.29$ (VER TABLA DE ANALISIS DE VARIANCIA</u>

<u>CABE OBSERVAR QUE TODOS LOS ESTIMADORES $\hat{\rho}_t$, $\hat{k}_i$, $(\hat{\rho k})_{ti}$ Y $\hat{\sigma}^2$ SON INSESGADOS</u>.

MODELO CON DIFERENTES TAMAÑOS DE MUESTRA

SE DESARROLLA LA SUMA DE CUADRADOS:

$$\sum_{t=1}^{r}\sum_{i=1}^{c}\sum_{j=1}^{n_{ti}}(X_{tij}-\bar{X}_{...})^2 = \sum_{t=1}^{r}\sum_{i=1}^{c}\sum_{j=1}^{n_{ti}}(\bar{X}_{t..}-\bar{X}_{...})^2 +$$

$$+\sum_{t=1}^{r}\sum_{i=1}^{c}\sum_{j=1}^{n_{ti}}(\bar{X}_{.i.}-\bar{X}_{...})^2 +$$

$$+\sum_{t=1}^{r}\sum_{i=1}^{c}\sum_{j=1}^{n_{ti}}(X_{tij}-\bar{X}_{ti.})^2 +$$

$$+\sum_{t=1}^{r}\sum_{r=1}^{c}\sum_{j=1}^{n_{ti}}(\bar{X}_{ti.}-\bar{X}_{t..}-\bar{X}_{.i.}+\bar{X}_{...})^2$$

SST = SSBR + SSBC + SSI + SSR

$$n_{t.} = \frac{\sum_{i=1}^{c} n_{ti}}{c} \; ; \quad n_{.i} = \frac{\sum_{t=1}^{r} n_{ti}}{r} \; ; \quad n_{..} = \frac{\sum_{t=1}^{r}\sum_{i=1}^{c} n_{ti}}{cr}$$

ASI:

$$SSBR = \sum_{t=1}^{r}\sum_{i=1}^{c}\sum_{j=1}^{n_{ti}}(\bar{X}_{t..}^2 - 2\bar{X}_{t..}\bar{X}_{...} + \bar{X}_{...}^2)$$

$$= c\sum_{t=1}^{r} n_{t.}\bar{X}_{t..}^2 - 2c\bar{X}_{...}\sum_{t}^{r} n_{t.}\bar{X}_{t..} + n_{..}cr\bar{X}_{...}^2$$

$$= c \sum_{t=1}^{r} n_{t.} \bar{X}_{t..}^{2} - rcn_{..} \bar{X}^{2}_{...}$$

$$SSBC = \sum_{t=1}^{r} \sum_{i=1}^{c} \sum_{j=1}^{n_{ti}} (\bar{X}_{.i.}^{2} - 2\bar{X}_{.i.} \bar{X}_{...} + \bar{X}_{...}^{2})$$

$$= r \sum_{i=1}^{c} n_{.i} \bar{X}^{2}_{.i.} - rcn_{..} \bar{X}_{...}^{2}$$

$$SSR = \sum_{t=1}^{r} \sum_{i=1}^{c} \sum_{j=1}^{n_{ti}} (X_{tij}^{2} - 2X_{tij} \bar{X}_{ti.}^{2})$$

$$= \sum_{t=1}^{r} \sum_{i=1}^{c} \sum_{j=1}^{n_{ti}} X_{tij}^{2} - \sum_{t=1}^{r} \sum_{i=1}^{c} n_{ti} \bar{X}_{ti.}^{2}$$

$$SST = \sum_{t=1}^{r} \sum_{i=1}^{c} \sum_{j=1}^{n_{ti}} (X_{tij}^{2} - 2X_{tij} \bar{X}_{...} + \bar{X}^{2}_{...})$$

$$= \sum_{t=1}^{r} \sum_{i=1}^{c} \sum_{j=1}^{n_{ti}} X_{tij}^{2} - rcn_{..} \bar{X}^{2}_{...}$$

$$SSI = SST - SSBR - SSBC - SSR$$

EJEMPLO:

TRATAMIENTOS	BLOQUES		
	1	2	3
1	10 15 8	12 9	5 18
2	7 12	13 11 10	9

$t = 1,r; \quad r = 2$

$i = 1,c; \quad c = 3$

$j = 1,n_{ti}$

$$n_{ti} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 1 \end{bmatrix}; \qquad n_{t.} = \begin{bmatrix} 2.33 \\ 2 \end{bmatrix}$$

$$n_{.i} = [2.5,\ 2.5,\ 1.5]; \quad n_{..} = 2.165$$

CALCULOS NECESARIOS:

$\bar{X}_{...} = 10.692; \quad \bar{X}^2_{...} = 114.325$

$\sum_t \sum_i \sum_j X^2_{ijk} = 1627$

$$\bar{X}_{ti.} = \begin{bmatrix} 11 & 10.5 & 11.5 \\ 9.5 & 11.33 & 9 \end{bmatrix}; \qquad \bar{X}_{t..} = \begin{bmatrix} 11 \\ 10.33 \end{bmatrix}$$

$\bar{X}_{.i.} = [10.4,\ 11,\ 10.66];$

$$\sum_t \sum_i n_{ti} \bar{X}_{ti.}^2 = 1494.606$$

$$\sum_t n_{t.} \bar{X}_{t..}^2 = 495.711$$

$$\sum_i n_{.i} \bar{X}_{.i.}^2 = 743.353$$

SST = 1627 - (2)(3)(2.166)(114.325) = 140.775

SSBR = (3)(495.711) - (2)(3)(2.166)(114.325) = 1.366

SSBC = (2)(743.353) - (2)(3)(2.166)(114.325) = 0.939

SSR = 1627 - 1494.606 = 132.394

SSI = 140.775 - 1.366 - 0.939 - 132.394 = 6.076

ANALISIS DE VARIANZA:

FUENTE:	SS	G.L.	MS	F	α =0.05
RENGLONES (BR)	1.366	r - 1 = 1	1.366	0.0722	5.59
COLUMNAS (BC)	.939	c - 1 = 2	0.4695	0.0248	4.74
INTERACCION (I)	6.076	(r-1)(c-1) = 2	3.038	0.1606	4.74
RESIDUAL (R)	132.394	rc(n..-1) = 6.996 ≅ 7	18.913		
TOTAL (T)	140.775				

∴ NO HAY EFECTO POR RENGLONES (TRATAMIENTOS).

NO HAY EFECTO POR COLUMNAS (BLOQUES).

NO HAY EFECTO POR LA INTERELACION ENTRE RENGLONES Y COLUMNAS

MODELO CON NIVELES CRUZADOS ALEATORIOS

ESTE MODELO SE OBTIENE A PARTIR DEL PARAMETRICO REEMPLAZADO ρ_t, k_i $(\rho k)_{ti}$ POR U_t, V_i, W_{ti}, RESPECTIVAMENTE DONDE LAS U's V's Y W's SON VARIABLES ALEATORIAS NORMALES, MUTUAMENTE INDEPENDIENTES CADA UNA CON VALOR ESPERADO CERO Y:

$$1) \quad \mathrm{Var}(U_t) = \sigma_u^2 \ \forall t$$

$$2) \quad \mathrm{Var}(V_i) = \sigma_v^2 \ \forall i$$

$$3) \quad \mathrm{Var}(W_{ti}) = \sigma_w^2 \ \forall t,i$$

CONSIDERAMOS SOLAMENTE EL CASO $n_{ti} = n \ \forall t,i$ O SEA IGUAL NUMERO DE ELEMENTOS EN CADA CELDA ti, CON LO CUAL EL MODELO SERA:

$$4) \quad X_{tij} = \xi + U_t + V_i + W_{ti} + Z_{tij}$$

DE DONDE:

$$5) \quad \bar{X}_{...} = \xi + \bar{U} + \bar{V} + \bar{W}_{..} + \bar{Z}_{...}$$

$$6) \quad \bar{X}_{t..} = \xi + U_t + \bar{V} + \bar{W}_{t.} + \bar{Z}_{t..}$$

$$7) \quad \bar{X}_{.i.} = \xi + \bar{U} + V_i + \bar{W}_{.i} + \bar{Z}_{.i.}$$

$$8) \quad \bar{X}_{ti.} = \xi + U_t + V_i + W_{ti} + \bar{Z}_{ti.}$$

6) -5):

$$9) \quad \bar{X}_{t..} - \bar{X}_{...} = (U_t - \bar{U}) + (\bar{W}_{t.} - \bar{W}_{..}) + (\bar{Z}_{t..} - \bar{Z}_{...})$$

7) -5):

10) $\bar{X}_{.i.} - \bar{X}_{...} = (V_i - \bar{V}) + (\bar{W}_{.i} - \bar{W}_{..}) + (\bar{Z}_{.i.} - \bar{Z}_{...})$

8) - 6) - 7) + 5)

11) $\bar{X}_{ti.} - \bar{X}_{t..} - \bar{X}_{.i.} + \bar{X}_{...} = (W_{ti} - \bar{W}_{t.} - \bar{W}_{.i} + \bar{W}_{..}) +$

$+ (\bar{Z}_{ti.} - \bar{Z}_{t..} - \bar{Z}_{.i.} + \bar{Z}_{...})$

DONDE:

12) $\bar{U} = \sum_{t=1}^{r} U_t/r$ 13)... $\bar{V} = \sum_{i=1}^{c} V_i/c$

14) $\bar{W}_{t.} = \sum_{i=1}^{c} W_{ti}/c$ 15)... $\bar{W}_{.i} = \sum_{t=1}^{r} W_{ti}/r$

16) $\bar{W}_{..} = \sum_{t=1}^{r} \sum_{i=1}^{c} W_{ti}/rc$

4) - 8):

17) $X_{tij} - \bar{X}_{ti.} = Z_{tij} - \bar{Z}_{ti.}$

NUEVAMENTE, PARA ANALIZAR LA FUENTE DE VARIABILIDAD DE LOS DATOS, LA SUMA DE CUADRADOS SE PUEDE DIVIDIR EN 2 PARTES:

18). $\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{...})^2 = \underbrace{\sum_t \sum_i n_{ti} (\bar{X}_{ti.} - \bar{X}_{...})^2}_{\text{ENTRE CELDAS}} + \underbrace{\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2}_{\text{DENTRO DE LAS CELDAS}}$

DE AQUI QUE:

E{SUMA DE CUADRADOS DENTRO DE LAS CELDAS} =

$$E(\sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2) = (N_{..} - rc)\sigma^2$$

POR LO CUAL LA ESTADISTICA VALOR MEDIO CUADRATICO DENTRO DE CELDAS O RESIDUAL (MSW O MSR)

$$19) \quad E(MSR) = \sum_t \sum_i \sum_j (X_{tij} - \bar{X}_{ti.})^2/(N_{..} - rc) = \sigma^2$$

SE USA NUEVAMENTE PARA ESTIMAR σ^2 O SEA LA VARIANZA DE CADA Z_{tij}.

DE LAS ECS. 9), 10) y 11) ENCONTRAMOS LOS SIGUIENTES VALORES ESPERADOS DE LOS VALORES MEDIOS CUADRATICOS:

ENTRE RENGLONES: $E\{MSBR\} = \sigma^2 + n\sigma_w^2 + nc\sigma_u^2$

ENTRE COLUMNAS: $E\{MSBC\} = \sigma^2 + n\sigma_w^2 + nr\sigma_v^2$

INTERACCION: $E\{MSI\} = \sigma^2 + n\sigma_w^2$.

LA SITUACION ES SIMILAR A LA DE LA CLASIFICACION DE DOS FACTORES NO CRUZADOS CUANDO UN MODELO ALEATORIO ES APROPIADO.

LA HIPOTESIS H_o: $\sigma_w^2 = 0$ PUEDE PROBARSE COMPARANDO EL VALOR MEDIO CUADRATICO DE LAS INTERACCIONES CON EL RESIDUAL ; ESTO ES:

$$F = MSI/MSR$$

POR OTRO LADO PARA PROBAR LA HIPOTESIS H_o: $\sigma_u^2 = 0$ DE IGUALDAD DE VARIANCIAS ENTRE RENGLONES DEBERA HACERSE LA COMPARACION DE:

$$F = MSBR/MSI$$

Y FINALMENTE; PARA PROBAR LA HIPOTESIS H_0: $\sigma^2_{v.} = 0$, DE IGUALDAD DE VARIANCIAS ENTRE COLUMNAS, DEBERA HACERSE LA COMPARACION DE:

$$F = MSBC/MSI$$

JUSTAMENTE, COMO EN EL CASO DE LA CLASIFICACION NO CRUZADA, TAMBIEN ES LA ALEATORIEDAD DEL TERMINO QUE REPRESENTA LA INTERACCION EN EL MODELO EL QUE TOMA LA DIFERENCIA ESENCIAL EN EL ANALISIS. EL PROCEDIMIENTO FORMAL DE PRUEBA NO SE AFECTA SI LOS EFECTOS ENTRE RENGLONES O COLUMNAS SE CAMBIAN DE PARAMETRICOS A TERMINOS ALEATORIOS O VICEVERSA (DANDO UN MODELO MEZCLADO).

ES UTIL RECORDAR QUE SI EL MSBR O MSBC SE COMPARA CON EL MSR CUANDO EL MODELO ALEATORIO ES APROPIADO, EL POSIBLE EFECTO DE UNA VARIANCIA $\sigma^2_w \neq 0$ DE INTERACCION PUEDE DEBERSE SOLAMENTE AL INCREMENTO DEL TAMAÑO MEDIO DE LA RELACION CON EL MSR.

EJEMPLO

SUPONGAMOS QUE UNA COMPAÑIA DISPONE DE n FUENTES DIFERENTES DE MATERIAS PRIMAS A_n Y m MAQUINAS DE DISTINTAS MARCAS B_m PARA PRODUCIR UN NUEVO PRODUCTO. SE SABE QUE LAS MARCAS DE MAQUINAS SON IGUALMENTE PRODUCTIVAS EN TERMINOS DE VELOCIDAD - EL NUMERO DE TIRADAS PRODUCIDAS POR HORA - PERO NO SE SABE SI TRABAJAN IGUALMENTE BIEN EN TERMINOS DEL NUMERO DE UNIDADES DEFECTUOSAS ELABORADAS ENTRE LAS PRODUCCIONES POR HORA.

ADEMAS, LA FIRMA DESCONOCE SI HAY DIFERENCIAS EN LA CALIDAD DE LAS MATERIAS PRIMAS PROVENIENTES DE LAS FUENTES. POR ULTIMO SE SOSPECHA QUE LA MATERIA PRIMA DE UNA FUENTE PUEDE PRESENTAR UN EFECTO ESPECIAL EN UNA MAQUINA PARTICULAR O VICEVERSA. POR CONSIGUIENTE, SE DESEA ESTABLECER SI LOS A_n SON DIFERENTES, SI LOS B_m SON DIFERENTES Y SI EXISTE ALGUN EFECTO CONJUNTO A x B. PARA RESPONDER A ESTAS PREGUNTAS SE SELECCIONARON AL AZAR 4 FUENTES: A_1, A_2, A_3 Y A_4 Y 3 MARCAS DE MAQUINAS B_1, B_2 Y B_3, Y SE HIZO OPERAR CADA MARCA DE MAQUINA EN IDENTICAS CONDICIONES CON CADA FUENTE DE MATERIAL DURANTE DOS HORAS Y SE REGISTRO EL NUMERO DE UNIDADES DEFECTUOSAS POR CADA HORA COMO SE INDICA EN LA TABLA. CON ESTOS DATOS, ¿A QUE CONCLUSION SE PUEDE LLEGAR?

MAQUINA	FUENTES DE MAT. PRIMA				TOTALES	$\bar{X}_{t..}$	$\bar{X}^2_{t..}$
	1	2	3	4			
1	7 9 49 5	6 6 36 6	6 8 36 4	6 5 36 7	50	6.5	39.06
2	3 4 9 2	4 5 16 3	4 2 16 6	3 1 9 5	28	3.5	12.5
3	8 10 64 6	8.5 8 72.5 9	7.5 7 56.25 8	7 9 49 5	62	7.75	60.06
TOTALES	36	37	35	32	140		111.62
$\bar{X}_{.i.}$	6	6.17	5.83	5.33		$\bar{X}_{...}$ = 5.83	
$\bar{X}^2_{.i.}$	36	38.03	34.03	28.44	136.5		$\bar{X}^2_{...}$ = 33.99

$r = 3$, $c = 4$, $n = 2$

$$\sum_t \sum_i \sum_j X^2_{tij} = 9^2 + 5^2 + 4^2 + 2^2 + \ldots.\ 5^2 + 9^2 + 5^2 = 952$$

TOTAL: $SST = \sum_t \sum_i \sum_j X^2_{tij} - ncr\bar{X}^2_{...} = 952 - 2 \times 3 \times 4 \times 33.99 =$

$= 952 - 815.73 = 136.27$

ENTRE RENGLONES:

$$SSBR = nc\sum_t \bar{X}^2_{t..} - ncr\bar{X}^2_{...} = 892.96 - 815.73 = 77.23$$

ENTRE COLUMNAS:

$$SSBC = nr\sum_i \bar{X}^2_{.i.} - nrc\bar{X}^2_{...} = 819 - 815.73 = 3.27$$

ENTRE CELDAS:

$$SSR = \sum_t \sum_i \sum_j X^2_{tij} - n\sum_t \sum_i \bar{X}^2_{ti.} = 952 - 2 \times 448.75 = 54.50$$

INTERACCION: $SSI = SST - SSBR - SSBC - SSR$

$= 136.27 - 77.23 - 3.27 - 54.50$

$= 1.27$

LA TABLA DE ANALISIS DE VARIANZA SERA:

. .IACION:

.	G. de l.	SS	MS	F estimada.		F crítica.
IAS	r-1=3-1=2	SSBR = 77.23	SSBR/(R-1) = 38.62	F_{ER} = 38.62/0.21 183.90	>	$F_{ER} = F_{0.01,2,6}$ = 10.92
TES	c-1=4-1=3	SSBC = 3.27	MSBC = SSBC/(c-1) = 1.09	F_{EC} = 1.09/0.21 5.19	<	$F_{CC} = F_{0.01,3,6}$ = 9.78
ON	(r-1)(c-1) = 6	SSI = 1.27	MSI = SSI/(r-1)(c-1) = 0.21	F_{EI}=0.21/4.54 = 0.05	<	$F_{CI} = F_{0.01,6,12}$ = 4.82
	rc(n-1)=12	SSR = 54.50	MSR = SSR/rc(n-1) = 4.54			
	rcn-1 = 23	SST = 136.27				

O ANTERIOR CONCLUIMOS QUE:

), QUE $F_{CR} < F_{ER}$ ⇒ SI HAY VARIABILIDAD ENTRE LAS DIFERENTES

ICAS DE MAQUINA.

MO $F_{CC} > F_{EC}$ ⇒ NO HAY EFECTO ENTRE LAS DIFERENTES FUENTES DE MATERIA PRIMA

FINALMENTE COMO:

$F_{CI} > F_{EI}$ ⇒ NO HAY EFECTO ENTRE LAS INTERACCIONES DE LAS MAQUINAS Y LAS FUENTES DE MATERIA PRIMA.

EJEMPLO

EN UNA INVESTIGACION SOBRE LA ACUMULACION DE UNA SUSTANCIA EN LOS DIENTES DE LAS JOVENES DE 18 A 20 AÑOS DE EDAD EN UNA LOCALIDAD, SE DISEÑO UN EXPERIMENTO COMPLETAMENTE ALEATORIZADO EN EL QUE SE SELECCIONARON AL AZAR TRES JOVENES, A CADA UNA DE LAS CUALES SE LES RASPO EL SARRO DE LA DENTADURA; EL SARRO DE CADA UNA SE DIVIDIO EN SEIS PARTES IGUALES Y SE LES ENTREGARON DOS PARTES A CADA UNO DE TRES ANALISTAS TOMADOS TAMBIEN AL AZAR, CON EL FIN DE QUE HICIERAN EL ANALISIS QUIMICO PARA DETERMINAR LA CANTIDAD DE LA SUSTANCIA DE INTERES CONTENIDA EN CADA PARTE. LAS CONCENTRACIONES, EN MICROGRAMOS OBTENIDAS SE PRESENTAN EN LA SIGUIENTE TABLA:

	MUJER		
ANALISTA	A	B	C
1	13.2	10.6	8.5
	12.3	9.8	8.9
2	12.5	9.6	7.9
	12.9	10.7	8.4
3	13.0	9.9	8.3
	12.4	10.3	8.6

SOLUCION

a) HACER EL ANALISIS DE VARIANCIA Y LAS ESTIMACIONES DE TODOS LOS PARAMETROS DE INTERES; TOME $\alpha = 0.05$. ESBOCE SUS CONCLUSIONES.

SE TRATA DE UN PROBLEMA DE NIVELES ALEATORIOS. PARA OBTENER LOS PARAMETROS NECESARIOS PARA EL CALCULO DE LAS ESTADISTICAS F, SE USARA LA SIGUIENTE TABLA.

ANALISTA	MUJER A		MUJER B		MUJER C		TOTALES	$\bar{X}_{t..}$	$\bar{X}^2_{t..}$
1	13.2 12.3	12.75 162.563	10.6 9.8	10.2 104.04	8.5 8.9	8.7 75.69	63.3	10.55	111.3025
2	12.5 12.9	12.7 161.29	9.6 10.7	10.15 103.023	7.9 8.4	8.15 66.422	62	10.333	106.7778
3	13.0 12.4	12.7 161.29	9.9 10.3	10.1 102.01	8.3 8.6	8.45 71.403	62.5	10.41667	108.5069
TOTALES	76.3		60.9		50.6		187.8		326.5872
$\bar{X}_{.i.}$	12.71667		10.15		8.433				$r = 3$
									$c = 3$
$\bar{X}^2_{.i.}$	161.71361		103.0225		71.1211		Σ=335.85722		$n = 2$

EN CADA CELDA SE INDICA:

X_{ti1}	$\bar{X}_{ti.}$
X_{ti2}	$\bar{X}^2_{ti.}$

DE LOS DATOS:

$$\Sigma\Sigma\Sigma\ X^2_{tij} = 13.2^2+12.3^2+12.5^2+12.9^2+13^2+12.4^2+10.6^2+9.8^2+9.6^2+10.7^2+9.9^2+ +10.3^2+8.5^2+8.9^2+7.9^2+8.4^2+8.3^2+8.6^2 = 2017.38$$

DE LA TABLA:

$$\Sigma\Sigma\Sigma\ \bar{X}^2_{ti.} = 162.563+161.29+161.29+104.4+103.023+102.01+75.69+66.422+71.403 = = 1007.73$$

$\bar{X}_{...} = \frac{187.8}{18} = 10.433$, $\bar{X}^2_{...} = 108.854$, $nrc\bar{X}^2_{...} = 2(3)(3)(108.854) = 1959.38$

$\sum_t \bar{X}^2_{t..} = 326.5872$, $\sum_i \bar{X}^2_{.i.} = 335.85722$

POR LO TANTO LAS SUMAS DE CUADRADOS VALDRAN:

$SST = \Sigma\Sigma\Sigma X^2_{tij} - nrc\bar{X}^2_{...} = 2017.38 - 1959.38 = 58$

$SSBR = nc\sum_t \bar{X}^2_{t..} - nrc\bar{X}^2_{...} = (2)(3)(326.58722) - 1959.38 = 0.14333$, CON $(r-1)$ G DE L.

$SSBC = nr\sum_i \bar{X}^2_{.i.} - nrc\bar{X}^2_{...} = (2)(3)(335.85722) - 1959.38 = 55.76332$, CON $(c-1)$ G. DE L.

$SSR = \Sigma\Sigma\Sigma X^2_{tij} - n\sum_{ti}\bar{X}^2_{ti.} = 2017.38 - (2)(1007.731) = 1.918$, CON $rc(n-1)$ G. DE L.

$SSI = SST - SSBR - SSBC - SSR = 58 - 0.1433 - 55.76332 - 1.918 = 0.1753467$, con $(r-1)(c-1)$ G. DE L.

EN LA SIGUIENTE TABLA SE RESUMEN LOS RESULTADOS DEL ANALISIS DE VARIANCIA; COMO SE TRATA DE UN MODELO DE NIVELES ALEATORIOS, LAS ESTADISTICAS F SE CALCULARAN COMO:

EFECTOS DE INTERACCION: $F = \frac{MSI}{MSR}$

EFECTOS "DEL ANALISTA" $F = \frac{MSBR}{MSI}$

EFECTOS DE "LA MUJER" $F = \frac{MSBC}{MSI}$

TABLA DE ANALISIS DE VARIANCIA

ORIGEN DE VARIACION	SS	G. DE L.	MS	F
ANALISTA	0.14333	2	0.071665	1.6347
MUJER	55.76332	2	27.88166	635.998
INTERACCION	0.1753467	4	0.0438392	0.2057
RESIDUO	1.918	9	0.213	
TOTAL	58			

LOS VALORES CRITICOS PARA LAS ESTADISTICAS ANTERIORES, CON $\alpha = 0.05$ SON:

ANALISTA: $F_{0.05,2,4} = 6.94$

MUJER $F_{0.05,2,4} = 6.94$

INTERACION $F_{0.05,4,9} = 3.63$

COMO: $6.94 > 1.6347$ EL EFECTO DEL ANALISTA NO ES SIGNIFICATIVO

$6.94 < 635.998$ EL EFECTO DE LA MUJER ES SIGNIFICATIVO

$3.63 > 0.2057$ EL EFECTO DE INTERACCION ANALISTA-MUJER NO ES SIGNIFICATIVO

COMO PUEDE VERSE DE LOS RESULTADOS ANTERIORES, EL UNICO EFECTO SIGNIFICATIVO ES EL DE LA MUJER; ES DECIR QUE LA CONCENTRACION DE LA SUSTANCIA DE INTERES SI DEPENDE DE LA MUJER DE QUE SE TRATE.

b) REALIZAR LO PEDIDO EN EL INCISO ANTERIOR CONSIDERANDO AHORA EL PROBLEMA COMO SI SE TRATARA DE PARAMETROS FIJOS. COMPARE Y COMENTE LOS RESULTADOS DE AMBOS INCISOS

EN ESTE CASO LAS ESTADISTICAS F ESTAN DADAS POR:

-ANALISTA: $F = \frac{MSBR}{MSR}$

-MUJER: $F = \frac{MSBC}{MSR}$

-INTERACCION: $F = \frac{MSI}{MSR}$

POR LO TANTO, LA TABLA DE ANALISIS DE VARIANCIA QUEDARIA:

ORIGEN DE VARIACION	SS	G. DE L.	MS	F
ANALISTA	0.14333	2	0.071665	0.33645
MUJER	55.76332	2	27.88166	131.236
INTERACCION	0.1753467	4	0.0438392	0.2058
RESIDUO	1.918	9	0.213	
TOTAL	58			

LOS VALORES CRITICOS, EN TABLAS, SON:

ANALISTA: $F_{0.05,2,\ 9} = 4.26$

MUJER: $F_{0.05,2,\ 9} = 4.26$

INTERACCION: $F_{0.05,4,9} = 3.63$

COMO:

$4.26 > 0.33645$ EL EFECTO DEL ANALISTA NO ES SIGNIFICATIVO

$4.26 < 131.236$ EL EFECTO DE LA MUJER ES SIGNIFICATIVO

$3.63 > 0.2058$ EL EFECTO DE INTERACCION NO ES SIGNIFICATIVO

COMPARANDO LOS RESULTADOS DE AMBOS MODELOS PODEMOS OBSERVAR QUE LOS RESULTADOS HAN SIDO IGUALES EN CUANTO A CONCLUSIONES; NO OBSTANTE LOS RANGOS DE LAS ZONAS DE ACEPTACION HAN SIDO ALTERADAS, ASI COMO LAS ESTADISTICAS F CALCULADAS, POR LO QUE CABRIA LA POSIBILIDAD DE QUE EN UN CASO CERCA DE LOS LIMITES DE ACEPTACION (VALORES CRITICOS), LA APLICACION DE UN MODELO U OTRO DERIVARA EN CONCLUSIONES DIFERENTES.

c) CALCULAR EL INTERVALO DE CONFIANZA DE LA DIFERENCIA DE LAS CONCENTRACIONES MEDIAS OBTENIDAS POR LOS ANALISTAS 2 Y 3.

	ANALISTA		ANALISTA
	2	3	1
	12.5	13.0	13.2
	12.9	12.4	12.3
	9.6	9.9	10.6
	10.7	10.3	9.8
	7.9	8.3	8.5
	8.4	8.6	8.9
PROMEDIO	10.333	10.417	10.55

EL INTERVALO DE CONFIANZA ESTA DADO POR:

$$\overline{X}_i - \overline{X}_j \pm t_{\nu,\,\alpha/2}\; S\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

LA SUMA DE CUADRADOS DENTRO DE GRUPOS, CONSIDERANDO LA TOTALIDAD DE LOS DATOS SERA:

$$(12.5-10.33)^2+(12.9-10.33)^2+(9.6-10.33)^2+(10.7-10.33)^2+(7.9-10.33)^2+(8.4-10.33)^2+$$

$$+\ (13-10.417)^2+(12.4-10.417)^2+(9.9-10.417)^2+(10.3-10.417)^2+(8.3-10.417)^2+(8.6-10.417)^2+$$

$$+\ (13.2-10.55)^2+12.3-10.55)^2+(12.5-10.55)^2+(12.9-10.55)^2+(13-10.55)^2+(12.4+10.55)^2$$

$$=\ 57.8567$$

ENTONCES $S^2 = \frac{57.8567}{N-k} = \frac{57.8567}{18-3} = 3.857,\ S = 1.9639$

EN TABLAS: $t_{15,0.025} = 2.132$

POR TANTO, EL INTERVALO DE CONFIANZA VALE:

$$10.417 - 10.333 \pm 2.132(1.9639)\sqrt{\frac{1}{6} + \frac{1}{6}} = 0.084 \pm 2.417$$

d) APLIQUE EL METODO DE TUKEY PARA REALIZAR LAS COMPARACIONES MULTIPLES DE LAS MEDIAS DE LOS RESULTADOS DE LAS MUJERES. DESSARROLLE Y APLIQUE A ESTE PROBLEMA LOS METODOS DE FISHER Y DE DUNCAN PARA COMPARACIONES MULTIPLES.

METODO DE TUKEY

EL MARGEN, DE ACUERDO AL METODO DE TUKEY, ESTA DADO POR LA ECUACION:

$$\frac{q_{k,\nu,\alpha/2}}{\sqrt{2}} \, s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

EN ESTE CASO: $n_i = n_j = cte = n = 6$

EL VALOR DE S SE OBTENDRA DE LA TOTALIDAD DE LOS DATOS, COMO $MSW = s^2$, PARA ESTO SE OBTENDRA MSW:

	MUJER		
	A	B	C
	13.2	10.6	8.5
	12.3	9.8	8.9
	12.5	9.6	7.9
	12.9	10.7	8.4
	13.0	9.9	8.3
	12.4	10.3	8.6
PROMEDIOS :	12.717	10.15	8.433

LA SUMA DE CUADRADOS DENTRO DE GRUPOS SERA:

$$(13.2-12.717)^2+(12.3-12.717)^2+(12.5-12.717)^2+(12.9-12.717)^2+(13-12.717)^2+$$
$$+(12.4-12.717)^2+(10.6-10.15)^2+(9.8-10.15)^2+(9.6-10.15)^2+(10.7-10.15)+$$
$$+(9.9-10.15)^2+(10.3-10.15)^2(8.5-8.433)^2+(8.9-8.433)^2+(7.9-8.433)^2+(8.4-8.433)^2$$
$$+(8.3-8.433)^2+(8.6-8.433)^2 = 2.23667$$

$$MSW = \frac{2.23667}{18-3} = 0.149, \quad S = 0.386$$

DE TABLAS, EL RANGO ESTUDENTIZADO ES; CON $k = 3$ Y $\nu = 15$: $q_{3,15,.025} = 3.67$

POR LO TANTO, EL MARGEN VALE: $\frac{3.67}{\sqrt{2}}\, 0.386\sqrt{\frac{1}{6} + \frac{1}{6}} = 0.578$

LAS DIFERENCIAS ENTRE MEDIAS SON LAS SIGUIENTES, INDICANDO LAS SIGNIFICATIVAS CON UN MARCO:

MUJER	A	B	C
MEDIA	12.717	10.15	8.433
DIFERENCIAS	*	$\boxed{2.567}$	$\boxed{4.284}$
		*	$\boxed{1.717}$
			*

TODAS LAS DIFERENCIAS SON SIGNIFICATIVAS

METODO DE DUNCAN

EL METODO DE DUNCAN, COMO EL DE TUKEY, SIRVE PARA EFECTUAR COMPARACIONES DE MEDIAS, NO OBSTANTE ESTE ES MAS CONSERVADOR QUE EL PRIMERO.

EL ERROR ESTANDAR DE CUALQUIER MEDIA ES: $S = \sqrt{\frac{MSW}{n}}$

DE LA TABLA DE DUNCAN PARA RANGOS SIGNIFICANTES OBTENEMOS $r_{\alpha}(p, f)$, DONDE α ES EL NIVEL DE SIGNIFICANCIA, $p = 2, 3, \ldots k$ SON LOS TRATAMIENTOS, CUYAS MEDIAS SE ORDENAN DE MENOR A MAYOR, f SON LOS GRADOS DE LIBERTAD DE SSW: (N-k). EL RANGO SE CALCULA COMO: $R_p = r_{\alpha}(p, f)S$, PARA $p = 2, 3, \ldots k$

PARA PROBAR LAS DIFERENCIAS, SE PRUEBA LA MAYOR CON LA MENOR, COMPARANDO CON EL MAYOR R_{α}, ASI SE CONTINUA COMPARANDO EL MAYOR CON LOS RESTANTES, EN ORDEN CRECIENTE ESTOS ULTIMOS. SE PROCEDE IGUALMENTE EN EL DE SEGUNDA IMPORTANCIA, ETC.

EN ESTE CASO, ORDENANDO LAS MEDIAS EN ORDEN CRECIENTE:

$$\overline{y}_C = 8.433,\ \overline{y}_B = 10.15,\ \overline{y}_A = 12.717$$

EL VALOR DE MSW ES = 0.149, POR LO QUE, EN CUALQUIER CASO: $S = \sqrt{\frac{0.149}{6}} = 0.1576$

EN TABLAS DEL METODO DE DUNCAN (DESIGN AND ANALYSIS OF EXPERIMENTS-MONTGOMERY-WILLEY INTERNATIONAL, 1976), CON $\alpha = 0.05$, $f = N-k=18-3=15$:

$$r_{0.05}(2,15) = 3.01,\ r_{.05}(3,15) = 3.16$$

POR LO TANTO, LOS MARGENES SERAN:

$$R_2 = 3.01(0.1576) = 0.474,\ R_3 = 3.16(0.1576) = 0.498$$

Y LAS COMPARACIONES DE MEDIAS SERAN:

VALORES CRITICOS EN LA PRUEBA DE DUNCAN DE RANGO MULTIPLE

Error df	α	p = NUMERO DE MEDIAS ADYACENTES 2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	.05	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
	.01	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	.05	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
	.01	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	.05	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
	.01	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3
4	.05	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
	.01	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5
5	.05	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
	.01	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8
6	.05	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
	.01	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.1	6.2	6.2	6.3	6.3
7	.05	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
	.01	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0
8	.05	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
	.01	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8
9	.05	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52	3.52
	.01	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7
10	.05	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48
	.01	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55
11	.05	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48
	.01	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39
12	.05	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48
	.01	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.23	5.26
13	.05	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47
	.01	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15
14	.05	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47
	.01	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07
15	.05	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00
16	.05	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47
	.01	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94
17	.05	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47
	.01	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89
18	.05	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47
	.01	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85
19	.05	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47
	.01	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82
20	.05	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47
	.01	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79
22	.05	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47
	.01	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75
24	.05	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47
	.01	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72
26	.05	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47
	.01	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69
28	.05	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47
	.01	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67
30	.05	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47
	.01	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65
40	.05	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47
	.01	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59
60	.05	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47
	.01	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53
100	.05	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47
	.01	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48
∞	.05	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47
	.01	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41

A VS. C: 12.717 - 8.433 = 4.284 > 0.498 (SIGNIFICATIVA)

A VS. B: 12.717 - 10.15 = 2.567 > 0.474 (SIGNIFICATIVA)

B VS. C: 10.15 - 8.433 = 1.717 > 0.474 (SIGNIFICATIVA)

COMO EN EL METODO DE TUKEY, TODAS LAS DIFERENCIAS SON SIGNIFICATIVAS.

METODO DE FISHER

PARA REALIZAR COMPARACIONES MULTIPLES ENTRE LAS MEDIAS DE DIVERSOS TRATAMIENTOS SE PUEDE USAR LA ESTADISTICA DE FISHER (ESTE METODO EN REALIDAD ES UNA MODIFICACION DE LA COMPARACION ENTRE MEDIAS CON LA t DE STUDENT).

LA DISTRIBUCION t SE DEFINE COMO:

$$t = \frac{y}{\sqrt{\frac{\mu}{\phi}}} \qquad \text{(a)}$$

DONDE y ES N(0,1) Y μ TIENE DISTRIBUCION χ^2 CON ϕ G. DE L.

SI QUEREMOS COMPARAR DOS MEDIAS:

$$y = \frac{(X_{1.} - X_{2.}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \text{(b)}$$

SI SE SUPONE $\sigma_1 = \sigma_2$:

$$y = \frac{(\overline{X}_{1.} - \overline{X}_{2.}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\overline{X}_{1.} - \overline{X}_{2.}) - (\mu_1 - \mu_2)}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Y

$$\mu = \sum^{n_1} \left(\frac{X_{1j} - \bar{X}_{1.}}{\sigma_1}\right)^2 + \sum^{n_2} \left(\frac{X_{2j} - \bar{X}_{2.}}{\sigma_2}\right)^2 + \ldots = \frac{1}{\sigma^2} \sum_i^k \sum_j^{n_i} (X_{ij} - \bar{X}_{i.})^2$$

QUE TIENE DISTRIBUCION χ^2 CON N-k G. DE L.

COMO $\quad SW = \sum^k \sum^{n_i} (X_{ij} - \bar{X}_{i.})^2$ CON N-k G. DE L.

$$E(MSW) = \sigma^2$$

SW ES LA VARIANCIA COMBINADA, ESTIMADOR INSESGADO DE σ^2, POR LO TANTO, EL DENOMINADOR DE (a) ES:

$$\frac{\mu}{\phi} = \frac{1}{\sigma^2} MSW \qquad \text{(c)}$$

SUSTITUYENDO (b) y (c) EN (a) SE OBTIENE, BAJO LA HIPOTESIS $\mu_1 = \mu_2$:

$$t = \frac{(\bar{X}_{1.} - \bar{X}_{2.}) - (\mu_1 - \mu_2)}{\sqrt{MSW}} \sqrt{\frac{n_1 \, n_2}{n_1 + n_2}}$$

Y COMO $F = t^2$ SE OBTIENE:

$$F_0 = \frac{(\bar{X}_{1.} - \bar{X}_{2.})^2}{MSW} \frac{n_1 n_2}{n_1 + n_2}$$

QUE COMPARADA CON F_c, CON 1 Y N-K G. DE L., NOS PERMITE SABER SI EXISTE DIFERENCIA SIGNIFICATIVA EN LAS MEDIAS.

PARA EFECTUAR CON MAYOR FACILIDAD COMPARACIONES MULTIPLES SE ACOSTUMBRA CALCULAR:

$$(\bar{X}_{1.} - \bar{X}_{2.})^2 = \frac{n_1 + n_2}{n_1 \, n_2} \, MSW \, F_0$$

Y COMPARAR CON EL TEORICO: $\frac{n_1 + n_2}{n_1\ n_2}$ MSW $F_{\alpha,1,N-K}$ (MARGEN)

CUANDO $(\bar{X}_{1.} - \bar{X}_{2.})^2$ ES MAYOR QUE EL MARGEN EXISTE UNA DIFERENCIA SIGNIFICATIVA.

EJEMPLO

EN EL EJEMPLO ANTERIOR, EL MARGEN EN CUALQUIER CASO VALE, CON $F_{0.05,1,15} = 4.54$

$$\frac{6 + 6}{6(6)} (0.149)(4.54) = 0.225$$

LAS DIFERENCIAS ENTRE MEDIAS SON LAS SIGUIENTES (AL CUADRADO), INDICANDO LAS SIGNIFICATIVAS CON UN MARCO .

MUJER	A	B	C
$\bar{X}$	12.717	10.15	8.433
$(\text{DIFERENCIAS})^2$	*	$\boxed{6.59}$	$\boxed{18.35}$
		*	$\boxed{2.95}$
			*

TODAS LAS DIFERENCIAS SON SIGNIFICATIVAS.

10. EXPERIMENTO DE CUADRADOS LATINOS

SUPONGAMOS QUE EL ENSAYO QUE SE LLEVA A CABO PARA DETERMINAR EL VALOR QUE TOMA CIERTA VARIABLE EN UNA UNIDAD DE EXPERIMENTACION (ESPECIMEN) TOMA UN TIEMPO RELATIVAMENTE LARGO, DIGAMOS UNA SEMANA, Y QUE CADA ANALISTA (EXPERIMENTADOR) SOLO PUEDE REALIZAR UN ENSAYO A LA VEZ.

SI SE USARA, POR EJEMPLO, UN EXPERIMENTO POR BLOQUES COMPLETAMENTE ALEATORIZADO CON TRES ANALISTAS Y TRES SEMANAS, PODRIA PRESENTARSE LA SIGUIENTE DISTRIBUCION DE LOS ENSAYES PARA LOS ESPECIMENES TIPOS A, B Y C:

SEMANA	ANALISTA		
	1	2	3
1	A	B	A
2	C	A	B
3	B	C	C

SI SE PROBARA LA HIPOTESIS NULA $H_0: \mu_A - \mu_B = 0$, EN CONTRA DE LA ALTERNATIVA $H_1: \mu_A \neq \mu_B$, Y SE RECHAZARA H_0, QUEDARIA LA DUDA DE SI EN ESTE RESULTADO INFLUIRIA EL HECHO DE QUE LA PRIMER SEMANA SE PROBARON DOS ESPECIMENES DE A Y SOLO UNO DE B, EN LA SEGUNDA UNO DE A Y UNO DE B Y, EN LA TERCERA, SOLO UNO DE B.

SI ESTA DUDA FUERA LEGITIMA, SERIA NECESARIO ELIMINAR (FILTRAR)

EL EFECTO DEL FACTOR "SEMANA". ADICIONALMENTE AL FILTRADO, ES NECESARIO RESTRINGIR NUESTRO PROCESO DE ALEATORIZACION DE TAL MANERA QUE QUEDE UN SOLO ESPECIMEN DE CADA TIPO EN CADA SEMANA, QUEDANDO UNA DISTRIBUCION DE LOS ENSAYES COMO LA SIGUIENTE

SEMANA	ANALISTA		
	1	2	3
1	A	B	C
2	B	C	A
3	C	A	B

EN ESTE CASO LA ALEATORIZACION CONSISTIRIA EN ASIGNAR AL AZAR CADA ESPECIMEN TIPO A, B O C A CADA PAREJA (SEMANA, ANALISTA) DE NIVELES DE LOS FACTORES.

A UN DISEÑO EXPERIMENTAL COMO ESTE SE LE DENOMINA "DISEÑO DE CUADRADOS LATINOS". SE USA CUANDO SE QUIEREN COMPARAR t MEDIAS DE TRATAMIENTOS, EN PRESENCIA DE DOS FUENTES EXTRAÑAS DE VARIABILIDAD, LAS CUALES SE BLOQUEAN EN t RENGLONES Y EN t COLUMNAS.

<u>DEFINICION</u>: UN DISEÑO EXPERIMENTAL DE CUADRADOS LATINOS txt, ES TAL QUE LOS t TRATAMIENTOS QUE SE DESEAN COMPARAR SE ASIGNAN AL AZAR ENTRE t RENGLONES Y t COLUMNAS, DE TAL FORMA QUE CADA TRATAMIENTO APARECE EN CADA RENGLON Y EN CADA COLUMNA.

EL MODELO PARA REPRESENTAR A CADA UNO DE LOS RESULTADOS,

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + Z_{ijk} \qquad (1)$$

DONDE Z_{ijk} SON VARIABLES ALEATORIAS NORMALES INDEPENDIENTES ENTRE SI CON MEDIA CERO Y VARIANCIA DESCONOCIDA, σ^2, CADA UNA. LOS TERMINOS α_i, β_j Y γ_k SON LOS EFECTOS DEL TRATAMIENTO i, EL RENGLON j Y LA COLUMNA k, RESPECTIVAMENTE, CON

$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$, Y μ ES LA MEDIA GLOBAL.

EN TAL CASO

$$E(X_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k \qquad (2)$$

$$\mathrm{Var}(X_{ijk}) = \sigma^2$$

LA DESCOMPOSICION DE LA SUMA DE CUADRADOS QUEDA EN LA FORMA SIGUIENTE:

$$TSS = SST + SSR + SSC + SSE \qquad (3)$$

DONDE TSS ES LA SUMA DE CUADRADOS TOTAL, SST LA DE LOS TRATAMIENTOS, SSC LA DE COLUMNAS, SSR LA DE RENGLONES Y SSE LA DEL ERROR. LAS ECUACIONES PARA CALCULAR A CADA UNA DE ELLAS SON:

$$TSS = \sum_i \sum_j (X_{ijk} - \overline{X}_{...})^2 = \sum_i \sum_j X^2_{ijk} - n\,\overline{X}^2_{...} \qquad (4)$$

$$SST = t \sum_i (\overline{X}_{i..} - \overline{X}_{...})^2 = t \sum_i \overline{X}^2_{i..} - n\,\overline{X}^2_{...} \qquad (5)$$

$$SSR = t \sum_j (\overline{X}_{.j.} - \overline{X}_{...})^2 = t \sum_j \overline{X}^2_{.j.} - n\,\overline{X}^2_{...} \qquad (6)$$

$$SSC = t \sum_k (\bar{X}_{..k} - \bar{X}_{...})^2 = t \sum_k \bar{X}^2_{..k} - n \bar{X}^2_{...} \qquad (7)$$

$$SSE = TSS - SST - SSR - SSC \qquad (8)$$

$$n = t^2$$

LA TABLA DEL ANALISIS DE VARIANCIA DE ESTE EXPERIMENTO ES

FUENTE	SS	G. DE L.	MS	F
TRATAMIENTOS	SST	t-1	MST=SST/(t-1)	MST/MSE
RENGLONES	SSR	t-1	MSR=SSR/(t-1)	MSR/MSE
COLUMNAS	SSC	t-1	MSC=SSC/(t-1)	MSC/MSE
ERROR	SSE	(t-1)(t-2)	MSE=SSE/(t-1)(t-2)	
TOTALES	TSS	t^2-1		

CON ESTAS ESTADISTICAS F SE PRUEBAN, RESPECTIVAMENTE, LAS HIPOTESIS:

a) $H_0: \alpha_i = 0 \quad i = 1, 2, \ldots, n$

H_1: AL MENOS UNA α_i NO ES CERO

b) $H_0: \beta_j = 0, \quad j = 1, 2, \ldots, t$

H_1: AL MENOS UNA β_j NO ES CERO

c) $H_0: \gamma_k = 0, \quad k = 1, 2, \ldots, t$

H_1: AL MENOS UNA γ_k NO ES CERO

ESTAS PRUEBAS DE HIPOTESIS SON TAMBIEN PARA EL CASO DE NIVELES

ALEATORIOS.

EJEMPLO

EN UN PROBLEMA DE INGENIERIA DE TRANSITO SE DESEAN COMPARAR LOS TIEMPOS EN QUE NO SE APROVECHA LA LUZ VERDE DEL SEMAFORO POR NO PASAR NINGUN VEHICULO, PARA 4 DISPOSITIVOS DE CONTROL AUTOMATICO DE SEMAFOROS EN 4 CRUCEROS DIFERENTES DE LA CIUDAD, LO SUFICIENTEMENTE DISTANTES ENTRE SI COMO PARA CONSIDERARSE INDEPENDIENTES. PARA ESTO, SE DISEÑO UN EXPERIMENTO EN EL QUE SE MIDIERON LOS TIEMPOS DE DESPERDICIO, EN MINUTOS, QUE SE TUVIERON EN CUATRO HORAS DIFERENTES DEL DIA, DOS HORAS "PICO", Y DOS HORAS "VALLE" DEL DIA, CON LO CUAL SE INTEGRO EL SIGUIENTE EXPERIMENTO DE CUADRADOS LATINOS 4x4:

INTERSECCION	HORA DEL DIA				TOTALES	$\bar{X}_{.j.}$
	A. M. PICO	A. M. VALLE	P. M. VALLE	P. M. PICO		
1	D(15.5)	B(33.9)	C(13.2)	A(29.1)	91.7	22.92
2	B(16.3)	C(26.6)	A(19.4)	D(22.8)	85.1	21.27
3	C(10.8)	A(31.1)	D(17.1)	B(30.3)	89.3	22.32
4	A(14.7)	D(34.0)	B(19.7)	C(21.6)	90.0	22.50
TOTALES	57.3	125.6	69.4	103.8	356.1	
$\bar{X}_{..k}$	14.33	31.40	17.35	25.95		

EN ESTA TABLA LAS CIFRAS ENTRE PARENTESIS SON MINUTOS DE DESPERDICIO POR HORA PARA LOS DISPOSITIVOS A, B, C Y D.

LOS PROMEDIOS PARA CADA DISPOSITIVO SON:

$\overline{X}_{A..} = 94.3/4 = 23.58$; $\overline{X}_{C..} = 72.2/4 = 18.05$

$\overline{X}_{B..} = 100.2/4 = 25.05$; $\overline{X}_{D..} = 89.4/4 = 22.35$

$\overline{X}_{...} = 356.1/16 = 22.26$; $16\overline{X}^2_{...} = 7925.45$

$\overline{X}_{.1.} = 91.7/4=22.92$; $\overline{X}_{.2.} = 85.1/4;21.27$; $\overline{X}_{.3.}=89.3/4=22.32$;

$\overline{X}_{.4.}=90.0/4=22.50$; $\overline{X}_{..1}=57.3/4=14.33$; $\overline{X}_{..2}=125.6/4=31.40$;

$\overline{X}_{..3}=69.4/4=17.35$; $\overline{X}_{..4}=103.8/4=25.95$.

$SST=4(23.58^2+25.05^2+18.05^2+22.35^2)-7925.45=$

$=4(555.78+627.50+325.80+499.52)-7925.45=8034.41-7925.45=108.96$

$SSR=4(22.92^2+21.27^2+22.32^2+22.5^2)-7925.45=$

$=4(525.56+452.63+498.41+506.25)-7925.45=7931.40-7925.45=5.95$

$SSC=4(205.21+985.96+301.02+673.40)-7925.45=8662.36-7925.45=736.91$

$TSS=15.5^2+16.3^2+10.8^2+14.7^2+33.9^2+...+21.6^2-7925.45=8801.05-7925.45=875.6$

$SSE=875.6-108.96-5.95-736.91=23.78$

FUENTE	SS	G. DE L.	MS	F
DISPOSITIVOS (TRATAMIENTOS)	108.96	3	36.32	9.17 > 4.76
RENGLONES (INTERSECCIONES)	5.95	3	1.98	0.50 < 4.76
COLUMNAS (HORAS DEL DIA)	736.91	3	245.64	61.87 > 4.76
ERROR	23.78	6	3.96	
TOTALES	875.60	15		

$F_{0.95,3,6}=4.76$

DE LO ANTERIOR SE CONCLUYE QUE SI HAY DIFERENCIAS SIGNIFICATIVAS ENTRE LOS DISPOSITIVOS Y ENTRE LAS HORAS DEL DIA, A UN 95 POR CIENTO DE NIVEL DE CONFIANZA.

EXPERIMENTOS DE
CUADRADOS LATINOS
CON REPLICAS

CON FRECUENCIA SE DISPONE DE TIEMPO Y RECURSOS PARA TENER VARIAS REPLICAS DE UN EXPERIMENTO DE CUADRADOS LATINOS, PRINCIPALMENTE CUANDO t ES PEQUEÑO. SUPONGAMOS QUE SE EJECUTAN $\underline{r}$ REPLICAS, EL MODELO MATEMATICO SERA, EN ESTE CASO:

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \rho_l + Z_{ijkl} \qquad (1)$$

EN DONDE ρ_l ES EL EFECTO DE LA L-ESIMA REPLICA, i, j Y k = 1, 2, ..., t, Y l = 1, 2, ..., r; LOS DEMAS TERMINOS TIENEN EL MISMO SIGNIFICADO QUE EN EL EXPERIMENTO SIN REPLICAS. LAS RESTRICCIONES DE LOS PARAMETROS SON

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = \sum_l \rho_l = 0 \qquad (2)$$

LA SUMA TOTAL DE CUADRADOS, EN ESTE CASO, SE DESCOMPONE DE LA SIGUIENTE MANERA:

$$TSS = SST + SSR + SSC + SSRe + SSE \qquad (3)$$

EN DONDE

$$TSS = \sum_i \sum_j \sum_l X^2_{ijkl} - N\bar{X}^2_{....}; \; N = t^2 r \qquad (4)$$

$$SST = rt \sum_j \bar{X}^2_{i...} - N\bar{X}^2_{....} \qquad (5)$$

$$SSR = rt \sum_j \bar{X}^2_{.j..} - N\bar{X}^2_{....} \qquad (6)$$

$$SSC = rt \sum_{k} \bar{X}^2_{..k.} - N\bar{X}^2_{....} \tag{7}$$

$$SSRe = t^2 \sum_{l} \bar{X}^2_{...l} - N\bar{X}^2_{....} \tag{8}$$

$$SSE = TSS - SST - SSR - SSC - SSRe \tag{9}$$

LA TABLA DEL ANALISIS DE VARIANCIA CORRESPONDIENTE A ESTE EXPERIMENTO ES:

FUENTE DE VARIABILIDAD	G. de l	SS	MS	F
TRATAMIENTOS	t - 1	SST	MST=SST/(t-1)	MST/MSE
RENGLONES	t - 1	SSR	MSR=SSR/(t-1)	MSR/MSE
COLUMNAS	t - 1	SSC	MSC=SSC/(t-1)	MSC/MSE
REPLICAS	r - 1	SSRe	MSRe=SSRe/(r-1)	MSRe/MSE
ERROR	g=(t-1)(rt+r-3)	SSE	MSE=SSE/g	
TOTAL	$rt^2 - 1$	TSS		

EJEMPLO

SE TIENE UN PROCESO DE FABRICACION EN EL CUAL SE RECUBRE UNA LAMINA CON UN CIERTO METAL. EXISTE LA DUDA DE SI EL ESPESOR DE ESE RECUBRIMIENTO CAMBIA EN LAS DIRECCIONES DEL ROLADO Y TRANSVERSAL A EL. PARA ESTUDIAR ESTO SE TOMO COMO VARIABLE AL PESO POR UNIDAD DE AREA QUE SE TENGA DE DICHO RECUBRIMIENTO. PARA ELIMINAR ESTAS DOS FUENTES DE VARIACION, CADA UNA DE 2 PLACAS FABRICADAS SE DIVIDIO EN 16 PARTES REPRESENTANDO 4 POSICIONES EN DIRECCION LONGITUDINAL Y 4 TRANSVERSALES AL ROLADO, Y LUEGO SE TOMARON 4 MUESTRAS DE CADA UNA Y SE MANDARON A LOS LABORATORIOS A, B, C Y D PARA DETERMINAR EL PESO DEL RECUBRIMIENTO, TENIENDOSE LOS SIGUIENTES RESULTADOS:

TRANSVERSAL

	FACTOR	2.1	2.2	2.3	2.4	2.1	2.2	2.3	2.4	TOTALES	$\bar{X}_{.j..}$
LONGITUDINAL	1.1	$^{B}0.29$	$^{A}0.25$	$^{C}0.18$	$^{D}0.28$	$^{C}0.20$	$^{A}0.24$	$^{D}0.20$	$^{B}0.27$	1.91	0.239
	1.2	$^{D}0.28$	$^{B}0.18$	$^{A}0.21$	$^{C}0.25$	$^{B}0.28$	$^{C}0.19$	$^{A}0.22$	$^{D}0.28$	1.89	0.236
	1.3	$^{C}0.28$	$^{D}0.23$	$^{B}0.20$	$^{A}0.28$	$^{D}0.34$	$^{B}0.23$	$^{C}0.21$	$^{A}0.28$	2.05	0.256
	1.4	$^{A}0.30$	$^{C}0.19$	$^{D}0.24$	$^{B}0.25$	$^{A}0.32$	$^{D}0.22$	$^{B}0.16$	$^{C}0.27$	1.95	0.244
										7.80	

1.1	$^{C}0.20$	$^{A}0.24$	$^{D}0.20$	$^{B}0.27$
1.2	$^{B}0.28$	$^{C}0.19$	$^{A}0.22$	$^{D}0.28$
1.3	$^{D}0.34$	$^{B}0.23$	$^{C}0.21$	$^{A}0.28$
1.4	$^{A}0.32$	$^{D}0.22$	$^{B}0.16$	$^{C}0.27$
TOTALES	2.29	1.730	1.620	2.160
$\bar{X}_{..k.}$	0.286	0.216	0.203	0.27

VERIFICAR LAS HIPOTESIS DE EFECTOS NULOS Y SI HAY ALGUNA QUE NO LA CUMPLA, HACER LA PRUEBA DE COMPARACIONES MULTIPLES.

SOLUCION

a) ANALISIS DE VARIANCIA

$$\bar{X}_{...1} = (0.29 + 0.28 + \dots + 0.28 + 0.25)/16 = 0.243$$
$$\bar{X}_{...2} = (0.20 + 0.28 + \dots + 0.28 + 0.27)/16 = 0.244$$
$$\bar{X}_{A...} = (0.25 + 0.21 + \dots + 0.28 + 0.32)/8 = 0.263$$
$$\bar{X}_{B...} = (0.29 + 0.18 + \dots + 0.23 + 0.16)/8 = 0.233$$
$$\bar{X}_{C...} = (0.18 + 0.25 + \dots + 0.21 + 0.27)/8 = 0.221$$
$$\bar{X}_{D...} = (0.28 + 0.28 + \dots + 0.34 + 0.22)/8 = 0.259$$
$$\bar{X}_{....} = \frac{0.29 + 0.28 + 0.28 + \dots + 0.28 + 0.28 + 0.27}{32} = 0.244$$

TOTALES: $TSS = \sum_j \sum_k \sum_\ell X^2_{ijk\ell} - N\bar{X}^2_{....} = 1.9628 - 32 \times 0.244^2 =$
$= 1.9628 - 1.905 = 0.058$

RENGLONES: $SSR = rt \sum_j \bar{X}^2_{.j..} - N\bar{X}^2_{....} = 2 \times 4 \times (0.239^2 + 0.236^2 +$
$+ 0.256^2 + 0.244^2) - 1.905 = 0.002$

COLUMNAS: $SSC = rt \sum_k \bar{X}^2_{..k.} - N\bar{X}^2_{....} = 8(0.286^2 + 0.216^2 + 0.203^2 +$
$+ 0.27^2) - 1.905 = 0.035$

REPLICA: $SSRe = t^2 \sum_l \bar{X}^2_{...l} - N\bar{X}^2_{....} = 16(0.243^2 + 0.244^2) - 1.905 =$
$= 0.008$

TRATAMIENTOS: $SST = rt \sum \bar{X}^2_{i...} - N\bar{X}^2_{....} = 8(0.263^2 + 0.233^2 +$
$+ 0.221^2 + 0.259^2) - 1.905 = 0.010$

ERROR: $SSE = TSS - SST - SSC - SSR - SSRe = 0.058 - 0.002 - 0.035 - 0.008$
$- 0.01 = 0.003$

TABLA DEL ANALISIS DE VARIANCIA

FUENTE	g. de l.	SS	MS	F_E		$F_{c(\alpha=0.05)}$
LONG. AL ROLADO	3	SSR=0.002	0.0007	5.00	>	3.07
TRANSV. AL ROLADO	3	SSC=0.035	0.0117	83.57	>	3.07
LABORATORIOS	3	SST=0.010	0.0033	23.57	>	3.07
REPLICAS	1	SSRe=0.008	0.008	57.14	>	4.32
ERROR	3(7)=21	SSE=0.003	0.00014			
TOTAL	31	TSS=0.058				

DE LO ANTERIOR SE CONCLUYE QUE:

1. SI HAY EFECTOS EN LA LONGITUD AL ROLADO
2. SI HAY EFECTOS ENTRE REPLICAS
3. SI HAY EFECTOS ENTRE LOS DIFERENTES LABORATORIOS
4. SI HAY EFECTOS EN LA DIRECCION TRANSVERSAL AL ROLADO

b) PRUEBA DE COMPARACIONES MULTIPLES

b-1) ENTRE LOS DIFERENTES LABORATORIOS:

LABORATORIO	C	B	D	A
MEDIA	0.221	0.233	0.259	0.263

DE LAS TABLAS PARA $\alpha = 0.05$, 21 G de L. y P = 2,3,4, TENEMOS (INTERPOLANDO)

p	2	3	4
r_p	2.9425	3.0925	3.1825

$$S_{\bar{x}} = \sqrt{\frac{MSE}{n}} = \sqrt{\frac{0.00014}{8}} = 0.00418$$

p	2	3	4
$R_p = r_p S_{\bar{x}}$	0.01231	0.01293	0.01330

EL RANGO PARA LAS 4 MEDIAS ES $R_4 = 0.042 > R_{4c}(0.01330)$, LO CUAL ERA DE ESPERARSE YA QUE LA PRUEBA F MOSTRO QUE SI HABIA EFECTO ENTRE LOS 4 TRATAMIENTOS.

LOS RANGOS PARA 3 MEDIAS ADYACENTES SON:

$$CBD = 0.259 - 0.221 = 0.038 > 0.01293$$

$$BDA = 0.263 - 0.233 = 0.030 > 0.01293$$

LOS RANGOS PARA PARES DE MEDIAS ADYACENTES SON:

$$CB = 0.233 - 0.221 = 0.012 < 0.01231$$

$$BD = 0.259 - 0.233 = 0.026 > 0.01231$$

$$DA = 0.263 - 0.259 = 0.0040 < 0.01231$$

POR LO TANTO TENDREMOS: $\underline{C \quad B} \quad \underline{D \quad A}$

DE DONDE CONCLUIMOS QUE LOS LABORATORIOS C Y B, ASI COMO D Y A TUVIERON RESULTADOS CONSISTENTES, MIENTRAS LOS LABORATORIOS

B Y D PRESENTARON RESULTADOS DIFERENTES ESTADISTICAMENTE Y, POR ENDE, NO HABRA CONSISTENCIA ENTRE B Y A Y C Y D

b-2) EN LOS NIVELES DE LA DIRECCION TRANSVERSAL AL ROLADO:

NIVELES	3	2	4	1
MEDIAS	0.203	0.216	0.270	0.286

DE LAS TABLAS PARA $\alpha = 0.05$; 21 G. de L., p = 2,3,4 Y $S_{\bar{x}} = 0.00418$ TENEMOS:

p	2	3	4
r_p	2.948	3.0925	3.1825
$R_p = r_p \times S_{\bar{x}}$	0.01231	0.01293	0.01330

EL RANGO PARA LAS 4 MEDIAS ES R4 = 0.286 - 0.203 = 0.0830 > Rcrítico (0.01330), LO CUAL RATIFICA EL RESULTADO DE LA PRUEBA F DE QUE SI HAY EFECTO ENTRE LOS 4 NIVELES DEL ROLADO TRANSVERSAL.

PARA LOS CONJUNTOS DE 3 MEDIAS ADYACENTES:

$$R_{324} = 0.27 - 0.203 = 0.0670 > 0.01293$$

$$R_{241} = 0.286 - 0.216 = 0.07 > 0.01293$$

POR LO QUE TAMBIEN HAY EFECTO SIGNIFICATIVO ENTRE LAS TRIPLETAS DE MEDIAS ADYACENTES. PARA LOS CONJUNTOS DE 2 MEDIAS ADYACENTES:

$R_{32} = 0.216 - 0.203 = 0.0130 > 0.01231$

$R_{24} = 0.27 - 0.216 = 0.0540 > 0.01231$

$R_{41} = 0.286 - 0.27 = 0.0160 > 0.01231$

DE DONDE CONCLUIMOS QUE:

- FACTOR 2: N3 N2 N4 N1

EN LA DIRECCION TRANSVERSAL DEL ROLADO NINGUNA PAREJA DE NIVELES DIO RESULTADOS CONSISTENTES.

b-3) APLICANDO EL METODO DE FISHER DE COMPARACIONES MULTIPLES TENEMOS:

b-3.1) TRATAMIENTOS: $LSD = t_{21,\alpha/2} \cdot \sqrt{\frac{2MSE}{n}} = t_{21,0.025} \sqrt{\frac{2 \times 0.00014}{8}}$

$= 2.080 \times 0.0059 = 0.01231$

LABORATORIOS	C	B	D	A
MEDIAS	0.221	0.233	0.259	0.263
	*	0.0120	$\boxed{0.038}$	$\boxed{0.042}$
		*	$\boxed{0.026}$	$\boxed{0.03}$
			*	0.004

<u>C B</u> <u>D A</u>, QUE COINCIDE CON EL RESULTADO ANTERIOR

b-3.2) A NIVELES DEL ROLADO TRANSVERSAL: 3 2 4 1, QUE COINCIDE CON EL RESULTADO ANTERIOR

NIVELES	3	2	4	1
MEDIAS	0.203	0.216	0.27	0.286
	*	$\boxed{0.013}$	$\boxed{0.067}$	$\boxed{0.083}$
		*	$\boxed{0.054}$	$\boxed{0.07}$
			*	$\boxed{0.016}$

EJEMPLO

PARA EL EJERCICIO QUE SE DESARROLLO EN LA CLASE SOBRE FUNDENTES TENEMO

a) APLICANDO DUNCAN:

PARA LOS METODOS:

METODO	C	B	A
MEDIA	11.0	14.4	14.6

PARA $\alpha = 0.01$, $\nu = 10$; $p = 2,3$ TENEMOS

p	2	3
r_p	4.48	4.67
$R_p = r_p \times S_{\bar{x}} =$	2.1485	2.2397

$$S_{\bar{x}} = \sqrt{\frac{MSE}{n}} = \sqrt{\frac{1.38}{6}} = 0.4795$$

EL RANGO PARA 3 MEDIAS ADYACENTES = $\bar{X}_A - \bar{X}_C = 14.6 - 11 =$ 3.6 > 2.397 LO QUE SE VERIFICO EN LA PRUEBA F.

LOS RANGOS PARA LOS CONJUNTOS DE 2 MEDIAS ADYACENTES SON

$\bar{X}_B - \bar{X}_C = 14.4 - 11 = 3.40 > 2.1485 \therefore$ SIGNIFICATIVO

$\bar{X}_A - \bar{X}_B = 14.6 - 14.4 = 0.2 < 2.1485 \therefore$ NO SIGNIFICATIVO

LO CUAL IMPLICA QUE C $\underline{B\ A}$; EL METODO C ES EL QUE PRODUCE EFECTOS ESTADISTICAMENTE SIGNIFICATIVOS.

b) PARA LAS FUNDENTES:

FUNDENTE	1	3	2
MEDIA	11.6	13	15.33

EL RANGO PARA LAS TRES MEDIAS ADYACENTES: $R_{132} = 15.33 - 11.6 =$ $= 3.73 > 2.2397$ O SEA QUE SI HAY EFECTO ENTRE LOS 3 FUNDENTES COMO SE HABIA VISTO EN LA PRUEBA F. PARA LOS CONJUNTOS DE 2 MEDIAS:

$$\bar{X}_3 - \bar{X}_1 = 1.40 < 2.1485 \therefore \text{ NO SIGNIFICATIVO}$$

$$\bar{X}_2 - \bar{X}_3 = 2.33 > 2.1485 \therefore \text{ SI SIGNIFICATIVO}$$

ENTONCES: FUNDENTES $\underline{1\ 3}$ 2, POR LO QUE EL FUNDENTE 2 PRODUCE EFECTOS DIFERENTES ESTADISTICAMENTE SIGNIFICATIVOS.

APLICANDO FISHER:

a) PARA LOS METODOS

$$LSD = t_{0.005,10}\sqrt{\frac{2 \times 1.38}{6}} = 3.169 \times 0.6782 = 2.1493$$

METODOS	C	B	A
MEDIAS	11.0	14.4	14.6
	*	$\boxed{3.4}$	$\boxed{3.60}$
		*	0.20

C $\underline{B\ A}$

PARA LOS FUNDENTES

FUNDENTES	1	3	2
MEDIAS	11.6	B	15.33
	*	1.4	$\boxed{3.73}$
		*	$\boxed{2.33}$

1 3 2

11. EXPERIMENTO DE CUADRADOS GRECO-LATINOS

EN OCASIONES SE CONSIDERA QUE EXISTEN NO SOLO DOS SINO TRES FACTORES EXTRAÑOS QUE PUEDEN INFLUIR EN LOS RESULTADOS DE UN TRATAMIENTO, COMO SUCEDE EN EL EXPERIMENTO DE CUADRADOS LATINOS; CUANDO ESTO SUCEDE, SE PUEDE FILTRAR O AISLAR EL EFECTO DEL TERCER FACTOR MEDIANTE EL EMPLEO DE UN EXPERIMENTO DE CUADRADOS GRECO-LATINOS $t \times t$.

EN ESTE TIPO DE EXPERIMENTO LOS t NIVELES DEL TERCER FACTOR SE REPRESENTAN USUALMENTE CON LETRAS GRIEGAS, LAS CUALES SE COMBINAN CON LAS LATINAS QUE REPRESENTAN LOS t NIVELES DEL TRATAMIENTO, DE TAL MANERA QUE CADA LETRA LATINA APARECE SOLO UNA VEZ EN CONJUNCION CON UNA GRIEGA EN CADA COLUMNA Y EN CADA RENGLON.

POR EJEMPLO, EN UN EXPERIMENTO DE CUADRADOS GRECO-LATINOS DE 4x4 LAS LETRAS SE COMBINAN DE LA SIGUIENTE MANERA:

FACTOR 1	FACTOR 2			
	1	2	3	4
1	$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
2	$B\delta$	$A\gamma$	$D\beta$	$C\alpha$
3	$C\beta$	$D\alpha$	$A\delta$	$B\gamma$
4	$D\gamma$	$C\delta$	$B\alpha$	$A\beta$

UN EJEMPLO EN EL QUE SE USARIA UN EXPERIMENTO DE ESTE TIPO SERIA EL CASO DEL PROBLEMA MENCIONADO EN LOS CUADRADOS LATINOS

SI ADEMAS DE LOS FACTORES "OPERARIO" Y "FUNDENTE", SE AGREGARA EL DE "TEMPERATURA" DE LA SOLDADURA.

EL MODELO MATEMATICO PARA REPRESENTAR A CADA RESULTADO DEL EXPERIMENTO ES UNA EXTENSION NATURAL DEL DE CUADRADOS LATINOS:

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \lambda_k + \gamma_l + Z_{ijkl} \qquad (1)$$

DONDE λ_k Y γ_l REPRESENTAN AHORA LOS EFECTOS DE LOS FACTORES REPRESENTADOS POR LAS LETRAS LATINAS Y GRIEGAS, RESPECTIVAMENTE.

POR SU PARTE, LA SEPARACION DE LA SUMA DE CUADRADOS QUEDA EN LA FORMA

$$TSS = SSR + SSC + SSL + SSG + SSE \qquad (2)$$

EN DONDE

$$TSS = \sum_i \sum_j X^2_{ijkl} - t^2\, \bar{X}^2_{....} \qquad (3)$$

$$SSR = t \sum_i \bar{X}^2_{i...} - t^2\, \bar{X}^2_{....} \qquad (4)$$

$$SSC = t \sum_j \bar{X}^2_{.j..} - t^2\, \bar{X}^2_{....} \qquad (5)$$

$$SSL = t \sum_k \bar{X}^2_{..k.} - t^2\, \bar{X}^2_{....} \qquad (6)$$

$$SSG = t \sum_l \bar{X}^2_{...l} - t^2\, \bar{X}^2_{....} \qquad (7)$$

$$SSE = TSS - SSR - SSC - SSL - SSG \quad (8)$$

DE ESTA MANERA LA TABLA DEL ANALISIS DE VARIANCIA CORRESPONDIENTE ES:

FUENTE	G. DE L.	SS	MS	F
FACTOR I (RENGLONES	t-1	SSR	MSR=SSR/(t-1)	MSR/MSE
FACTOR II (COLUMNAS)	t-1	SSC	MSC=SSC/(t-1)	MSC/MSE
FACTOR III (LETRAS LATINAS)	t-1	SSL	MSL=SSL/(t-1)	MSl/MSE
FACTOR IV (LETRAS GRIEGAS)	t-1	SSG	MSG=SSG/(t-1)	MSG/MSE
ERROR O RESIDUAL	(t-1)(t-3)	SSE	MSE=SSE/(t-1)(t-3)	
TOTAL	t^2-1			

EN ESTE EXPERIMENTO LAS ESTADISTICAS F TIENEN t-1 Y (t-1)(t-3) GRADOS DE LIBERTAD EN EL NUMERADOR Y EN EL DENOMINADOR, RESPECTIVAMENTE.

PUESTO QUE EL ERROR TIENE (t-1)(t-3) GRADOS DE LIBERTAD, PARA t=3 SE TIENE G. DE L.=0, POR LO CUAL NO SE PUEDE HACER EL ANALISIS DE VARIANCIA.

EJEMPLO

EN UN PROBLEMA DE LA INDUSTRIA QUIMICA SE SOSPECHO QUE EN LOS RESULTADOS DE UN ENSAYE INFLUIAN CUATRO FACTORES: CONCENTRACION

DE LA SUBSTANCIA, VOLUMEN USADO, TAMAÑO DE ESPECIMEN Y TIEMPO DE LA REACCION, POR LO QUE SE DISEÑO UN EXPERIMENTO DE CUADRADOS GRECO-LATINOS PARA VERIFICAR ESTADISTICAMENTE CUALES DE ELLOS EFECTIVAMENTE INFLUIAN DE MANERA DIFERENTE AL CAMBIAR SUS RESPECTIVOS NIVELES. LOS RESULTADOS QUE SE OBTUVIERON TOMANDO 5 NIVELES DE LOS FACTORES FUERON LOS SEÑALADOS EN LA TABLA SIGUIENTE (LAS LETRAS LATINAS SON LOS NIVELES DEL FACTOR TAMAÑO):

FACTOR I (CONCENTRACION)	FACTOR II (VOLUMEN)					TOTALES	$\bar{X}_{i...}$
	1	2	3	4	5		
1	$A\alpha$ 65	$B\gamma$ 82	$C\varepsilon$ 108	$D\beta$ 101	$E\delta$ 126	482	96.4
2	$B\beta$ 84	$C\delta$ 109	$D\alpha$ 73	$E\gamma$ 97	$A\varepsilon$ 83	446	89.2
3	$C\gamma$ 105	$D\varepsilon$ 129	$E\beta$ 89	$A\delta$ 89	$B\alpha$ 52	464	92.8
4	$D\delta$ 119	$E\alpha$ 72	$A\gamma$ 76	$B\varepsilon$ 117	$C\beta$ 84	468	93.8
5	$E\varepsilon$ 97	$A\beta$ 59	$B\delta$ 94	$C\alpha$ 78	$D\gamma$ 106	434	86.8
TOTALES	470	451	440	482	451	2294	
$\bar{X}_{.j..}$	94.0	90.2	88.0	96.4	90.2	$\bar{X}_{....} = \frac{2294}{25} = 91.76$	

$\Sigma X_{..A.} = 372$, $\Sigma X_{..B.} = 429$, $\Sigma X_{..C.} = 484$, $\Sigma X_{..D.} = 528$, $\Sigma X_{..E.} = 481$

$\bar{X}_{..A.} = \frac{372}{5} = 74.4$, $\bar{X}_{..B.} = \frac{429}{5} = 85.8$, $\bar{X}_{..C.} = \frac{484}{5} = 96.8$,

$\bar{X}_{..D.} = \frac{528}{5} = 105.6$, $\bar{X}_{..E.} = \frac{481}{5} = 96.2$

$\Sigma X_{...\alpha} = 377,\ \Sigma X_{...\beta} = 398,\ \Sigma X_{...\gamma} = 466,\ \Sigma X_{...\delta} = 537,\ \Sigma X_{...\epsilon} = 534$

$\bar{X}_{...\alpha} = \frac{377}{5} = 75.4,\ \bar{X}_{...\beta} = \frac{398}{5} = 79.6,\ \bar{X}_{...\gamma} = \frac{466}{5} = 93.2,$

$\bar{X}_{...\delta} = \frac{537}{5} = 107.4,\ \bar{X}_{...\epsilon} = \frac{534}{5} = 106.8,\ t^2\,\bar{X}^2_{....} = 25x91.76^2 = 210{,}497.44$

$SSR = 5(96.4^2+89.2^2+92.8^2+93.6^2+86.8^2)-210{,}497.44 = 227.76$

$SSC = 5(94.0^2+90.2^2+88.0^2+96.4^2+90.2^2)-210{,}497.44 = 285.76$

$TSS = 65^2+82^2+108^2+...+106^2-210{,}497.44 = 9880.56$

$SSL = 5(74.4^2+85.8^2+96.8^2+105.6^2+96.2^2)-210{,}497.44 = 2867.76$

$SSG = 5(75.4^2+79.6^2+93.2^2+107.4^2+106.8^2)-210{,}497.44 = 5536.56$

$SSE = 9880.56-227.76-285.76-2867.76-5536.76 = 962.72$

LA TABLA DEL ANALISIS DE VARIANCIA RESULTANTE ES:

FUENTE	SS	G. DE L.	MS	F
CONCENTRACION	227.76	4	56.94	0.47 < 3.84
VOLUMEN	285.76	4	71.44	0.59 < 3.84
TAMAÑO	2867.76	4	716.94	5.96 > 3.84
TIEMPO	5536.76	4	1384.14	11.50 > 3.84
ERROR	962.72	8	120.34	
TOTAL	9880.56	24		

$F_{0.95,4,8} = 3.84$ (PARA $\alpha = 0.05$)

DEL ANALISIS DEL EXPERIMENTO ANTERIOR SE CONCLUYE QUE LOS FACTORES "CONCENTRACION" Y "VOLUMEN" NO INFLUYEN SIGNIFICATIVAMENTE EN LOS RESULTADOS A UN 95 POR CIENTO DE NIVEL DE CONFIANZA Y, EN CAMBIO, LOS FACTORES "TAMAÑO" Y "TIEMPO" SI INFLUYEN.

EJEMPLO

LOS FOCOS DE UNAS CAMARAS FOTOGRAFICAS FUERON COMPARADAS CON 5 CAMARAS, 5 TIPOS DE PELICULA Y 5 TIPOS DE FILTROS (DENOTADOS $\alpha, \ldots, \varepsilon$). DOS DUPLICADOS FUERON TOMADOS PARA CADA COMBINACION DE LOS 4 FACTORES OBTENIENDOSE LOS SIGUIENTES DATOS:

PELICULA	CAMARA 1	CAMARA 2	CAMARA 3	CAMARA 4	CAMARA 5	$\bar{X}_{i....}$
1	(Aα) 0.64, 0.66, $\bar{X}_{ij...}$ = 0.65	(Bγ) 0.70, 0.74, 0.72	(Cε) 0.73, 0.69, 0.71	(Dβ) 0.66, 0.66, 0.66	(Eδ) 0.66, 0.64, 0.65	0.6780
2	(Bβ) 0.62, 0.64, 0.63	(Cδ) 0.63, 0.61, 0.62	(Dα) 0.69, 0.67, 0.68	(Eγ) 0.70, 0.72, 0.71	(Aε) 0.78, 0.76, 0.77	0.6820
3	(Cγ) 0.65, 0.64, 0.645	(Dε) 0.72, 0.73, 0.725	(Eβ) 0.68, 0.68, 0.68	(Aδ) 0.64, 0.65, 0.645	(Bα) 0.74, 0.70, 0.72	0.6830
4	(Dδ) 0.64, 0.63, 0.635	(Eα) 0.73, 0.72, 0.725	(Aγ) 0.68, 0.70, 0.69	(Bε) 0.74, 0.74, 0.74	(Cβ) 0.72, 0.75, 0.735	0.7050
5	(Eε) 0.74, 0.74, 0.74	(Aβ) 0.73, 0.71, 0.725	(Bδ) 0.67, 0.66, 0.665	(Cα) 0.74, 0.75, 0.745	(Dγ) 0.78, 0.78, 0.78	0.73
$\bar{X}_{.j...}$	0.66	0.702	0.685	0.70	0.731	

a) DETERMINE LA VARIANCIA RESIDUAL

b) QUE EFECTOS SON SIGNIFICANTES? (NOTA: LOS DUPLICADOS SE CORRIERON AL MISMO TIEMPO. ENTONCES ESTOS PUEDEN NO SER UNA MEDICION VERDADERA DEL ERROR).

c) DETERMINE UN INTERVALO DE CONFIANZA DEL 95% PARA LA DENSIDAD MEDIA DE LA CAMARA # 5

SOLUCION

$\bar{X}_{.....} = 34.78/50 = 0.6956; \quad \bar{X}^2_{.....} = 0.483859$

TOTALES: $TSS = \sum_{ijklr}\sum\sum\sum\sum X^2_{ijklr} - t^2 r\bar{X}^2_{.....} = (0.64^2 + 0.66^2 + 0.62^2 +$

$+ 0.64^2 + ... + 0.72^2 + 0.75^2 + 0.78^2 + 0.78^2) -$

$- 5^2 \times 2 \times 0.483859 = 24.298400 - 2 \times 25 \ (34.78/50)^2$

$= 24.298400 - 12.096484 \times 2 = 0.105432$

FACTOR I (PELICULAS): $SSR = (t \sum_i \bar{X}^2_{i....} - t^2 \bar{X}^2_{.....})r$

$= [5(0.459684 + 0.465124 + 0.466489 +$

$0.497025 + 0.532900) - 12.096484 \times 2$

$= 5 \times 2.421222 - 12.096484] \times 2 =$

$= (0.009626) \times 2 = 0.019252$

FACTOR II (CAMARAS): $SSC = (t \sum_j \bar{X}^2_{.j...} - t^2\bar{X}^2_{.....})r$

$SSC = [5(0.4356+0.492804+0.469225+0.49+0.534361) -$

$- 12.096484] \times 2 = 2(5\times2.421990-12.096484) =$

$= (12.109950-12.096484) \times 2 = (0.013466) \times 2$

$= 0.026932$

FACTOR III (LETRAS LATINAS (FOCOS))

$SSL = (5 \sum_k \bar{X}^2_{..k..} - t^2\bar{X}^2_{.....})r$

$= [5(0.483025+0.483025+0.477481$

$\bar{X}_{..A..} = 0.695000$

$\bar{X}_{..B..} = 0.695000$

$\bar{X}_{..C..} = 0.691000$

$\bar{X}_{..D..} = 0.696000$

$= \quad + 0.484416 + 0.491401)$

$- 12.096484] \times 2$

$= (5 \times 2.419348 - 12.096484) \times 2$

$= (12.096740 - 12.096484) \times 2 = 0.000256 \times 2 = 0.000512$

$\bar{X}_{..E..} = 0.701000$

FACTOR IV (LETRAS GRIEGAS (FILTROS)):

$$SSG = (t\sum_{l}\bar{X}^2_{...l.} - t^2\bar{X}^2_{.....})r$$

$$= [5(0.495816+0.469225+0.502681+0.413449+0.543169) - 12.096484] \times 2$$

$$= (5 \times 2.424140 - 12.096484) \times 2$$

$$= (12.1207 - 12.096484) \times 2$$

$$= 0.024216 \times 2 = 0.048432$$

$\bar{X}_{...\alpha.} = 0.704$

$\bar{X}_{...\beta.} = 0.685$

$\bar{X}_{...\gamma.} = 0.709$

$\bar{X}_{...\delta.} = 0.643$

$\bar{X}_{...\varepsilon.} = 0.737$

RESIDUAL (DUPLICADOS):

$$SSRes = \sum_{ijklr}\sum\sum\sum\sum X^2_{ijklr} - r\sum_{ij}\sum\bar{X}^2_{ij...}$$

$$= 24.2984-2(0.65^2+0.72^2+0.71^2+...+0.665^2+0.745^2 + 0.78^2)$$

$$= 24.2984 - 2 \times 12.1467 = 0.005000$$

INTERACCIONES: $SSI = TSS - SSR - SSC - SSL - SSG - SSRes$

$$= 0.105432-0.019252-0.026932-0.000512-0.048432-0.0050 = 0.005304$$

CON LO ANTERIOR PODEMOS FORMULAR LA SIGUIENTE TABLA DE ANALISIS DE VARIANCIA:

FUENTE DE VARIACION	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	MEDIOS CUADRATICOS	F_{CALC}	$F_c=F_{\alpha,\nu_1\nu_2}$
FACTOR I (PELICULAS)	$t-1=5-1$ $=4$	SSR 0.019252	MSR=0.019252/4 =0.004813	F_I=MSR/MSRe =24.065	$F_I=F_{0.99,4,25}$ 4.18
FACTOR II (CAMARAS)	$t-1=5-1$ $=4$	SSC 0.026932	MSC=$\frac{0.026932}{4}$ =0.006733	F_{II}=MSC/MSRe = 33.665	$F_{II}=F_{0.99,4,25}$ 4.18
FACTOR III (BULBOS)	$t-1=4$	SSL 0.000512	MSL=0.000512/4 =0.000128	F_{III} =MSL/MSRe = 0.64	$F_{III}=F_{0.99,4,25}$ 4.18
FACTOR IV (FILTROS)	$t-1=4$	SSG 0.048432	MSG=$\frac{0.048432}{4}$ =0.012108	F_{IV}=MSG/MSRe =60.54	$F_{IV}=F_{0.99,4,25}$ 4.18
INTERACCIONES	$(t-1)(t-3)=$ $4\times 2=8$	SSI 0.005304	MSI=0.005304/8 =0.000663	F_{IN}=MSI/MSRe =3.3150	$F=F_{0.99,8,25}$ 3.32
RESIDUAL (DUPLICADOS)	$=49-16-8$ $=25$	SSRe 0.0050	MSRe=0.0050/25 =0.00020		
OTAL	rt^2-1 $2\times 25-1$ $=49$	0.105432			

a) EL ESTIMADOR INSESGADO DE LA VARIANCIA RESIDUAL σ^2 ES $\hat{\sigma}^2$ = MSRes = 0.00020

b) DE LA TABLA DE ANALISIS DE VARIANCIA SE OBSERVA QUE LAS PELICULAS, LAS CAMARAS Y LOS TIPOS DE FILTROS PRODUCEN EFECTOS SIGNIFICATIVOS.

c) EL INTERVALO DE CONFIANZA PARA LA CAMARA # 5 SERA (α = 0.05):

$$\bar{X}_{.5...} \pm t_{.025,25}\sqrt{\frac{MSRes}{5}} = 0.731 \pm 2,060\sqrt{\frac{0.00020}{5}} =$$

$$= 0.731 \pm 2.060 \times 0.006325 = 0.731 \pm 0.013029 = (0.717971,\ 0.744029)$$

d) COMPARACIONES MULTIPLES:

d.1) ENTRE LAS PELICULAS:

PELICULA	1	2	3	4	5
$\overline{X}_{i....}$	0.6780	0.682	0.683	0.705	0.73
	*	0.0040	0.0050	0.0270	0.0520
		*	0.001	0.023	0.048
			*	0.022	0.047
				*	0.025

DUNCAN: 1 2 3 4 5

p	2	3	4
q'	2.915	3.065	3.145
W_p	0.013	0.0137	0.014

DONDE

$$W_p = q' \sqrt{\frac{0.00020}{10}}$$

$$q' = q'_{0.05,(r,25)}$$

FISHER: $LSD = t_{\alpha/2,\nu}\sqrt{\frac{2MSR_{es}}{rt}} = t_{0.01/2,25}\sqrt{\frac{2x6.00020}{2x5}} =$

$= 2.060 \times 0.0063 = 0.013$

DE LA TABLA OBSERVAMOS QUE LAS PELICULAS 4 Y 5 PRESENTAN EFECTOS SIGNIFICATIVOS

METODO DE TUKEY:

$$W = q_{\alpha(t,v)}\sqrt{\frac{MSR_{es}}{rt}} = q_{0.05,}(5,25)\sqrt{\frac{0.00020}{10}} = 4.1583x0.0045 \doteq 0.0$$

OBSERVANDO LOS RESULTADOS DE FISHER SE LLEGA A LA MISMA CONCLUSION (VER TABLA).

d.2) ENTRE CAMARAS

CAMARAS	1	3	4	2	5
$\overline{X}_{.j...}$	0.66	0.685	0.70	0.702	0.731
	*	0.0250	0.04	0.042	0.071
		*	0.015	0.017	0.0450
			*	0.002	0.031
				*	0.029

FISHER: LSD = 2.060x0.0063

= 0.013

TUCKEY: W = 0.0186

DUNCAN:

p	2	3	4	5
q'	2.915	3.065	3.145	3.215
W_p	0.013	0.0137	0.014	0.0144

OBSERVAMOS EN ESTE CASO QUE FISHER Y DUNCAN COINCIDEN EN RESULTADOS: μ_1, μ_3 Y μ_5 SON SIGNIFICATIVAMENTE DIFERENTES MIENTRAS μ_4 Y μ_2 SON MENOS SIGNIFICATIVOS; EL METODO DE TUCKEY DIFIERE EN LO REFERENTE A μ_3 DE DONDE SE INFIERE QUE LAS CAMARAS 1 Y 5 SON LAS QUE DIFIEREN.

d.3) PARA LOS FILTROS:

FILTROS	δ	β	α	γ	ε
$\overline{X}_{....i}$	0.643	0.685	0.704	0.709	0.737
	*	0.042	0.0610	0.066	0.094
		*	0.019	0.024	0.052
			*	0.005	0.033
				*	0.028

FISHER: LSD = 0.013

TUCKEY: W = 0.0186

DUNCAN:

p	2	3	4	5
W_p	0.013	0.0137	0.014	0.0144

EN ESTE CASO LOS FILTROS α Y γ SON MENOS SIGNIFICATIVOS EN LOS EFECTOS QUE LOS FILTROS RESTANTES δ, β Y ε (OBSERVESE LA COINCIDENCIA DE RESULTADOS POR LOS 3 METODOS.

12. BLOQUES ALEATORIZADOS INCOMPLETOS

ES USUAL QUE AL PLANEAR UN EXPERIMENTO SE PRESENTA LA SITUACION DE QUE LOS BLOQUES NO SON LO SUFICIENTEMENTE GRANDES COMO PARA ACOMODAR UNA REPLICA COMPLETA.

POR EJEMPLO, SI EN UN DIA SOLO SE PUEDEN REALIZAR 3 ENSAYES Y SI HAY 4 NIVELES DEL "TRATAMIENTO", ENTONCES EN UN SOLO DIA NO SE PUEDEN REALIZAR LOS ENSAYES PARA OBSERVAR LOS CUATRO NIVELES EN UN SOLO BLOQUE (DIA). EN ESTE CASO EL DISEÑO EXPERIMENTAL QUEDARIA CON 4 BLOQUES CON TRES RESULTADOS SOLAMENTE CADA UNO, DE LA SIGUIENTE MANERA:

BLOQUES			
I	II	III	IV
B	A	C	B
A	B	A	D
C	D	D	C

EN EL QUE EL ORDEN DE APARICION DE CADA TRATAMIENTO EN CADA BLOQUE HA SIDO ALEATORIZADO.

UN DISEÑO EXPERIMENTAL COMO ESTE SE DENOMINA DE <u>BLOQUES ALEATORIZADOS INCOMPLETOS</u> O <u>BLOQUES INCOMPLETOS BALANCEADOS</u> (BIB). EL TERMINO <u>BALANCEADO</u> NO SOLO SIGNIFICA QUE TODOS LOS BLOQUES SON DEL MISMO TAMAÑO Y QUE CADA NIVEL DEL TRATAMIENTO APARECE EL MISMO NUMERO DE VECES, SINO TAMBIEN QUE CADA PAREJA DE NIVELES DEL TRATAMIENTO APARECE JUNTA (EN EL MISMO BLOQUE) EL

MISMO NUMERO DE VECES; EN EL EJEMPLO ANTERIOR ESTO SUCEDE 2 VECES.

PARA DESCRIBIR UN EXPERIMENTO BIB SE UTILIZAN LOS SIGUIENTES TERMINOS:

t = NUMERO DE NIVELES DEL TRATAMIENTO

b = NUMERO DE BLOQUES

k = NUMERO DE NIVELES DEL TRATAMIENTO EN CADA BLOQUE

r = NUMERO DE REPLICAS DE CADA NIVEL DEL TRATAMIENTO

λ = NUMERO DE BLOQUES EN LOS CUALES APARECE CADA PAREJA DE NIVELES DEL TRATAMIENTO

UNA FORMA ALTERNATIVA DE EXPRESAR EL EXPERIMENTO ANTERIOR ES MEDIANTE LA SIGUIENTE TABLA:

TRATAMIENTOS	BLOQUES			
	I	II	III	IV
A	X	X	X	
B	X	X		X
C	X		X	X
D		X	X	X

$t=4$

$b=4$

$k=3$

$r=3$

$\lambda=2$

OTRO EJEMPLO ES EL SIGUIENTE:

TRATA-MIENTOS	BLOQUES I	II	III	IV	V	VI	VII	VIII	IX	X	XI
A	X				X	X	X		X	X	
B		X				X	X	X		X	X
C	X		X				X	X	X		X
D	X	X		X				X	X	X	
E		X	X		X				X	X	X
F	X		X	X		X				X	X
G	X	X		X	X		X				X
H	X	X	X		X	X		X			
I		X	X	X		X	X		X		
J			X	X	X		X	X		X	
K				X	X	X		X	X		X

EN ESTE EJEMPLO: $t = 11$, $b = 11$, $k = 6$, $r = 6$ y $\lambda = 3$.

EN EL LIBRO DE COCHRAN Y COX, "EXPERIMENTAL DESIGNS", SE PRESENTAN UNA LISTA DE DISEÑOS BIB.

EL MODELO MATEMATICO PARA REPRESENTAR AL DISEÑO BIB ES

$$X_{ij} = \mu + \beta_i + \tau_j + Z_{ij} \qquad (1)$$

DONDE LAS β_i SON LOS EFECTOS DE LOS BLOQUES, Y LAS τ_j LOS EFECTOS DE LOS TRATAMIENTOS, CON $\sum_{i=1}^{b} \beta_i = \sum_{j=1}^{t} \tau_j = 0$.

EN ESTOS EXPERIMENTOS SE PRESUME QUE <u>NO HAY INTERACCION</u> ENTRE LOS DOS FACTORES.

LA DIFERENCIA DEL EXPERIMENTO BIB Y EL DE BLOQUES COMPLETOS ALEATORIZADOS, ES QUE EN EL PRIMERO NO ESTAN PRESENTES TODAS LAS POSIBLES COMBINACIONES DE i Y j.

CONSIDEREMOS UN NIVEL PARTICULAR DEL TRATAMIENTO, q; LA SUMA DE TODAS LAS OBSERVACIONES DE ESTE NIVEL ES, UTILIZANDO LA EC (1):

$$X_{.q} = \sum_{i(q)} X_{iq} = r\mu + \sum_{i(q)} \beta_i + r\tau_q + \sum_{i(q)} Z_{iq} \qquad (2)$$

DONDE $\sum_{i(q)}$ DENOTA LA SUMATORIA SOBRE TODOS LOS BLOQUES (r) QUE CONTIENEN EL q-ESIMO TRATAMIENTO. SIMILARMENTE:

$$X_{i.} = \sum_{j(i)} X_{ij} = k\mu + k\beta_i + \sum_{j(i)} \tau_j + \sum_{j(i)} Z_{ij} \qquad (3)$$

DONDE $\sum_{j(i)}$ DENOTA LA SUMATORIA SOBRE TODOS LOS TRATAMIENTOS INCLUIDOS EN EL i-ESIMO BLOQUE.

SUMANDO LA EC (3) SOBRE TODOS LOS BLOQUES QUE CONTIENEN EL q-ESIMO TRATAMIENTO SE OBTIENE:

$$\sum_{i(q)} \sum_{j(i)} X_{ij} = rk\mu + k \sum_{i(q)} \beta_i + \sum_{i(q)} \sum_{j(i)} \tau_j + \sum_{i(q)} \sum_{j(i)} Z_{ij} \qquad (4)$$

EL TERCER TERMINO DEL MIEMBRO DERECHO DE ESTA ECUACION VALE:

$$\sum_{i(q)} \sum_{j(i)} \tau_j = r\tau_q + \lambda \sum_{j \neq q} \tau_j = (r - \lambda)\tau_q \tag{5}$$

YA QUE $\sum_{j=1}^{t} \tau_j = 0 = \tau_q + \sum_{j \neq q} \tau_j$, POR LO QUE $\sum_{j \neq q} \tau_j = -\tau_q$

SUSTRAYENDO EL RESULTADO DE LA EC (4) PREVIA SUSTITUCION DE LA EC (5) AL DE LA EC (2) MULTIPLICADO POR k SE OBTIENE

$$k \sum_{i(q)} X_{iq} - \sum_{i(q)} \sum_{j(i)} X_{ij} = (kr - r + \lambda)\tau_q + k \sum_{i(q)} Z_{iq} - \sum_{i(q)} \sum_{j(i)} Z_{ij} \tag{6}$$

POR TANTO, Y CONSIDERANDO QUE $E(Z_{ij}) = 0$ Y QUE LA RELACION $\lambda = r(k - 1)/(t - 1)$ ES VALIDA, DE LA EC (6) SE OBTIENE QUE UN ESTIMADOR INSESGADO DE τ_q ES

$$\hat{\tau}_q = \frac{1}{\lambda t} \{k \sum_{i(q)} X_{iq} - \sum_{i(q)} \sum_{j(i)} X_{ij}\} \tag{7}$$

O

$$\hat{\tau}_q = \frac{k}{\lambda t} \{ \sum_{i(q)} X_{iq} - \bar{X}_{i.} \} = \frac{k}{\lambda t} \{X_{.q} - \sum_{i(q)} \bar{X}_{i.}\} \tag{8}$$

DONDE $\bar{X}_{i.} = \sum_j X_{ij}/k$ = PROMEDIO ARITMETICO MARGINAL DE LAS OBSERVACIONES DEL BLOQUE i

$X_{.q}$ = SUMA DE TODAS LAS OBSERVACIONES DEL q-ESIMO TRATAMIENTO

SUMANDO LA EC (1) SOBRE TODAS LAS OBSERVACIONES SE ENCUENTRA QUE EL PROMEDIO GLOBAL

$$\bar{X}_{..} = \sum_i \sum_j X_{ij}/(kb) \tag{9}$$

ES UN ESTIMADOR INSESGADO DE μ. POR TANTO, UN ESTIMADOR INSESGADO DEL EFECTO DEL q-ESIMO TRATAMIENTO ES $\bar{X}_{..} + \hat{\tau}_q$, EL CUAL TIENE COMO VARIANCIA A

$$\text{Var}\ (\bar{X}_{..} + \hat{\tau}_q) = \frac{\sigma^2}{r}\left(\frac{1}{t} + \frac{k(t-1)^2}{(k-1)t^2}\right) \tag{10}$$

DE IGUAL MANERA, LA DIFERENCIA DE EFECTOS ENTRE LOS TRATAMIENTOS q Y q' SE ESTIMA CON $\hat{\tau}_q - \hat{\tau}_{q'}$, CON LO CUAL SE TIENE UNA VARIANCIA DE LA ESTIMACION

$$\text{Var}(\hat{\tau}_q - \hat{\tau}_{q'}) = \sigma^2\ \frac{2k}{\lambda t} \tag{11}$$

LA TABLA PARA EL ANALISIS DE VARIANCIA DE ESTE EXPERIMENTO ES

FUENTE	G. DE L.	SS	MS	F
BLOQUES (SIN AJUSTAR)	b - 1	SSB	MSB = SSB/(b - 1)	
TRATAMIENTOS (AJUSTADO)	t - 1	SST	MST = SST/(t - 1)	MST/MSE
ERROR O RESIDUAL	bk-t-b+1	SSE	MSE=SSE/(bk-t-b-1)	
TOTAL	bk - 1	TSS		

DONDE

$$SSB = \sum_{i=1}^{b} X_{i.}^{2}/k - bk\,\bar{X}_{..}^{2} \tag{12}$$

$$SST = \frac{1}{k\lambda t}\sum_{j=1}^{t}\left(kX_{.j} - \sum_{i(j)} X_{i.}\right)^{2} \tag{13}$$

$$TSS = \sum_{i}\sum_{j} X_{ij}^{2} - bk\,\bar{X}_{..}^{2} \tag{14}$$

$$SSE = TSS - SSB - SST \tag{15}$$

ES NECESARIO MENCIONAR QUE EL SSB CALCULADO CON LA EC (12) SOLO SIRVE EN ESTE CASO COMO AUXILIAR PARA CALCULAR SSE CON LA EC (15), PERO NO PARA HACER LA PRUEBA DE HIPOTESIS DE EFECTOS DE LOS BLOQUES; LA RAZON DE ESTO ES QUE EN ESTE CASO, AL USAR LA EC (1) PARA CALCULAR SSB SE ENCUENTRA QUE DEPENDE DE β_i Y DE τ_j; PARA QUE SE PUEDA HACER PRUEBA DE EFECTOS DE BLOQUES SE REQUIERE DISEÑAR UN EXPERIMENTO DE BLOQUES INCOMPLETOS BALANCEADO Y SIMETRICO, EL CUAL SE ESTUDIARA MAS ADELANTE.

EJEMPLO

EN LA PRODUCCION DE UN COMPONENTE DE UNA MAQUINA, SE TIENE QUE EL DIAMETRO INTERIOR DE UN TUBO ES UNA DIMENSION CRITICA. ESTOS COMPONENTES SE FABRICAN CON 7 MAQUINAS Y 7 ALEACIONES DIFERENTES.

PARA DETERMINAR LOS EFECTOS DE LAS ALEACIONES SE DISEÑO UN EXPERIMENTO BIB, EN EL QUE LOS BLOQUES FUERON LAS MAQUINAS Y

LOS TRATAMIENTOS FUERON LAS ALEACIONES, Y SE TOMARON MUESTRAS DE 10 DIAMETROS EN CADA CASO. EN LA SIGUIENTE TABLA SE PRESENTAN LAS DIFERENCIAS DEL PROMEDIO DE LOS DIEZ DATOS Y LA DIMENSION NOMINAL, EN MM.

TRATAMIENTOS (ALEACIONES)	MAQUINAS (BLOQUES)							TOTALES ($X_{\cdot j}$)
	1	2	3	4	5	6	7	
A	5	4	9					18
B			12	9	9			30
C	7			6		8		21
D			7			5	3	15
E	4				6		5	15
F		10			12	9		31
G		4		4			3	11
TOTALES ($X_{i\cdot}$)	16	18	28	19	27	22	11	141
$\bar{X}_{i\cdot}$	5.33	6.00	9.33	6.33	9.00	7.33	3.67	

EN ESTE CASO SE TIENE QUE $b=t=7$, $k=r=3$, $\lambda=1$, $\bar{X}_{\cdot\cdot}=\frac{141}{21}=6.7143$

$$SSB=\frac{1}{3}\sum_{i=1}^{7} X_{i\cdot}^2-7x3x\,\bar{X}_{\cdot\cdot}^2=\frac{1}{3}(16^2+18^2+28^2+19^2+27^2+22^2+11^2)-946.7143=72.96$$

$$SST=\frac{1}{3x1x7}\sum_{j=1}^{7}(3X_{\cdot j}-\sum_{i(j)}X_{i\cdot})^2=\frac{1}{21}\{(3x18-(16+18+28))^2+$$

$$+(3x30-(28+19+27))^2+(3x21-(16+19+22))^2+$$

$$+(3x15-(28+22+11))^2+(3x15-(16+27+11))^2+$$

$$+(3x31-(18+27+22))^2+(3x11-(18+19+11))^2\}=75.90$$

$$TTS=\sum_i\sum_j X_{ij}^2-bk\,\bar{X}_{\cdot\cdot}^2=5^2+4^2+9^2+12^2+\ldots+3^2-946.7143=156.29$$

$SSE = 156.29 - 72.96 - 75.90 = 7.43$

$MST = 75.90/6 = 12.65$, $MSE = 7.43/8 = 0.929$, $F_T = \frac{12.65}{0.929} = 13.62$

$F_{0.99,6,8} = 6.37 < 13.62$, POR LO QUE SE CONCLUYE QUE CON UN 99% DE NIVEL DE CONFIANZA SI HAY EFECTO DEBIDO A LA ALEACION QUE SE UTILIZA PARA FABRICAR EL COMPONENTE.

TAREA: ESTIMAR LOS τ_i

PARA EL EJEMPLO DE LOS DIAMETROS INTERNOS DE LOS TUBOS, CALCULAR LOS VALORES ESTIMADOS DE τ_j'S Y HACER COMPARACIONES MULTIPLES:

PARA ESTIMAR LOS EFECTOS DE CADA TRATAMIENTO PODEMOS USAR LA FORMULA ALTERNATIVA:

$$\hat{\tau}_q = \frac{k}{\lambda t}\left[\sum_{i(q)} X_{iq} - \sum_{i(q)} \overline{X}_{i.}\right]$$

$\hat{\tau}_A = \frac{3}{1x7}[18 - 20.66] = -1.143$ $\qquad$ $\hat{\tau}_E = -1.287$

$\hat{\tau}_B = 0.429[30 - 24.66] = 2.288$ $\qquad$ $\hat{\tau}_F = 3.718$

$\hat{\tau}_C = 0.429[21 - 19] = 0.858$ $\qquad$ $\hat{\tau}_G = -2.145$

$\hat{\tau}_D = 0.429[15 - 20.33] = -2.288$

COMPARACIONES MULTIPLES:

TRATAMIENTO	D	G	E	A	C	B	F
$\overline{X}_{..} + \hat{\tau}_q$	4.4263	4.5693	5.4273	5.5713	7.5723	9.0023	10.4323
	*	0.143	1.0010	1.1450	3.146	4.576	6.006
		*	0.858	1.002	3.003	4.433	5.863
			*	0.144	2.145	3.575	5.005
				*	2.001	3.431	4.861
					*	1.43	2.86
						*	1.43

<u>FISHER</u>: $LSD = t_{\alpha/2,\nu}\sqrt{\frac{2k\ MSE}{\lambda t}} = t_{0.05,8}\sqrt{\frac{2x3x0.929}{1x7}} = 0.061$

TUCKEY: $W = q_{0.05}\ (7,8)\ \frac{MSE}{t} = 5.4\ \frac{0.929}{7} = 1.967$

DUNCAN:

p	2	3	4	5	6	7
q'	3.26	3.39	3.47	3.52	3.55	3.56
W_p	1.88	1.235	1.264	1.282	1.293	1.297

DONDE $\quad W_p = q'_{0.05},\ (p,8)\ \frac{0.929}{7}$

DE DONDE CONCLUIMOS QUE LOS TRATAMIENTOS D, G, E Y A SON SIGNIFICATIVAMENTE MENORES QUE C, B Y F.

EJEMPLO

UNA FABRICA DESEA COMPARAR LA COMODIDAD QUE OFRECEN 8 TIPOS NUEVOS DE ALMOHADAS Y UNO QUE YA ESTA EN EL MERCADO. PARA ESTO SE DISEÑO EL SIGUIENTE EXPERIMENTO DE BLOQUES INCOMPLETOS BALANCEADO:

PARA REDUCIR EL PROBLEMA QUE TENDRIA UNA PERSONA AL ASIGNAR UNA CALIFICACION AL GRADO DE COMODIDAD SI SE TUVIERAN LOS 9 TIPOS DE ALMOHADA JUNTOS, SE DECIDIO AGRUPARLAS EN 12 BLOQUES DE 3, Y A CADA BLOQUE SE LE ASIGNARON AL AZAR LOS TIPOS DE ALMOHADA LOS CUALES, A SU VEZ, SE IDENTIFICARON CON LAS LETRAS DE LA A A LA I (LAS LETRAS NO SE PUSIERON VISIBLES). LA PRUEBA CONSISTIO EN SELECCIONAR AL AZAR A 20 PERSONAS PARA QUE CALIFICARAN CON NUMEROS DEL 1 AL 5 EL GRADO DE COMODIDAD; EL DATO QUE SE ANOTO EN CADA CASO FUE LA SUMA DE LAS CALIFICACIONES DE LAS 20 PERSONAS, HABIENDOSE OBTENIDO LOS SIGUIENTES RESULTADOS:

BLOQUE	TRATAMIENTO (TIPO DE ALMOHADA)			TOTAL
1	A59	B26	C38	123
2	D85	E92	F69	246
3	G74	H52	I27	153
4	A62	D70	G68	200
5	B27	E98	H59	184
6	C31	F60	I35	126
7	A63	E85	I30	178
8	B22	F73	G75	170
9	C45	D74	H51	170
10	A52	F76	H43	171
11	B18	D79	I41	178
12	C41	E84	G81	206
				2065

$t = 9,\ b = 12,\ k = 3,\ r = 4,\ \lambda = 1.$

OTRA FORMA DE PRESENTAR LOS DATOS ANTERIORES ES:

TRATAMIENTO (TIPO DE AL- MOHADA)	BLOQUE 1	2	3	4	5	6	7	8	9	10	11	12	TOTALES $(X_{.j})$
A	59			62			63			52			236
B	26				27			22			18		93
C	38					31			45			41	155
D		85		70					74		79		308
E		92			98		85					84	359
F		69				60		73		76			278
G			74	68				75				81	298
H			52		59				51	43			205
I			27			35	30				41		133
TOTALES $(X_{i.})$	123	246	153	200	184	126	178	170	170	171	138	206	2065

$\overline{X}_{..} = 2065/(4x9) = 57.361111,\ 36\ \overline{X}^2_{..} = 36x57.3611^2 = 118,450.69$

$$SSB = \frac{1}{3}(123^2+246^2+153^2+200^2+184^2+126^2+178^2+170^2+170^2+$$

$$+171^2+178^2+206^2) - 118,450.69 =$$

$$= \frac{368,991.00}{3} - 118,450.69 = 4,546.31$$

$$SST = \frac{1}{3x1x9}\ \{(3x236-(123+200+178+171))^2+$$

$$+(3x93-(123+184+170+138\ ))^2+$$

$$+(3x155-(123+126+170+206))^2+$$

$$+(3x308-(246+200+170+138))^2+$$

$$+(3x359-(246+184+178+206))^2+$$

$$+(3x278-(246+126+170+171))^2+$$

$$+(3x298-(153+200+170+206))^2+(3x205-(153+$$

$$+\ 184+170+171))^2+(3x133-(153+126+178+138))^2\} =$$

$$= 322,122.00/27 = 11,930.07$$

$TSS = 59^2+62^2+63^2+52^2+26^2+27^2+ \ldots + 41^2-118{,}450.69 =$

$= 135{,}435.00-118{,}450.69 = 16{,}984.31$

$SSE = 16{,}984.31-4{,}546.31-11{,}930.07 = 507.93$

LA TABLA DEL ANALISIS DE VARIANCIA ES:

FUENTE	G. DE L.	SS	MS	F
BLOQUES	11	4,546.31	---	
TRATAMIENTOS	8	11,930.07	1491.26	46.97 > 2.59
ERROR	16	507.93	31.75	
TOTAL	35	16,984.31		

PUESTO QUE $F_{0.95,8,16} = 2.59 < 46.97$, SE CONCLUYE QUE SI HAY DIFERENCIA SIGNIFICATIVA ENTRE LOS NUEVE TIPOS DE ALMOHADA. VEAMOS, POR TANTO, CUALES TIPOS SON LOS QUE DIFIEREN DE LOS DEMAS, PARA LO CUAL ESTIMAREMOS LOS EFECTOS, τ_q, DE CADA NIVEL.

$$\hat{\tau}_q = \frac{k}{\lambda t}\,\{X_{.q} - \sum_{i(q)} \overline{X}_{i.}\}$$

$$\hat{\tau}_1 = \frac{3}{9}(236 - \frac{123+200+178+171}{3}) = \frac{1}{3}(236 - 224.00) = 4$$

$$\hat{\tau}_2 = \frac{1}{3}(93 - \frac{123+184+170+138}{3}) = \frac{1}{3}(93 - 205) = -37.33$$

$$\hat{\tau}_3 = \frac{1}{3}(155 - \frac{123+126+170+206}{3}) = \frac{1}{3}(155 - 208.33) = -17.78$$

$$\hat{\tau}_4 = \frac{1}{3}(308 - \frac{246+200+170+138}{3}) = 18.89$$

$$\hat{\tau}_5 = \frac{1}{3}(359 - \frac{246+184+178+206}{3}) = 29.22$$

$$\hat{\tau}_6 = \frac{1}{3}\left(278 - \frac{246 + 126 + 170 + 171}{3}\right) = 13.44$$

$$\hat{\tau}_7 = \frac{1}{3}\left(298 - \frac{153 + 200 + 170 + 106}{3}\right) = 18.33$$

$$\hat{\tau}_8 = \frac{1}{3}\left(205 - \frac{153 + 184 + 170 + 171}{3}\right) = -7.00$$

$$\hat{\tau}_9 = \frac{1}{3}\left(133 - \frac{153 + 126 + 178 + 138}{3}\right) = -21.78$$

LA TABLA DE ESTIMACIONES DE LOS EFECTOS DE LOS NIVELES DEL TRATAMIENTO SON:

TRATAMIENTO	A	B	C	D	E	F	G	H	I
$\bar{X}_{..} + \hat{\tau}_q$	61.36	20.03	39.58	76.25	86.58	70.80	75.69	50.36	35.

USANDO MSW = MSE = 31.75, CON 16 GRADOS DE LIBERTAD, LA MINIMA DIFERENCIA SIGNIFICATIVA ENTRE DOS MEDIAS ES, CON $\alpha = 0.05$:

$$LSD = t_{\alpha/2}\sqrt{\frac{2k(MSE)}{\lambda t}} = 2.12\sqrt{\frac{2 \times 3 \times 31.75}{1 \times 9}} = 9.75$$

LAS ESTIMACIONES $\hat{\tau}_q$ ORDENADAS EN FORMA CRECIENTE SON LAS QUE SE MUESTRAN EN LA TABLA SIGUIENTE, EN LA CUAL SE HAN ANOTADO TAMBIEN LAS DIFERENCIAS QUE HAY ENTRE ELLAS:

B	I	C	H	A	F	G	D	E
20.3	35.58	39.58	50.36	61,36	70.80	75.69	76.25	86.58
*	15.28							
	*	$\boxed{4.00}$	14,78					
		*	10.78					
			*	11.00				
				*	$\boxed{9.44}$	14.33		
					*	$\boxed{4.89}$	$\boxed{5.45}$	15.78
						*	$\boxed{0.56}$	10.89
							*	10.33

LAS MEDIAS QUE RESULTARON SER ESTADISTICAMENTE IGUALES SON LAS SUBRAYADAS A CONTINUACION CON LINEA COMUN:

B	I	C	H	A	F	G	D	E
20.3	35.58	39.58	50.36	61.36	70.80	75.69	76.25	86.58

BLOQUES INCOMPLETOS

BALANCEADOS SIMETRICOS

SI EL NUMERO DE BLOQUES ES IGUAL AL DE TRATAMIENTOS ($b = t$), ENTONCES $r = k$. EN ESTE CASO SE DICE QUE EL EXPERIMENTO ES DE BLOQUES INCOMPLETOS BALANCEADOS SIMETRICOS (SBIB), Y ES PO SIBLE HACER PRUEBA DE HIPOTESIS PARA LOS EFECTOS DE LOS BLO- QUES EN UNA MANERA SIMILAR QUE PARA LOS TRATAMIENTOS, MEDIAN

TE LA SIGUIENTE TABLA DEL ANALISIS DE VARIANCIA, EN LA CUAL SE NOTA QUE HAY SUMAS DE CUADRADOS AJUSTADOS PARA CADA UNO DE LOS DOS FACTORES.

FUENTE	SS	G. de L.	MS	F
BLOQUES	SSB			
TRATAMIENTOS (AJUSTADA)	$\tilde{SST}$	t - 1	$MST=\tilde{SST}/(t-1)$	MST/MSE
TRATAMIENTOS	SST			
BLOQUES (AJUSTADA)	$\tilde{SSB}$	b - 1	$MSB=\tilde{SSB}/(b-1)$	MSB/MSE
ERROR	SSE	bk-b-t-1		
TOTAL	TSS	bk - 1		

EN ESTA TABLA SSB, $\tilde{SST}$, SSE Y TSS SE CALCULAN CON LAS MISMAS FORMULAS QUE EN EL EXPERIMENTO BIB; LAS OTRAS SE CALCULAN CON LAS SIGUIENTES EXPRESIONES:

$$\tilde{SST} = \frac{1}{r} \sum_{j=1}^{t} X^2_{.j} - bk\bar{X}^2_{..}$$

$$\tilde{SSB} = \frac{1}{kt\lambda} \sum_{i=1}^{b} \left(rX_{i.} - \sum_{j(i)} X_{.j}\right)^2$$

<u>EJEMPLO</u>

EL PROBLEMA PRESENTADO ANTERIORMENTE, DE LAS MAQUINAS Y ALEACIONES, ES UN EXPERIMENTO SBIB, YA QUE EN EL $t = b = 7$. PROBAR LA HIPOTESIS DE QUE $\beta_i = 0$ PARA TODA i, A UN 95% DE NIVEL

DE CONFIANZA.

$$SST = \frac{1}{3}\sum_{j=1}^{7} X_{.j}^2 - bk\bar{X}_{..}^2 = \frac{1}{3}(18^2+30^2+21^2+15^2+15^2+31^2+11^2) -$$

$$- 946.71 = 118.96$$

$$\bar{SSB} = \frac{1}{3\times7\times1}\sum_{i=1}^{7}(3X_{i.} - \sum_{j(i)} X_{.j})^2 = \frac{1}{21}[(3\times16-(18+21+15))^2 +$$

$$+(3\times18-(18+31+11))^2+(3\times28-(18+30+15))^2+$$

$$+(3\times19-(30+21+11))^2+(3\times27-(30+15+31))^2+$$

$$+(3\times22-(21+15+31))^2+(3\times11-(15+15+11))^2] = 29.90$$

PARA VERIFICAR, CALCULEMOS $SSE = TSS - SST - \bar{SSB} =$

$156.29 - 118.96 - 29.90 = 7.43 = TSS - \bar{SST} - SSB$

LA TABLA DEL ANALISIS DE VARIANCIA ES:

FUENTE	G. de L.	SS	MS	F
MAQUINAS		72.96		
ALEACIONES (AJUSTADA)	6	75.90	12.65	13.62>3.58
MAQUINAS (AJUSTADA)	6	29.90	4.98	5.36>3.58
ALEACIONES		118.96		
ERROR	8	7.43	0.929	
TOTAL	20	156.29		

$F_{0.95,6,8} = 3.58$

POR LO ANTERIOR SE CONCLUYE QUE SI HAY DIFERENCIAS SIGNIFICATIVAS ENTRE LOS NIVELES TANTO DE LAS ALEACIONES COMO DE LAS MAQUINAS.

TAREA: ESTIMAR LAS MEDIAS PARA CADA NIVEL DE BLOQUES Y TRATAMIENTOS

EJEMPLO

DIEZ ESPECIMENES DE HULE SE ENVIARON A UN LABORATORIO PARA UNA PRUEBA DE RESISTENCIA A LA FLEXION. HAY CINCO TIEMPOS DE CURADO. SIN EMBARGO CADA ESPECIMEN ES SUFICIENTE SOLAMENTE PARA DOS MUESTRAS. ENTONCES SE PROPUSO UN DISEÑO BIB. LOS ESPECIMENES SE CONSIDERARON COMO BLOCKS Y LOS TIEMPOS DE CURADO COMO TRATAMIENTOS. INVESTIGUE EL EFECTO DEL TIEMPO DE CURADO SOBRE LA RESISTENCIA A LA FLEXION, USANDO LOS DATOS CODIFICADOS DE ABAJO.

(BLOQUES)	TIEMPOS DE CURADO				(TRAT)	TOTALES	
ESPECIMENES	1	2	3	4	5	$X_{i.}$	$\bar{X}_{i.}$
1	25				6	31	15.5
2	10		3			13	6.5
3	3			16		19	9.5
4	15	11				26	13
5			0		6	6	3
6				14	11	25	12.5
7		6			17	23	11.5
8			10	27		37	18.5
9		10	5			15	7.5
10		7		21		28	14
TOTALES $X_{.j}$	53	34	18	78	40	223	
$\bar{X}_{.j}$	13.25	8.5	4.5	19.5	10		$\bar{X}_{..} = 11.15$

EN ESTE CASO TENEMOS: b = # BLOQUES = 10; t = # TRATAMIENTOS = 5; r = # REPLICAS = 4; k = # NIV. DE TRAT/BLOQUE = 2; λ = # BLOQUES C/PAREJAS IGUALES = 1

PARA LOS BLOQUES: $SSB = k^{-1} \sum_{i=1}^{b} X_{i.}^2 - (bk)^{-1} X_{..}^2$

$$= \frac{1}{2}(31^2 + 13^2 + \ldots + 15^2 + 28^2) - \frac{1}{10 \times 2} 223^2$$

$$= 2867.5 - 2486.45 = 381.05$$

PARA LOS TRATAMIENTOS:

$$SST = \frac{t-1}{Nk(k-1)} \sum_{j=1}^{t} \left[kX_{.j} - \sum_{i(j)} X_{i.} \right]^2$$

$$SST = \frac{5-1}{20 \times 2(1)} \{ [2 \times 53 - (31 + 13 + 19 + 26)]^2 + [2 \times 34 - (26 + 23 + 15 + 28)]^2 + [2 \times 18 - (13 + 6 + 37 + 15)]^2 + [2 \times 78 - (19 + 25 + 37 + 28)]^2 + [2 \times 40 - (31 + 6 + 25 + 23)]^2 \} = \frac{1}{10} \{ (17)^2 + (-24)^2 + (-35)^2 + (47)^2 + (-5)^2 \} = \frac{1}{10}(289 + 576 + 1225 + 2209 + 25) = 432.4$$

TOTALES: $TSS = \sum_i \sum_j X_{ij}^2 - \frac{X_{..}^2}{bk}$

$$= 25^2 + 10^2 + 3^2 + 15^2 + \ldots + 6^2 + 6^2 + 11^2 + 17^2 - 2486.45$$

$$= 3503 - 2486.45 = 1016.55$$

ERROR: $SSE = TSS - SST - SSB$

$$= 1016.55 - 432.4 - 381.05 = 203.10$$

DONDE:

FUENTE	g. de l.	SS	MS	F	$F_c = F_{0.05,4,6}$
ESPECIMENES (BLOQUES S/AJUST)	b - 1 = 10 - 1 = 9	SSB=381.05	MSB=SSB/(b-1) = 42.34	NO SE PUEDE	
TIEMPO DE CURADO (AJUSTADOS)	t - 1 = 5 - 1 = 4	SST=432.4	MST=SST/(t-1) =108.10	MST/MSE =108.10/33.85 = 3.19	< 4.53
ERROR	bk-t-b+1= 20-5-10+1= 6	SSE=203.10	MSE=SSE/bk-t-b+1 = 33.85		
TOTAL	bk - 1 = 10 x 2-1=19	TSS=1016.55			

DADO QUE F CALCULADA (3.19) < F CRITICA ($F_{0.05,4,6}$ = 4.53) ENTONCES CONCLUIMOS QUE LAS RESISTENCIAS A LA FLEXION DE LOS ESPECIMENES DE HULE NO SE AFECTAN POR LOS TIEMPOS DE CURADO, O SEA, POR LOS TRATAMIENTOS.

b) ESTIMACION DE LOS EFECTOS DE LOS TRATAMIENTOS:

$$\hat{\tau}_q = \frac{kr}{\lambda t}\left[\bar{x}_{.q} - r^{-1}\sum_{i(q)}\bar{x}_{i.}\right]$$

$$\hat{\tau}_1 = \frac{2 \times 4}{1 \times 5}\left[13.25 - \frac{15.5 + 6.5 + 9.5 + 13}{4}\right] = 3.40$$

$$\hat{\tau}_2 = \frac{8}{5}\left[8.5 - \frac{13 + 11.5 + 7.5 + 14}{4}\right] = -4.80$$

$$\hat{\tau}_3 = \frac{8}{5}\left[4.5 - \frac{6.5 + 3 + 18.5 + 7.5}{4}\right] = -7.00$$

$$\hat{\tau}_4 = \frac{8}{5}\left[19.5 - \frac{9.5 + 12.5 + 18.5 + 14}{4}\right] = 9.40$$

$$\hat{\tau}_5 = \frac{8}{5}\left[10 - \frac{15.5 + 3 + 12.5 + 11.5}{4}\right] = -1.00$$

c) AUNQUE EN ESTE CASO LA PRUEBA DE ANALISIS DE VARIANCIA INDICO

INDEPENDENCIA ENTRE LOS TIEMPOS DE CURADO (TRATAMIENTOS) HAREMOS LA PRUEBA DE COMPARACIONES MULTIPLES PARA VERIFICAR QUE NO DIFIEREN DICHOS TRATAMIENTOS.

USANDO EL CRITERIO LSD $= t_{\alpha/2,\nu}\sqrt{\frac{2k(MSE)}{\lambda t}} =$

$$t_{0.05/2,6}\sqrt{\frac{2 \times 2 \times 33.85}{5}} = 2.447\sqrt{27.08} = 12.73$$

TIEMPOS DE CURADO	3	2	5	1	4
$\bar{X}_{..} + \hat{\tau}_q$	4.15	6.35	10.15	14.55	20.55
	*	2.2	6.0	10.4	16.4
		*	3.8	8.20	14.20
			*	4.4	10.40
				*	6.0

13 CUADRADOS DE YUDEN

EL EXPERIMENTO DE CUADRADOS DE YUDEN ES UN TIPO DE CUADRADOS LATINOS INCOMPLETO. SI EL FACTOR I ES EL DE LOS RENGLONES, EL II EL DE LAS COLUMNAS, Y EL III EL DE LAS LETRAS LATINAS, Y SI SE CUMPLE QUE LOS FACTORES I Y III TIENEN EL MISMO NUMERO DE NIVELES ($t = b$), ENTONCES LOS CUADRADOS DE YUDEN QUEDAN EN FORMA SEMEJANTE A LOS DOS SIGUIENTES EJEMPLOS 7 x 3 Y 7 x 4:

FACTOR I	FACTOR II		
	1	2	3
1	G	A	C
2	A	B	D
3	B	C	E
4	C	D	F
5	D	E	G
6	E	F	A
7	F	G	B

FACTOR I	FACTOR II			
	1	2	3	4
1	D	F	G	A
2	E	G	A	B
3	F	A	B	C
4	G	B	C	D
5	A	C	D	E
6	B	D	E	F
7	C	E	F	G

ESTE DISEÑO EXPERIMENTAL SE PUEDE VER TAMBIEN COMO UN BIB CON UN FACTOR ADICIONAL (EL II), EN CUYO CASO LA TABLA DE DATOS TENDRIA LA SIGUIENTE PRESENTACION, QUE EJEMPLIFICA EL CASO 7 x 4 ANTERIOR:

TRATAMIENTOS (FACTOR III)	FACTOR I						
	1	2	3	4	5	6	7
A	(4)	(3)	(2)		(1)		
B		(4)	(3)	(2)		(1)	
C			(4)	(3)	(2)		(1)
D	(1)			(4)	(3)	(2)	
E		(1)			(4)	(3)	(2)
F	(2)		(1)			(4)	(3)
G	(3)	(2)		(1)			(4)

EN ESTA TABLA LOS NUMEROS EN PARENTESIS SON LOS NIVELES DEL FACTOR II; EN ELLA: $t=7$, $b=7$, $r=4$, $k=4$ y $\lambda=2$.

EL MODELO MATEMATICO PARA ESTUDIAR ESTE EXPERIMENTO ES

$$X_{ijl} = \mu + \beta_i + \tau_j + \gamma_l + Z_{ijl} \qquad (1)$$

DONDE $i = 1,2,\ldots, b$; $j = 1, 2,\ldots, t = b$; $l = 1, 2,\ldots, k(<t)$, Y $\Sigma\beta_i = \Sigma\tau_j = \Sigma\gamma_l = 0$.

LA TABLA DEL ANALISIS DE VARIANCIA DE ESTE EXPERIMENTO ES LA SIGUIENTE:

FUENTE	G. DE L.	SS	MS	F
BLOQUES		SSB		
TRATAMIENTOS (AJUSTADA)	t-1	$\tilde{SST}$	MST=$\tilde{SST}$/(t-1)	MST/MSE
TRATAMIENTOS		SST		
BLOQUES (AJUSTADA)	b-1	$\tilde{SSB}$	MSB=$\tilde{SSB}$/(b-1)	MSB/MSE
FACTOR II	k-1	SS2	MS2=SS2/(k-1)	MS2/MSE
ERROR	bk-2b-k+2	SSE	MSE=SSE/(bk-2b-k+2)	
TOTAL	bk-1	TSS		

EN ESTA TABLA:

$$SSB = k^{-1} \sum_{i=1}^{b} X_{i..}^{2} - bk\,\bar{X}_{...}^{2} \tag{2}$$

$$\tilde{SST} = (k\lambda t)^{-1} \sum_{j=1}^{t} (kX_{.j.} - \sum_{i(j)} X_{i..})^{2} \tag{3}$$

$$SST = k^{-1} \sum_{j=1}^{t} X_{.j.}^{2} - bk\,\bar{X}_{...}^{2} \tag{4}$$

$$\tilde{SSB} = (k\lambda t)^{-1} \sum_{i=1}^{b} (rX_{i..} - \sum_{j(i)} X_{.j.})^{2} \tag{5}$$

$$SS2 = b^{-1} \sum_{l=1}^{k} X_{..l}^{2} - bk\,\bar{X}_{...}^{2} \tag{6}$$

$$TSS = \sum \sum X_{ijl}^{2} - bk\,\bar{X}_{...}^{2} \tag{7}$$

$$SSE = TSS - SSB - \tilde{SST} - SS2 \tag{8}$$

EJEMPLO

EN LA DETERMINACION DEL NUMERO DE OCTANOS DE UNA GASOLINA, UN METODO USA UNA GASOLINA BASE Y SE TIENEN 6 ADITIVOS COMO CANDITATOS PARA FORMAR UNA NUEVA MARCA. EL EXPERIMENTO ES UNO DE CUADRADOS DE YUDEN 7x3: A CADA COMBUSTIBLE SE LE DAN 2 MINUTOS EN EL MOTOR Y EL RESULTADO SE REGISTRA EN UN INSTRUMENTO ESPECIAL, EL CUAL SE LEE A LOS 60, 90 Y 120 SEG PARA VERIFICAR LA ESTABILIDAD; UNA MARCADA DIFERENCIA EN LA LECTURA A LOS 90 Y 120 SEG ES CAUSA DE ALARMA; LOS BLOQUES SON GRUPOS DE 3 LECTURAS DE 2 MINUTOS. LOS RESULTADOS FUERON:

FACTOR III (TRATAMIENTOS O GASOLINAS)	FACTOR I (BLOQUES)							TOTAL ($X_{.j.}$)
	1	2	3	4	5	6	7	
A	(1) 43				(3) 44		(2) 41	128
B	(2) 34	(1) 36				(3) 32		102
C		(2) 32	(1) 33				(3) 27	92
D	(3) 47		(2) 47	(1) 44				138
E		(3) 46		(2) 40	(1) 41			127
F			(3) 43		(2) 35	(1) 36		114
G				(3) 33		(2) 32	(1) 33	98
TOTAL ($X_{i..}$)	124	114	123	117	120	100	101	799

$\bar{X}_{...} = 799/3x7 = 38.0476;\ 3x7x38.0476^2 = 30,400.05$

$$SSB = \frac{1}{3}(124^2+114^2+123^2+117^2+120^2+100^2+101^2)-30,400.05 =$$

$$= 30,597-30,400.05 = 196.95$$

$$S\hat{S}T = \frac{1}{3x1x7}[\{3x128-(124+120+101)\}^2+\{3x102-(124+114+100)\}^2+$$
$$+\{3x92-(114+123+101)\}^2+\{3x138-(124+123+117)\}^2+$$
$$+\{3x127-(114+117+120)\}^2+\{3x114-(123+120+100)\}^2+$$
$$+\{3x98-(117+100+101)\}^2] = 493.62$$

$$\dot{S}ST = \frac{1}{3}(128^2+102^2+92^2+138^2+127^2+114^2+98^2)-30,400.05 = 608.29$$

$$\hat{S}SB = \frac{1}{3x1x7}[\{3x124-(128+102+138)\}^2+\{3x114-(102+92+127)\}^2+$$
$$+\{3x123-(92+138+114)\}^2+\{3x117-(138+127+98)\}^2+$$
$$+\{3x120-(128+127+114)\}^2+\{3x100-(102+114+98)\}^2+$$
$$+\{3x101-(128+92+98)\}^2] = \frac{1}{21}\{4^2+21^2+...+(-15)^2\} = 82.29$$

$X_{..1} = 43 + 36 + 33 + 44 + 41 + 36 + 33 = 266$

$X_{..2} = 34 + 32 + 47 + 40 + 35 + 32 + 41 = 261$

$X_{..3} = 44 + 32 + 27 + 47 + 46 + 43 + 33 = 272$

$$SS2 = \frac{1}{7}(266^2 + 261^2 + 272^2) - 30{,}400.05 =$$

$$= \frac{1}{7}(70{,}756 + 68{,}121 + 73{,}984) - 30{,}400.05 =$$

$$= 30{,}408.71 - 30{,}400.05 = 8.66$$

$$TSS = 43^2 + 44^2 + 41^2 + 34^2 + \ldots + 33^2 - 30{,}400.05 = 706.95$$

$$SSE = TSS - SSB - SST - SS2 = 7.72$$

$$F_{0.01,6,6} = 8.47, \quad F_{0.01,2,6} = 10.90$$

LA TABLA DEL ANALISIS DE VARIANCIA ES:

FUENTE	G. de L.	SS	MS	F
ORDEN (BLOQUES)		196.95		
GASOLINA (AJUSTADA)	6	493.62	82.27	64.27 > 8.47
GASOLINA		608.29		
ORDEN (AJUSTADA)	6	82.29	13.72	10.72 > 8.47
TIEMPO	2	8.66	4.33	3.39 < 10.90
ERROR	6	7.72	1.28	
TOTAL	20	706.95		

14. ANALISIS DE EXPERIMENTOS FACTORIALES 2^k

EL EXPERIMENTO 2^k ES UN EXPERIMENTO DE k FACTORES CON DOS NIVELES CADA UNO.

CONSIDERESE UN EXPERIMENTO CON 2 FACTORES A Y B, CADA UNO CON 2 NIVELES, A LOS CUALES LLAMAREMOS "ALTO" Y "BAJO". DESIGNEMOS CON MAYUSCULAS A "LOS EFECTOS" Y CON MINUSCULAS A LAS COMBINACIONES DE LOS NIVELES DE LOS TRATAMIENTOS POSIBLES.

POR EJEMPLO, LAS CUATRO COMBINACIONES PARA ESTABLECER LOS TRATAMIENTOS PARA UN EXPERIMENTO 2^2 SON LAS QUE SE MUESTRAN EN LA TABLA SIGUIENTE. EL METODO DE DESIGNAR ESTOS TRATAMIENTOS ES INCLUYENDO LA LETRA MINUSCULA SI EL FACTOR ESTA AL NIVEL ALTO Y EXCLUYENDOLA EN CASO CONTRARIO, SI TODOS LOS FACTORES ESTAN AL NIVEL "BAJO" SE USA EL SIBOLO (1). POR CONVENIENCIA A_0 = NIVEL INFERIOR Y A_1 = NIVEL SUPERIOR DE A (DE MANERA SIMILAR PARA LOS OTROS FACTORES). LOS SIMBOLOS a, b, ab Y (1) REPRESENTAN LAS OBSERVACIONES (O SU SUMA SI HAY REPLICAS), PARA LAS COMBINACIONES NIVEL-TRATAMIENTO CORRESPONDIENTES.

COMBINACIONES NIVEL-TRATAMIENTO
EN UN EXPERIMENTO 2^2

	A_0	A_1
B_0	(1)	a
B_1	b	ab

EL <u>EFECTO PROMEDIO</u> DE A PARA ESTE EXPERIMENTO 2^2 PUEDE ESTIMARSE COMO: $A = \frac{1}{2}\{(ab-b) + [a-(1)]\}$, SIENDO ESTA LA DIFERENCIA PROMEDIO DEL NIVEL SUPERIOR E INFERIOR DE A, TOMANDO PRIMERO EL NIVEL SUPERIOR DE B Y DESPUES EL INFERIOR. OCASIONALMENTE SE OMITE EL COEFICIENTE 1/2, CON LO CUAL SE ESTIMA EL <u>EFECTO TOTAL</u> DE A.

DE MANERA SIMILAR, AL EFECTO PROMEDIO DE B SERA:

$$B = \frac{1}{2}\{(ab-a) + [b-(1)]\}$$

LA <u>INTERACCION AB</u> SE DEFINE COMO LA <u>DIFERENCIA PROMEDIO</u>; ESTO ES, EL EFECTO DE A AL NIVEL SUPERIOR DE B MENOS EL EFECTO DE A AL NIVEL INFERIOR DE B:

$$AB = \frac{1}{2}\{(ab-b) - [a-(1)]\}$$

ESTAS RELACIONES PUEDEN GENERARSE COMO SIGUE (CONSIDERANDO LOS EFECTOS TOTALES Y REEMPLAZANDO (1) POR 1)

$$A : (a-1)(b+1) = ab - b + a - (1)$$

$$B : (a+1)(b-1) = ab - a + b - (1)$$

$$AB : (a-1)(b-1) = ab - a - b + (1)$$

PARA DETERMINAR CUANDO EL RENDIMIENTO DE UN FACTOR PARTICULAR SE SUMA O SE RESTA, SE FORMA EL PRODUCTO DE BINOMIOS FORMADOS POR CADA UNA DE LAS LETRAS MENOS 1 SI EL FACTOR ESTA INCLUIDO EN LA INTERACCION (O EFECTO), O MAS 1 SI EL FACTOR NO ESTA INCLUIDO.

EJEMPLO

EN UN PROBLEMA DE TRES FACTORES A, B Y C (2^3), LAS EXPRESIONES PARA LOS EFECTOS E INTERACCIONES TOTALES, (SIN CONSIDERAR EL FACTOR MULTIPLICATIVO) SON:

$$A : (a-1)\ (b+1)\ (c+1) = abc + ab + ac - bc + a - b - c - (1)$$
$$B : (a+1)\ (b-1)\ (c+1) = abc + ab - ac + bc - a + b - c - (1)$$
$$C : (a+1)\ (b+1)\ (c-1) = abc - ab + ac + bc - a - b + c - (1)$$

$$AB : (a-1)\ (b-1)\ (c+1) = abc + ab - ac - bc - a - b + c + (1)$$
$$AC : (a-1)\ (b+1)\ (c-1) = abc - ab + ac - bc - a + b - c + (1)$$
$$BC : (a+1)\ (b-1)\ (c-1) = abc - ab - ac + bc + a - b - c + (1)$$
$$ABC : (a-1)\ (b-1)\ (c-1) = abc - ab - ac + bc + a + b + c - (1)$$

COMBINACIONES DE TRATAMIENTOS
DE UN EXPERIMENTO 2^3

	A_o		A_1	
	B_o	B_1	B_o	B_1
c_o	(1)	b	a	ab
c_1	c	bc	ac	abc

NOTACION PARA CALCULAR LOS EFECTOS

LA TABLA QUE SE REPRESENTA MAS ADELANTE SIRVE PARA CALCULAR LOS EFECTOS DE CADA FACTOR, EN LAS COLUMNAS SE TIENEN LOS EFECTOS PRINCIPALES Y LAS INTERACCIONES (I INDICA EL TOTAL PRODUCIDO POR EL EXPERIMENTO PARA CADA TRATAMIENTO); LOS

RENGLONES TIENEN LAS COMBINACIONES DE LOS TRATAMIENTOS.

EL CUERPO DE LA TABLA SE HACE CON SIGNOS + Y -. PARA CADA EFECTO LOS SIGNOS INDICAN COMO SE COMBINA CADA TRATAMIENTO. POR EJEMPLO, ABAJO DE I HAY PUROS "+", EL CUAL ESTABLECE QUE EL GRAN TOTAL ES LA SUMA DE TODOS LOS RENDIMIENTOS. EL EFECTO A TIENE EN SUS 8 RENGLONES UN SIGNO "+" DONDE EL TRATAMIENTO INCLUYE LA LETRA "a" (O SEA EL NIVEL SUPERIOR), Y "-" DONDE LA "a" NO ESTA INCLUIDA.

CUANDO LOS SIGNOS DE LOS EFECTOS PRINCIPALES SE HAN INCLUIDO EN LA TABLA, LOS SIGNOS DE LAS COLUMNAS RESTANTES SE OBTIENEN MEDIANTE LA MULTIPLICACION ALGEBRAICA DE ALGUNAS DE LAS COLUMNAS PRECEDENTES. POR EJEMPLO, LOS SIGNOS DE AB SON EL PRODUCTO DE LOS SIGNOS DE A Y B, RENGLON POR RENGLON.

SIGNOS ALGEBRAICOS PARA CALCULAR LOS EFECTOS

	EFECTO							
TRATAMIENTO	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

PROPIEDADES DE LA TABLA

1. A EXCEPCION DE LA COLUMNA I, EL NUMERO DE SIGNOS "+" Y "-" ES EL MISMO EN CADA COLUMNA.

2. LA SUMA DE PRODUCTOS DE SIGNOS DE DOS COLUMNAS CUALESQUIERA ES CERO; ENTONCES, EL PRODUCTO TIENE IGUAL NUMERO DE SIGNOS MAS Y MENOS.

3. EL PRODUCTO DE DOS COLUMNAS CUALESQUIERA GENERA UNA COLUMNA INCLUIDA EN LA TABLA. POR EJEMPLO, $AB \times B = A$; $ABC \times AB = C$, ETC.

ESTAS PROPIEDADES ESTAN IMPLICADAS POR LA ORTOGONALIDAD (QUE INDICA QUE SI UNA INTERACCION ES NULA ENTONCES LOS EFECTOS SON INDEPENDIENTES).

NOTESE QUE LOS PRODUCTOS $AB \times B + AB^2 = A$

$$ABC \times BC = AB^2C^2 = A, \text{ ETC.}$$

TENIENDOSE PRODUCTOS MODULO 2, O SEA, EL EXPONENTE PUEDE SER SOLAMENTE 0 O 1; SI PASA DE 2 SE HACE 0.

ALGORITMO DE YATES

LOS CALCULOS Y LAS PRUEBAS PARA OBTENER LOS EFECTOS TOTALES Y LAS INTERACCIONES ENTRE LOS FACTORES, SE PUEDEN HACER CON UN PROCEDIMIENTO DESARROLLADO POR FRANK YATES; ESTE SERA ILUSTRADO MEDIANTE UN EJEMPLO

EJEMPLO

LA SIGUIENTE TABLA MUESTRA LAS COSECHAS OBTENIDAS (EN KGS), EN PARCELAS EXPERIMENTALES PARA EL CULTIVO DE PAJA, LOS CUALES RECIBIERON TRES TIPOS DE FERTILIZANTES MEZCLADOS CON NITRATO (n), FOSFATO (p) Y POTASIO (k). EN EL EXPERIMENTO SE TOMARON 3 REPLICAS DE LAS 8 COMBINACIONES POSIBLES DE LOS FERTILIZANTES, DANDO UN TOTAL DE 24 PARCELAS EN TOTAL.

PLAN EXPERIMENTAL Y GENERACIONES OBTENIDAS

				TOTALES/BLOQUE
pk 36.9	k 31.4	nk 43.6	n 33.8	
np 43.3	(1) 28.1	p 31.9	npk 41.8	290.8
npk 41.0	(1) 31.8	pk 36.5	p 33.0	
nk 42.8	np 35.2	k 35.9	n 35.4	291.6
np 35.0	k 29.6	pk 38.0	nk 36.5	
p 32.1	n 38.3	(1) 34.2	npk 41.5	285.2
			GRAN TOTAL	867.6

LOS TOTALES POR TRATAMIENTO SE DAN EN LA SIGUIENTE TABLA

GENERACIONES DE PAJA

	N_0		N_1	
	P_o	P_1	P_o	P_1
K_o	94.1	97.0	107.5	113.5
K_1	96.9	111.4	122.9	124.3

EL PRIMER PASO ES ESTIMAR LOS EFECTOS DE LOS TRATAMIENTO A PARTIR DE LAS PRODUCCIONES. EN LA SIGUIENTE TABLA SE HAN ARREGLADO LAS PRODUCCIONES TOTALES (COLUMNA 1) POR TRATAMIENTO. EL ORDEN DE LAS COMBINACIONES DE LOS TRATAMIENTOS DEBE MANTENERSE SIEMPRE DE MANERA QUE CADA FACTOR INTRODUCIDO SE SIGUE CON TODAS LAS COMBINACIONES DE ÉL Y DE LOS FACTORES PREVIAMENTE INTRODUCIDOS.

ALGORITMO DE YATES PARA UN EXPERIMENTO 2^3

TRATAMIENTO	PRODUCCION	(1)	(2)	(3)	EFECTO	MEDIA	SS
(1)	94.1	201.6	412.1	867.6	TOTAL		
n	107.5	210.5	455.5	68.8	N	5.73	197.2
p	97.0	219.8	29.9	24.8	P	2.07	25.6
np	113.5	235.7	38.9	-10.0	NP	-0.83	4.2
k	96.9	13.4	8.9	43.4	K	3.62	78.5
nk	122.9	16.5	15.9	9.0	NK	0.75	3.4
pk	111.4	26.0	3.1	7.0	PK	0.58	2.0
npk	124.3	12.9	-13.1	-16.2	NPK	-1.35	10.9
						TOTAL =	321.8

LA COLUMNA DE PRODUCCIONES SE USA PARA CALCULAR LA COLUMNA (1), ESTA A SU VEZ PARA CALCULAR LA (2), Y ASI SUCESIVAMENTE. LOS CUATRO PRIMEROS TERMINOS DE (1) SE ENCUENTRAN SUMANDO POR PAREJAS, DE ARRIBA A ABAJO, LAS PRODUCCIONES. POR EJEMPLO, $201.6 = 94.1 + 107.5$. LOS CUATRO ULTIMOS TERMINOS DE LA MISMA

COLUMNA SE ENCUENTRAN CALCULANDO LA DIFERENCIA POR PAREJAS DE LAS GENERACIONES, RESTANDO EL NUMERO SUPERIOR DEL INFERIOR EN CADA CASO; POR EJEMPLO, 107.5 - 94.1 = 13.4, ETC. DE MANERA IDENTICA SE ENCUENTRAN LOS VALORES DE LAS COLUMNAS (2) Y (3). DEBERAN DESARROLLARSE TANTAS COLUMNAS DE ESTAS COMO NUMERO DE FACTORES HAY EN EL EXPERIMENTO (3 EN NUESTRO EJEMPLO). LA COLUMNA (3) DA EL EFECTO TOTAL DEL FACTOR (O INTERACCION DESIGNADO CON LA LETRA MINUSCULA. PARA OBTENER EL EFECTO PROMEDIO DIVIDIMOS LOS ELEMENTOS DE (3) ENTRE EL NUMERO DE DIFERENCIAS QUE HAY EN CADA EFECTO TOTAL (4 EN ESTE CASO) POR EL NUMERO DE REPLICAS $2^{n-1}r$ (3 EN ESTE CASO), O SEA 3 x 4 = 12 (QUE ES EQUIVALENTE A LA MITAD DEL NUMERO DE PARCELAS). ESTOS VALORES SE MUESTRAN EN LA CUARTA COLUMNA.

HAY VERIFICACIONES PARA LOS CALCULOS:

a) LA SUMA DE LA COLUMNA (i) ES IGUAL A 2^i VECES LA GENERACION TOTAL DE LOS TRATAMIENTOS QUE TENGAN LOS PRIMEROS i FACTORES AL NIVEL "ALTO"; POR EJEMPLO, LA SUMA DE LA COLUMNA (3) ES 8 VECES EL TOTAL GENERADO DE npk, ES DECIR, 8 x 124.3 = 994.4; LA SUMA DE LA COLUMNA (2) ES 4 VECES EL TOTAL GENERADO POR np Y npk, O SEA, 951.2 = (113.5 + 124.3) x 4, ETC.

b) EL TERMINO QUE ENCABEZA LA COLUMNA (3) ES EL GRAN TOTAL

c) LA SUMA DE CUADRADOS DE LOS OTROS TERMINOS DE LA COLUMNA (3) DIVIDIDA ENTRE EL NUMERO DE PARCELAS (24) DA LA SUMA DE CUADRADOS DE LOS TRATAMIENTOS:

$$SST = (68.8^2 + 24.8^2 + \ldots + 7.0^2 + 16.2^2)/24 = 321.9$$

DE LOS RESULTADOS ANTERIORES PUEDEN DERIVARSE LAS SIGUIENTES CONCLUSIONES:

1. LOS EFECTOS N, P Y K SON TODOS POSITIVOS.

2. LOS EFECTOS NK Y PK SON POSITIVOS, INDICANDO QUE LA APLICACION DE POTASIO TIENDE A INCREMENTAR LOS EFECTOS DEL NITRATO Y DEL FOSFATO.

3. EL EFECTO NP ES NEGATIVO, MOSTRANDO QUE LA PRESENCIA DE NITRATO REDUCE EL EFECTO DEL FOSFATO. DE HECHO, EN PRESENCIA DE NITRATO EL EFECTO MEDIO DEL FOSFATO SE REDUCE A $2.07 - 0.83 = 1.24$.

4. LA INTERACCION NPK ES NEGATIVA, INDICANDO QUE CUANDO EL POTASIO ESTA PRESENTE LA INTERACCION NP SE REDUCE Y QUE EL EFECTO MEDIO DEL FOSFATO SE REDUCE AUN MAS. EL EFECTO MEDIO DEL FOSFATO EN PRESENCIA DE NITRATO Y POTASIO ES $2.07 - 0.83 + 0.58 - 1.35 = 0.47$.

5. LA CONCLUSION SOBRE TODO ESTO ES QUE EL NITRATO Y EL POTASIO DAN EFECTOS BENEFICOS, ESPECIALMENTE CUANDO SE APLICAN JUNTOS; POCO SE GANA APLICANDO FOSFATO SI EL NITRATO ESTA PRESENTE Y ESPECIALMENTE SI EL NITRATO ESTA TAMBIEN PRESENTE.

6. POSIBLEMENTE SE HUBIERA LLEGADO A ESTAS MISMAS CONCLUSIONES

- INSPECCIONANDO LAS PRODUCCIONES MEDIAS, PERO PARA MAS DE TRES FACTORES ESTA CONCLUSION ES MAS DIFICIL, AUN CUANDO LA INSPECCION DE LOS EFECTOS E INTERACCIONES MEDIAS SEA AUN POSIBLE.

ES IMPORTANTE CONOCER CUALES DE LOS EFECTOS E INTERACCIONES MEDIOS SON SIGNIFICATIVOS; ES DECIR, QUE TAN CONFIABLES SON ESAS CARACTERISTICAS DEL EXPERIMENTO. PARA ESTO SE REQUIERE CALCULAR ERRORES ESTANDAR (A PESAR DE QUE LA MAGNITUD RELATIVA DE LOS EFECTOS E INTERACCIONES CASI SIEMPRE DA UNA BUENA GUIA DE SU CONFIABILIDAD), Y LA TABLA DE ANALISIS DE VARIANCIA. ESTE ES UN TIPO DE ANALISIS DE BLOQUES ALEATORIZADOS CUYA TABLA ANOVA ES LA SIGUIENTE

TABLA DEL ANALISIS DE VARIANCIA

FUENTE	G. DE L.	SS	MS
BLOQUES	2	3.0	
TRATAMIENTOS	7	321.9	
ERROR	14	124.6	8.90
TOTAL	23	449.5	

LOS ERRORES ESTANDAR DE LOS EFECTOS DE LOS TRATAMIENTOS PUEDEN CALCULARSE COMO SIGUE: SI S^2 ES LA VARIANCIA RESIDUAL POR UNIDAD, ENTONCES LOS ERRORES ESTANDAR PARA LOS EFECTOS TOTALES Y MEDIOS SE DEFINEN ASI:

PARA LOS EFECTOS TOTALES: $S_t = \sqrt{2^n r S^2}$

PARA LOS EFECTOS MEDIOS: $S_m = \sqrt{\frac{S^2}{2^{n-2} r}}$

DONDE n = NUMERO DE FACTORES (3 EN NUESTRO CASO) Y

r = NUMERO DE REPLICAS (3 EN NUESTRO CASO).

PARA EL EJEMPLO ANTERIOR:

$$S_m = \sqrt{\frac{8.90}{2^{3-2} \times 3}} = \pm 1,22$$

USANDO LA DISTRIBUCION t CON 14 G. DE L. PARA NIVELES DE SIGNIFICANCIA DE 5 Y 1%.

$$t_{\alpha=0.05} = 2.14 \Rightarrow N.S_1 = 1.22 \times 2.14 = \pm 2.61$$

$$t_{\alpha=0.01} = 2.98 \Rightarrow N.S_2 = 1.22 \times 2.98 = \pm 3.64$$

COMPARANDO ESTOS VALORES CON LOS EFECTOS MEDIOS, SE OBSERVA QUE PARA α = 0.05 N Y K SON SIGNIFICATIVOS, MIENTRAS QUE PARA α = 0.01 N ES SIGNIFICATIVO Y K LO ES LIGERAMENTE; NINGUN OTRO EFECTO ES SIGNIFICATIVO.

OTRA FORMA DE LLEGAR A ESTAS CONCLUSIONES ES CALCULANDO LA SUMA DE CUADRADOS PARA CADA EFECTO SEPARADAMENTE. ESTO SE LOGRA ELEVANDO AL CUADRADO CADA COMPONENTE DE LA COLUMNA (3) DE LA TABLA DEL ALGORITMO DE YATES Y DIVIDIENDO ENTRE EL TOTAL DE OBSERVACIONES; POR EJEMPLO, PARA N TENEMOS $68.8^2/24 = 197.2$,

ETC. ESTOS VALORES ESTAN ANOTADOS EN LA ULTIMA COLUMNA DE ESA TABLA.

CON ESTO SE TIENE PARTICION DE LA SUMA DE CUADRADOS DE LOS TRATAMIENTOS. CON ESTOS VALORES SE PUEDE INTEGRAR LA TABLA ANOVA SIGUIENTE PARA HACER EL ANALISIS DE SIGNIFICANCIA.

TABLA ANOVA

FUENTE	G. DE L.	SS	MS	F CALCULADAS
BLOQUES	2	3.0	1.5	0.17
n	1	157.2	197.2	22.16
p	1	25.6	25.6	2.88
np	1	4.2	4.2	0.47
k	1	78.5	78.5	8.82
nk	1	3.4	3.4	0.38
pk	1	2.0	2.0	0.22
npk	1	10.9	10.9	1.22
ERROR	14	124.6	8.9	
TOTAL	23	449.5		

$F_{\alpha=0.05} = 4.60$, $F_{\alpha=0.05} = 3.74$

$F_{\alpha=0.01} = 8.86$, $F_{\alpha=0.01} = 6.51$

COMPARANDO LAS F TEORICAS CON LAS CALCULADAS SE LLEGA A LAS MISMAS CONCLUSIONES ANTERIORES.

COMO PASO FINAL PARA LA PRESENTACION DE RESULTADOS DEBERAN PREPARARSE TABLAS DE MEDIAS Y ERRORES ESTANDAR. LAS TABLAS DE MEDIAS PUEDEN CONSTRUIRSE DE LAS PRODUCCIONES DIRECTAMENTE O DE LOS EFECTOS CALCULADOS, PREFIRIENDOSE ESTO ULTIMO CUANDO HAY MUCHOS FACTORES INVOLUCRADOS.

EN EL EJEMPLO QUE SE VIENE DESARROLLANDO LA PRODUCCION MEDIA TOTAL ES

$$\bar{x} = \frac{867.6}{24} = 36.15$$

CON ESTO SE TIENE:

PRODUCCION MEDIA CON NITRATO (n) $= \bar{x} + 1/2\ N = 36.15 + 1/2(5.73) = 39.02$

PRODUCCION MEDIA SIN NITRATO $= \bar{x} - 1/2\ N = 33.28$

DE MANERA SIMILAR, PARA CONSTRUIR UNA TABLA DE DOS DIRECCIONES QUE MUESTRE LA INTERACCION DEL NITRATO Y POTASIO SE TIENE:

PRODUCCION MEDIA CON n Y k $= \bar{x} + 1/2\ (N + K + NK) = 41.20$

PRODUCCION MEDIA CON n Y SIN k $= \bar{x} + 1/2\ (N - K - NK) = 36.83$

PRODUCCION MEDIA SIN n Y CON k $= \bar{x} + 1/2\ (-N + K - NK) = 34.72$

PRODUCCION MEDIA SIN n O k $= \bar{x} + 1/2\ (-N - K + NK) = 31.85$

TABLA DE MEDIAS PARA EL NITRATO Y POTASIO

	SIN n	CON n	MEDIA
SIN k	31.85	36.83	34.34
CON k	34.72	41.20	37.96
MEDIA	33.29	39.20	36.15

RESUMEN

EL DISEÑO FACTORIAL 2^k PRUEBA k FACTORES A DOS NIVELES CADA UNO, TIENE 2^k COMBINACIONES DE POSIBLES TRATAMIENTOS Y PUEDEN HACERSE $2^k - 1$ COMPARACIONES EN FORMA DE EFECTOS PRINCIPALES E INTERACCIONES; POR EJEMPLO, CON CINCO FACTORES A, B, C, D, E; SE REQUIEREN $2^5 = 32$ COMBINACIONES DE TRATAMIENTOS Y PUEDEN HACERSE 31 COMPARACIONES COMO SIGUE:

EFECTOS PRINCIPALES, A, B, C, D, E	5
INTERACCIONES DE PRIMER ORDEN AB, AC, ETC.	10
INTERACCIONES DE SEGUNDO ORDEN ABC, ABD, ETC.	10
INTERACCIONES DE TERCER ORDEN, ABCD, ABCE, ETC.	5
INTERACCIONES DE CUARTO ORDEN, ABCDE	1
TOTAL	31

ES IMPORTANTE SEÑALAR QUE LA INTERPRETACION DE LAS INTERACCIONES DE TERCERO Y MAYOR ORDEN ES COMPLICADA Y NECESITA CONSIDERARSE CUIDADOSAMENTE A LA LUZ DE LAS OTRAS INTERACCIONES QUE PAREZCAN IMPORTANTES. USUALMENTE TALES INTERACCIONES NO REFLEJAN EFECTOS REALES.

RESULTA TAMBIEN IMPORTANTE EL COMENTARIO DE YATES (1937) AL RESPECTO: "EL EXPERIMENTADOR... DEBE EVITAR DAR ENFASIS EXAGERADO A ALGUNAS INTERACCIONES AISLADAS DE ALTO ORDEN ESTADISTICAMENTE SIGNIFICATIVAS QUE NO TENGAN SIGNIFICADO FISICO APARENTE. SI SE ESTA USANDO UN NIVEL DE SIGNIFICANCIA DE 1 EN 20 (0.05), UNO DE CADA VEINTE EFECTOS PRINCIPALES E INTERACCIONES SERA EN PROMEDIO ESTADISTICAMENTE SIGNIFICANTE, AUN CUANDO LOS TRATAMIEN-

TOS NO PRODUZCAN EFECTOS EN TODOS. TALES RESULTADOS ANOMALOS JUNTO CON LOS EFECTOS NO SIGNIFICATIVOS DEBERAN ANOTARSE Y RESERVARSE EL JUICIO HASTA QUE SE ACUMULE MAS INFORMACION".

EL ANALISIS DEL EXPERIMENTO FACTORIAL 2^k SIGUE LAS LINEAS INDICADAS EN EL EJEMPLO ANTERIOR SIENDO LOS PASOS PRINCIPALES:

a) EL ALGORITMO DE YATES SE DESARROLLA HASTA k PASOS, LOS VALORES FINALES DIVIDIDOS ENTRE LA MITAD DEL NUMERO DE OBSERVACIONES (N/2) DAN LOS EFECTOS DE LOS TRATAMIENTOS Y LAS INTERACCIONES. ESTOS PUEDEN EXAMINARSE DIRECTAMENTE.

b) EL ERROR ESTANDAR DE LOS EFECTOS Y LAS INTERACCIONES SE CALCULA CON $4S^2/N$, DONDE S^2 SE OBTIENE DEL ANALISIS DE VARIANCIA DEL EXPERIMENTO. ESTE PUEDE USARSE PARA PROBAR LA SIGNIFICANCIA DE LOS EFECTOS. SI SE DESEA UN PROCEDIMIENTO ALTERNATIVO, LA SUMA DE CUADRADOS DE LOS TRATAMIENTOS PUEDE PARTIRSE ENTRE LOS COMPONENTES CORRESPONDIENTES A LOS EFECTOS PRINCIPALES E INTERACCIONES.

c) EL ANALISIS TERMINA CONSTRUYENDO LAS TABLAS DE MEDIAS PARA LOS EFECTOS SIGNIFICATIVOS, LAS CUALES PUEDEN CONSTRUIRSE DIRECTAMENTE O USANDO LOS EFECTOS ESTIMADOS.

EJEMPLO

EL DESARROLLO DE UN PROCESO DE FERMENTACION INDUSTRIAL USUALMENTE COMIENZA CON UN ESTUDIO DE LABORATORIO DE LOS REQUERIMIENTOS FISIOLOGICOS DE LOS MICROORGANISMOS INMISCUIDOS. EN UNO DE TALES ESTUDIOS SE ENCONTRO QUE UNA SUSTANCIA UTIL LA SEGREGA UNA

ESPECIE DE MOHO CUANDO CRECE EN UN MEDIO DE CULTIVO LIQUIDO POR LO QUE SE DESEO INCREMENTAR LA PRODUCCION. PARA LA FORMACION DE LA SUSTANCIA SE SABÍA QUE DEPENDIA PRINCIPALMENTE DE LOS NIVELES DE DOS INGREDIENTES EN EL MEDIO DE CULTIVO, Y DE LA TEMPERATURA, LA AEREACION, EL PH, Y LA EDAD EN QUE EL CULTIVO ERA LOGRADO.

SE SOSPECHO QUE CUATRO DE ESOS SEIS FACTORES PODIAN SER INDEPENDIENTES. PARA PROBAR ESTO SE DESARROLLO UN EXPERIMENTO FACTORIAL 2^4 CON DOS INGREDIENTES EN EL MEDIO DE CULTIVO (X_1,X_2) Y DOS FACTORES AMBIENTALES (X_4,X_5); PARA CADA TRATAMIENTO SE PREPARARON DUPLICADOS. LOS DATOS PRESENTADOS EN LA SIGUIENTE TABLA ESTAN CODIFICADOS. LOS EFECTOS SE REPORTARON COMO UNIDADES PRODUCIDAS (UP) POR UNIDAD DE DISEÑO (UD). HAY 2 REPLICAS PARA CADA UNA DE LAS COMBINACIONES DE LOS FACTORES.

EXPERIMENTO DE FERMENTACION 2^4

		X_1	- 1		+ 1	
X_4	X_5	X_2	- 1	+ 1	- 1	+ 1
- 1	- 1		32.7	50.4	70.6	115
			19.3	89.8	84.5	108.6
	+ 1		20.2	94.1	76.1	133.6
			29.9	96.5	73.3	131.6
+ 1	- 1		50.0	72.6	104.2	81.3
			52.1	76.9	103.4	88.2
	+ 1		50.5	91.8	78.6	108.3
			49.1	86.9	74.1	108.3

ALGORITMO DE YATES PARA EL PROBLEMA DE LA FERMENTACION

TRATAMIENTO	GENERACION					EFECTO MEDIO		G. DE L.		
(1)	(2)	(3)	(4)	(5)	(6)	(7) = (6)/16	$(6)^2/32$ = (8)		F CAL.	***
(1)	52	207.1	610.9	1239.6	2542.5					
X_1	155.1	403.8	628.7	1302.9	536.9	33.56	9008.2	1	448.17	***
X_2	180.2	309.7	655.3	272.0	605.3	37.83	11449.6	1	569.63	***
X_1X_2	223.6	319.0	647.6	264.9	-185.1	-11.57	1070.7	1	53.27	***
X_4	102.1	199.5	146.5	206.0	10.1	0.63	3.2	1	0.16	
X_1X_4	207.6	455.8	125.5	399.3	-103.9	-6.49	337.4	1	16.79	***
X_2X_4	149.5	252.3	173.9	-145.2	-300.7	-18.79	2825.6	1	140.58	***
$X_1X_2X_4$	169.5	395.3	91.0	-39.9	-16.3	-1.02	8.3	1	0.41	
X_5	50.1	103.1	196.7	17.8	63.3	3.96	125.2	1	6.23	*
X_1X_5	149.4	43.4	9.3	-7.7	-7.1	-0.44	1.6	1	0.08	
X_2X_5	190.6	105.5	256.3	-21.0	193.3	12.08	1167.7	1	58.9	***
$X_1X_2X_5$	265.2	20.0	143.0	-82.9	105.3	6.58	346.5	1	17.22	***
X_4X_5	99.6	99.3	-59.7	-187.4	-25.5	-1.59	20.3	1	1.01	***
$X_1X_4X_5$	152.7	74.6	-85.5	-113.3	-61.9	-3.87	119.7	1	5.96	**
$X_2X_4X_5$	178.7	53.1	-24.7	- 25.8	74.1	4.63	171.6	1	8.54	
$X_1X_4X_2X_5$	216.6	37.9	-15.2	9.5	35.3	2.21	38.9	1	1.94	
RESIDUAL							20.10	16		

ANALISIS DE VARIANCIA

FUENTE	G. DE L.	S.S	MS	F_c	$F_{0.05,15,16}$
BLOQUES	$\cancel{1} \Rightarrow 0$	0.20			
TRATAMIENTOS	15.	26694.50	1779.63	88.28* >	2.35
RESIDUAL	16	321.58	20.10		
TOTAL	31	27016.08			

DE TABLAS:

$F_{0.95,1,16} = 4.49$; $F_{0.99,1,16} = 8.53$; $F_{0.999,1,16} = 16.12$

ERROR ESTANDAR = $\sqrt{\frac{20.10}{4 \times 2}} = \pm 1.59$

$t_{16,0.95} = 2.12$, $t_{16,0.99} = 2.92$, $t_{16,0.999} = 4.01$, DE DONDE

$N.S_{0.95} = 2.12 \times 1.59 = \pm 3.37$

$N.S_{0.99} = 2.92 \times 1.59 = \pm 4.64$

$N.S_{0.999} = 4.01 \times 1.59 = \pm 6.38$

COMPARANDO LOS EFECTOS MEDIOS CON ESTOS NIVELES DE SIGNIFICANCIA Y LAS ESTADISTICAS F CALCULADAS CON LAS F TEORICAS, SE OBSERVA LA COINCIDENCIA DE RESULTADOS PARA LOS EFECTOS SIGNIFICATIVOS INDICADOS PARA LOS ASTERISCOS SITUADOS EN LA ULTIMA COLUMNA DE LA TABLA.

LAS CONCLUSIONES A LAS QUE SE LLEGA SON;

a) LOS DOS INGREDIENTES EN EL MEDIO DE CULTIVO (x_1 Y x_2) ACTUANDO SEPARADAMENTE FAVORECEN LA REPRODUCCION DE LA SUBSTANCIA; SIN EMBARGO, UNO EN PRESENCIA DEL OTRO LA REDUCEN.

b) SE OBSERVA QUE LOS EFECTOS PRINCIPALES DE x_1 Y x_2 SE TOMAN EN CUENTA EN LA MAYORIA DE LAS DIFERENCIAS ENTRE LAS PREPARACIONES.

c) LA INTERACCION MAS NEGATIVA ES POSIBLE, CIERTOS REQUERIMIENTOS NUTRICIONALES DEL MOHO PUEDEN ALIMENTARSE POR CUALQUIERA DE LOS INGREDIENTES.

d) ES SORPRENDENTE ENCONTRAR QUE LOS FACTORES AMBIENTALES x_4 Y x_5 TIENEN POCO EFECTO DIRECTO, PERO EJERCEN SU INFLUENCIA A TRAVES DE SU INTERACCION CON X_2 DE MANERA INVERSA.

e) IDEM QUE d) PERO EN MENOS GRADO CON X_1

f) NINGUNO DE LOS CUATRO FACTORES ES INDEPENDIENTE DE LOS OTROS, EN EL SENTIDO DE AFECTAR LA GENERACION DE MANERA PURAMENTE ADITIVA.

EJEMPLO

EN UNA PLANTA PILOTO SE OBTUVIERON LOS SIGUIENTES DATOS:

PRUEBA No.	TEMPERATURA °C	CONCENTRACION %	CATALIZADOR A o B	RESULTADO gramos
1	160	20	A	60
2	180	20	A	72
3	160	40	A	54
4	180	40	A	68
5	160	20	B	52
6	180	20	B	83
7	160	40	B	45
8	180	40	B	80

A. CALCULAR LOS EFECTOS PRINCIPALES Y LAS INTERACCIONES

B. REALICE EL ANALISIS DE VARIANCIA

LOS DATOS ANTERIORES SE PUEDEN REESCRIBIR EN LA SIGUIENTE TABLA:

	FACTOR A			
FACTOR C	CATALIZADOR A		CATALIZADOR B	
TEMPERATURA	FACTOR B		FACTOR B	
	CONC.=20%	CONC.=40%	CONC.=20%	CONC.=40%
160°	60 (1)	54 b	52 a	45 ab
180°	72 c	68 bc	83 ac	80 abc

POR LO TANTO, SE TIENE UN EXPERIMENTO FACTORIAL 2^3. APLICANDO LA ECUACION GENERAL, LOS EFECTOS PRINCIPALES E INTERACCIONES ESTAN DADAS POR:

$$\begin{aligned}
\text{EFECTO}\quad A&: (a-1)(b+1)(c+1) = abc+ab+ac-bc+a-b-c-(1)\\
B&: (a+1)(b-1)(c+1) = abc+ab-ac+bc-a+b-c-(1)\\
C&: (a+1)(b+1)(c-1) = abc-ab+ac+bc-a-b+c-(1)\\
AB&: (a-1)(b-1)(c+1) = abc+ab-ac-bc-a-b+c+(1)\\
AC&: (a-1)(b+1)(c-1) = abc-ab+ac-bc-a+b-c+(1)\\
BC&: (a+1)(b-1)(c-1) = abc-ab-ac+bc+a-b-c+(1)\\
ABC&: (a-1)(b-1)(c-1) = abc-ab-ac-bc+a+b+c-(1)
\end{aligned}$$

DONDE LAS COMBINACIONES DE TRATAMIENTOS SE INDICAN EN LA MISMA TABLA ANTERIOR. SUSTITUYENDO SE TIENE QUE:

$$\begin{aligned}
\text{EFECTO}\quad A&: 80+45+83-68+52-54-72-60 = 6\\
B&: 80+45-83+68-52+54-72-60 = -20\\
C&: 80-45+83+68-52-54+72-60 = 92\\
AB&: 80+45-83-68-52-54+72+60 = 0\\
AC&: 80-45+83-68-52+54-72+60 = 40\\
BC&: 80-45-83+68+52-54-72+60 = 6\\
ABC&: 80-45-83-68+52+54+72-60 = 2
\end{aligned}$$

Y, POR LO TANTO, LAS SUMAS DE CUADRADOS CORRESPONDIENTES SERAN:

$$SSX = \frac{(\text{efecto } X)^2}{n2^k}$$

ES DECIR:

$$SSA = \frac{6^2}{8} = 4.5$$

$$SSB = \frac{(-20)^2}{8} = 50$$

$$SSC = \frac{92^2}{8} = 1058$$

$$SSAB = \frac{0}{8} = 0$$

$$SSAC = \frac{40^2}{8} = 200$$

$$SSBC = \frac{6^2}{8} = 4.5$$

$$SSABC = \frac{2^2}{8} = 0.5$$

POR OTRA PARTE, LA SUMA DE CUADRADOS TOTAL SERA:

$$SST = \Sigma\Sigma\Sigma\, X_{ijk}^2 - \frac{(\Sigma\Sigma\Sigma X_{ijk})^2}{n2^k}$$

$$\Sigma\Sigma\Sigma\, X_{ijk} = 60+54+52+45+72+68+83+80 = 514$$

$$n2^k = 8; \quad (\Sigma\Sigma\Sigma\, X_{ijk})^2/nk = 514^2/8 = 33,024.50$$

$$\Sigma\Sigma\Sigma(X_{ijk})^2 = 60^2+54^2+52^2+45^2+72^2+68^2+83^2+80^2 = 34,342$$

POR TANTO:

$$SST = 34,342 - 33024.50 = 1,317.5$$

Y:

SSE = SST-SSA-SSB-SSC-SSAB-SSAC-SSBC-SSABC

SSE = 1,317.5-4.5-50-1,058-0-200-4.5-0.5 = 0

LO CUAL COMPRUEBA LOS RESULTADOS ANTERIORES, PUESTO QUE EN UN EXPERMIENTO 2^k CON UNA SOLA REPLICA ES IMPOSIBLE CALCULAR UN VALOR DE MSE (YA QUE SSE = 0 Y LOS GRADOS DE LIBERTAD $n2^k(n-1) = 2^k(1-1) = 0$).

SE ACOSTUMBRA, EN ESTE CASO, CONSIDERAR LA SUMA DE CUADRADOS RESIDUAL COMO LA SUMA DE LAS INTERACCIONES; PARA UN EXPERIMENTO 2^3 COMO ESTE, DONDE LOS EFECTOS DE ESTAS PUEDEN CONSIDERARSE NO SIGNIFICATIVOS, SE TOMAN TODAS LA INTERACCIONES, <u>SI SUS SUMAS DE CUADRADOS NO SON MUY GRANDES.</u>

DE LA INSPECCION DE LAS SUMAS DE CUADRADOS DE LAS INTERACCIONES, RESULTA CLARO QUE LA SSAC ES UN ORDEN DE MAGNITUD COMPARABLE A LOS DE LOS EFECTOS PRINCIPALES, POR LO QUE <u>SE CONSIDERA CONVENIENTE NO INCLUIRLA EN LA SUMA DE CUADRADOS RESIDUAL.</u>

DE ACUERDO A LO ANTERIOR EL ANALISIS DE VARIANCIA SERIA:

SSE = SSAB+SSBC+SSABC = 0+4.5+0.5 = 5

LA TABLA DE ANALISIS DE VARIANCIA QUEDARIA:

FUENTE DE VARIACION	SS	G. DE L.	MS	F
A	4.5	1	4.5	2.7
B	50	1	50	30
C	1058	1	1058	634.8
AC	200	1	200	120
RESIDUAL	5	3	1.6667	
TOTAL	1317.5			

COMO $F_{0.05,1,3} = 10.13$, RESULTAN SIGNIFICATIVOS, CON $\alpha = 5\%$, LOS EFECTOS DEL FACTOR B(CONCENTRACION), LOS DEL C(TEMPERATURA) Y LA INTERACCION AC.

COMPROBACION CON EL ALGORITMO DE YATES.

APLICANDO EL ALGORITMO DE YATES SE OBTIENE LA SIGUIENTE TABLA:

COMBINACION TRATAMIENTOS	DATOS (1)	(2)	(3)	(4)	EFECTO PROMEDIO (4)÷4	SUMA CUADRADOS $(4)^2 \div 8$
(1)	60	112	211	514	I:128.5	—
a	52	99	303	6	A:1.5	4.5
b	54	155	−17	−20	B: −5	50
ab	45	148	23	0	AB:0	0
c	72	−8	−13	92	C: 23	1058
ac	83	−9	−7	40	AC: 10	200
bc	68	11	−1	6	BC:1.5	4.5
abc	80	12	1	2	ABC:0.5	0.5
TOTAL	514					

OBSERVANDO LAS SUMAS DE CUADRADOS SE COMPRUEBAN LAS OBTENIDAS CON EL PROCEDIMIENTO NORMAL; EL RESTO DE LOS CALCULOS SE EFECTUARIA IGUAL.

EJEMPLO

CONSIDEREMOS EL EXPERIMENTO 2^4, CON UNA SOLA REPLICA, INDICADO EN LA SIGUIENTE TABLA:

	A_o				A_1			
	B_o		B_1		B_o		B_1	
	C_o	C_1	C_o	C_1	C_o	C_1	C_o	C_1
D_o	45 (1)	68 c	48 b	80 bc	71 a	60 ac	65 ab	65 abc
D_1	43 d	75 dc	45 db	70 dcb	100 ad	86 adc	104 dab	96 dacb

SOLUCION

DE ACUERDO A LAS EXPRESIONES GENERALES; LOS EFECTOS PRINCIPALES ESTARAN DADOS POR:

$$SSA = \frac{1}{n2^4}\left[(a-1)(b+1)(c+1)(d+1)\right]^2$$

$$= \frac{1}{16}\left[abcd - cbd + acd - cd - d + ad - bd + abd + abc - cb + ac - \ldots\right.$$
$$\left.\ldots - c - 1 + a - b + ab\right]^2$$

SUSTITUYENDO VALORES:

$$SSA = \frac{1}{16}\left[96 - 70 + 86 - 75 - 43 + 100 - 45 + 104 + 65 - 80 + 60 - 68 - 45 + \ldots\right.$$
$$\left.+ 71 - 48 + 65\right]^2 = (173)^2/16 = 1870.56$$

SIMILARMENTE SE OBTIENEN:

SSB = 39.06, SSC = 390.06, SSD = 855.56

PARA LAS INTERACCIONES DE 2° ORDEN:

$$SSAB = \frac{1}{16}\left[(a-1)(b-1)(c+1)(d+1)\right]^2$$

$$= \left[abcd-bcd-acd+cd+abd-bd-ad+d+abc-bc-ac+c+ab-b-a+1\right]^2/16$$

$$= \left[96-70-86+75+104-45-100+43+65-80-60+68+65-48-71+45\right]^2/16$$

$$SSAB = (1)^2/16 = 0.06$$

SIMILARMENTE:

SSAB=0.06, SSAC=1314.06, SSAD=1105.56, SSBC=22.56, SSBD=0.56, SSCD=5.06

SE DESPRECIARAN EN ESTE CASO EFECTOS DE ORDEN MAYOR.

POR OTRA PARTE, EL PROMEDIO GLOBAL VALE:

$$\bar{X}_{...} = \frac{\sum\sum\sum X_{ijk}}{n2^k} = \frac{1}{16}\left[45+68+48+\dots+104+96\right] = 1121/16 = 70.06$$

$$n2^k\bar{X}_{...}^{\,2} = 78534.458$$

POR TANTO:

$$SST = (45^2+68^2+48^2+\dots+104^2+96^2) - 78534.458 = 5730.94$$

$$SSE = 5730.94 - 1870.56 - 39.06 - 390.06 - 855.56 - 0.06 - 1314.06 -$$

$$- 1105.56 - 22.56 - 0.56 - 5.06 = 127.84$$

LOS GRADOS DE LIBERTAD TOTALES SON: $n2^k - 1 = 16 - 1 = 15$

COMO SE CONSIDERAN 4 EFECTOS PRINCIPALES Y 6 INTERACCIONES, EL ERROR DEBE TENER 15 − 10 = 5 g. DE l.

LA TABLA RESUMEN DE ANALISIS DE VARIANCIA ES:

FUENTE DE VARIACION	SS	G. DE L.	MS	F
A	1870.56	1	1870.56	73.15
B	39.06	1	39.06	1.53
C	390.06	1	390.06	15.25
D	855.56	1	855.56	33.46
AB	0.06	1	0.06	0.002
AC	1314.06	1	1314.06	51.39
AD	1105.56	1	1105.56	43.24
BC	22.56	1	22.56	0.88
BD	0.56	1	0.56	0.02
CD	5.06	1	5.06	0.198
ERROR	127.84	5	25.57	
TOTAL	5730.94	15		

CALCULO USANDO EL ALGORITMO DE YATES.

COMBINACION DE TRATAM.	DATOS (1)	(2)	(3)	(4)	(5)	EFECTO	(6) EFECTO PROMEDIO (5) ÷ 8	(7) SS $(6)^2 \div 16$
(1)	45	116	229	502	1127	1		
a	71	113	273	619	169	A	21.125	1785.06
b	48	128	292	16	25	B	3.125	39.06
ab	65	145	327	153	1	AB	0.125	0.0625
c	68	143	43	14	79	C	9.875	390.06
ac	60	149	-23	11	-145	AC	-18.125	1314.06
bc	80	161	116	-16	19	BC	2.375	22.563
abc	65	166	37	17	15	ABC	1.875	14.062
d	43	26	-3	44	117	D	14.625	855.56
ad	100	17	17	35	137	AD	17.125	1173.06
bd	45	-8	6	-66	-3	BD	0.375	0.5625
abd	104	-15	5	-79	33	ABD	4.125	68.06
cd	75	57	-9	20	-9	CD	1.125	5.063
acd	86	59	-7	-1	-13	ACD	1.625	10.563
bcd	70	11	2	2	-21	BCD	2.625	27.563
abcd	96	26	15	13	11	ABCD	1.375	7.5625

15. PRUEBAS DE HIPOTESIS E INTERVALOS DE CONFIANZA EN REGRESION LINEAL

SI EL MODELO QUE RELACIONA A Y CON X ES LINEAL, ENTONCES

$$Y = \beta X + \alpha$$

SI NO SE CONOCEN β Y α, ES NECESARIO ESTIMARLOS CON BASE EN UNA MUESTRA, CON LO CUAL SE OBTIENE

$$\tilde{Y} = mX + b$$

EN DONDE m ES EL ESTIMADOR DE M, Y b, EL DE B. SEA $\sigma^2_{y|x}$ LA VARIANCIA DE LA ESTIMACION DE Y CON BASE EN X.

SE PUEDE DEMOSTRAR QUE, SI SE CONOCE $\sigma^2_{y|x}$, ENTONCES:

$$\mathrm{Var}(m) = \sigma^2_m = \sigma^2_{y|x} / \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sigma^2_{y|x}/nS^2_x$$

$$\mathrm{Var}(b) = \sigma^2_b = \sigma^2_{y|x}/n + \frac{\bar{x}^2 \sigma^2_{y|x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sigma^2_{y|x}\left(\frac{1}{n} + \frac{\bar{x}^2}{nS^2_x}\right)$$

$$\mathrm{Var}(mX + b) = \sigma^2_{y|x}/n + \frac{\sigma^2_{y|x}(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sigma^2_{\tilde{y}} = \sigma^2_{y|x}\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{nS^2_x}\right)$$

SI $\sigma^2_{y|x}$ NO SE CONOCE, SE PUEDE OBTENER UNA ESTIMACION INSESGADA DE ELLA MEDIANTE LA ECUACION

$$S^2_{y|x} = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

INTERVALOS DE CONFIANZA: $\sigma_{y|x}$ CONOCIDA

a. PARA LA ORDENADA EN EL ORIGEN, α,

$$b \pm z_c \sigma_b$$

DONDE $z_c = P(Z < z_c) = 1 - \alpha/2$; α = NIVEL DE SIGNIFICANCIA

b. PARA LA PENDIENTE, M:

$$m \pm z_c \sigma_m$$

c. PARA LA PREDICCION, Y_i;

$$\tilde{Y}_i \pm z_c \sigma_{\tilde{Y}}$$

EN CASO DE QUE $\sigma_{y|x}$ SEA DESCONOCIDA (ES LO USUAL), DEBE ESTIMARSE A PARTIR DE LA MUESTRA MEDIANTE $S_{y|x}$. EN TAL CASO LOS INTERVALOS DE CONFIANZA CAMBIAN A:

a. PARA LA ORDENADA EN EL ORIGEN, α: $b \pm t_c \sigma_b$

$$b \pm t_c S_{y|x}\sqrt{\frac{\bar{x}^2}{nS_x^2}}$$

DONDE t_c ES EL VALOR CRITICO DE UN NIVEL DE SIGNIFICANCIA α, CORRESPONDIENTE A UNA DISTRIBUCION t DE STUDENT CON $\nu = n - 2$ GRADOS DE LIBERTAD, Y S_x^2 ES LA VARIANCIA (SESGADA) DE LA MUESTRA DE x.

b. PARA LA PENDIENTE, β: $m \pm t_c \sigma_m$

$$m \pm t_c S_{y|x}/\sqrt{nS_x^2} \qquad \text{O} \qquad m \pm t_c \frac{S_{y|x}}{S_x\sqrt{n}}$$

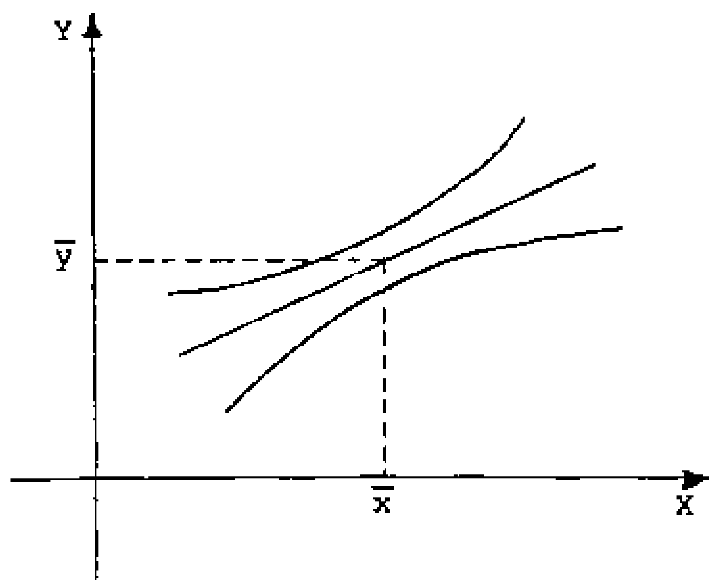
c. PARA LA PREDICCION, Y_i: $\bar{Y}_i \pm t_c \sigma_{\bar{y}}$

$$\bar{Y}_i \pm t_c\, S_{y|x} \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{nS_x^2}}$$

SI x_i ESTA DENTRO DEL RANGO DE LA MUESTRA, O

$$\bar{Y}_i \pm t_c\, S_{y|x} \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{nS_x^2}}$$

SI x_i ESTA FUERA DEL RANGO.



EJEMPLO

LA FORMACION DEL ALCOHOL EN UN PROCESO DE FERMENTACION SE RELACIONA CON LA TEMPERATURA. EN UNA SERIE DE SEIS MEDICIONES A DISTINTAS TEMPERATURAS SE OBTUVO LO SIGUIENTE:

TEMPERATURA, x, °C	35	40	45	50	55	60
ALCOHOL, lt	20.2	23.1	23.2	23.6	25.8	26.3

SI SE AJUSTA UNA RECTA POR MINIMOS CUADRADOS SE OBTIENE

$$\tilde{Y} = 0.225\ X + 13.01$$

$(\bar{x} = 47.5,\ \bar{y} = 23.7)$

1. INTERVALOS DE CONFIANZA CON $\sigma_{y|x} = 0.8$ (CONOCIDA); $\alpha = 0.05$.

$$\sigma_b = \sqrt{\frac{0.8}{6} + \frac{0.8 \times 47.5^2}{437.5}} = 1.845$$

DONDE $\sum_{i=1}^{6} (x_i - \bar{x})^2 = 437.5$

$$b \pm z_c \sigma_b = 13.01 \pm 1.96 \times 1.845 = (9.39,\ 16.63)$$

$$\sigma_m = \frac{0.8}{437.5} = \frac{0.8}{20.92} = 0.0382$$

$$m \pm z_c \sigma_m = 0.225 \pm 1.96 \times 0.0382 = 0.225 \pm 0.075 =$$

$$= (0.150,\ 0.300)$$

INTERVALOS DE CONFIANZA CON $\sigma_{y|x}$ DESCONOCIDA.

EN ESTE CASO $S^2_{y|x} = \frac{1}{n-2}\sum_{i=1}^{6} (y_i - 0,225x_i - 13.01)^2 = \sum_{i=1}^{6} (y_i-\tilde{y}_i)^2/(n-2)$

TEMP, x, °C	ALCOHOL, y, lts	$\tilde{y}_i$	$y_i-\tilde{y}_i$	$(y_i-\tilde{y}_i)^2$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
35	20.2	20.9	0.7	0.49	-12.5	156.2
40	23.1	22.0	1.1	1.21	- 7.5	56.2
45	23.2	23.1	0.1	0.01	- 2.5	6.2
50	23.6	24.3	-0.7	0.49	2.5	6.2
55	25.8	25.4	0.4	0.16	7.5	46.2
60	26.3	26.5	-0.2	0.04	12.5	156.2
Σ=285				Σ=2.40		Σ=437.4

$\bar{x} = \frac{285}{6} = 47.5;\quad S_x^2 = \frac{437.4}{6} = 72.9$

SABEMOS QUE $\tilde{y} = 0.225\,x + 13.01$; POR TANTO:

$\tilde{y}(35) = 0.225(35) + 13.01 = 20.9,$

$y(40) = 0.225(40) + 13.01 = 22.0,$ etc.

<u>INTERVALOS DE CONFIANZA:</u>

a) <u>PARA α</u> : $\quad 13.01 \pm t_c\, S_{y|x}\sqrt{\frac{\bar{x}^2}{nS_x^2} + \frac{1}{n}}$

$t_c = t_{0.975,4} = 2.776,\qquad S^2_{y|x} = \frac{1}{n-2}\Sigma(y_i-\tilde{y}_i)^2 = \frac{1}{4}\,2.4 = 0.6,$

$S_{y|x} = \sqrt{0.6} = 0.77$

$$13.01 \pm 2.776 \times 0.77\sqrt{\frac{47.5^2}{6(72.9)} + \frac{1}{6}} = 13.01 \pm 4.93 = (8.08, 17.94)$$

b) PARA β: $$0.225 \pm t_c \frac{S_{y|x}}{\sqrt{nS_x^2}} = 0.25 \pm 2.776 \frac{0.77}{\sqrt{6(72.9)}} =$$

$$= 0.225 \pm 0.102 = (0.123, 0.327)$$

c) PARA y_i $(x=50)$: $y_i(50)=24.3$

$$24.3 \pm t_c S_{y|x}\sqrt{\frac{1}{n} + \frac{(x_i-\bar{x})^2}{n\, S_x^2}} = 24.3 \pm 2.776 \times 0.77\sqrt{\frac{1}{6} + \frac{(50-47.5)^2}{6(72.9)}} =$$

$$= 24.3 \pm 0.9 = (23.4, 25.2)$$

PRUEBAS DE HIPOTESIS

a. PARA LA ORDENADA EN EL ORIGEN

SE DEMUESTRA QUE
$$\frac{\alpha - b_0}{S_{y|x}\sqrt{\dfrac{\overline{x^2}}{nS_x^2}}} = \frac{\alpha - b_0}{\dfrac{S_{y|x}}{S_x}\sqrt{\dfrac{\overline{x^2}}{n}}} = T$$

TIENE DISTRIBUCION t DE STUDENT CON $\nu = n - 2$ GRADOS DE LIBERTAD.

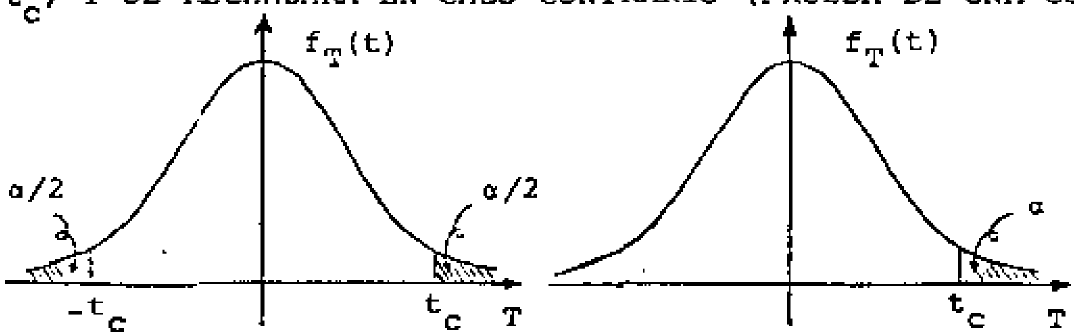
SI SE DESA PROBAR LA HIPOTESIS

$$H_0 : \alpha = b_0$$
$$H_1 : \alpha \neq b_0$$

BASTA SUSTITUIR A $\alpha = b_0$ EN LA ECUACION ANTERIOR Y EVALUAR $T = t$, ES DECIR,

$$t = \frac{b - b_0}{\dfrac{S_{y|x}}{S_x}\sqrt{\dfrac{\overline{x^2}}{n}}}$$

SE ACEPTARA H_0 SI $|t|<|t_c|$; EN CASO CONTRARIO SE RECHAZARA (PRUEBA DE DOS COLAS). SI H_1 FUERA $B > b_0$, SE ACEPTARA SI $t < t_c$, Y SE RECHAZARA EN CASO CONTRARIO (PRUEBA DE UNA COLA)



PRUEBA DE DOS COLAS PRUEBA DE UNA COLA

b. PARA LA PENDIENTE, β

ANALOGAMENTE, PARA β, LA ESTADISTICA

$$\frac{\beta - m_0}{S_{y|x}/\sqrt{nS_x^2}} = \frac{\beta - m_0}{\dfrac{S_{y|x}}{S_x\sqrt{n}}} = T,$$

DONDE m_0 = VALOR DE M BAJO LA HIPOTESIS NULA H_0 : $\beta = m_0$,

TAMBIEN TIENE DISTRIBUCION t DE STUDENT CON $\nu = n - 2$ GRADOS DE LIBERTAD:

$$t = \frac{m - m_0}{\dfrac{S_{y|x}}{S_x\sqrt{n}}}.$$

EJEMPLO

CONSIDERE LOS DATOS SIGUIENTES:

x	0	1	2	3	4	5	6	7	8	9
y	0.16	0.09	0.08	0.23	0.60	0.39	0.55	0.75	0.81	0.85

$m = 0.093$, $b = 0.032$, $S^2_{y|x} = 0.01258$

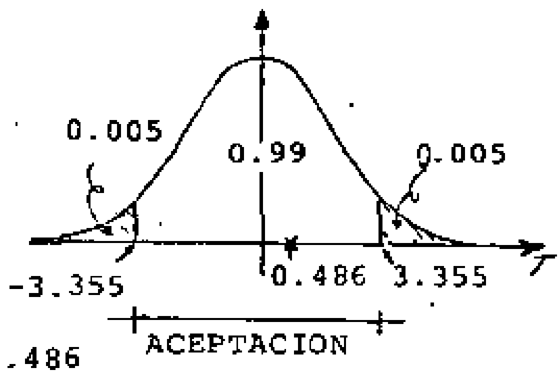
$$S_x^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 82.50/10 = 8.25;\ \Sigma\, x_i^2 = 285,\ \overline{x^2} = \frac{285}{10} = 28.5$$

a. PROBAR LA HIPOTESIS DE QUE $\alpha = 0$

b. PROBAR LA HIPOTESIS DE QUE $\beta = 0.1$

CON $\alpha = 0.01$ Y $S_{y|x}$ DESCONOCIDA.

$\therefore\ H_0 : \alpha = 0;\quad H_1 : \alpha \neq 0$



$$t = \frac{b - b_0}{S_{y|x}\sqrt{\dfrac{\overline{x^2}}{nS_x^2}}} = \frac{0.032 - 0}{\sqrt{0.01258}\sqrt{\dfrac{28.5}{10 \times 8.25}}} = 0.486$$

$t_c = t_{0.995,\ 8} = 3.355 > 0.486\ \therefore$ SE ACEPTA H_0.

b. $H_0 : \beta = 0.1; \qquad H_1 : \beta \neq 0.1$

$$t = \frac{m - m_0}{\dfrac{S_{y|x}}{S_x\sqrt{n}}} = \frac{0.093 - 0.1}{\dfrac{\sqrt{0.01258}}{\sqrt{8.25 \times 10}}} = 0.567 < 3.355$$

SE ACEPTA H_0 CON 99% DE NIVEL DE CONFIANZA.

PRUEBA DE HIPOTESIS PARA EL COEFICIENTE DE CORRELACION, ρ_{xy}

PRUEBA

$H_0 : \rho_{xy} = 0$; $H_1 : \rho_{xy} \gtrless \neq 0$

SE DEMUESTRA QUE EN CASO DE QUE X Y Y SON INDEPENDIENTES ($\rho = 0$), LA ESTADISTICA

$$T = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}}$$

TIENE DISTRIBUCION t DE STUDENT CON $n - 2$ GRADOS DE LIBERTAD.

EJEMPLO

EN BASE A UNA MUESTRA ALEATORIA DE 30 DATOS SOBRE LA TEMPERATURA MEDIA DURANTE UN MES, X, Y EL PESO MEDIO DE LOS TOMATES PISCADOS, Y, SE OBTUVO UN COEFICIENTE DE CORRELACION $r_{xy} = 0.931$.

PROBAR LA HIPOTESIS DE QUE $\rho_{xy} = 0$. USAR $\alpha = 0.05$.

$H_0 : \rho_{xy} = 0$; $H_1 : \rho_{xy} \neq 0$

$$t = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0.931\sqrt{\frac{30-2}{1-0.931^2}} = 13.448$$

$t_c = t_{0.975,\ 28} = 2.048 < 13.448$

$\therefore$ SE RECHAZA H_0 A UN NIVEL DE CONFIANZA DEL 95%.

16. ANALISIS DE VARIANCIA EN REGRESION LINEAL

EN EL CAPITULO DE REGRESION LINEAL SE TENIA QUE LA ECUACION $\hat{Y} = mX + b$ ESTIMABA A LA ECUACION ENTRE LAS VARIABLES Y Y X, SIENDO m UN ESTIMADOR DE LA PENDIENTE, β, DE LA RECTA, Y b UN ESTIMADOR DE LA ORDENADA EN EL ORIGEN, α. ASIMISMO, SE TENIA QUE LA VARIANCIA SESGADA TOTAL ERA

$$\underset{\cdot}{S}^2(Y) = S^2_{Y|X} + m^2 S^2(X) \qquad (1)$$

POR LO QUE LA SUMA TOTAL DE CUADRADOS SERIA

$$\Sigma(y_i - \bar{y})^2 = \Sigma(y_i - \hat{y}_i)^2 + m^2\Sigma(x_i - \bar{x})^2 \qquad (2)$$

LA PRIMERA SUMA DE CUADRADOS DEL MIEMBRO DERECHO DE ESTA ECUACION ES LA <u>INEXPLICADA</u>, <u>ALEATORIA</u> O <u>RESIDUAL</u> Y, LA SEGUNDA, ES LA <u>EXPLICADA</u>.

EL MODELO LINEAL ES $Y_i = \alpha + \beta X_i + Z_i$ DONDE Z_i SON VARIA-

BLES ALEATORIAS QUE SATISFACEN LAS CONDICIONES DEL ANALISIS DE VARIANCIA. EN TAL CASO, $E(m) = \beta$, $E(b) = \alpha$, $Var(\bar{Y}) = \sigma^2/n$, $Var\ (m) = \sigma^2/\Sigma(x_i - \bar{x})^2$ Y $cov\ (\bar{y},m) = 0$

PUESTO QUE $E(m) = \beta$, SE OBTIENE QUE LA ESPERANZA DE LA SUMA DE CUADRADOS EXPLICADA ES

$$E\left[m^2\Sigma(x_i-\bar{x})^2\right] = \sigma^2 + \beta^2\Sigma(x_i-\bar{x})^2 \qquad (3)$$

SE OBSERVA QUE ESTA SUMA DE CUADRADOS TIENE UN GRADO DE LIBERTAD.

POR OTRA PARTE LA ESPERANZA DE LA SUMA DE CUADRADOS RESIDUAL ES

$$E\left[(y_i-\tilde{y}_i)^2\right] = (n-2)\sigma^2 \qquad (4)$$

PARA LO QUE ESTE TIENE n-2 GRADOS DE LIBERTAD.

OBSERVANDO LAS ECS (3) y (4) SE CONCLUYE QUE LA PRUEBA DE HIPOTESIS DE INDEPENDENCIA DE Y y X, O SEA DE $\beta = 0$, SE PUEDE HACER FORMULANDO UNA ESTADISTICA CON EL COCIENTE DE LAS SUMAS DE LOS CUADRADOS (4) ENTRE (3) CON $\beta = 0$:

$$F = \frac{(n-2)m^2\Sigma(x_i-\bar{x})^2}{\Sigma(y_i-mx_i-b)^2} \qquad (5)$$

ESTA ESTADISTICA TIENE DISTRIBUCION F CON 1 Y n-2 GRADOS DE LIBERTAD.

PARA PROBAR LA HIPOTESIS DE QUE $\beta=\beta_o$ SE REMPLAZA EN LA EC (5) A $\bar{m}$ POR $m-\beta_o$.

LA TABLA DEL ANALISIS DE VARIANCIA RESULTANTE ES:

FUENTE	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	VALOR MEDIO CUADRATICO
EXPLICADA	1	$m^2\Sigma(x_i-\bar{x})^2$	$m^2\Sigma(x_i-\bar{x})^2$
RESIDUAL	n-2	$\Sigma(y_i-mx_i-b)^2$	$\dfrac{\Sigma(y_i-mx_i-b)^2}{n-2}$
TOTAL	n-1	$\Sigma(y_i-\bar{y})^2$	

17. CLASIFICACION EN UNA DIRECCION. OBSERVACION DE DOS VARIABLES

SI SE MIDEN DOS CARACTERISTICAS, X Y Y, EN CADA SUJETO DE EXPERIMENTACION EN UN EXPERIMENTO CON CLASIFICACION EN UNA DIRECCION, NECESITAMOS CONSIDERAR TANTO LA RELACION QUE HAY ENTRE ELLAS COMO LA POSIBLE VARIACION DE ESTA DE GRUPO A GRUPO.

SI SE TIENE QUE ES ACEPTABLE UNA RELACION LINEAL DE Y CON BASE EN X PERO QUE PUDIERA VARIAR DE UN GRUPO A OTRO, UN MODELO APROPIADO SERIA:

$$Y_{ti} = \alpha_t + \beta_t X_{ti} + Z_{ti}; \quad t = 1,2,\ldots,k; \quad i = 1,2,\ldots,n_t \tag{1}$$

UN PROBLEMA NATURAL SERIA VERIFICAR SI ES POSIBLE USAR UN SOLO MODELO $Y = \alpha + \beta X$ PARA CADA UNO DE LOS GRUPOS. ESTO IMPLICARIA PROBAR LA HIPOTESIS DE QUE $\alpha_1 = \alpha_2 = \ldots = \alpha_k$ Y DE QUE $\beta_1 = \beta_2 = \ldots = \beta_k$.

PARA PROBAR ESTA HIPOTESIS CONVIENE SEPARAR EL PROBLEMA EN TRES PARTES, CADA UNA DE LAS CUALES PUEDE PROBARSE POR SEPARADO:

a. $H_o^{(1)}$: LAS LINEAS DE REGRESION SON PARALELAS, ESTO ES,

$$\beta_1 = \beta_2 = \ldots = \beta_k.$$

b. $H_o^{(2)}$: LAS MEDIAS DE LOS GRUPOS CAEN EN UNA LINEA RECTA, ESTO ES, LOS PUNTOS $(\bar{x}_{t.}, \alpha_t + \beta_t \bar{x}_{t.})$ SE ALINEAN EN UNA RECTA.

c. $H_o^{(3)}$: LA PENDIENTE DE LA LINEA ANTERIOR ES IGUAL AL COMUN, β_c, DE $\beta_1, \beta_2, \ldots, \beta_k$

PARA HACER LO ANTERIOR SE CALCULAN PRIMERO LAS RECTAS DE REGRESION PARA CADA GRUPO POR SEPARADO, CON LO CUAL SE OBTIENEN LAS ESTIMACIONES

$$E(Y|X) = A_t + B_t X; \quad t = 1,2,\ldots,k \tag{2}$$

LA ESPERANZA DE B_t ES β_t Y SU VARIANCIA ES

$$\mathrm{Var}(B_t) = \sigma^2 / \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_{t.})^2 = \sigma^2 / w_t \tag{3}$$

DONDE

$$w_t = \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_{t.})^2 \tag{4}$$

LOS ANALISIS DE VARIANCIA DE LA REGRESION LINEAL EN CADA GRUPO SE BASAN EN LAS IDENTIDADES ALGEBRAICAS

$$\sum_{i=1}^{n_t} (Y_{ti} - \bar{Y}_{t.})^2 = w_t B_t^2 + \sum_{i=1}^{n_t} \{Y_{ti} - \bar{Y}_{t.} - B_t (x_{ti} - \bar{x}_{t.})\}^2; \; t = 1,2,\ldots,k \tag{5}$$

SI SUMAMOS ESTAS k IDENTIDADES SE OBTIENE:

$$\sum_{t=1}^{k}\sum_{i=1}^{n_t}(Y_{ti} - \bar{Y}_{t.})^2 = \sum_{t=1}^{k} w_t B_t^2 + \sum_{t=1}^{k}\sum_{i=1}^{n_t}\{Y_{ti} - \bar{Y}_{t.} - B_t(x_{ti} - \bar{x}_{t.})\}^2$$

$$= \sum_{t=1}^{k} w_t B_t^2 + S_R \qquad (6)$$

DONDE S_R ES LA SUMA DE CUADRADOS RESIDUAL CON $\sum_{t=1}^{k}(n_t - 2) = N - 2k$ GRADOS DE LIBERTAD, DONDE $N = \Sigma n_t$, ES DECIR,

$$E(S_R) = (N - 2k)\sigma^2 \qquad (7)$$

SI $B_c = \sum_{t=1}^{k} w_t B_t / w_c$, DONDE $w_c = \sum_{t=1}^{k} w_t$ ES UN PROMEDIO PESADO DE LAS B_t, ENTONCES LA DESVIACION CUADRATICA TOTAL DE LAS B_t RESPECTO A B_c ES

$$S_w = \sum_{t=1}^{k} w_t (B_t - B_c)^2 = \sum_{t=1}^{k} w_t B_t^2 - w_c B_c^2 \qquad (8)$$

DESPEJANDO DE ESTA ECUACION A $\Sigma w_t B_t^2$ SE OBTIENE

$$\Sigma w_t B_t^2 = w_c B_c^2 + S_w \qquad (9)$$

LA ESPERANZA DE S_w ES

$$E(S_w) = (k - 1)\sigma^2 + \sum_{t=1}^{k} w_t (\beta_t - \beta_c)^2 \qquad (10)$$

DONDE $\beta_c = \sum_{t=1}^{k} w_t \beta_t / w_c$

ES LA PENDIENTE COMUN (PROMEDIO) DENTRO DE LOS GRUPOS.

POR SU PARTE, LA VARIANCIA DE B_c ES

$$Var(B_c) = \sigma^2/w_c \quad (11)$$

CON LO ANTERIOR LA SUMA DE CUADRADOS TOTAL SERA:

$$\sum_{t=1}^{k}\sum_{i=1}^{n_t}(Y_{ti} - \bar{Y}_{..})^2 = \sum_{t=1}^{k} n_t(\bar{Y}_{t.} - \bar{Y}_{..})^2 + \sum_{t=1}^{k}\sum_{i=1}^{n_t}(Y_{ti} - \bar{Y}_{t.})^2$$

$$= \Sigma n_t(\bar{Y}_{t.} - \bar{Y}_{..})^2 + w_c B_c^2 + S_R + S_w \quad (12)$$

DONDE S_w SE DENOMINA LA SUMA DE CUADRADOS DE LAS PENDIENTES ENTRE GRUPOS.

ANALIZANDO LAS ECS. (7) y (10) SE PUEDE VER QUE LA HIPOTESIS $H_o^{(1)}$: $\beta_1 = \beta_2 = \dots = \beta_k$ SE PUEDE PROBAR MEDIANTE LA ESTADISTICA

$$F = \frac{S_w/(k-1)}{S_R/(N-2k)} \quad (13)$$

QUE TIENE DISTRIBUCION F CON (k-1) Y (N-2k) GRADOS DE LIBERTAD, YA QUE BAJO LA HIPOTESIS NULA EL SEGUNDO TERMINO DEL MIEMBRO DERECHO DE LA EC. (10) ES CERO.

PARA REALIZAR LA PRUEBA $H_o^{(2)}$ PRIMERO AJUSTAMOS LA RECTA QUE PASA POR LOS PROMEDIOS $(\bar{X}_{t.}, \bar{Y}_{t.})$ CON FACTORES DE PESO n_t. CON ESTO SE OBTIENE LA SUMA DE CUADRADOS ENTRE GRUPOS:

$$\sum_{t=1}^{k} n_t(\bar{Y}_{t.} - \bar{Y}_{..})^2 = w_m B_m^2 + \sum_{t=1}^{k} n_t\{\bar{Y}_{t.} - \bar{Y}_{..} - B_m(\bar{x}_{t.} - \bar{x}_{..})\}^2$$

$$= w_m B_m^2 + S_G \qquad (14)$$

DONDE

$$B_m = \frac{\sum_{t=1}^{k} n_t(\bar{x}_{t.} - \bar{x}_{..})(\bar{Y}_{t.} - \bar{Y}_{..})}{\Sigma n_t(\bar{x}_{t.} - \bar{x}_{..})^2} \qquad (15)$$

Y

$$w_m = \Sigma n_t(\bar{x}_{t.} - \bar{x}_{..})^2 \qquad (16)$$

LA VARIANCIA DE B_m Y LA ESPERANZA DE S_G SON:

$$\mathrm{Var}(B_m) = \sigma^2/w_m \qquad (17)$$

$$E(S_G) = (k-2)\sigma^2 + \sum_{t=1}^{k} n_t(\alpha_t - \alpha_m - \beta_m \bar{x}_{...})^2 \qquad (18)$$

DONDE

$$\alpha_m = \Sigma n_t \alpha_t / N \qquad (19)$$

$$\beta_m = \Sigma n_t(\bar{x}_{t.} - \bar{x}_{..})(\alpha_t + \beta_t \bar{x}_{t.})/w_m \qquad (20)$$

POR CONSIGUIENTE, LA HIPOTESIS $H_o^{(2)}$ SE PUEDE PROBAR FORMULANDO LA ESTADISTICA

$$F = \frac{S_G/(k-2)}{S_R/(N-2k)} \qquad (21)$$

QUE TIENE DISTRIBUCION F CON (k-2) Y (N-k) GRADOS DE LIBERTAD.

FINALMENTE, PARA PROBAR $H_o^{(3)}$ USAREMOS LA SUMA DE LOS DOS TERMINOS $w_c B_c^2$ Y $w_m B_m^2$:

$$w_c B_c^2 + w_m B_m^2 = w_o B_o^2 + \frac{w_c w_m}{w_o} (B_c - B_m)^2 = w_o B_o^2 + S_{WG} \tag{22}$$

DONDE

$$w_o = w_c + w_m = \sum_t \sum_i (x_{ti} - \bar{x}_{..})^2 \tag{23}$$

$$B_o = \frac{w_c B_c + w_m B_m}{w_o} \tag{24}$$

DONDE B_o ES LA PENDIENTE GLOBAL QUE SE OBTENDRIA SI TODOS LOS PUNTOS SE AJUSTARAN A UNA SOLA RECTA, SIN DISTINCION DE GRUPOS. LA ESPERANZA DE S_{WG} ES

$$E(S_{WG}) = \sigma^2 + \frac{w_c w_m}{w_o} (\beta_o - \beta_m)^2 \tag{25}$$

POR LO TANTO, LA HIPOTESIS $H_o^{(3)}$ SE PUEDE PROBAR CON LA ESTADISTICA

$$F = \frac{S_{WG}}{S_R/(N-2k)} \tag{26}$$

QUE TIENE DISTRIBUCION F CON 1 Y N-2k GRADOS DE LIBERTAD

LAS PRUEBAS ANTERIORES SE PUEDE RESUMIR EN LA TABLA DE ANALISIS DE VARIANCIA SIGUIENTE:

Fuente	G. de L.	Suma de cuadrados SS	Esperanzas de MS
Pendiente global	1	$S_o = w_o B_o^2$	$\sigma^2 + w_o \beta_o^2$
Pendiente de las medias de los grupos vs promedio de las pendientes dentro de grupos	1	$S_{WG} = \frac{w_c w_m}{w_o}(B_c - B_m)^2$	$\sigma^2 + \frac{w_c w_m}{w_o}(\beta_c - \beta_m)^2$
Acerca de la línea de regresión de las medias de los grupos	$k-2$	$S_G = \sum_{i=1}^{k} n_i\left[\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} - B_m(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})\right]^2$	$\sigma^2 + (k-2)^{-1}\sum_{i=1}^{k} n_i[\alpha_i - \alpha_m - \beta_m \bar{x}_{i\cdot}]^2$
Pendientes entre grupos	$k-1$	$S_W = \sum_{i=1}^{k} w_i(B_i - B_c)^2$	$\sigma^2 + (k-1)^{-1}\sum_{i=1}^{k} w_i(\beta_i - \beta_c)^2$
Residual	$N-2k$	$S_R = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[Y_{ij} - \bar{Y}_{i\cdot} - B_i(x_{ij} - \bar{x}_{i\cdot})\right]^2$	σ^2
Total	$N-1$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	

ANALISIS DE COVARIANCIA EN UNA DIRECCION

EL ANALISIS DE COVARIANCIA SE UTILIZA PARA PROBAR SI LAS DIFERENCIAS EN LA RESPUESTA MEDIA DE UN GRUPO A OTRO PUEDEN SER EXPLICADAS POR UNA REGRESION LINEAL CON UNA VARIABLE DE CONTROL. EL PLANTEAMIENTO DEL ANALISIS DE COVARIANCIA DEPENDE DEL MODELO QUE SE UTILICE; PARA CLASIFICACION DE GRUPOS EN UNA DIRECCION SE PUEDEN USAR LOS SIGUIENTES MODELOS:

$$\text{I.} \quad Y_{ti} = \alpha_t + \beta(X_{ti} - \bar{X}_{..}) + z_{ti} \tag{1}$$

$$\text{II.} \quad Y_{ti} = \alpha_t + \beta_t(X_{ti} - \bar{X}_{..}) + z_{ti} \tag{2}$$

PARA AMBOS MODELOS SE PRETENDE PROBAR LA HIPOTESIS

$$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_k \tag{3}$$

CONTRA H_1: NO TODAS LAS α_t SON IGUALES

LAS TABLAS DEL ANALISIS SON:

MODELO	FUENTE	G. de L.	SS
I	GRUPOS (AJUSTADA)	$k - 1$	SWG + SG
	RESIDUAL	$N - k - 1$	SR + SW
II	GRUPOS (AJUSTADA)	$k - 1$	$S_0 + SWG + SG + SW - w_0 \sum_{t=1}^{k} B_t'^2$
	RESIDUAL	$N - 2k$	SR

DONDE SWG, SG, SR, SW, S_0 Y W_0 SE CALCULAN CON LAS FORMULAS DEL CAPITULO DE OBSERVACION DE DOS VARIABLES, Y

$$B_t' = \frac{\sum_{i=1}^{N_t} (X_{ti} - \bar{X}_{..})(\bar{Y}_{ti} - \bar{Y}_{..})}{\sum_{i=1}^{N_t} (X_{ti} - \bar{X}_{..})^2} \tag{4}$$

LOS VALORES ESTIMADOS DE LAS α_t SON

MODELO I: $\bar{Y}_{t.} - B_c(\bar{X}_{t.} - \bar{X}_{..})$ (5)

MODELO II: $\bar{Y}_{t.} - B_t(\bar{X}_{t.} - \bar{X}_{..})$ (6)

SI UNO ESTA BASTANTE SEGURO DE QUE $\beta_1 = \beta_2 = \ldots \beta_k$, ENTONCES EL MODELO I ES MEJOR, YA QUE DA MAS GRADOS DE LIBERTAD EN EL RESIDUO.

TAREA

EN UN EXPERIMENTO, A 40 SEÑORES SE LES SUJETO A UNA PRUEBA (TRATAMIENTO) PARA DETERMINAR QUE TAN CERCA PODIAN CAMINAR HACIA UN OBJETO PELIGROSO (EN ESTE CASO UNA VIBORA), ANTES DE SENTIRSE ANSIOSOS; PARA ESTO, CADA SUJETO SE SITUO ALEATORIAMENTE EN UNO DE CUATRO GRUPOS, CADA UNO CON DIEZ SUJETOS; CON CADA GRUPO SE EMPLEO DIFERENTE TIPO DE VIBORA. DESPUES DEL TRATAMIENTO A CADA SEÑOR SE LE SUJETO DE NUEVO AL MISMO TRATAMIENTO (POSTRATAMIENTO). LOS RESULTADOS DEL TRATAMIENTO SON LAS X_{ti} Y LOS DEL POSTRATAMIENTO SON LAS Y_{ti}, LOS CUALES SE PRESENTAN EN LA TABLA SIGUIENTE

SUJETO	GRUPOS (X_{ti},Y_{ti})			
	1	2	3	4
1	25,25	17,11	32,24	10,8
2	13,25	9,9	30,18	29,17
3	10,12	19,16	12,2	7,8
4	25,30	25,17	30,24	17,12
5	10,37	6,1	10,2	8,7
6	17,25	23,12	8,0	30,26
7	9,31	7,4	5,0	5,8
8	18,26	5,3	11,1	29,29
9	27,28	30,26	5,1	5,29
10	17,29	19,20	25,10	13,0

a) CALCULAR LAS RECTAS DE REGRESION PARA CADA GRUPO, PARA LOS PROMEDIOS Y PARA TODOS LOS PUNTOS JUNTOS. EN UNA MISMA GRAFICA DIBUJAR LOS PUNTOS Y LAS RECTAS CALCULADAS.

b) ESTIMAR LOS EFECTOS α_t

c) PROBAR LA HIPOTESIS DE IGUALDAD DE PENDIENTES H_0: $\beta_1 = \beta_2 \dots = \beta_K$

d) PROBAR LA HIPOTESIS DE IGUALDAD DE MEDIAS DE LAS Y_{ti} DE LOS CUATRO GRUPOS, DESPUES DE AJUSTAR POR LA REGRESION CON X_{ti}, O SEA, PROBAR H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$.

SOLUCION

CALCULO DE LAS RECTAS DE REGRESION PARA CADA GRUPO:

$$\hat{Y}_t = a_t + b_t X$$

DONDE

$$b_t = \left(\frac{n \sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}\right)_t ; \quad a_t = (\bar{y} - b \bar{x})_t$$

SE TIENE PARA CADA GRUPO:

	1	2	3	4
$\sum_i x_i =$	171	160	168	153
$\sum_i y_i =$	268	119	82	144
$\sum_i x_i y_i =$	4,611	2,482	2,338	2,695
$\sum_i x_i^2 =$	3,331	3,256	3,928	3,303
$\bar{x}_{t.} =$	17.1	16.0	16.8	15.3
$\bar{y}_{t.} =$	26.8	11.9	8.2	14.4

POR LO TANTO:

$$b_1 = \frac{10(4611) - (171)(268)}{10(3,331) - (171)^2} = 0.0693; \quad a_1 = 26.8 - 0.0693(17.1) = 25.61$$

$$b_2 = \frac{10(2,482) - (160)(119)}{10(3,256) - (160)^2} = 0.8305; \quad a_2 = 11.9 - 0.8305(16) = -1.39$$

$$b_3 = \frac{10(2,338) - (168)(82)}{10(3,982) - (168)^2} = 0.8687; \quad a_3 = 8.2 - 0.8687(16.8) = -6.39$$

$$b_4 = \frac{10(2,695) - (153)(144)}{10(3,303) - (153)^2} = 0.5112; \quad a_4 = 14.4 - 0.5112(15.3) = 6.58$$

POR LO QUE LAS RECTAS DE REGRESION SON, PARA CADA UNO DE LOS GRUPOS:

$\tilde{Y}_1 = 25.61 + 0.07X$

$\tilde{Y}_2 = -1.39 + 0.83X$

$\tilde{Y}_3 = -6.39 + 0.87X$

$\tilde{Y}_4 = 6.58 + 0.51X$

CALCULO DE LA RECTA QUE SE AJUSTA A LOS PROMEDIOS:

(17.1, 26.8), (16.0, 11.9), (16.8, 8.2) (15.3, 14.4)

$\Sigma X_i = 65.2, \Sigma Y_i = 61.3, \Sigma X_i Y_i = 1,006.76, \Sigma X_i^2 = 1,064.74$

$\bar{x} = 16.3, \ \bar{y} = 15.325$

$$b_p = \frac{4(1,006.76) - (65.2)(61.3)}{4(1,064.74) - (165.2)} = 3.8232$$

$a_p = 15.325 - (3.8232)\ 16.3 = -46.9937$

LA RECTA DE REGRESION PARA LOS PROMEDIOS ES:

$$\tilde{y}_p = -46.99 + 3.82X$$

CALCULO DE LA RECTA PARA TODOS LOS PUNTOS JUNTOS:

$\Sigma X_i = 171 + 160 + 168 + 153 = 652, \ \bar{x} = 16.3$

$\Sigma Y_i = 268 + 119 + 82 + 144 = 613, \ \bar{y} = 15.325$

$\Sigma X_i Y_i = 4,611 + 2,482 + 2,338 + 2,695 = 12,126$

$\Sigma X_i^2 = 3,331 + 3,256 + 3,928 + 3,303 = 13,818$

LA RECTA DE REGRESION RESULTANTE ES

$$b_r = \frac{40(12,126) - (652)(613)}{40(13,818) - (652)^2} = 0.6689, \quad a_r = 15.325 - (0.6689)\ 16.3 = 4.4217$$

$\hat{y}_r = 4.42 + 0.67X$

b) ESTIMAR LOS EFECTOS α_t

COMO

$E(a_t) = \alpha_t$; a_t ES UN ESTIMADOR INSESGADO DE α_t Y :

$\hat{\alpha}_1 = 25.61$; $\hat{\alpha}_2 = -1.39$; $\hat{\alpha}_3 = -6.39$; $\hat{\alpha}_4 = 6.58$

c) PROBAR LA HIPOTESIS DE IGUALDAD DE PENDIENTES

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$; H_1: NO TODAS LAS β_i SON IGUALES

$$W_t = \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_{t.}) = \sum_{i=1}^{n_t} x_{ti}^2 - n_t \bar{x}_{t.}^2$$

$$W_1 = 3{,}331 - 10(17.1)^2 = 406.9, \qquad B_1 = 0.0693$$
$$W_2 = 3{,}256 - 10(16.0)^2 = 696, \qquad B_2 = 0.8305$$
$$W_3 = 3{,}928 - 10(16.8)^2 = 1{,}105.6, \qquad B_3 = 0.8687$$
$$W_4 = 3{,}303 - 10(15.3)^2 = \underline{962.1,} \qquad B_4 = 0.5112$$
$$3{,}170.6$$

$$S_w = \sum_{t=1}^{n_t} W_t\, B_t^2 - W_c\, B_c^2 -$$

$$W_c = \sum_{t=1}^{n_t} W_t = 3{,}170.6$$

$$B_c = \frac{1}{W_c} \sum_{t=1}^{n_t} W_t\, B_t = \frac{1}{3{,}170.6}\,(2058.4864) = 0.6492$$

$$S_w = 1{,}567.7572 - (3{,}170.6)(0.6492)^2 = 231.30$$

AHORA, DE LA ECUACION (12) DE LOS APUNTES:

$$S_R = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_{..})^2 - \sum_{t=1}^{k} n_t(\bar{y}_{t.} - \bar{y}_{..})^2 - (W_c\, B_c^2 + S_w)$$

$$= \left(\sum_{t=1}^{k} \sum_{i=1}^{n_t} y_{ti}^2 - N\bar{y}_{..}^2\right) - \left(\sum_{t=i}^{k} n_t\, \bar{y}_{t.}^2 - N\bar{y}^2\right) - (W_c\, B_c^2 + S_w)$$

CON $\bar{y}_{..} = 15.325$ SE OBTIENE

$$S_R = 14{,}161 - 11{,}344.5 - 1{,}567.7572 = 1{,}248.74$$

EN CONSECUENCIA $\quad F = \dfrac{S_w/(k-1)}{S_R/(N-2k)} = \dfrac{231.30/3}{1{,}248.74/32} = 1.98 < F_{0.05,3,32} = 2.0$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE LAS PENDIENTES SON IGUALES, CON 5% DE NIVEL DE SIGNIFICANCIA.

d) PROBAR LA HIPOTESIS $H_o : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$

DE LOS RESULTADOS DEL INCISO ANTERIOR ES RAZONABLE SUPONER QUE TODAS LAS β_i SON IGUALES, POR LO QUE EL MODELO CORRESPONDIENTE ES:

$$y_{ti} = \alpha_t + \beta(x_{ti} - \bar{x}_{..}) + z_{ti}$$

ENTONCES:

$$S_R + S_W = 1248.74 + 231.30 = 1,480.04$$

$$w_m = \sum_{t=1}^{k} n_t(\bar{x}_{t.} - \bar{x}_{..})^2 = n_t \Sigma \bar{x}_t^2 - kn_t \bar{x}_{..}^2 = 10(1,064.74) - (40)(16.3)^2$$

$$= 19.8$$

$$B_m = \sum_{t=1}^{4} \frac{10(\bar{x}_{t.} - 16.3)(\bar{y}_{t.} - 15.325)}{19.8} = \frac{10(9.18+1.0275-3.5625+0.9250)}{19.8}$$

$$= 3.8232$$

$$S_G = \sum_{t=1}^{k} n_t \bar{y}_{t.}^2 - N\bar{y}_{..}^2 - w_m B_m^2 = 11,344.5 - 40(15.325)^2 + 19.8(3.8232)^2 = 2,239.68$$

$$S_{WG} = \frac{w_c w_m}{w_o}(B_c - B_m)^2 = \frac{(3,170.6)(19.8)}{3170.6+19.8}(0.6492-3.8232)^2 = 198.23$$

$$S_{WG} + S_G = 2437.91$$

POR TANTO :

$$F = \frac{2437.91/3}{1480.04/35} = 19.22 > 2.81 = F_{0.05,3,35}$$

POR LO QUE SE RECHAZA LA HIPOTESIS H_o DE QUE TODAS LAS α_t SON IGUALES ENTRE SI.

19. BIBLIOGRAFIA

1. Johnson, N.L. y Leone, F.C.,"Statistics and experimental design in ingineering and the physical sciences", Vol II, 2a ed., J. Wiley (1977)

2. Lee, W., "Experimental design and analysis", Freeman (1975)

3. Ogawa, J., "Statistical theory of the analysis of experimental designs", Ed. Dakker (1974)

4. Biles, W.E. y Swain, J.J., "Optimization and insdustrial experimentation", J. Wiley (1978)

5. Box, G.E.P., Hunter N.G. y Hunter, J.S. "Statistics for experimenters", J. Wiley (1978)

6. Cochran, W. G. y Cox, G.M., "Experimental designs", J. Wiley

7. Kirk, R., "Experimental design: procedures for the behavioral sciences"

8. Winer, B. J., "Statistical principles in experimental design"

9. Afifi, A.A y Asen, S. P., "Statistical Analysis", 2a Ed., Academic Press.

DIVISION DE EDUCACION CONTINUA
FACULTAD DE INGENIERIA U.N.A.M.

DISEÑO ESTADISTICO DE EXPERIMENTOS

INFERENCIA ESTADISTICA

M. EN I. AUGUSTO VILLARREAL ARANDA

ABRIL, 1982

Palacio de Minería Calle de Tacuba 5 primer piso Deleg. Cuauhtemoc 06000 México, D.F. Tel.: 521-40-20 Apdo. Postal M-2285

INFERENCIA ESTADISTICA

Por: M en I Augusto Villarreal Aranda

1. Introducción

La parte de la estadística que proporciona las reglas para inferir ciertas características de una población a partir de muestras extraídas de ella, junto con indicaciones probabilísticas de la veracidad de tales inferencias, se llama *inferencia estadística*.

En la inferencia estadística se estudian las relaciones existentes entre una población, las muestras obtenidas de ella, y las técnicas para estimar parámetros, tales como la media y la variancia, o bien para determinar si las diferencias entre dos muestras son debidas al azar, etc.

2. Distribuciones muestrales

Si se consideran todas las muestras posibles de tamaño n que pueden extraerse de una población, y para cada una se calcula el valor del promedio aritmético, este seguramente variará de una muestra a otra, ya que depende de los valores de los datos que se hayan obtenido en cada muestra. Por lo tanto, el promedio aritmético es en sí una variable aleatoria, como también lo son, por la misma razón, el rango y la variancia de la muestra.

A todo elemento que es función de los valores de los datos que se tienen en una muestra se le denomina *estadística*: toda estadística es, entonces, una variable aleatoria cuya distribución de probabilidades se conoce como *distribución muestral*. Si, por ejemplo, la estadística considerada es la variancia de la muestra, su densidad de probabilidades se llama *distribución muestral de la variancia*.

En forma similar se pueden obtener las distribuciones muestrales de la desviación estándar, del rango, etc., cada una de las cuales tendrá sus propios parámetros, lo que permite hablar de la media y la desviación estándar de la variancia, etc.

3. Muestreo con y sin remplazo

Cuando se efectúa un muestreo en una población de tal manera que cada elemento de la misma se pueda escoger más de una vez, se dice que el muestreo es *con remplazo*; en caso contrario, el muestreo es *sin remplazo*. Si de una urna se quiere extraer una muestra de bolas de colores, se puede proceder de dos maneras: se saca al azar una bola, se anota su color y se regresa a la urna antes de obtener otra, y así sucesivamente; en este caso el muestreo es *con remplazo*. La segunda forma consiste en extraer

al azar todas las bolas que constituyen la muestra sin regresarlas a la urna, siendo entonces un muestreo *sin remplazo*.

4. Distribución muestral del promedio aritmético

Supóngase que se extraen sin remplazo todas las muestras posibles de tamaño n de una población finita de tamaño $N_p > n$. Si la media y la desviación estándar de la distribución muestral del promedio aritmético se denotan con $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$, y la media y la desviación estándar de la población con μ y σ, respectivamente, entonces es posible demostrar que se cumplen las siguientes ecuaciones

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

Además, si la población es infinita (o el muestreo es con remplazo), los resultados anteriores se reducen a

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

puesto que

$$\lim_{N_p \to \infty} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}} = \frac{\sigma}{\sqrt{n}}$$

Para valores grandes de n ($n \geq 30$) se demuestra, empleando el teorema del límite central, que la distribución muestral del promedio aritmético es aproximadamente una distribución normal con media $\mu_{\bar{X}}$ y desviación estándar $\sigma_{\bar{X}}$, independientemente de cuál sea la densidad de probabilidades de X, la variable aleatoria asociada a la población. Si esta variable tiene distribución normal, la distribución muestral del promedio aritmético también es normal, aun para valores pequeños de n ($n < 30$).

Ejemplo 4.1

Supóngase que se tiene una población finita formada por los datos 1,2,3,4,5. Se desea conocer la media y la desviación estándar de la distribución muestral del promedio aritmético, considerando las muestras de tamaño 3 obtenidas sin remplazo.

Primer procedimiento.

Siendo la población finita y el muestreo sin remplazo, es posible obtener la distribución muestral correspondiente para calcular después sus parámetros, considerando que el número total de muestras distintas de tamaño 3 que pueden obtenerse a partir de una población de 5 elementos es

$$\frac{5!}{3!(5-3)!} = 10$$

Dichas muestras son las siguientes, junto con sus promedios aritméticos correspondientes:

	$\bar{x}_i$		$\bar{x}_i$
1, 2, 3	6/3	3, 4, 5	12/3
1, 2, 4	7/3	3, 4, 1	8/3
1, 2, 5	8/3	4, 5, 1	10/3
2, 3, 4	9/3	4, 5, 2	11/3
2, 3, 5	10/3	5, 1, 3	9/3

Para calcular la media y la desviación estándar, se emplea la siguiente tabla

$\bar{x}_i$	6/3	7/3	8/3	8/3	9/3	9/3	10/3	10/3	11/3	12/3
$\bar{x}_i^2$	36/9	49/9	64/9	64/9	81/9	81/9	100/9	100/9	121/9	144/9

$$\sum_{i=1}^{10} \bar{x}_i = 90/3 \qquad\qquad \sum_{i=1}^{10} \bar{x}_i^2 = 840/9$$

$$\mu_{\bar{x}} = \bar{\bar{x}} = \frac{1}{10}\sum_{i=1}^{10} \bar{x}_i = \frac{1}{10}\cdot\frac{90}{3} = 3$$

$$\sigma_{\bar{x}}^2 = \frac{1}{10}\sum_{i=1}^{10} \bar{x}_i^2 - \bar{\bar{x}}^2 = \frac{1}{10}\cdot\frac{840}{9} - (3)^2 =$$

$$= 9.333 - 9.000 = 0.333 \Rightarrow \sigma_{\bar{x}} = \sqrt{0.333} = 0.577$$

Es decir, $\mu_{\bar{x}} = 3$ y $\sigma_{\bar{x}} = 0.577$

Segundo procedimiento.

Por tratarse de una población finita, se verifica que

$$\mu_{\bar{x}} = \mu \qquad \text{y} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

en donde $N_p = 5$, $n = 3$ y $\mu = 3$.

El valor de σ^2 de la población es

$$\sigma^2 = \frac{1+4+9+16+25}{5} - (3)^2 = \frac{55}{5} - 9 = 11-9 = 2$$

Por lo tanto, $\sigma = \sqrt{2} = 1.4145$ y

$$\sigma_{\bar{x}} = \frac{1.4145}{\sqrt{3}}\sqrt{\frac{5-3}{5-1}} = (0.8164)(0.7071) = 0.577$$

Es decir, $\mu_{\bar{x}} = 3$ y $\sigma_{\bar{x}} = 0.577$

Comparando los resultados, se puede observar que ambos procedimientos conducen a la obtención de los mismos valores de $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$ para la distribución muestral del promedio aritmético.

Ejemplo 4.2

En una bodega se tienen cinco mil varillas de acero; el valor medio del peso, X, de cada varilla es de 5.02 kg, y la desviación estándar 0.3 kg. Hallar la probabilidad de que una muestra de cien varillas, escogida al azar, tenga un peso total

a. entre 496 y 500 kg

b. de más de 510 kg.

Para la distribución muestral del promedio, se tiene que $\mu_{\bar{X}} = \mu = 5.02$ kg y, por tratarse de una población finita,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} = \frac{0.30}{\sqrt{100}} \sqrt{\frac{5000 - 100}{5000 - 1}} = 0.027$$

a. El peso total de la muestra estará entre 496 y 500 kg si el peso promedio de las cien varillas se encuentra entre 4.96 y 5.00 kg. Puesto que la muestra es mayor de 30 elementos se puede considerar como aproximadamente normal a la distribución muestral, y los valores estándar correspondientes a $\bar{X} = 4.96$ y a $\bar{X} = 5.00$ se obtienen mediante la transformación

$$z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

es decir,

$$z_1 = \frac{4.96 - 5.02}{0.027} = -2.22$$

$$z_2 = \frac{5.00 - 5.02}{0.027} = -0.74$$

En la fig 4.1 se puede apreciar que

$$P[496 < X < 500] = P[-2.22 < z < -0.74] =$$

$$= P[-2.22 < z < 0] - P[-0.74 < z < 0]$$

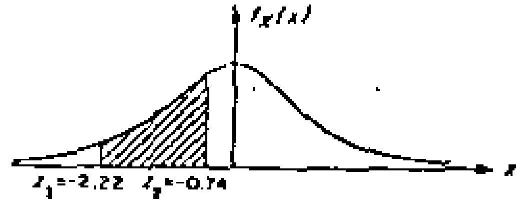


Fig 4.1 *Distribución normal correspondiente al ejemplo*

Recurriendo a la tabla de áreas bajo la curva normal estándar entre 0 y z queda finalmente

$$P[496 \leq X < 500] = 0.4868 - 0.2704 = 0.2164$$

b. El peso total de la muestra excederá de 510 kg si el peso promedio de las cien varillas pasa de 5.10 kg.

Estandarizando dicho valor, queda

$$z_3 = \frac{5.10 - 5.02}{0.027} = 2.96$$

Calculando el área bajo la curva normal a la derecha de este valor (fig 4.2), se tiene que

$$P[X > 510] = P[z > 2.96] = P[z > 0] - P[0 \leq z \leq 2.96] =$$

$$= 0.5 - 0.4985 = 0.0015$$

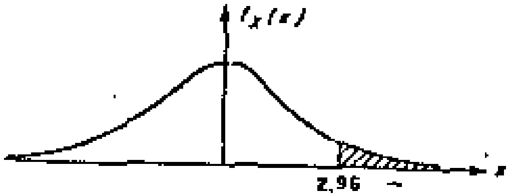


Fig 4.1 Distribución normal correspondiente al ejemplo

5. Distribución muestral de diferencias de promedios aritméticos

Con frecuencia se presenta el caso en el que se tienen datos de dos poblaciones con variables aleatorias asociadas X y Y, respectivamente, surgiendo la duda de si estas se pueden considerar como una sola, es decir, si X = Y. Para probar estadísticamente esta hipótesis (como se verá más adelante), es necesario obtener las distribuciones muestrales de la diferencia de los promedios y de las variancias de las muestras de ambas variables.

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de muestras aleatorias de tamaño n_X y n_Y de dos poblaciones con características X y Y, respectivamente. Se puede demostrar que la distribución muestral de la diferencia de los promedios correspondientes a poblaciones infinitas con medias μ_X y μ_Y y desviaciones estándar σ_X y σ_Y, tiene los siguientes parámetros:

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X} + \frac{\sigma^2_Y}{n_Y}}$$

si las muestras son independientes.

Esta distribución también es aplicable a poblaciones finitas si el muestreo es con remplazo. Para el caso de poblaciones finitas en las cuales el muestreo se hace sin remplazo, los parámetros de la distribución muestral de la diferencia de los promedios aritméticos son

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X}\,\frac{N_X - n_X}{N_X - 1} + \frac{\sigma^2_Y}{n_Y}\,\frac{N_Y - n_Y}{N_Y - 1}}$$

suponiendo que las muestras sean independientes.

Ejemplo 5.1

Considérese que de una población X se obtienen tres muestras posibles, cuyos correspondientes promedios aritméticos son 3, 7 y 8. De otra población Y se extraen dos muestras posibles, con promedios 2 y 4, respectivamente. Se deben obtener los parámetros de la distribución muestral de las diferencias de los promedios aritméticos.

Primer procedimiento

Todas las posibles diferencias de promedios aritméticos de X con los de Y serían

$$\begin{matrix} 3-2 & 7-2 & 8-2 \\ 3-4 & 7-4 & 8-4 \end{matrix} \longrightarrow \begin{matrix} 1 & 5 & 6 \\ -1 & 3 & 4 \end{matrix}$$

Es decir,

$$\mu_{\bar{X}-\bar{Y}} = \frac{-1+1+3+4+5+6}{6} = \frac{18}{6} = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{(-1-3)^2 + (1-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 + (6-3)^2}{6} =$$

$$= \frac{34}{6} = \frac{17}{3}$$

Segundo procedimiento

Se sabe que

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} \ ; \ \sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}$$

Por ello,

$$\mu_{\bar{X}} = \frac{3+7+8}{3} = \frac{18}{3} = 6$$

$$\mu_{\bar{Y}} = \frac{2+4}{2} = \frac{6}{2} = 3$$

$$\sigma^2_{\bar{X}} = \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3} = \frac{14}{3}$$

$$\sigma^2_{\bar{Y}} = \frac{(2-3)^2 + (4-3)^2}{2} = \frac{2}{2} = 1$$

$$\mu_{\bar{X}-\bar{Y}} = 6 - 3 = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{14}{3} + 1 = \frac{17}{3}$$

Se observa que ambos procedimientos conducen a los mismos resultados.

Ejemplo 5.2

Las varillas de acero que fábrica una compañía A tienen un peso medio de 6.5 kg y una desviación estándar de 0.4, en tanto que las producidas por una empresa B tienen un peso medio de 6.3 kg y una desviación estándar de 0.3 kg. Si se toman muestras aleatorias de 100 varillas de cada fábrica, ¿cuál es la probabilidad de que las de la compañía A tengan un peso promedio de por lo menos:

a. 0.35 kg

b. 0.10 kg

mayor que el de la compañía B?

Se puede suponer en este caso que las distribuciones muestrales involucradas son normales, en virtud de que el tamaño de ambas muestras es mayor de 30 elementos. También se puede suponer que ambas poblaciones son infinitas, y siendo $\bar{X}_A$ y $\bar{X}_B$ los pesos promedios de las muestras de las fábricas A y B, respectivamente, entonces

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_{\bar{X}_A} - \mu_{\bar{X}_B} = 6.5 - 6.3 = 0.20 \text{ kg}$$

$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} = 0.05 \text{ kg}$$

La variable estandarizada de la diferencia de los promedios es

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - \mu_{\bar{X}_A - \bar{X}_B}}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 0.20}{0.05}$$

a. Estandarizando la diferencia de 0.35 kg se llega a

$$Z_1 = \frac{0.35 - 0.20}{0.05} = \frac{0.15}{0.05} = 3$$

La probabilidad deseada es el área bajo la curva normal a la derecha de $Z = 3$, es decir

$$P[\bar{X}_A > \bar{X}_B + 0.35] = P[Z > 3] = 0.500 - 0.4987 = 0.0013$$

b. Al estandarizar la diferencia de 0.10 kg, la variable Z resulta

$$Z_2 = \frac{0.10 - 0.20}{0.05} = \frac{-0.1}{0.05} = -2$$

La probabilidad requerida es el área bajo la curva normal a la derecha de $Z = -2$, es decir

$$P[\bar{X}_A > \bar{X}_B + 0.10] = P[Z > -2] = 0.5 + 0.4772 = 0.9772$$

6. Teoría estadística de la estimación

En la práctica profesional a menudo resulta necesario inferir información acerca de una población mediante el uso de muestras extraídas de ella; una parte básica de dicha inferencia consiste en *estimar* los valores de los parámetros de la población (media, variancia, etc.) a partir de las estadísticas correspondientes de la muestra, como se explica a continuación.

7. Estimadores puntuales. Clasificación

Si un estimador de un parámetro de la población consiste en un solo valor de una estadística, se le conoce como *estimador puntual* del parámetro.

Cuando la media de la distribución muestral de una estadística es igual al parámetro que se está estimando de la población, entonces la estadística se conoce como *estimador insesgado* del parámetro; si no sucede así, entonces se denomina *estimador sesgado*. Ambos estimadores son puntuales, y sus valores correspondientes se llaman estimaciones insesgadas o sesgadas, respectivamente. Dicho de otra manera, si S es una estadística cuya distribución muestral tiene media μ_S, y el parámetro correspondiente de la población es θ, se dice que S es un estimador insesgado de θ si

$$\mu_S = \theta$$

Por otra parte, si la estadística S_n de la muestra tiende a ser igual al parámetro θ de la población a medida que se hace más grande el tamaño de la muestra, entonces la estadística recibe el nombre de *estimador consistente* del parámetro.

Empleando símbolos, si

$$\lim_{n\to\infty} S_n = \theta$$

resulta que la estadística S_n es un estimador consistente. Por ejemplo, el promedio aritmético es un estimador insesgado y consistente de la media, y la variancia de la muestra es un estimador sesgado y consistente de la variancia de la población.

Si las distribuciones muestrales de varias estadísticas tienen el mismo valor de la media, se dice que la estadística que cuenta con la menor variancia es un *estimador eficiente* de dicha media, en tanto que las estadísticas restantes se conocen como *estimadores ineficientes* del parámetro.

Por ejemplo, las distribuciones muestrales del promedio aritmético y de la mediana cuentan con medias que son, en ambos casos, iguales a la media de la población. Sin embargo, la variancia de la distribución muestral del promedio aritmético es menor que la de la distribución de la mediana, por lo que el promedio aritmético obtenido de una muestra aleatoria proporciona un estimador eficiente de la media de la población, en tanto que la mediana obtenida de la muestra proporciona un estimador ineficiente de dicho parámetro.

9. Estimación de intervalos de confianza para los parámetros de una población

La estimación de un parámetro de una población mediante un par de números entre los cuales se encuentra, con cierta probabilidad, el valor de dicho parámetro, se llama estimación del intervalo del mismo.

Sea S una estadística obtenida de una muestra de tamaño n para estimar el valor del parámetro θ, y sea σ_S la desviación estándar (conocida o estimada) de su distribución muestral. La probabilidad, $1 - \alpha$, de que el valor de θ se localice en el intervalo de $S - z_c\,\sigma_S$ a $S + z_c\,\sigma_S$, donde z_c es una constante, se escribe en la forma

$$P[S - z_c\,\sigma_S \le \theta \le S + z_c\,\sigma_S] = 1 - \alpha$$

Si se fija el valor de $1 - \alpha$, se puede obtener el valor de z_c necesario para que se satisfaga la ecuación anterior, con lo cual queda definido el *intervalo de confianza* del parámetro θ, $(S - z_c\,\sigma_S,\ S + z_c\,\sigma_S)$, correspondiente al nivel de *confianza* $1 - \alpha$.

La constante z_c que fija el intervalo de confianza se conoce como *valor crítico*. Si la distribución de S es normal, el valor de z_c correspondiente a uno de α se obtiene de la tabla de áreas bajo la curva normal o de la tabla 9.1 siguiente.

TABLA 8.1 VALORES DE z_c PARA DISTINTOS NIVELES DE CONFIANZA

Nivel de confianza, en porcentaje	z_c
99.73	3.00
99.00	2.58
98.00	2.33
96.00	2.05
95.45	2.00
95.00	1.96
90.00	1.64
80.00	1.28
68.27	1.00
50.00	0.674

Ejemplo 8.1

Sea el promedio aritmético $\bar{X}$ una estadística con distribución normal. Las probabilidades o niveles de confianza de que $\mu_{\bar{X}}$ (o μ de la población) se encuentre localizada entre los límites $\bar{X} \pm \sigma_{\bar{X}}$, $\bar{X} \pm 2\ \sigma_{\bar{X}}$ y $\bar{X} \pm 3\ \sigma_{\bar{X}}$ son 68.26, 95.44 y 99.73%, respectivamente, obteniéndose dichos valores de la tabla de áreas bajo la curva normal. Lo anterior significa que el intervalo $\bar{X} \pm 3\ \sigma_{\bar{X}}$ contendrá a $\mu_{\bar{X}}$ en el 99.73 por ciento de las muestras de tamaño n, por lo que los intervalos de confianza de 68.26, 95.44 y 99.73 por ciento para estimar a μ son $(\bar{X} - \sigma_{\bar{X}}, \bar{X} + \sigma_{\bar{X}})$ $(\bar{X} - 2\ \sigma_{\bar{X}}, \bar{X} + 2\ \sigma_{\bar{X}})$ y $(\bar{X} - 3\ \sigma_{\bar{X}}, \bar{X} + 3\ \sigma_{\bar{X}})$, lo cual se aprecia en la *fig* 8.1 siguiente.

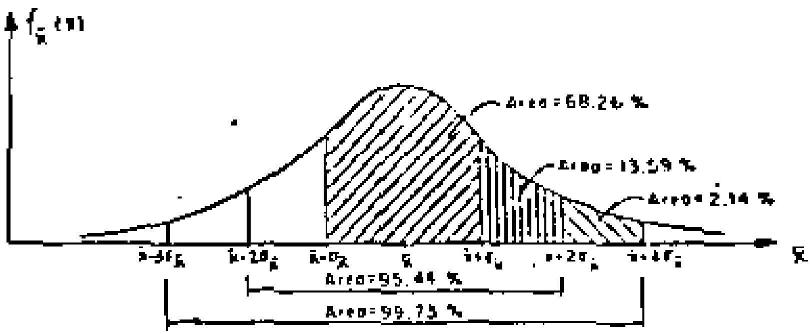


Fig 8.1

9. Estimación de intervalos de confianza para la media

Los límites de confianza para la media de una población con variable aleatoria X asociada están dados por

$$\bar{X} \pm z_c \, \sigma_{\bar{X}}$$

en donde z_c depende del nivel de confianza deseado. Si $\bar{X}$ tiene distribución normal, z_c puede obtenerse en forma directa de la tabla 8.1. Por ejemplo, los límites de confianza de 95 y 99 por ciento para estimar la media, μ, de la población son $\bar{X} \pm 1.96\sigma_{\bar{X}}$ y $\bar{X} \pm 2.58\,\sigma_{\bar{X}}$, respectivamente. Al obtener estos límites hay que usar el valor calculado de $\bar{X}$ para la muestra correspondiente.

Entonces, los límites de confianza para la media de la población quedan dados por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}}$$

en caso de que el muestreo se haga a partir de una población infinita o de que se efectúe con remplazo a partir de una población finita, o por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

si el muestreo es sin remplazo a partir de una población finita de tamaño N_p.

Ejemplo 9.1

Las mediciones de los diámetros de una muestra aleatoria de 100 tubos de albañal mostraron una media de 32 cm y una desviación estándar de 2 cm. Obténganse los límites de confianza de

a. 95 por ciento
b. 97 por ciento

para el diámetro medio de todos los tubos.

a. De la tabla 8.1, los límites de confianza del 95 por ciento son

$$\bar{X} \pm 1.96\sigma/\sqrt{n} = 32 \pm 1.96(2/\sqrt{100}) = 32 \pm 0.392 \text{ cm}$$

o sea 31.608 y 32.392, en donde se ha empleado el valor de S_x para estimar el de σ de la población, puesto que la muestra es suficientemente grande (mayor de 30 elementos). Esto significa

que con una probabilidad de 95 por ciento, el valor de μ_x se encuentra entre 31.608 y 32.392 cm.

b. Si $Z = z_c$ es tal que el área bajo la curva normal a la derecha de z_c es el 1.5 por ciento del área total, entonces el área entre 0 y z_c es 0.5 - 0.015 = 0.485, por lo que de la tabla de áreas bajo la curva normal se obtiene z_c = 2.17. Por lo tanto, los límites de confianza del 97 por ciento son:

$$\bar{x}\pm 2.17\sigma/\sqrt{n} = 32\pm 2.17(2/\sqrt{100}) = 32\pm 0.434 \text{ cm}$$

y el intervalo de confianza respectivo es (31.566 cm, 32.434 cm).

Ejemplo 9.2

Una muestra aleatoria de 50 calificaciones de cierto examen de admisión tiene un promedio aritmético de 72 puntos, con desviación estándar igual a 10. Si el examen se aplicó a 1018 personas, obtener

a. El intervalo de confianza del 95% para la media del total de calificaciones.

b. El tamaño de muestra necesario para que el error en la estimación de la media no exceda de 2 puntos, considerando el mismo nivel de confianza.

c. El nivel de confianza para el cual la media de la población sea 72 ± 1 puntos.

a. Si se estima a σ de la población con S_x de la muestra y se considera que la población es finita, los límites de confianza son, puesto que $\bar{x}$ = 72, Z_c = 1.96, S_x = 10, N_p = 1018 y n = 50,

$$72 \pm 1.96 \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm 1.96\ (1.4142)\ (0.9755)$$

$$72 \pm 2.704$$

y el intervalo de confianza respectivo es

(69.296, 74.704)

b. Puesto que el error en la estimación de la media es, para población finita,

$$\text{Error en la estimación} = Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en este caso se tendría

$$Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} < 2$$

o sea, para un nivel de confianza de 95%,

$$1.96 \frac{10}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

$$\frac{19.6}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

Elevando al cuadrado la desigualdad, queda

$$\frac{384.16}{n}\ \frac{1018-n}{1017} < 4$$

o sea

$$87.85 < n$$

Por lo cual, se requieren al menos 88 elementos en la muestra para que el error en la estimación no exceda de 2 puntos, para $1-\alpha = 0.95$.

c. Los límites de confianza son, en este caso

$$72 \pm Z_c \frac{10}{\sqrt{50}}\sqrt{\frac{1018-50}{1018-1}}$$

$$72 \pm Z_c\,(1.4142)(0.9755)$$

o sea

$$72 \pm 1.3795\, Z_c$$

Puesto que se desea que el valor de la media sea 72 ± 1 puntos, se verifica que

$$1 = 1.3795\, Z_c$$

Es decir

$$Z_c = \frac{1}{1.3795} = 0.725$$

El área bajo la curva normal estándar entre 0 y $Z_c = 0.725$ es, por interpolación lineal, igual a 0.2657. Por lo tanto, el nivel de confianza es igual al doble del área anterior, es decir, 2(0.2657) = 0.5314 (o 53.14%), tal como se muestra en la fig 9.1.

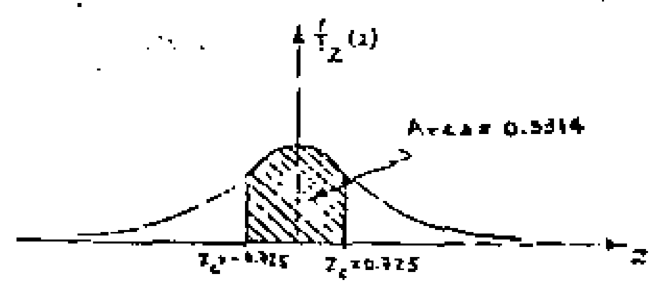


Fig 9.1

10. Intervalos de confianza para diferencias de medias

Los límites de confianza para la diferencia de las medias cuando las poblaciones X y Y son infinitas, o cuando el muestreo se realiza con reemplazo de poblaciones finitas, se encuentran dados por

$$\bar{X} - \bar{Y} \pm Z_c\,\sigma_{\bar{X}-\bar{Y}} = \bar{X} - \bar{Y} \pm Z_c\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}$$

en donde $\bar{X}$, n_X y $\bar{Y}$, n_Y son los respectivos promedios aritméticos y tamaños de las dos muestras extraídas de las poblaciones, y σ_X y σ_Y las desviaciones estándar de estas últimas.

En el caso de que las poblaciones X y Y sean finitas y el muestreo sin remplazo, los límites de confianza son

$$\bar{X} - \bar{Y} \pm Z_c \sigma_{\bar{X}-\bar{Y}} = \bar{X} - \bar{Y} \pm Z_c \sqrt{\frac{\sigma_X^2}{n_X} \frac{N_X - n_X}{N_X - 1} + \frac{\sigma_Y^2}{n_Y} \frac{N_Y - n_Y}{N_Y - 1}}$$

en donde N_X y N_Y son los tamaños de las poblaciones X y Y, respectivamente.

Las dos ecuaciones anteriores son válidas únicamente si las muestras aleatorias seleccionadas son independientes.

Ejemplo 10.1

Para el ejemplo de las varillas tratado anteriormente (5.2), encontrar el intervalo de confianza del 95.45% para las diferencias de las medias de las poblaciones.

Siendo $\bar{X}_A = \mu_A = 6.5$ kg, $\sigma_A = 0.4$ kg, $\bar{X}_B = \mu_B = 6.3$ kg, $\sigma_B = 0.3$ kg y $n_A = n_B = 100$, los límites de confianza para la diferencia de las medias son, empleando la tabla 8.1

$$\bar{X}_A - \bar{X}_B \pm Z_c \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 6.5 - 6.3 \pm 2 \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} =$$

$$= 0.2 \pm 0.1$$

Por lo tanto, el intervalo de confianza respectivo es (0.1, 0.3).

Ejemplo 10.2

Se tienen en una bodega 3000 focos de marca X, y 5000 de marca Y. Se extrae una muestra aleatoria de 150 focos de la marca X, y se obtiene una duración promedio de 1400 horas, con desviación estándar igual a 120 horas. Otra muestra aleatoria de 200 focos de la marca Y tuvo una duración promedio de 1200 horas, con desviación estándar igual a 80 horas. Obtener intervalos de confianza de

a. 95%

b. 99%

para la diferencia de los tiempos medios de duración de los focos de ambas marcas.

a. Puesto que se trata de poblaciones finitas y $\bar{X} = 1400$ h, $S_X = 120$ h, $N_X = 3000$, $n_X = 150$, $\bar{Y} = 1200$ h, $S_Y = 80$ h, $N_Y = 5000$ y $n_Y = 200$, se obtiene, estimando a σ_X y σ_Y con S_X y S_Y, respectivamente

$$1400 - 1200 \pm 1.96 \sqrt{\frac{(120)^2}{150} \frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200} \frac{5000 - 200}{5000 - 1}}$$

$$200 \pm 1.96\ (11.04)$$

$$200 \pm 21.638$$

o sea, (178.362, 221.638), puesto que de la tabla 8.1, para un nivel de confianza de 95%, $Z_c = 1.96$.

b. En este caso, al emplear la tabla 8.1 se obtiene

$Z_c = 2.58$ para un nivel de confianza de 99%, por lo cual los límites son

$$1400 - 1200 \pm 2.58 \sqrt{\frac{(120)^2}{150}\,\frac{3000-150}{3000-1} + \frac{(80)^2}{200}\,\frac{5000-2000}{5000-1}}$$

$$200 \pm 2.58\ (11.04)$$

$$200 \pm 28.483$$

y el intervalo de confianza es

$$(171.517,\ 228.483)$$

11. Pruebas de hipótesis

Supóngase que una empresa armadora de automóviles está en la disyuntiva de emplear una nueva marca de bujías en sus unidades o la que regularmente utiliza, y que su departamento de control de calidad debe decidir, con base en la información de las muestras de las dos marcas distintas. Las decisiones de este tipo, es decir, que se basan en estudios estadísticos, reciben el nombre de *decisiones estadísticas*, y a los procedimientos que permiten decidir si se acepta o rechaza una hipótesis se les llama *pruebas de hipótesis*, *pruebas de significancia* o *reglas de decisión*.

Al tomar decisiones estadísticas, es necesario postular las diversas alternativas o cursos de acción que pueden adoptarse.

En el caso particular de una *prueba de hipótesis* solamente se tienen dos cursos de acción posibles, los que se denotarán como H_0 y H_1. A la acción H_0 se le llama *hipótesis nula*, y a la H_1, *hipótesis alternativa*. Por ejemplo, si la hipótesis nula establece que $\mu_1 = \mu_2$, la hipótesis alternativa puede ser una de las siguientes:

$$\mu_1 > \mu_2,\ \mu_1 < \mu_2 \text{ o } \mu_1 \neq \mu_2$$

Al realizar una prueba de hipótesis, se prueba siempre la verdad de la hipótesis nula H_0, aun cuando de antemano se desee rechazarla.

12. Errores de los tipos I y II. Nivel de significancia

En muchas ocasiones se presenta el caso de que se rechaza una hipótesis nula cuando en realidad debería ser aceptada; cuando esto sucede se dice que se ha cometido un *error de tipo I*. En otras ocasiones se acepta una hipótesis nula siendo en realidad falsa; en este caso se dice que se ha cometido un *error de tipo II*.

Al probar una hipótesis nula, a la máxima probabilidad con la que se está dispuesto a cometer un error del tipo I se le llama *nivel de significancia*, α, de la prueba, el cual dentro de la práctica se acostumbra establecer de 5 por ciento (0.05) o 10 por ciento (0.1). El complemento del nivel de significancia, 1 − α, se conoce como *nivel de confianza*.

Si, por ejemplo, al realizar una prueba de hipótesis se escoge un nivel de significancia de 10 por ciento, significa que existen 10 posibilidades en 100 de que se rechace ésta cuando debería ser aceptada; es decir, que se rechaza a un nivel de significancia del 10 por ciento, y que la probabilidad de que la decisión haya sido errónea es de 0.1.

13. Comportamiento de los errores tipos I y II

Supóngase que se trata de probar la hipótesis nula de que la media, μ_S, de la distribución muestral de la estadística S es μ_1, en contra de la hipótesis alternativa que establece que $\mu_S = \mu_2$, donde $\mu_2 > \mu_1$, es decir

$$H_0 : \mu_S = \mu_1$$

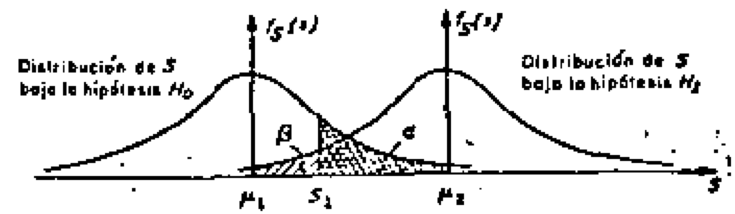
$$H_1 : \mu_S = \mu_2$$

En la fig 13.1 se muestra en forma gráfica la relación entre los errores tipos I y II en el caso en el que la regla de decisión para aceptar o rechazar H_0 es la siguiente:

Si el valor de la estadística S obtenido de una muestra excede de cierto valor crítico S_1, recházese H_0; en caso contrario, acéptese.

Es evidente que si H_0 es verdadera, entonces α (área con rayado doble) es la probabilidad de que $S > S_1$, o sea la de rechazar a H_0 siendo verdadera (error tipo I). Por otro lado, si H_1 es verdadera, entonces β (área con rayado sencillo) es la probabilidad de que $S < S_1$, o sea la de aceptar H_0 siendo falsa (error tipo II).

Obsérvese que si se aumenta el valor de S_1 se reduce la probabilidad α, pero se incrementa la β; lo contrario sucede si se disminuye el valor de S_1.



$$P[S > S_1] = \alpha \text{ (error tipo I)}$$
$$P[S < S_1] = \beta \text{ (error tipo II)}$$

Fig 13.1 Probabilidades de los errores tipos I y II en pruebas de hipótesis.

En realidad, la única forma posible en la cual se pueden minimizar simultáneamente los errores de tipos I y II es aumentando el tamaño de la muestra, para hacer más "picudas" las distribuciones muestrales de la estadística bajo las hipótesis H_0 y H_1.

Al observar la fig 13.2 siguiente, es posible concluir que el tamaño de los errores I y II es menor para un tamaño de muestra igual a 100 que para un tamaño igual a 50, considerando la misma regla de decisión anterior.

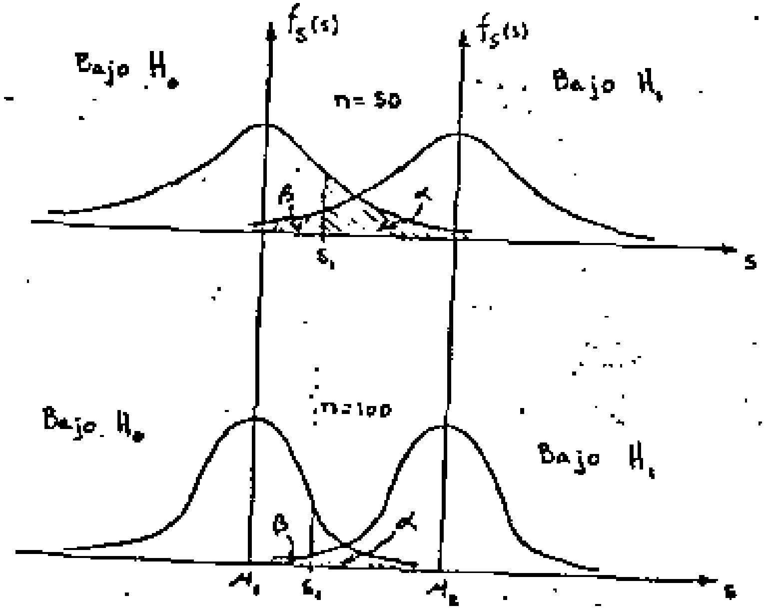


Fig 13.2

Sin embargo, esta técnica de reducción simultánea de ambos tipos de errores no siempre puede ponerse en práctica, debido a razones de costo, tiempo, etc.

14. Regiones críticas, de rechazo ó de significancia. Regiones de aceptación.

Cuando una hipótesis nula no se acepta se dice que se *rechaza a un nivel de significancia del α por ciento*, o que el valor estandarizado de la estadística involucrada es *significativo a un nivel de significancia* α.

Al conjunto de los valores de la estadística en el que se rechaza la hipótesis nula se le denomina *región crítica, de rechazo, o de significancia*. Por el contrario, al conjunto de los valores de la estadística en que se acepta la hipótesis, se le llama *región de aceptación*.

Considérese que la distribución muestral de la estadística S es normal con desviación estándar σ_S, que la variable Z resulta de estandarizar a S, que la hipótesis nula, H_0, es que la media de S vale μ_S, y que la hipótesis alternativa H_1 es que dicha media es diferente de μ_S, es decir, que

$$Z = \frac{S - \mu_S}{\sigma_S}$$

H_0: media de la distribución muestral de $S = \mu_S$

H_1: media de la distribución muestral de $S \neq \mu_S$

Si se adopta la regla de decisión de aceptar la hipótesis H_0, si el valor de Z cae dentro del intervalo central que encierra al 99 por ciento del área de la distribución de probabilidades, entonces H_0 se aceptará en el caso en que

$$-2.58 \le Z \le 2.58$$

empleando la tabla de áreas bajo la curva normal estándar. Pero si el valor estandarizado de la estadística se encuentra fuera de dicho intervalo, se concluye que el evento puede ocurrir con probabilidad de 0.01 si la hipótesis H_0 es verdadera (área rayada total de la fig 14.1). En tal caso, el valor Z de la variable estándar difiere *significativamente* del que se podría esperar de acuerdo con la hipótesis nula, lo cual inclina a rechazarla a un nivel de confianza del 99 por ciento.

De lo anterior se deduce que el área total rayada de la fig 14.1 es el nivel de significancia α de la prueba, y representa la probabilidad de cometer un error del tipo I. Por ello, la región de aceptación de H_0 es $-2.58 \le Z \le 2.58$, y la de rechazo es $Z > 2.58$ y $Z < -2.58$.

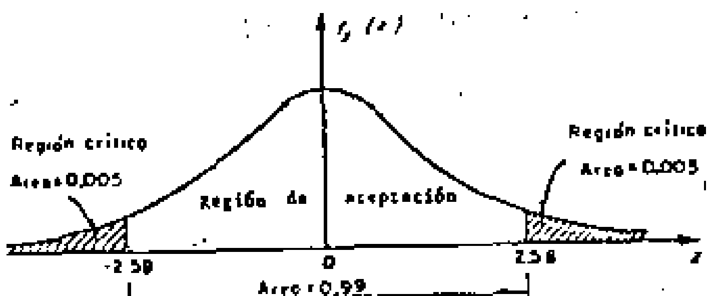


Fig 14.1 Región de significancia

En la tabla 14.1 se presentan los valores de la variable estandarizada, Z, que limitan las regiones de aceptación y de rechazo para el caso en el que la estadística involucrada en la prueba tenga distribución muestral normal. Cuando en alguna prueba de hipótesis se consideren niveles de significancia diferentes a los que aparecen en la tabla mencionada, resulta necesario emplear la de áreas bajo la curva normal estándar.

TABLA 14.1 VALORES CRITICOS DE z

Nivel de significancia, α	Valores de z para pruebas de una cola	Valores de z para pruebas de dos colas
0.1	−1.281 o 1.281	−1.645 y 1.645
0.05	−1.645 o 1.645	−1.960 y 1.960
0.01	−2.326 o 2.326	−2.575 y 2.575
0.005	−2.575 o 2.575	−2.810 y 2.810

15. Pruebas de una y de dos colas

En la prueba de hipótesis del ejemplo anterior, la región de rechazo de la hipótesis nula quedó en ambos extremos (colas) de la distribución muestral de la estadística involucrada en la prueba; a las pruebas de este tipo se les denomina *pruebas de dos colas*. Cuando la región de rechazo se encuentra solamente en un extremo de la distribución muestral en cuestión, se les llama *pruebas de una cola*.

Las pruebas de dos colas se presentan cuando en la hipótesis alternativa aparece el signo ≠ (diferente de), como en el siguiente caso

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S \neq \mu_1$$

en donde μ_S es la media de la estadística S, y μ_1 es un valor fijo.

En los casos

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S < \mu_1$$

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S > \mu_1$$

las pruebas resultan de una cola.

16. Pruebas de hipótesis para la media

Para el caso de una población infinita (o finita en que se muestree con remplazo), cuya desviación estándar σ se conoce o se puede estimar adecuadamente, si se tiene que la estadística S obtenida de la muestra es el promedio aritmético, entonces la media de su distribución muestral es $\mu_S = \mu_{\bar{X}} = \mu$, y su desviación estándar es $\sigma_S = \sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ son, respectivamente, la media y la desviación estándar de la variable aleatoria X asociada a la población, y n es el tamaño de la muestra. En tal caso, si $\bar{X}$ tiene distribución *normal*, la variable estandarizada correspondiente será

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Para el caso de muestreo sin remplazo de población finita, se tiene que $\sigma_S = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$, en donde N_p es el tamaño de la población, por lo que la variable estandarizada será

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}}$$

En los dos casos anteriores, el valor de Z correspondiente al de $\bar{X}$ de la muestra es el que se debe comparar con el valor crítico correspondiente al nivel de significancia fijado, para así aceptar o no la hipótesis nula (prueba de una cola). Si se trata de una prueba de dos colas, el valor de Z se debe comparar con los dos valores críticos que corresponden al valor de α seleccionado. En cualquiera de los casos anteriores, el valor o valores críticos se pueden obtener de la tabla 14.1, para valores comunes de α.

Ejemplo 16.1

Se sabe que el promedio de calificaciones de una muestra aleatoria de tamaño 100 de los estudiantes de tercer año de ingeniería civil es de 7.6, con una desviación estándar de 0.2. Si μ denota la media de la población de esas calificaciones, X, y si se supone que $\bar{X}$ tiene distribución normal, probar la hipótesis

$\mu = 7.65$ en contra de la hipótesis alternativa $\mu \neq 7.65$, usando un nivel de significancia de

a. 0.05

b. 0.01

Para la solución se deben considerar las hipótesis

$$H_0 : \mu = 7.65$$

$$H_1 : \mu \neq 7.65$$

Puesto que $\mu \neq 7.65$ incluye valores menores y mayores de 7.65, se trata de una prueba de dos colas.

La estadística bajo consideración es el promedio aritmético, $\bar{X}$, de la muestra, que se supone extraída de una población infinita. La distribución muestral de $\bar{X}$ tiene media $\mu_{\bar{X}} = \mu$, y desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ denotan, respectivamente, la media y la desviación estándar de la población de calificaciones.

Bajo la hipótesis H_0 (considerándola verdadera), se tiene que

$$\mu_{\bar{X}} = 7.65 = \mu$$

y utilizando la desviación estándar de la muestra como una estimación de σ, lo cual se supone razonable por tratarse de una muestra grande,

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.2/\sqrt{100} = 0.2/10 = 0.02$$

a. Para la prueba de dos colas a un nivel de significancia de 0.05 se establece la siguiente regla de decisión

Aceptar H_0 si el valor Z correspondiente al valor del promedio de la muestra se encuentra dentro del intervalo de -1.96 a 1.96 (tabla 14.1). En caso contrario, rechazar H_0.

Puesto que

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{7.6 - 7.65}{0.02} = -2.5$$

se encuentra fuera del rango de -1.96 a 1.96, se rechaza la hipótesis H_0 a un nivel de significancia de 0.05.

b. Si el nivel de significancia es 0.01, el intervalo de -1.96 a 1.96 de la regla de decisión del inciso a se remplaza por el de -2.58 a 2.58 tabla(14.1). Entonces, puesto que el valor muestral $Z = -2.5$ se encuentra dentro de este intervalo, se acepta la hipótesis H_0 a un nivel de significancia de 0.01.

Ejemplo 16.2

La resistencia media a la ruptura de cables de acero fabricados por la empresa X es de 905 kg. Una empresa consultora sugiere a X que cambie su proceso de manufactura, con lo cual incrementará la resistencia de sus cables. Se prueba el nuevo proceso, y se extrae una muestra aleatoria de 50 cables, obteniéndose para ellos una resistencia promedio de 926 kg, con desviación estándar igual a 42 kg. ¿Se puede considerar que el nuevo proceso realmente incrementa la resistencia, con un nivel de confianza de 99%?

En este caso, se debe plantear una prueba de hipótesis de una cola, para la cual

$$H_0 : \mu = 905 \text{ kg}$$

$$H_1 : \mu > 905 \text{ kg}$$

Puesto que el tamaño de la muestra es suficientemente grande, se puede aproximar la distribución muestral de la resistencia promedio mediante una normal, y estimar el valor de σ de la población mediante S_X de la muestra.

Considerando a la población infinita, y suponiendo como verdadera a H_0, se tiene que

$$\mu_{\bar{X}} = \mu = 905 \text{ kg}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{50}} = 5.94$$

Para la prueba de una cola a un nivel de significancia de $\alpha = 1 - (1 - \alpha) = 1 - 0.99 = 0.01$, la regla de decisión es

Aceptar H_0 si el valor estandarizado de $\bar{X}$ de la muestra es menor o igual a $Z_c = 2.326$ (tabla T4.1); en caso contrario, rechazar H_0.

En virtud de que

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{926 - 905}{5.94} = 3.535$$

es mayor de 2.326, se rechaza H_0 a un nivel de significancia de 1%, concluyéndose que en realidad el nuevo proceso sí incrementa la resistencia de los cables.

17. Pruebas de diferencias de medias

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de dos muestras de tamaños n_X y n_Y, extraídas respectivamente de dos poblaciones con medias μ_X y μ_Y, y desviaciones estándar σ_X y σ_Y. Se trata de probar la hipótesis nula, H_0, de que no existe diferencia entre las medias, es decir, que $\mu_X = \mu_Y$. Si n_X y n_Y son suficientemente grandes (>30), la distribución muestral de las diferencias de los promedios es aproximadamente normal. Dicha distribución muestral es rigurosamente normal si las variables aleatorias X y Y asociadas a la población tienen distribución normal, aunque n_X y n_Y sean menores de 30. Para esta distribución muestral, la variable estandarizada Z, que se compara con los valores críticos correspondientes, se encuentra dada por

$$Z = \frac{\bar{X} - \bar{Y} - \mu_{\bar{X}-\bar{Y}}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y} - 0}{\sigma_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}}$$

con la cual se puede probar la hipótesis nula H_0 en contra de otras hipótesis alternativas, H_1, a un nivel apropiado de significancia.

Ejemplo 17.1

En el laboratorio de pruebas de una empresa fabricante de aparatos electrónicos se ensayaron dos marcas de transistores, A y B, de características similares, con objeto de comprobar su ganancia de voltaje. Se tomaron muestras aleatorias de 100 transistores de cada marca, arrojando una ganancia promedio de 31 decibeles, con desviación estándar de 0.3 decibeles para la marca A, y 30.9 decibeles de ganancia promedio, con desviación estándar de 0.4 decibeles para la otra. ¿Existe una diferencia significativa entre las ganancias en voltaje de los transistores a un nivel de significancia de

a. 0.05

b. 0.01?

Si μ_A y μ_B son las medias respectivas de las dos poblaciones infinitas a las que corresponden las muestras, la prueba de hipótesis adopta la forma siguiente:

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

Entonces, el valor de Z es, bajo la hipótesis H_0:

$$Z = \frac{\bar{X}_A - \bar{X}_B}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{31 - 30.9}{\sqrt{\frac{(0.3)^2}{100} + \frac{(0.4)^2}{100}}} = 2$$

a. Puesto que se trata de una prueba de dos colas a un nivel de significancia de 0.05, la diferencia es significativa si el valor de Z se encuentra fuera del intervalo de -1.96 a 1.96. Como este es el caso, puede concluirse que efectivamente existe diferencia significativa en la ganancia en voltaje de los transistores.

b. Si la prueba es a un nivel de significancia de 0.01, la diferencia es significativa si Z se encuentra fuera del rango de -2.58 a 2.58. Partiendo del hecho de que Z = 2, la diferencia entre las ganancias es producto del azar, y se acepta la hipótesis de que ambos tipos de transistores tienen igual ganancia media en voltaje a un nivel de confianza de 99 por ciento.

Ejemplo 17.2

La estatura promedio de 50 estudiantes varones tomados al azar que participan en actividades deportivas es de 173 cm., con desviación estándar de 6.3 cm. Otra muestra aleatoria de 50 estudiantes varones que no participan en ese tipo de actividades tiene promedio de estatura igual a 171 cm. con desviación estándar igual a 7.1 cm. Probar la hipótesis de que los estudiantes varones que practican deportes son más altos que los que no lo hacen, a un nivel de significancia de 0.05.

Se debe decidir entre las hipótesis

$$H_0 : \mu_X = \mu_Y$$

$$H_1 : \mu_X > \mu_Y$$

siendo X la variable aleatoria asociada a la población infinita de estaturas de alumnos que practican deportes, y Y la asociada a la de estudiantes que no lo hacen, que también es infinita.

Bajo la hipótesis H_0 , se tiene que

$$\mu_{\bar{X}-\bar{Y}} = 0$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} = \sqrt{\frac{(6.3)^2}{50} + \frac{(7.1)^2}{50}} = 1.3424$$

Entonces, el valor de Z es

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{173 - 171}{1.3424} = \frac{2}{1.3424} = 1.489$$

Puesto que se trata de una prueba de hipótesis de una cola, a un nivel $\alpha = 0.05$, se rechazaría H_0 si el valor de Z muestral fuera mayor del valor crítico para dicho nivel, el cual es $Z_c = 1.645$. Puesto que $Z < Z_c$, en este caso se concluye que la diferencia en las estaturas de ambos grupos de estudiantes se debe únicamente al azar.

tiene ordenadas mayores de cero en el lado de las abscisas negativas. De hecho, la estadística S_x^2 se puede estudiar si se consideran muestras aleatorias de tamaño n extraídas de una población normal con desviación estándar σ_x y si para cada muestra se calcula el valor de la estadística.

$$\chi^2 = \frac{n\,S_x^2}{\sigma^2} \qquad (3.14)$$

donde S_x^2 es la varianza de la muestra.

El número de grados de libertad, ν, de una estadística se define como

$$\nu = n - k$$

siendo n el tamaño de la muestra y k el número de parámetros de la población que deben estimarse a partir de ella.

La distribución muestral de la estadística χ^2 está dada por la ecuación

$$f(\chi^2) = U\,\chi^{\nu-2}\,e^{-\frac{1}{2}\chi^2}$$

en la que U es una constante que hace que el área total bajo la curva resulte igual a uno, y $\nu = n - 1$ es el número de grados de libertad. Esta distribución se llama *Ji cuadrada*, misma que se presenta en la fig 21 para distintos valores de ν.

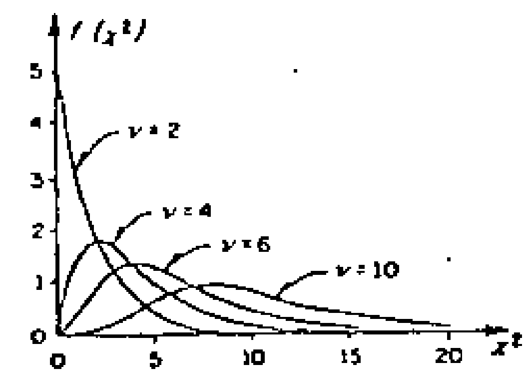


Fig 21. Distribución Ji cuadrada para distintos valores de ν

3.4 Muestras pequeñas

Como ya se indicó, para muestras grandes ($n > 30$) las distribuciones muestrales de muchas estadísticas son aproximadamente normales, siendo tanto mejor la aproximación cuanto mayor es el tamaño de n. Sin embargo, cuando se trata de muestras en las que $n < 30$, llamadas *muestras pequeñas*, la aproximación no es suficientemente buena, por lo que resulta necesario introducir una teoría apropiada para su estudio.

Al estudio de las distribuciones muestrales de las estadísticas para muestras pequeñas se le llama *teoría estadística de las muestras pequeñas*. Existen al respecto tres distribuciones importantes: *Ji cuadrada*, *F* y *t de Student*.

3.4.1 Distribución *Ji* cuadrada (χ^2)

Hasta ahora solo se ha tratado la distribución muestral de la media. En esta sección se verá lo concerniente a la distribución muestral de la variancia, S_X^2, para muestras aleatorias extraídas de poblaciones normales. Puesto que S_X no puede ser negativa, es de esperarse que su distribución muestral no sea una curva normal, ya que esta

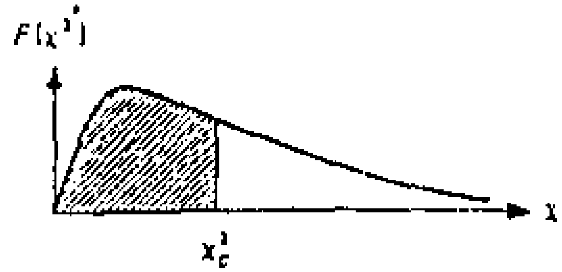


TABLA 8. VALORES CRITICOS χ_c^2

ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.016	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.2	12.8	11.15	9.2	6.63	4.35	2.67	1.61	1.15	.831	.554	.413
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.18	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.35	7.57	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.2	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.7	30.6	27.5	25.1	22.3	18.2	14.3	11.0	8.55	7.26	6.25	5.22	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.45	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.8	35.6	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.02
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.5	13.15	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.5	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.7	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	43.0	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.32	82.4	77.9	74.2	70.1	67.3

No obstante que la distribución Ji cuadrada solo se ha presentado en el estudio de las muestras pequeñas, cabe aclarar que es válida para aquellas mayores de 30 si la variable aleatoria involucrada tiene distribución normal.

3.4.1.1 Intervalo de confianza para la varianza

Tal como se hizo para la distribución normal, se pueden establecer intervalos de confianza para la varianza de la población en términos de la varianza de una muestra extraída de ella, a un nivel de confianza dado $1 - \alpha$, si se hace uso de los valores críticos χ_c^2 de la tabla 8. Por lo tanto, un intervalo de confianza para la estadística χ^2, estaría dado por

$$\chi_{c'}^2 < \frac{n S_X^2}{\sigma^2} < \chi_c^2$$

donde $\chi_{c'}^2$ y χ_c^2 son los valores críticos para los cuales el $(1 - \alpha)/2$ por ciento del área se encuentra en los extremos izquierdo y derecho de la distribución, respectivamente.

Con base en lo anterior, se concluye que

$$\frac{n S_X^2}{\chi_c^2} < \sigma^2 < \frac{n S_X^2}{\chi_{c'}^2}$$

es un intervalo de confianza para estimar a σ^2 a un nivel de confianza $1 - \alpha$.

3.4.1.2 Prueba de hipótesis para la varianza

La prueba de hipótesis para la varianza de una población normal se efectúa calculando el valor de la estadística χ^2 y estableciendo las hipótesis H_0 y H_1 apropiadas, es decir, se adoptan reglas de decisión similares a las usadas para la estadística Z.

Ejemplo

La varianza del tiempo de elaboración de cierto producto es igual a 40 min; sin embargo, su proceso de manufactura se modifica y se toma una muestra de veinte tiempos, para la cual la varianza resulta ser igual a 62 min. ¿Es significativo el aumento del tiempo de elaboración a un nivel de significancia de

a) 0.05

b) 0.01?

Se debe decidir de entre las hipótesis

$$H_0 : \sigma^2 = 40 \text{ min}$$
$$H_1 : \sigma^2 > 40 \text{ min}$$

Suponiendo que la hipótesis nula es correcta, el valor de la estadística χ^2 para la muestra considerada es

$$\chi^2 = \frac{n S_X^2}{\sigma^2} = \frac{(20)(62)}{40} = 31$$

a) Como se trata de una prueba de una cola, la hipótesis H_0 se rechazaría si el valor de la estadística χ^2 fuera mayor que el de χ^2 para un nivel de significancia igual a 0.05, el cual, para $\nu = 20 - 1 = 19$ grados de libertad resulta ser 30.1 (tabla 8). Como $31 > 30.1$, H_0 se rechaza a un nivel de significancia de 0.05.

b) En este caso, el valor de χ^2 para un nivel de significancia de 0.01 y 19 grados de libertad es igual a 36.2. Puesto que $31 < 36.2$, se acepta H_0 a un nivel de significancia de 0.01.

3.4.2 Distribución F

Al efectuar la prueba de hipótesis de igualdad de medias para muestras pequeñas, en la siguiente sección se supondrá que las variancias de las poblaciones a las que corresponden tales muestras son iguales. Por lo tanto, es necesario probar antes si tal suposición es correcta. Para ello, debe considerarse que si S_X^2, n_X y S_Y^2, n_Y son respectivamente la variancia y el tamaño de dos muestras extraídas de poblaciones normales que tienen igual variancia, entonces

$$F = \frac{S_X^2}{S_Y^2} \qquad (3.15)$$

TABLA 9. VALORES F_c PARA α = 0.01

v_2 = Grados de libertad del denominador	v_1 = Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.50	99.00	99.20	99.20	99.30	99.30	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.50	99.50	99.50	99.50	99.50	99.50
3	34.10	30.80	29.50	28.70	28.20	27.90	27.70	27.50	27.30	27.20	27.10	26.90	26.70	26.60	26.50	26.40	26.30	26.20	26.10
4	21.20	18.00	16.70	16.00	15.50	15.20	15.00	14.80	14.70	14.50	14.40	14.20	14.00	13.90	13.80	13.70	13.70	13.60	13.50
5	16.30	13.30	12.10	11.40	11.00	10.70	10.50	10.30	10.20	10.10	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.70	10.90	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.20	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.30	8.65	7.59	7.01	6.63	6.37	6.17	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.60	8.02	6.99	6.42	6.06	5.81	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.00	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.68	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.03	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.95	2.87
16	8.53	6.23	5.29	4.77	4.43	4.20	4.03	3.89	3.78	3.69	3.55	3.40	3.26	3.18	3.10	3.02	2.93	2.84	2.76
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.06	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.36	4.04	3.81	3.64	3.50	3.41	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.46	3.30	3.17	3.07	2.98	2.84	2.71	2.54	2.47	2.39	2.29	2.20	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

resulta ser el valor de una variable aleatoria (estadística) que tiene distribución F, con parámetros $\nu_X = n_X - 1$ y $\nu_Y = n_Y - 1$. Esta distribución (fig 22) cuenta con dos parámetros, ν_X y ν_Y, que son los grados de libertad que corresponden a la variancia del numerador y del denominador de la ec 3.15, respectivamente. Cuando se hace referencia a una distribución F en particular, siempre se dan primero los grados de libertad para la variancia del numerador; es decir, $F(\nu_X, \nu_Y)$. En la tabla 9 se presentan los valores críticos F_c para distintos valores de ν_X y ν_Y y un nivel de significancia de 0.01. Cuando los grados de libertad ν_X o ν_Y no se encuentren en dicha tabla, el valor de F se puede obtener mediante interpolación lineal. Si se desea probar la hipótesis a otros niveles de significancia, es factible emplear las tablas de la distribución F (refs 9 y 11).

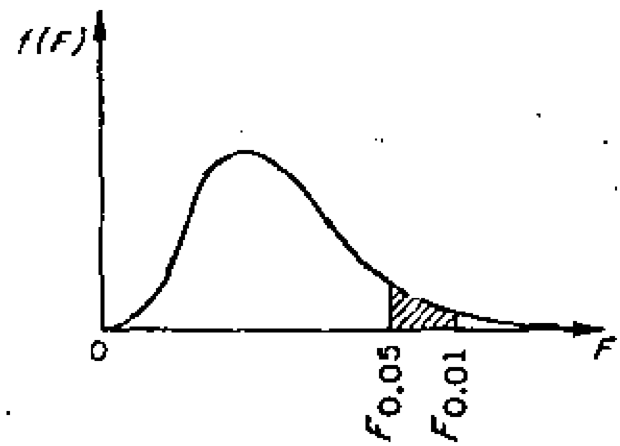


Fig 22. Distribución F.

De acuerdo con lo anterior, se puede probar la hipótesis nula

$$H_0: \ \sigma_X^2 = \sigma_Y^2$$

en contra de alguna hipótesis alternativa adecuada haciendo uso del hecho de que el cociente S_X^2/S_Y^2 es una estadística que tiene distribución F.

Ejemplo

Una empresa manufacturera de cartón prensado va a decidir acerca del empleo de una prensadora A o una B a fin de obtener un grosor determinado en su producto. El problema estriba en que ambas prensadoras proporcionan grosores muy similares, es decir, que la variancia de los grosores para las dos máquinas es la misma. Para decidir acertadamente, se toma una muestra aleatoria de 31 cartones prensados por la máquina A y otra de 41 por la B. Como las variancias del grosor para los cartones de las muestras resul-

tan ser de 12 y de 5 micras, respectivamente, se establecen las hipótesis

$$H_0:\ \sigma_A^2 = \sigma_B^2$$

$$H_1:\ \sigma_A^2 > \sigma_B^2$$

con objeto de probarlas a un nivel de significancia de 0.01.

El valor de la estadística F resulta

$$F = \frac{S_A^2}{S_B^2} = \frac{12}{5} = 2.4$$

Puesto que $\nu_A = 31 - 1 = 30$ y $\nu_B = 41 - 1 = 40$, en la tabla 9 se puede ver que para un nivel de significancia de 0.01 el valor, F_c, de $F(30, 40)$ es 2.11. De acuerdo con estos valores, la hipótesis H_0 se rechazaría si el valor de F fuera mayor que F_c (30, 40).

Puesto que lo anterior resulta ser cierto, se rechaza H_0, concluyéndose que la prensadura B sería la mejor elección.

3.4.3 Distribución t de Student

Si se consideran muestras de tamaño n extraídas de una población normal con media μ y variancia desconocida, para cada muestra se puede calcular la estadística T definida mediante la fórmula

$$T = \frac{\overline{X} - \mu}{S_X} \sqrt{n - 1} \tag{3.16}$$

donde $\overline{X}$ es el promedio y S_X la desviación estándar de la muestra.

La distribución muestral de T (fig 23) está dada por la ecuación

$$f(t) = \frac{U}{\left(1 + \frac{t^2}{\nu}\right)^{(\nu + 1)/2}}$$

$(\nu+1)/2$ va de exponente

en la que U es una constante que hace que el área bajo la curva sea igual a uno, y $\nu = n - 1$ es el número de grados de libertad.

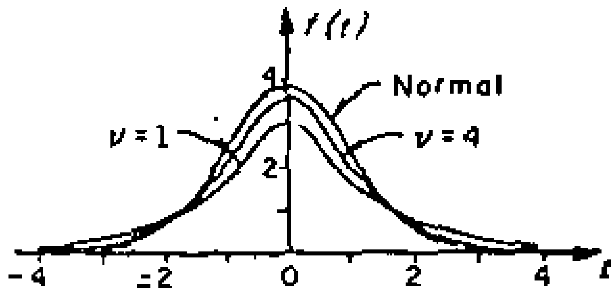


Fig 23. Distribución t de Student para distintos valores de ν

En la fig 23 se aprecia que conforme ν (o n, el tamaño de la muestra) aumenta, la distribución de $f(t)$ se aproxima a la distribución normal.

3.4.3.1 Límites e intervalos de confianza

De manera similar a como se hizo con la distribución normal, es posible estimar los límites de confianza de la media, μ, de una población mediante los *valores críticos*, t_c, de la distribución t, que dependen del tamaño de la muestra y del nivel de confianza deseado, encontrándose dichos valores en la tabla 10.

Así pues,

$$- t_c < \frac{\bar{X} - \mu}{S_X} \sqrt{n-1} < t_c$$

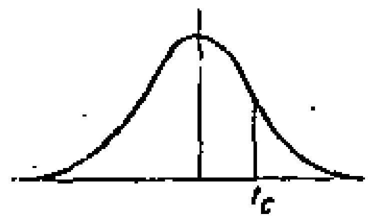
representa un intervalo de confianza para t, a partir del cual se puede estimar que μ se encuentra dentro del intervalo

$$\bar{X} - t_c \frac{\sigma_X}{\sqrt{n-1}} < \mu < \bar{X} + t_c \frac{\sigma_X}{\sqrt{n-1}}$$

En términos generales, los límites de confianza para la media de la población se representan como

$$\bar{X} \pm t_c \frac{\sigma_X}{\sqrt{n-1}}$$

TABLA 10. VALORES t_c PARA LA DISTRIBUCION t DE STUDENT



ν	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.70}$	$t_{.60}$	$t_{.55}$
1	63.66	31.82	12.71	6.31	3.07	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.275	.138
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.04	3.36	2.58	2.02	1.48	.920	.727	.560	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.91	1.43	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.36	.871	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.693	.537	.258	.128
15	2.95	2.61	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.128
19	2.87	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.256	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.05	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.71	1.31	.855	.683	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.30	.853	.683	.530	.256	.127
40	2.70	2.43	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.528	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

3.4.3.2 Pruebas de hipótesis

La prueba de hipótesis para la media de una población se puede efectuar con muestras pequeñas en forma análoga a la de muestras de tamaño mayor de 30 si en lugar de utilizar a la estadística Z se emplea la T. Entonces, si se consideran dos muestras aleatorias cuyos tamaños, desviaciones estándar y promedios son n_X, S_X, $\bar{X}$ y n_Y, S_Y, $\bar{Y}$, respectivamente, extraídas de poblaciones normales de igual variancia ($\sigma_X^2 = \sigma_Y^2$), se puede probar la hipótesis, H_0, de que las muestras provienen de una misma población, es decir, de que también sus medias son iguales, utilizando la estadística T definida por

$$T = \frac{\bar{X} - \bar{Y}}{\epsilon \sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \tag{3.17}$$

donde

$$\epsilon = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \tag{3.18}$$

cuya distribución es la t de Student, con $\nu = n_X + n_Y - 2$ grados de libertad.

Ejemplo

Conforme al plan de desarrollo agrícola de una región, se probó un nuevo fertilizante para maíz. Para ello se escogieron 24 ha de terreno, aplicándose dicho producto a la mitad de ellas. El promedio de producción de maíz en la zona que se usó fertilizante fue de 5.3 ton, con una desviación estándar de 0.40 ton, en tanto que en la otra zona el promedio fue de 5.0 ton, con desviación estándar de 0.36 ton.

De acuerdo con los resultados, ¿se puede concluir que existe un aumento significativo en la producción de maíz al usar fertilizante, si se utiliza un nivel de significancia de

a) 0.01

b) 0.05?

Solución

Para probar la hipótesis de igualdad de medias es indispensable saber primero si las muestras provienen de dos poblaciones normales de igual variancia. En ese caso, si σ_X^2 y σ_Y^2 denotan a las variancias de la producción de maíz en la zona tratada y en la no tratada, respectivamente, se debe probar la hipótesis nula H_0: $\sigma_X^2 = \sigma_Y^2$ en contra de la hipótesis alternativa H_1: $\sigma_X^2 > \sigma_Y^2$ a los dos niveles de significancia establecidos.

El valor de la estadística F es, de la ec 3.15,

$$F = \frac{S_X^2}{S_Y^2} = \frac{(0.40)^2}{(0.36)^2} = 1.27 .$$

y el valor crítico de F (11, 11), obtenido de la tabla 9 mediante interpolación lineal, resulta 4.47. Por lo tanto, como $1.27 < 4.47$, se acepta la hipótesis nula a un nivel de significancia de 0.01.

El valor crítico de F (11, 11) a un nivel de significancia de 0.05 (ref 9) es 2.82, de ahí que como $1.27 < 2.82$, también se acepta la hipótesis H_0.

Con base en lo anterior, se debe decidir entre las hipótesis

H_0: $\mu_X = \mu_Y$ (la diferencia en los promedios se debe al azar)

H_1: $\mu_X > \mu_Y$ (el fertilizante mejora la producción)

Bajo la hipótesis H_0, se tiene que

$$\epsilon = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{12(0.40)^2 + 12(0.36)^2}{12 + 12 - 2}} = 0.397$$

por lo cual

$$\frac{5.3 - 5.0}{0.397 \sqrt{\frac{1}{12} + \frac{1}{12}}} = 1.85$$

a) Puesto que se trata de una prueba de una cola a un nivel de significancia de 0.01, se rechaza la hipótesis H_0 si t es mayor que el valor crítico, t_c, correspondiente a dicho nivel, el cual para $\nu = n_X + n_Y - 2 = 12 + 12 - 2 = 22$ grados de libertad, se obtiene de la tabla 8 como $t_c = 2.51$. Como $t < t_c$, la hipótesis H_0 no se puede rechazar a un nivel de significancia de 0.01.

b) Si el nivel de significancia de la prueba es de 0.05, se rechaza H_0 si t es mayor que el valor t_c respectivo que para 22 grados de libertad es $t_c = 1.72$, por lo que de acuerdo con lo anterior, H_0 se rechaza a un nivel de significancia de 0.05.

DIVISION DE EDUCACION CONTINUA
FACULTAD DE INGENIERIA U.N.A.M.

DISEÑO ESTADISTICO DE EXPERIMENTOS

ANALISIS DE VARIANCIA EN REGRESION LINEAL

Ing. Bernardo Frontana de la Cruz

MAYO, 1982

Palacio de Minería Calle de Tacuba 5 primer piso Deleg. Cuauhtemoc 06000 México, D.F. Tel.: 521-40-20 Apdo. Postal M-2285

8. ANALISIS DE VARIANCIA EN REGRESION LINEAL

8.1 Asociación entre variable

En el análisis estadístico se pueden tener datos UNIVARIANTES y MULTIVARIANTES. Los primeros corresponden a una única observación de cada unidad elemental de una muestra de la población (una sola variable). Las estadísticas muestrales calculadas con estos datos se utilizan para hacer inferencias acerca de los parámetros correspondientes a la población univariante relacionada. Cuando cada unidad elemental de una población puede dar dos o más medidas, referidas a una caracterización específica tenemos una población MULTIVARIANTE; por ejemplo, los gastos de consumo medio se pueden asociar con una variedad de factores tales como el ingreso disponible, tamaño y distribución de efectivo, edades, etc. En particular, una POBLACION BIVARIANTE es la que contiene dos medidas en cada unidad elemental; por ejemplo, podemos observar la altura y el peso de cada individuo de una población adulta.

La técnica de estimación por asociación es, en realidad, un método de predicción, siendo la predicción la función central de las ciencias. La tarea principal de cualquier estudio científico es descubrir las relaciones generales entre las variables observadas y expresar la naturaleza de tales relaciones en forma matemáticamente precisa de manera que pueda predecirse el valor de una con base en otra (u otras). La toma de decisiones por asociación en estadística comercial y económica permite, entre otras cosas:

a) reducir los costos en la toma de decisiones

b) encontrar una variable de explicación suficientemente consistente cuando restringimos nuestra investigación al análisis bivariante

c) Aumentar la precisión

Existen dos aspectos distintos pero complementarios en el estudio de la asociación entre variables. El primero llamado ANALISIS DE REGRESION trata de establecer "la naturaleza de la relación entre las variaibles"; esto es, se estudia la relación funcional entre las variables a fin de predecir el valor de una con base en las otras. Convencionalmente la predicha se llama VARIABLE DEPENDIENTE y las variables básicas de la predicción son las VARIABLES INDEPENDIENTES.

El segundo aspecto del análisis por asociación se conoce como ANALISIS DE CORRELACION y trata de determinar "el grado de relación entre las variables".

De lo anterior puede observarse que el análisis de asociación puede clasificarse en ANALISIS DE ASOCIACION SIMPLE para cuando hay una sola variable independiente y ANALISIS DE ASOCIACION MULTIPLE para cuando hay más de una variable independiente. Además conforme a la relación funcional entre las variables, el análisis de asociación puede diferenciarse entre LINEAL y NO LINEAL.

2 Variancia explicada e inexplicada

Los cálculos necesarios para ajustar ecuaciones de regresión lineal ya han sido discutidos con algún detalle. Aquí consideraremos como tratar estos problemas vía los métodos de análisis de variancia.

Recordemos que si ajustamos una regresión de la forma $E[y/x] = a + b\,x$ usando n parejas de valores $(X_j\ Y_j)$ (j= 1,2,...,n) el estimador de b es;

$$B = \frac{\Sigma\,(X_j - \bar X)\,(Y_j - \bar Y)}{\Sigma\,(X_j - \bar X)^2}$$

y el de a: $A = \bar y - B\,\bar x$

El modelo de regresión puede escribirse como $Y_j = a + b\,x_j + z_j$ donde z_j satisfacen las condiciones del análisis de variancia.

En la figura 1 la línea de regresión ajustada $E[y/x] = A + Bx$ pasa, como se explicó, por el punto $G(\bar x, \bar y)$ que es "el centro de gravedad" del conjunto de puntos observados, de los cuales $P_j(X_j, X_j)$ es uno

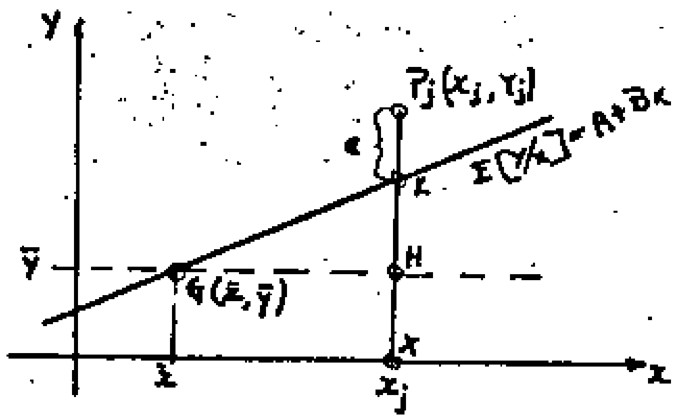


Fig. 1 Una recta de regresión ajustada

Obsérvese que la perpendicular al eje x desde P_j establece los puntos K en la intersección con la recta de regresión, H en la intersección con la recta $\bar y$ y X al intersectarse con el eje x; donde k y h tienen por coordenadas $K(X_j, A+B\,X_j)$ y $H(X_j, \bar y)$.

El segmento P_jX puede dividirse en P_jK, KH y HX o bien en términos algebráicos:

1)... $Y_j = \bar y + (A+BX_j - \bar y) + (Y_j - A - BX_j)$

Como 2)... $\bar y = A + B\bar X$, sustituyendo en el primer paréntesis de

1) tenemos: 3)... $Y_j = \bar y + B\,(X_j - \bar X) + (Y_j - A - B\,X_j)$

De aquí podemos calcular la suma total de cuadrados como:

$$\Sigma(Y_j - \bar y)^2 = \Sigma\left[B\,(X_j - \bar X) + (Y_j - A - B\,X_j)\right]^2$$

4)... $\underline{\Sigma(Y_j - \bar Y)^2 = B^2\,\Sigma(X_j - \bar X)^2 + \Sigma(Y_j - A - BX_j)^2}$

ya que:

$$2\,B\,\Sigma(X_j - \bar X)\,(Y_j - A - BX_j) = 2B\,\Sigma(X_j - \bar X)\left[(Y_j - \bar Y) - B\,(X_j - \bar X)\right]$$

$=$

$$= 2B\left[\Sigma(X_j - \bar X)\,(Y_j - \bar Y) - B(X_j - \bar X)^2\right] = 2B\left[\Sigma(X_j - \bar X)(Y_j - \bar Y) - (X_j - \bar X)(Y_j - \bar Y)\right]$$

$= 0$

La expresión 4) muestra que la suma total de cuadrados $\Sigma(Y_j - \bar Y)^2$ está dividida en dos partes; la primera:

$$B^2\,\Sigma(X_j - \bar X)^2 = \Sigma(Yc - \bar y) = \Sigma(A + BX_j - A - B\bar X) = B^2\,\Sigma(X_j - \bar X)^2$$

Es entonces proporcional a RH y mide la cantidad de variación de las Y's "explicada" por la recta de regresión ajustada; por lo tanto, se le llama "la suma de cuadrados debida a la regresión lineal de Y sobre X o más brevemente "suma de cuadrados debida a la regresión".

como sabemos $\sigma^2 = E(\hat{B}^2) - E^2(\hat{B})$ entonces:

5)... $E\left[B^2 \Sigma(X_j-\bar{X})^2\right] = \sigma^2 + b^2 \Sigma(X_j-\bar{X})^2$

y el número de grados de libertad de esta suma de cuadrados es 1 (el coeficiente de σ^2).

La segunda suma de cuadrados 4) si observamos la figura, corresponde a la de las desviaciones de los valores observados Y_j respecto a los valores predichos para la regresión. En otras palabras, es la suma de los cuadrados de los errores "no explicados" debidos a la aleatorización. Por tanto a esta suma de cuadrados se le llama "alrededor de la regresión" o suma de cuadrados "residual". En efecto

$$Y_j - A - Bx_j = a + bx_j + E_j - A - Bx_j$$

$$= E_j - (A - a) - (B - b)\, X_j$$

Dado que $E(A) = a$ y $E(B) = b$ y A y B no dependen en otro sentido a y b, se sigue que

$\Sigma(Y_j - A - Bx_j)^2 = \Sigma(E^2)$ y su valor esperado es un múltiplo de σ^2 o sea:

$E\left[\Sigma(Y_j - A - Bx_j)^2\right] = \lambda\sigma^2$ podemos encontrar λ calculando el valor esperado de 4):

$$E\left[\Sigma(Y_j - \bar{Y})^2\right] = E\left[B^2\Sigma(X_i - \bar{X})^2\right] + E\left[\Sigma(Y_j - A - Bx_j)^2\right]$$

como λ no depende de b podemos hacer $b = 0$

$$(n-1)\,\sigma^2 = \sigma^2 + \lambda\,\sigma^2$$

de donde

$$\lambda = n-2$$

luego entonces la suma de cuadrados residual tiene n-2 grados de libertad. Podemos resumir los resultados obtenidos en la tabla de análisis de variancia siguiente:

TABLA I. Análisis de variancia de la regresión lineal

F u e n t e	G. de l.	S.S.	MS
regresión lineal	1	$B^2\Sigma(X_j - \bar{X})^2$	$B^2\Sigma(X_j - \bar{X})^2$
residual (alrededor de la regresión)	n-2	$\Sigma(Y_j - A - Bx_j)^2$	$[\Sigma(Y_j - A - Bx_j)^2]/(n-2)$
T o t a l	n-1	$\Sigma(Y_j - \bar{Y})^2$	

La estadística $F = \dfrac{MSR}{MSE} = \dfrac{(n-2)\,B^2\,\Sigma(X_j - \bar{X})^2}{\Sigma(Y_j - A - Bx_j)^2}$

Se compara con la distribución $F_{1,n-2}$ para probar la hipótesis Ho: $b=0$ contra la alternativa H1: $b \neq 0$ (independencia entre X y Y en la población).

8.3 Ejemplo 1 Un fabricante de soldaduras de puntos de aluminio de alta resistencia al esfuerzo cortante desea predecir la resistencia al esfuerzo cortante por los diámetros de la soldadura de punto en lugar de destruir el producto con ese propósito. Una muestra de diez soldaduras, escogidas para establecer la relación entre las dos variables dio los siguientes resultados:

Diámetro de la soldadura (cm)	Resistencia al esfuerzo cortante (1000 kg)
2.4	7.0
1.8	5.3
1.6	4.2
1.0	3.3
1.2	3.8
1.1	6.6
2.8	8.5
1.6	6.6
1.5	4.5
2.3	8.8

La estimación de la ecuación de regresión poblacional resultó ser

$$Yc = 1.481 + 2.531\ X$$

Par probar la independencia entre las variables X y Y de la población establecemos la hipótesis:

$$Ho:\ b = 0$$

$$Hi:\ b \neq 0$$

Para probar dicha hipótesis construyemos nuestra tabla de análisis de variancia:

$$\bar{x} = \frac{2.4 + 1.8 + \ldots + 1.5 + 23}{10} = 1.73$$

$$\Sigma (X_j - \bar{X})^2 = (2.4 - 1.73)^2 + (1.8 - 1.73)^2 + \ldots + (1.5 - 1.73)^2 + (2.3 - 1.63)^2 = 3.22$$

$$B^2\ \Sigma (X_j - \bar{X})^2 = 2.531^2 X\ 3.22 = 20.6272$$

Observamos que $\Sigma (Y_j - A - BX_j)^2 = \Sigma (Y_j - Y_C)^2$ donde Y_C se obtiene para los valores de X por la recta de regresión. Con esto:

valores observados diámetro X	res. al cortante Y	resistencia calculada Y_C	$Y-Y_C$	$(Y-Y_C)^2$
2.4	7.0	7.56	−0.56	0.3136
1.8	5.3	6.04	−0.74	0.5476
1.6	4.2	5.53	−1.33	1.7689
1.0	3.3	4.01	−0.71	0.5041
1.2	3.8	4.52	−0.72	0.5184
1.1	6.6	4.26	+2.34	5.4756
2.8	8.5	8.57	−0.07	0.0049
1.6	6.6	5.53	+1.07	1.1449
1.5	4.5	5.28	−0.78	0.6084
2.3	8.8	7.30	+1.50	2.2500
17.3	58.6	58.60	0	13.1364

n = 10

La tabla ANOVA será:

Tabla 2. ANOVA para la regresión lineal de resistencias al cortante entre los diámetros de soldadura

Fuente	G. de l.	S.S.	MS	Fc
regresión lineal	1	20.6272	20.6272	12.5619
residual (alrededor de la regresión)	8	13.1364	1.6421	
Total	9	33.7636		

Para un nivel de significancia $\alpha = 0.05$, $F_{0.05,\ 1,8} = 3.46$

Como F teórica < F calculada (3.46 < 12.5619) entonces rechazamos Ho implicando que sí hay dependencia entre los diámetros de la soldadura y la resistencia al esfuerzo cortante con una significancia estadística del 95%. Dicha dependencia se explica con la relación funcional Y = 1.481 + 2.531 X.

8.3 Análisis de variancia en regresión lineal múltiple. Recordemos que para encontrar los coeficientes de regresión lineal con dos variables independientes habrá que resolver el sistema normal

$$n a_{1.23} + b_{12.3}\,\Sigma X_2 + b_{13.2}\,\Sigma X_3 = \Sigma Y$$

1)... $a_{1.23}\,\Sigma X_2 + b_{12.3}\,\Sigma X_2^2 + b_{13.2}\,\Sigma X_2X_3 = \Sigma YX_2$

$$a_{1.23}\,\Sigma X_3 + b_{12.3}\,\Sigma X_2X_3 + b_{13.2}\,\Sigma X_3^2 = \Sigma X_3 Y$$

puesto que:

$$\Sigma(y - \bar{y}) = \Sigma(X_2 - \bar{x}_2) = \Sigma(X_3 - \bar{X}_3) = 0$$

si hacemos la transformación:

$$y' = y - \bar{y} \; ; \; X_2' = X_2 - \bar{x} \;\; y \;\; X_3' = X_3 - \bar{X}$$

cambiamos el origen de las ecuaciones normales de (0,0,0) a $(\bar{y}, \bar{X}_2, \bar{X}_3)$ reduciendo 1) a 2 ecuaciones en términos de las desviaciones alrededor de las medias:

$$b_{12.3}\,\Sigma x_2' + b_{13.2}\,\Sigma x_2'^2 x_3' = y' x_2'$$

2)... $b_{12.3}\, x_2' x_3' + b_{13.2}\, x_3'^2 = y' x_3'$

que resolviendo obtenemos los coeficientes de regresión parciales $b_{12.3}$ y $b_{13.2}$ y al tercero lo encontramos de

3)... $a_{1.23} = \bar{y} - b_{12.3}\,\bar{X}_2 - b_{13.2}\,\bar{X}_3$

Para comentar el análisis de variancia para este caso consideremos el siguiente ejemplo:

La compañía de cigarros FIPA comenzará su XI año de operaciones y se considera una empresa próspera en la industria. A fin de programar su producción requiere un pronóstico de las ventas totales. Se sospecha que éstas dependen, entre otros factores, de la publicidad de su producto y del índice comparativo de precios (el precio de su producto comparado con el precio medio de otras marcas similares en porciento). Se dispone de datos históricos de la década pasada para estos factores, los cuales se muestran junto con los porcentajes correspondientes:

Tabla III Datos históricos de la compañía FIPA

	Datos originales			Datos originales expresados como porcentaje		
Año	Y	X_2	X_3	Y	X_2	X_3
1	24	4	80	6.80	6.06	8.25
2	27	4	80	7.65	6.06	8.25
3	31	5	90	8.78	7.58	9.28
4	29	5	100	8.22	7.58	10.31
5	33	6	100	9.35	9.09	10.31
6	38	7	110	10.76	10.61	11.34
7	37	8	120	10.48	12.12	12.37
8	40	8	100	11.33	12.12	10.31
9	45	9	90	12.75	13.64	9.28
10	49	10	100	13.38	15.15	10.31
Total	353	66	970	100	100	100

En la tabla: y = ventas anuales en millones de pesos

X_2 = gastos anuales de publicidad en millones de pesos

X_3 = P ó índice comparativo de precios

Con los datos tenemos lo siguiente:

$$\bar{Y} = \bar{x}_2 = \bar{x}_3 = 100/10 = 10$$

$\Sigma y^2 = 1046.03$ $\quad \Sigma X_2^2 = 1092.91$ $\quad \Sigma X_3^2 = 1015.18$

$\Sigma X_2 y = 1064.11$ $\quad \Sigma yX_3 = 1011.73$ $\quad \Sigma X_2X_3 = 1020.28$

con lo cual tenemos:

$$\Sigma y'^2 = \Sigma(y-\bar{y})^2 = \Sigma y^2 - n(\bar{y})^2 = 1046.03 - 10(10)^2 = 46.03$$

$$\Sigma x_2'^2 = 1092.91 - 10(10)^2 = 92.91$$

$$\Sigma x_3'^2 = 1015.18 - 10(10)^2 = 15.18$$

$$\Sigma\, y'\, x_3' = \Sigma\, Y\, X_2 - n(\bar{y})\;(\bar{X}_2) = 1054.11 - 10(10) = 64.11$$
$$\Sigma\, y'\, x_3' = 1011.73 - 10(10)(10) = 11.73$$
$$\Sigma\, x_2'\, x_3' = 1020.28 - 10(10)(10) = 20.28$$

sustituyendo en 2) y 3) obtenemos los coeficientes de regresión y la ecuación de regresión estimada es:

$$\bar{y}\ 1.23 = 4.7452 + 0.73595\ X_2 - 0.21047\ X_3$$

Este resultado indica que la publicidad incrementa las ventas y que los aumentos en los precios relativos las disminuyen. Esto es, el valor 0.73595 indica que si los gastos en publicidad aumentan en 1% las ventas aumentarán en 0.74% mientras que -0.21047 revela que al aumentar el precio relativo en 1% las ventas caerán en 0.21%.

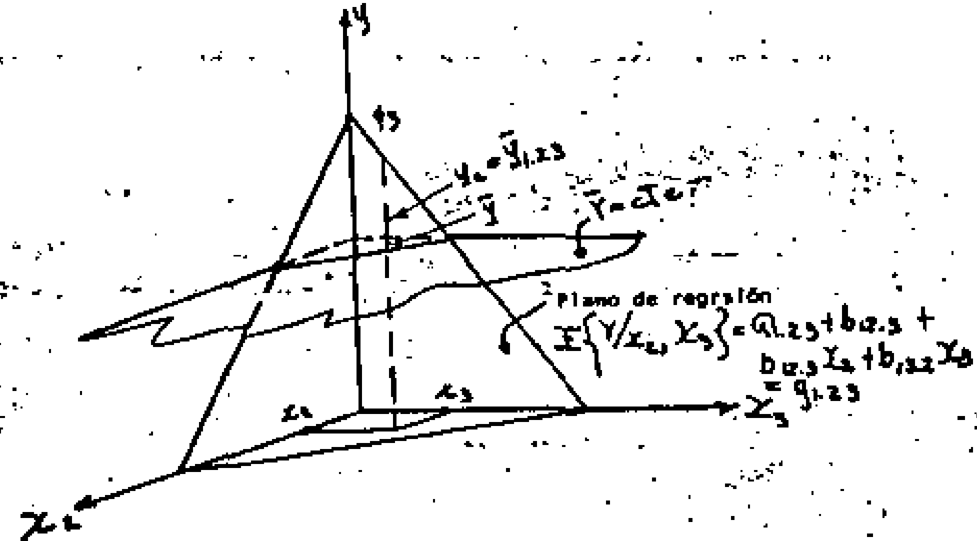
8.3.1 Significado de los coeficientes de regresión parciales. Nuestro principal interés en esta parte del curso se centra en saber:

a) ¿Qué tan significativos son los valores de los coeficientes de regresión parciales? O sea, si encontramos como en nuestro ejemplo que $b\ 12.3 \neq 0$ y $b\ 13.2 \neq 0$ ¿podemos considerar también que los correspondientes coeficientes de la población toman valores distintos de cero?

b) ¿Hay diferencia relativa entre los efectos de las variables independientes y el valor de la variable dependiente? Dicho en otras palabras, estamos interesados en determinar la contribución neta de cada variable independiente a la dependiente.

Estas preguntas se contestan con pruebas estadísticas basadas en el análisis de la varianza; veamos como.

Dada la ecuación de regresión muestral y por tanto, para este caso, el plano de regresión, podemos pensar en las desviaciones totales de los valores Y con relación a la media estimada como las desviaciones verticales con relación al plano de regresión ajustado.



Dividiendo esta variación, como en el caso bivariante en dos partes independientes: una parte mide la variación en Y que ha sido "explicada" por la regresión $(Y_C - \bar{y})$ y la otra mide la variación "no explicada" debida a la aleatoriedad, siendo por tanto la residual $(Y - Y_C)$

Lo anterior en términos algebraicos será:

$$(Y - \bar{Y}) = (Y - Y_C) + (Y_C - \bar{Y})$$

cuya suma cuadrada es:

$$4)\ldots \quad \Sigma(Y-\bar{Y})^2 = \Sigma(Y-Y_C)^2 + \Sigma(Y_C-\bar{Y})^2$$

recordemos que el doble producto $\Sigma(Y-Y_C)(Y_C-\bar{Y}) = 0$

en 4) se tiene

SST = suma de cuadrados totales = $\Sigma(Y-\bar{Y})^2 = \Sigma Y^2 - n(\bar{Y})^2 = \Sigma Y'^2$

SSR = suma de cuadrados de la regresión = $\Sigma(Y_C-\bar{Y})^2 = b12.3\ \Sigma x_2'\, y' + b13.2\ \Sigma x_3'\, y'$

con k grados de libertad (k = número de coeficiente de regresión parcial en la ecuación de regresión muestral. Finalmente

SSE = suma de cuadrados del error = $SST - SSR = \Sigma(Y_c)^2$ con n-k-1 grados de libertad.

Resumimos lo anterior en el cuadro ANOVA USUAL

Tabla IV Tabla ANOVA para la regresión trivariante

F u e n t e	G. de l.	S.S.	MS
regresión	k = 2	SSR	MSR = SSR/k
residual	n-k-1	SSE	MSE = SSE/(n-k-1)
T o t a l	n-1	SST	

donde el error medio cuadrático $MSE = \frac{SSE}{n-k-1} = \frac{(y-y_c)}{n-k-1} = \hat{\sigma}^2_{1.23}$

es la variancia muestral del plano de regresión ajustado y es una estimación insesgada de la variancia de la población $\sigma^2_{1.23}$

si las subpoblaciones de Y están normalmente distribuidas MSE mide la precisión del ajuste

La estadística $F = \frac{MSR}{MSE}$ se distribuye como $F_{k,\ n-k-1}$ y puede emplearse para efectuar una prueba general de hipótesis:

$Ho : B_2 = B_3 = 0 \qquad H_1 : B_2 \neq 0,\ B_3 \neq 0$ (B_i = coefs. poblacionales)

si la hipótesis nula es falsa o sea que sí existe regresión significativa los valores Y_c diferirán significativamente de $\bar{Y}$ y SSR será grande. Como resultado los residuos tenderán a ser pequeños. Esto supone que el valor de F será grande indicando una regresión importante. Cuando los residuos son relativamente grandes o la mejora provocada por el plano de regresión es pequeña entonces F será pequeño aceptando en consecuencia Ho. Con esto contestamos la primera pregunta planteada al inicio de este punto.

Para nuestro ejemplo tenemos:

$SST = \Sigma y'^2 = 46.03$

$SSR = b_{12.3}\ \Sigma X_2' y + b_{13.2}\ \Sigma X_3' y' = (0.7359)(64.11)+(0.21047)(11.73)$

$= 44.71$

y $SSE = SST - SSR = 46.03 - 44.71 = 1.32$

nuestra tabla ANOVA será:

Tabla V ANOVA para el problema de la Cía. FIPA

F u e n t e	SS	G. de l.	MS
regresión X_2,X_3	44.71	2	MSR = 44.71/2 = 22.355
residual	1.32	7	MSE = 1.32/7 = 0.1886
T o t a l	46.02	9	

para la prueba de hipótesis $Ho : B_2 = B_3 = 0 \ ; \ H_1 : B_2 \neq 0 \ ; \ B_3 \neq 0$

$F = \frac{22.355}{0.1886} = 118.50 \ ; \ F_{\alpha = 0.01,2,7} = 9.55$

como observamos hay una alta asociación significativa o regresión entre las ventas, la publicidad y el índice relativo de precios.

Para contestar la segunda pregunta planteada primero calculamos los coeficientes de regresión simple (con (2)):

b_{12} = coeficiente de X_2 en regresión simple de Y sobre X_2

$= \frac{X_2' Y'}{\Sigma X_2'^2} = \frac{64.11}{92.91} = 0.690$

de manera similar

$b_{13} = \frac{\Sigma X_3' Y'}{\Sigma X_3'^2} = \frac{11.73}{15.18} = 0.773$

Las sumas de cuadrados explicadas debidas a X_2 y X_3 solas son:

$SSR\ (X_2) = b_{12}\ \Sigma x_2^{\prime}\cdot y^{\prime} = (0.690)(64.11) = 44.24$

$SSR\ (X_3) = b_{13}\ \Sigma x_3^{\prime}\ y^{\prime} = (0.773)(11.73) = 9.07$ teniendo

Tabla VI . ANOVA para la aportación de X_2

Fuente	SS	G. de l.	MS
regresión X_2	44.24	1	44.24
adición de X_3	0.47	1	0.47
X_2 y X_3	44.71	2	
residuo	1.32	7	0.1886
Total	46.02	9	

para probar la significancia de X_2 sola calculamos la SSE (X_2)

como $SSE\ (X_2) = SCT - SSR\ (X_2) = 46.03 - 44.24 = 1.79$

con G. de l. = 10-1-1 = 8 ; por tanto

$$MSE(X_2) = \frac{1.79}{8} = 0.224$$

El estadístico F será $F(X_2) = \dfrac{MSR\ (X_2)}{MSE\ (X_2)} = \dfrac{44.24}{0.224} = 197.5$

que es altamente significativo, por lo tanto rechazamos la hipótesis

$Ho : B_2 = 0.$

El efecto adicional de X_3 sobre Y puede comprobarse con la estadística

$$F = \frac{MSR(X_3)}{MSE} = \frac{0.47}{0.1886} = 2.492$$ que comparado

con $F_{0.05,1,7} = 3.59$ resulta no significativo

Alternativamente podemos elaborar el cuadro VII con SSR (X_3).

Como debemos esperar de resultados anteriores, el efecto directo de X_3 es estadísticamente insignificante mientras que el de X_2 es altamente significativo. Finalmente los resultados de estas pruebas concuerdan apreciablemente con la interpretación hecha de los mismos coeficientes de regresión parciales.

Tabla VII

Fuente	SS	G. de l.	MS
regresión X_3	9.07	1	9.07
adición de X_2	35.64	1	35.64
X_2 y X_3	44.71	2	
residual	1.32	7	0.1886
Total	46.02	9	

Problema No. 1

SUJETO	GRUPOS (X_{ti}, Y_{ti})			
	1	2	3	4
1	25,25	17,11	32,24	10,8
2	13,25	9,9	30,18	29,17
3	10,12	19,16	12,2	7,8
4	25,30	25,17	30,24	17,12
5	10,37	6,1	10,2	8,7
6	17,25	23,12	8,0	30,26
7	9,31	7,4	5,0	5,8
8	18,26	5,3	11,1	29,29
9	27,28	30,26	5,1	5,29
10	17,29	19,20	25,10	13,0

a) calcular las rectas de regresión para cada grupo:

$Y = a_t + b_t X$ donde
$$\begin{cases} b = \dfrac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} \\ a = \bar{y} - b\bar{x} \end{cases}$$

Tenemos entonces para cada grupo:

	1	2	3	4
$\sum_1 xi =$	171	160	168	153
$\sum_1 yi =$	268	119	82	144
$\sum_1 x_i y_i =$	4,611	2,482	2,338	2,695
$\sum_1 xi^2 =$	3,331	3,256	3,928	3,323
$\bar{x} =$	17.1	16.0	16.8	15.3
$\bar{y} =$	26.8	11.9	8.2	14.4

De donde:

$$b_1 = \frac{10(4611) - (171)(268)}{10(3,331) - (171)^2} = \underline{0.9693} \; ; \quad a_1 = 26.8 - 0.0693(17.1) = \underline{25.6149}$$

$$b_2 = \frac{10(2,482) - (160)(119)}{10(3,256) - (160)^2} = \underline{0.8305} \; ; \quad a_2 = 11.9 - 0.8305(16) = \underline{1.3874}$$

$$b_3 = \frac{10(2,338) - (168)(82)}{10(3,982) - (168)^2} = \underline{0.0687} \; ; \quad a_3 = 8.2 - 0.0687(16.8) = \underline{6.3936}$$

$$b_4 = \frac{10(2,695) - (153)(144)}{10(3,303) - (153)^2} = \underline{0.5112} \; ; \quad a_4 = 14.4 - 0.5112(15.3) = \underline{6.5790}$$

por lo que las rectas de regresión son, para cada uno de los grupos:

1) $\underline{y = 25.61 + (0.07)\, X}$

2) $\underline{y = 1.39 + (0.83)\, X}$

3) $\underline{y = 6.39 + (0.07)\, X}$

4) $\underline{y = 6.58 + (0.51)\, X}$

- para los promedios:

$$\begin{cases} (17.1,\ 26.8) \\ (16.0,\ 11.9) \\ (16.8,\ 8.2) \\ (15.3,\ 14.4) \end{cases}$$

$\Sigma\, Xi = 65.2$
$\Sigma\, Yi = 61.3$
$\Sigma\, XiYi = 1,006.76$
$\Sigma\, X_i^2 = 1,064.74$
$\bar{X} = 16.3$
$\bar{y} = 15.325$

$$b_p = \frac{4(1,006.76) - (65.2)(61.3)}{4(1,064.74) - (165.2)} = \underline{3.82}$$

$$a_p = 15.325 - (3.8232)\ 16.3 = \underline{46.993}$$

y la recta de regresión para los promedios es:

5) $\underline{y = -46.99 + (3.82)\, X}$

para todos los puntos juntos: $\Sigma X_i = 171 + 160 + 168 + 153 = 652$, $\bar{X} = 16.3$

$\Sigma Y_i = 268 + 119 + 82 + 144 = 613$, $\bar{Y} = 15.325$

$\Sigma X_i Y_i = 4,611 + 2,482 + 2,338 + 2,695 = 12,126$

$\Sigma X_i^2 = 3,331 + 3,256 + 3,928 + 3,303 = 13,818$

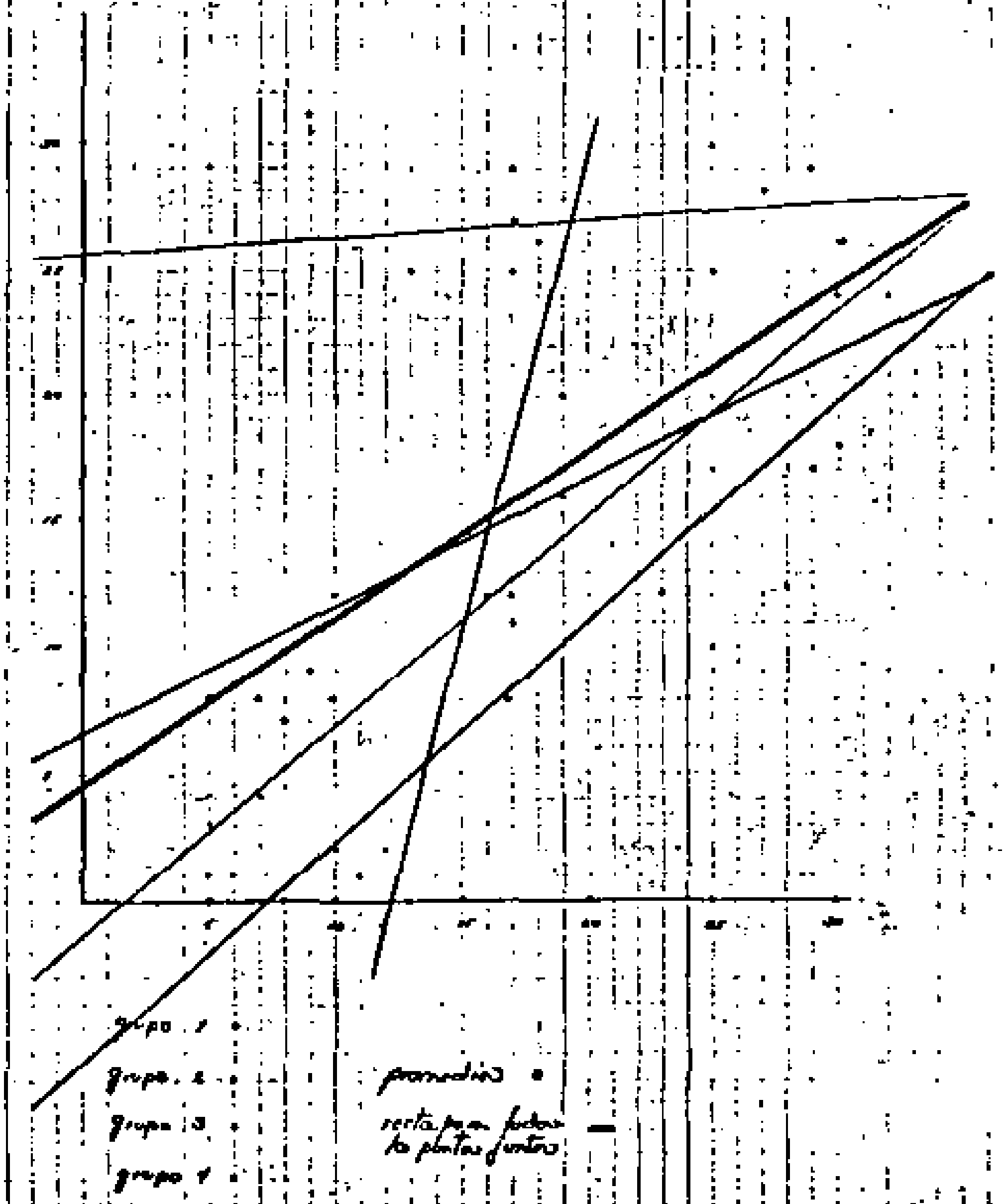
y la recta de regresión;

6) $\underline{y = 4.42 + (0.67)\ X}$

$$b_T = \frac{40(12,126) - (652)(613)}{40(13,818) - (652)^2} = \underline{0.6689}$$

$$a_T = 15.325 - (0.6689)\ 16.3 = \underline{4.4217}$$

a) *Gráfica de los puntos y las rectas calculadas*



b) Estimar los efectos α_t

como

$E(a_t) = \alpha_t$; a_t es un estimador insesgado de α_t y :

$\hat{\alpha}_1 = 25.61$; $\hat{\alpha}_2 = -1.39$ $\hat{\alpha}_3 = -6.39$ $\hat{\alpha}_4 = 6.58$

c) probar la hipótesis de igualdad de pendientes

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$

$$W_t = \sum_{i=1}^{n_t} (x_{ti} - \bar{x}_{t.}) = \sum_{i=1}^{n_t} x_{ti}^2 - n_t \bar{x}_{t.}^2$$

$W_1 = 3{,}331 - 10(17.1)^2 = 406.9$ $B_1 = 0.0693$

$W_2 = 3{,}256 - 10(16.0)^2 = 696$ $B_2 = 0.6305$

$W_3 = 3{,}928 - 10(16.8)^2 = 1{,}105.6$ $B_3 = 0.8687$

$W_4 = 3{,}303 - 10(15.3)^2 = 962.1$ $B_4 = 0.5112$

$$S_W = \sum_{t=1}^{n_t} W_t B_t^2 - W_c B_c^2$$

$$W_c = \sum_{t=1}^{n_t} W_t = \underline{3170.6} \quad , \quad B_c = \frac{1}{W_c} \sum_{t=1}^{n_t} W_t B_t = \frac{1}{3170.6}(2058.4814) = \underline{0.6492}$$

$$\therefore S_W = 1567.7572 - (3170.6)(0.6492)^2 = 231.30$$

Ahora, de la ecuación (12) : $\bar{Y}_{..} = 15.325$

$$S_E = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_{..})^2 - \sum_{t=1}^{k} n_t (\bar{y}_{t.} - \bar{y}_{..})^2 - (W_c B_c^2 + S_W)$$

$$= \left(\sum_{t=1}^{k} \sum_{i=1}^{n_t} y_{ti}^2 - N \bar{Y}_{..}^2 \right) - \left(\sum_{t=1}^{k} n_t \bar{y}_{t.}^2 - N \bar{Y}^2 \right) - (W_c B_c^2 + S_W)$$

$$S_E = 14{,}161 - 11{,}344.5 - 1{,}567.7572 = \underline{1{,}248.74}$$

En consecuencia $F = \dfrac{S_W/k-1}{S_E/N-2k} = \dfrac{231.30/3}{1{,}248.74/32} = 1.98 < 2.90$ $F_{.05,\,3,\,32}$

Por lo que se acepta que las pendientes son iguales con 5% de nivel de significancia: $\beta_1 = \beta_2 = \beta_3 = \beta_4$

d) Probar la hipótesis $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$

De los resultados del inciso anterior, es razonable asumir que las β_t son iguales; por lo que usamos el modelo I:

$$y_{ti} = \alpha_t + \beta(x_{ti} - \bar{x}_{..}) + \varepsilon_{ti}$$

tenemos entonces:

$$S_E + S_W = 1{,}248.74 + 231.30 = \underline{1{,}480.04}$$

$$W_m = 10(1{,}044.74) - (40)(11.3)^2 = 19.8$$

$$B_m = \sum_{t=1}^{4} 10 \frac{(\bar{x}_{t.} - 16.3)(\bar{y}_{t.} - 15.325)}{19.8} = \frac{10(9.18/10275 - 11.2570.925)}{19.8} =$$

$$\frac{-0.925}{19.8} = -0.0467 \; ; \; W_m B_m^2 = 0.0432$$

$$S_G = \sum_{t=1}^{4} n_t \bar{y}_{t.}^2 - N \bar{Y}_{..}^2 - W_m B_m^2 = 11{,}344.5 - 40(15.325)^2 + 0.0432 = 1530.32$$

$$S_{WG} = \frac{W_c W_m}{W_c + W_m}(B_c - B_m)^2 = \frac{(3{,}170.6)(19.8)}{3170.6 + 19.8}(0.6492 + 0.0467) = 13.69$$

$$\therefore S_{WG} + S_G = \underline{1544.01}$$

En consecuencia: $F = \dfrac{1544.01/3}{1480.04/35} = 15.48 > 2.88$ $(F_{.05,\,3,\,35})$

y se rechaza la hipótesis $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ con un nivel de confianza del 5%

Fuente	G. de L.	Suma de cuadrados SS	Esperanzas de MS
Pendiente global	1	$S_1 = w_0 B_0^2$	$\sigma^2 + w_0\beta_0^2$
Pendiente de las medias de los grupos vs promedio de las pendientes dentro de grupos	1	$S_{wm} = \frac{w_w w_m}{w_0}(B_m - B_w)^2$	$\sigma^2 + \frac{w_w w_m}{w_0}(\beta_m - \beta_w)^2$
Acerca de la línea de regresión de las medias de los grupos	$k-2$	$S_m = \sum_{i=1}^{k} n_i\left[\bar{Y}_i - \bar{Y} - B_m(\bar{X}_i - \bar{X})\right]^2$	$\sigma^2 + (k-2)^{-1}\sum_{i=1}^{k} n_i(\alpha_i - \alpha - \beta_m x_i)^2$
Pendientes entre grupos	$k-1$	$S_b = \sum_{i=1}^{k} w_i(B_i - B_w)^2$	$\sigma^2 + (k-1)^{-1}\sum_{i=1}^{k} w_i(\beta_i - \beta_w)^2$
Residual	$N-2k$	$S_r = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[Y_{ij} - \bar{Y}_i - B_i(X_{ij} - \bar{X}_i)\right]^2$	σ^2
Total	$N-1$	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2$	

ANALISIS DE COVARIANCIA

EN UNA DIRECCION

EL ANALISIS DE COVARIANCIA SE UTILIZA PARA PROBAR SI LAS DIFERENCIAS EN LA RESPUESTA MEDIA DE UN GRUPO A OTRO PUEDEN SER EXPLICADAS POR UNA REGRESION LINEAL CON UNA VARIABLE DE CONTROL. EL PLANTEAMIENTO DEL ANALISIS DE COVARIANCIA DEPENDE DEL MODELO QUE SE UTILICE; PARA CLASIFICACION DE GRUPOS EN UNA DIRECCION SE PUEDEN USAR LOS SIGUIENTES MODELOS:

$$\text{I.} \quad Y_{ti} = \alpha_t + \beta(X_{ti} - \bar{X}_{..}) + \varepsilon_{ti} \qquad (1)$$

$$\text{II.} \quad Y_{ti} = \alpha_t + \beta_t(X_{ti} - \bar{X}_{..}) + \varepsilon_{ti} \qquad (2)$$

PARA AMBOS MODELOS SE PRETENDE PROBAR LA HIPOTESIS

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k \qquad (3)$$

CONTRA H_1: NO TODAS LAS α_t SON IGUALES

LAS TABLAS DEL ANALISIS SON:

MODELO	FUENTE	G. de L.	SS
I	GRUPOS (AJUSTADA)	$k - 1$	$SWG + SG$
	RESIDUAL	$N - k - 1$	$SR + SW$
II	GRUPOS (AJUSTADA)	$k - 1$	$S_0 + SWG + SG + SW - w_0\sum_{t=1}^{k} B_t'^2$
	RESIDUAL	$N - 2k$	SR

DONDE SMG, SG, SR, SW, B_0 Y W_0 SE CALCULAN CON LAS FORMULAS DEL CAPITULO DE OBSERVACION DE DOS VARIABLES, Y

$$B'_t = \frac{\sum_{i=1}^{N_t} (X_{ti} - \bar{X}_{..})(\bar{Y}_{ti} - \bar{Y}_{..})}{\sum_{i=1}^{N_t} (X_{ti} - \bar{X}_{..})^2} \qquad (4)$$

LOS VALORES ESTIMADOS DE LAS α_t SON

MODELO I: $\bar{Y}_{t.} - B_c(\bar{X}_{t.} - \bar{X}_{..})$ (5)

MODELO II: $\bar{Y}_{t.} - B_t(\bar{X}_{t.} - \bar{X}_{..})$ (6)

SI UNO ESTA BASTANTE SEGURO DE QUE $\beta_1 = \beta_2 = \ldots \beta_k$, ENTONCES EL MODELO I ES MEJOR, YA QUE DA MAS GRADOS DE LIBERTAD EN EL RESIDUO.

TAREA

EN UN EXPERIMENTO, A 40 SEÑORES SE LES SUJETO A UNA PRUEBA (TRATAMIENTO) PARA DETERMINAR QUE TAN CERCA PODIAN CAMINAR HACIA UN OBJETO PELIGROSO (EN ESTE CASO UNA VIBORA), ANTES DE SENTIRSE ANSIOSOS; PARA ESTO, CADA SUJETO SE SITUO ALEATORIAMENTE EN UNO DE CUATRO GRUPOS, CADA UNO CON DIEZ SUJETOS; CON CADA GRUPO SE EMPLEO DIFERENTE TIPO DE VIBORA. DESPUES DEL TRATAMIENTO A CADA SEÑOR SE LE SUJETO DE NUEVO AL MISMO TRATAMIENTO (POSTRATAMIENTO). LOS RESULTADOS DEL TRATAMIENTO SON LAS X_{ti} Y LOS DEL POSTRATAMIENTO SON LAS Y_{ti}, LOS CUALES SE PRESENTAN EN LA TABLA SIGUIENTE

SUJETO	GRUPOS (X_{ti}, Y_{ti})			
	1	2	3	4
1	25,25	17,11	32,24	10,8
2	13,25	5,9	30,18	29,17
3	10,12	19,16	12,2	7,8
4	25,30	25,17	30,24	17,12
5	10,37	6,1	10,2	8,7
6	17,25	23,12	9,0	30,26
7	9,31	7,4	5,0	5,8
8	18,26	5,3	11,1	29,29
9	27,28	30,26	5,1	5,29
10	17,29	19,20	25,10	13,0

a) CALCULAR LAS RECTAS DE REGRESION PARA CADA GRUPO, PARA LOS PROMEDIOS Y PARA TODOS LOS PUNTOS JUNTOS. EN UNA MISMA GRAFICA DIBUJAR LOS PUNTOS Y LAS RECTAS CALCULADAS.

b) ESTIMAR LOS EFECTOS α_t

c) PROBAR LA HIPOTESIS DE IGUALDAD DE PENDIENTES H_0: $\beta_1 = \beta_2 \ldots = \beta_k$

d) PROBAR LA HIPOTESIS DE IGUALDAD DE MEDIAS DE LAS Y_{ti} DE LOS CUATRO GRUPOS, DESPUES DE AJUSTAR POR LA REGRESION CON X_{ti}, O SEA, PROBAR H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$.

DIVISION DE EDUCACION CONTINUA
FACULTAD DE INGENIERIA U.N.A.M.

DISEÑO ESTADISTICO DE EXPERIMENTOS

ANALISIS DE EXPERIMENTOS 2^k

Ing. Bernardo Frontana

MAYO, 1982

7.- ANALISIS DE EXPERIMENTOS FACTORIALES 2^k

7.1.- Principios involucrados en la experimentación.

(a) El primer paso importante en la planeación de un experimento es estar conciente de que no puede lograrse la perfección a partir de un número limitado de observaciones (muestras) y por ende deberán emplearse diseños y métodos que permitan la reproducibilidad de los resultados que deben determinarse.

(b) Las conclusiones derivadas de un experimento deben tener validez.- Para asegurar la ausencia de errores sistemáticos es necesario asignar aleatoriamente los tratamientos a los especímenes o material experimental. Con esto las estimaciones encontradas de los efectos de los tratamientos en un gran número de repeticiones del experimento tenderán a un promedio resultante de los verdaderos efectos de los tratamientos. Esto es, la aleatorización asegura la obtención de estimadores insesgados de los efectos de los tratamientos. El experimento válido será aquel que está planeado de manera tal que las conclusiones están libres de sesgos o parcialidades, sea conciente o inconcientemente del experimentador. La aleatorización es un seguro para el experimentador.

(c) Las conclusiones derivadas de un experimento deben tener PRECISION.-Si los errores sistemáticos se evitan mediante la aleatorización entonces la estimación de los efectos de los tratamientos deferirá de sus valores verdaderos solamente por la variación aleatoria. Un experimento verdadero es aquel que proporciona una medida de esta variación. Una de tales medidas será mediante la replicación o repetición de algunos o todos los tratamientos, de manera tal que un estimador de un error experimental pueda obtenerse por una comparación de unidades experimentales similares: es decir, unidades similares con respecto a los efectos controlados concientemente. En suma, la replicación permite la reproducibilidad de los resultados a determinarse.

(d) Los resultados de las conclusiones experimentales deben tener ancho rango de aplicación.- La precisión del experimento no solamente depende del tamaño del mismo como se refleja con el número de réplicas sino también en la variabilidad inherente de las unidades experimentales. El error experimental será más pequeño si las unidades (especímenes) son más homogeneas; sin embargo, para lograr una ancha cobertura de los resultados se tendrá que usar unidades heterogeneas en el experimento. Existen algunas técnicas disponibles para lograr un equilibrio; es decir, incrementar la precisión sin un excesivo sacrificio de cobertura.

7.2.- El problema del diseño de experimentos: Elegir un diseño para estimar los efectos de los tratamientos tan precisamente como sea posible.

7.3.- Primeros pasos en la planeación de un experimento.- El primer y más importante paso en la planeación de un experimento es decir, qué experimento se propone uno a realizar. Esto no es tan fácil como parece ya que además de establecer lo que se va a probar también se necesita especificar claramente la población a la cual se aplicarán las conclusiones del experimento. Resulta evidente que la población total posible consiste de todas las variedades de especímenes y rangos de condiciones bajo las cuales serán tratados, también deben considerarse las limitaciones puestas al experimento.

El experimentador deberá elegir a que ancho de la población se referirán su conclusiones.

El segundo paso en su experimento es medir la exactitud probable de los resultados que se obtendrán. Para esto es necesario medir la variabilidad de las observaciones individuales del experimento y determinar el número de réplicas necesarias para una diferencia de magnitud dada y tener límite de confianza predeterminados conforme a la rigurosidad del experimento. En resumen, para determinar cuando un experimento ha de ser bastante largo se requiere :

(a) La estimación del porcentaje de variación en las observaciones que no puede asignarse a ninguno de los factores del experimento. Esta cantidad se llama COEFICIENTE DE VARIACION.

(b) El valor de la exactitud deseada en el efecto del tratamiento expresado como un porcentaje de la media global. Por ejemplo puede desearse medir el efecto de un tratamiento al 5% porque efectos más pequeños ya no tiene importancia práctica.

(c) La probabilidad de que los valores verdaderos de las diferencias caigan dentro de límites asignados. El nivel de probabilidad que se usa depende de las consecuencias posibles que se derivan de las concluciones. Aún si aquellas llevan a acciones costosas e irrevocables, entonces se requiere un nivel de probabilidad tal que haga las pruebas más rigoristas.

7.4.- Métodos para mejorar la exactitud de un experimento.-

(a) Limitar la población a la cual serán aplicable las conclusiones del experimento.

(b) Usando material uniforme se mejora la exactitud del experimento.

(c) Mejorando los métodos de aplicación de los tratamientos y de medición de los efectos.

(d) Utilizando métodos estadísticos:

(d-1) Estratificando los tratamientos de *bloques* (lo más homogéneo posible) generando así una variedad de diseños experimentales. Estos eliminan automáticamente muchas de las variaciones en las observaciones de las comparaciones de los tratamientos.

(d-2) Si pueden tomarse series de observaciones para explicar algo de las variabilidades en las mediciones finales puede efectuarse un análisis de COVARIANCIA para *eliminar* variabilidad. Por ejemplo, los pesos finales de *animales* después de terminar un experimento pueden ajustarse usando sus pesos iniciales antes de comenzar el experimento. De esta manera se elimina la variabilidad debida a las diferencias iniciales de tamaño y posiblemente a la *habilidad* inherente al crecimiento. Debe notarse que las observaciones usadas de esta tarea pueden no reflejar los efectos del tratamiento.

7.5 Elección del Diseño.- Los tres pasos principales para la elección de un diseño experimental son:

(1) Cuando se ha decidido si el diseño es unifactor o factorial

(2) Cuando se ha decidido que agrupando las observaciones se *eliminan* 1, 2 ó más causas de variación; por ejemplo, si se desea eliminar simultáneamente los efectos de tiempo y *días de tomar* las observaciones, un diseño de cuadrados latinos puede ayudar.

(3) Cuando se ha visto que el número de tratamientos o combinaciones de tratamientos es lo bastante grande para tener una réplica total ajustada convenientemente en un bloque, teniendo así un *diseño* por "bloque incompleto".

La tabla 1 *indica* los tipos de diseño que pueden usarse para experimentos unifactoriales o factoriales, en bloques completos e incompletos, eliminando una o dos causas de variación. Estos mismos diseños aparecen en la tabla 2 listando sus propiedades relevantes para su selección o rechazo.

7.6 El propósito de los *experimentos* factoriales.- La principal característica de los experimentos factoriales consiste en que se pueden obtener amplios resultados variando las condiciones básicas o tratamientos dentro del experimento. Por ejemplo, en el estudio del incremento en peso de los animales logrado por diferentes dietas, podemos usar *diseños factoriales* en donde :

TABLA 1.- CLASIFICACION DE LOS PRINCIPALES DISEÑOS

Bloques		Unifactor	Factorial
Completos	Una agrupación dos agrupaciones	Bloque aleatorizados cuadrados latinos	
Incompletos	Una agrupación	-Bloque incompletos balanceados -Diseños cíclicos -Diseños parcialmente balanceados	-Diseños perturbados. -Replicaciones fraccionales
	Dos agrupaciones	-Cuadros de Youden -Cuadros Lattice	-Cuadros Quasi-latinos

TABLA 2.- PROPIEDADES DE LOS PRINCIPALES DISEÑOS

DISEÑOS	PROPIEDADES
1.- Bloques aleatorizados	Fácil de desarrollar, fácil de ajustarse por observaciones perdidas y otras complicaciones experimentales, pueden usarse cualquier número de tratamientos y réplicas.
2.- Cuadrados latinos	Relativamente fácil desarrollar, pequeña dificultad en correcciones por observaciones perdidas, el número de réplicas debe ser un múltiplo del número de tratamientos; es decir con 8 tratamientos tendrán que usarse 8, 16, 24... réplicas, es desventajoso si el número de tratamientos es grande, útil para trabajar con hasta 10 tratamientos.

3.- Bloques incompletos balanceados	Más difícil de desarrollar, difícil de ajustarse por observaciones perdidas y otras complicaciones experimentales, el número de réplica necesita ser bastante grande; sin embargo existen algunos diseños que requieren pocas réplicas.
4.- Diseños cíclicos y parcialmente balanceados.	Se usan cuando no existen bloques balanceados incompletos o cuando ciertas comparaciones entre tratamientos son de especial interés, difícil de desarrollar y de ajustar por observaciones perdidas y otras complicaciones experimentales, permite una considerable flexibilidad en la elección del número de tratamientos y de bloques, muchos de estos diseños pueden usarse cuando hay dos agrupaciones.
5.- Diseños perturbados confounded)	Pueden usarse para cualquier arreglo factorial pero es más útil para diseños 2^k, 3^k o 4^k, se necesita cuidado en la aplicación de las combinaciones de los tratamientos y si se usa una sola réplica, el número es muy difícil para ajustarse por observaciones perdidas, pueden usarse cualquier número de réplica.
6.- Réplica fraccional	La mitad, tercera o cuarta parte de las réplicas son comúnmente usadas, permite al experimentador planear su investigación como una secuencia de pequeños experimentos, difícil de ajustarse por observación perdida.
7.- Diseño Split-Plot	Permite que algunos efectos e interacciones sean estimados con más exactitud a expensas de la exactitud de otros, particularmente útil donde algunos de los factores en el experimento requiere grandes cantidades de material experimental mientras otros factores pueden usarse económicamente en pequeñas cantidades de material.
8.- Cuadrados de Youden	Más útil para menos de 40 tratamientos, más difícil de desarrollar, difícil de ajustar por observaciones perdidas y otras complicaciones experimentales, el número de réplicas debe ser igual al número de tratamientos por bloque, el número de tratamientos debe ser igual al número de bloques.
9.- Cuadrados Celosía (Lattice square)	Útil para tratar con 16-49 tratamientos, más difícil de desarrollar, difícil de ajustar por complicaciones experimentales y observaciones perdidas. El número de tratamientos debe ser P^2 donde el número de réplicas es P + 1, o si P es par posiblemente $\frac{1}{2}(P + 1)$
10.- Cuadrados Quasi-Latinos	Más útil para diseños 2^5, 2^6, 3^3, 3^4, 4^3, más difícil de desarrollar, difícil de ajustar por observación y otras complicaciones experimentales, el número de observaciones debe ser un cuadrado perfecto o múltiplo de un cuadrado perfecto, el número de réplica es usualmente pequeño, se necesita cuidados en la aleatorización de este diseño.

hagamos intervenir animales de ambos sexos y de diferentes razas, alimentándolos con diferentes métodos. Usando cada método de alimentación y cada dieta a ambos sexos y razas, esto es un diseño factorial, se pueden determinar los mejores métodos, dietas y razas. Además, tal vez ⟶ lo más característico de este tipo de diseños, es posible estudiar cuando el mejor método de alimentación varía de dieta a dieta o cuando el método y la dieta dependen del sexo o raza del animal.

Consecuentemente con un diseño factorial podemos estudiar la manera en que pueden variar los efectos con los cambios en otros factores experimentales; es decir, LA INTERACCION de los factores experimentales. El diseño factorial, por el uso de cada combinación de una serie de tratamientos y condiciones experimentales, proporciona los efectos medios, y sus interacciones con algún otro pueden estimarse simultáneamente. Si no hay interacción entre los factores, pueden usarse todas las observaciones para hacer comparaciones entre tratamientos; sin embargo, cuando las hay deberá restringirse la atención a las combinaciones particulares. La existencia de interacciones puede verificarse solamente por el uso de un experimento factorial y la determinación simultánea de interacciones significantes se facilita grandemente. El reconocimiento de qué interacciones son relevantes permite enfocar la atención sobre éstas. Por ejemplo, si encontramos que el mejor método de alimentación depende de la dieta pero no del sexo o raza del animal, podemos considerar métodos diferentes de alimentación para cada dieta separadamente pero promediados sobre todos los sexos y razas.

En resumen, el diseño factorial está interesado con el análisis simultáneo de un número básico de tratamientos o factores, cada uno de los cuales toma un número posible de formas o niveles. Una combinación particular de los niveles de los factores determina un tratamiento.

El término "factor" se usa aquí para indicar cualquier característica que está bajo el control del experimentador y que puede ser variado de prueba a prueba.

7.7. El análisis de experimentos factoriales 2^k

En este punto discutiremos el experimento 2^k que es un experimento de k factores cada uno con dos niveles.

Considérese un experimento con 2 factores A y B, cada uno con 2 niveles. Designemos con mayúsculas a "los efectos" y con minúsculas a las combinaciones de los niveles de los tratamientos posibles. "A" se referirá entonces al efecto del factor A y "a" al nivel "alto" de A que aparece en algunas combinaciones de un tratamiento. Arbitrariamente nos referimos a los dos niveles de cada factor como los niveles "alto" y "bajo" (pudiendo ser alto y bajo sobre alguna escala).

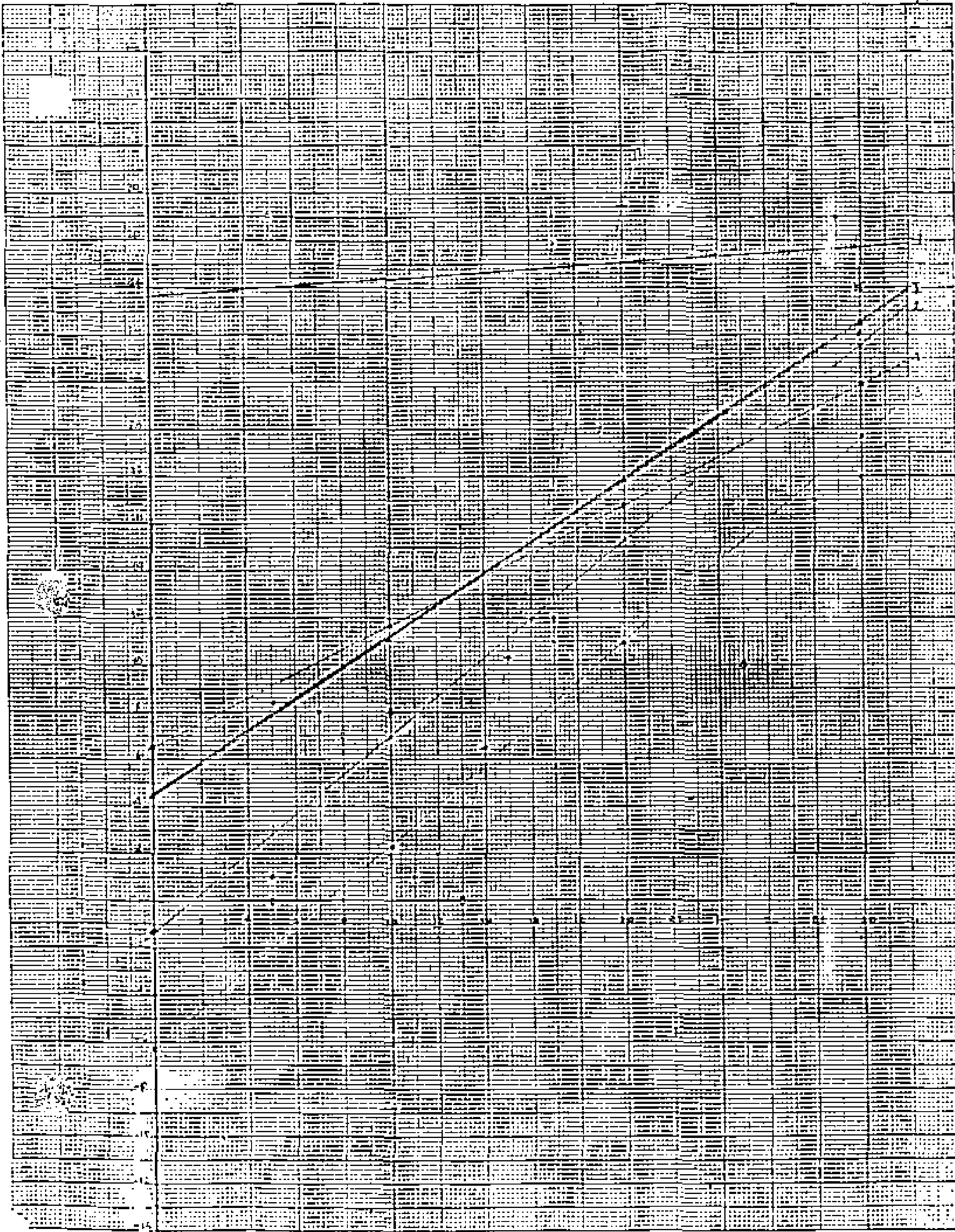
Las cuatro combinaciones para establecer los correspondientes tratamientos para este experimento 2^2 son como se muestra en la Tabla III: (1), a, b, ab. El método de designar estos tratamientos es incluyendo la letra minúscula si el factor está al nivel alto y excluyéndola en caso contrario.

Tabla III Combinaciones nivel-tratamiento en un experimento 2^2

	A_0	A_1
B_0	(1)	a
B_1	b	ab

Como se observa, si todos los factores están al nivel "bajo" se usa el símbolo (1). Por conveniencia A_0 = nivel inferior y A_1 = nivel superior de A (de manera similar para los otros factores). Los subíndices 0 y 1 serán ventajosos en discusiones posteriores. [los símbolos a, b, ab y (1)

7/13



c). Probar la hipótesis de igualdad de pendientes

$H_0: \beta_1 = \beta_2 = \cdots = \beta_k$

Fuente	G. de L.	SS	MS	F
Pendiente global	1	S_0 = 1927.4717	1927.4717	36.566
Pendiente de las medias de los grupos vs promedio de las pendientes dentro de grupos	1	S_{GW} = 158.7303	158.2503	5.0771
Alrededor de la línea de regresión de las medias de los grupos	2	S_G = 1660.8563	830.478	21.2372
// Pendiente entre grupos	3	S_W = 230.9943	76.9981	1.9723
Residual	32	S_E = 1249.222	39.0381	
Total	39	TSS = 4766.775		

// $F_{0.05,3,32} \doteq 2.90$, $F_{0.01,3,32} = 4.46$ $>$ $F_c = 1.0723$ Se acepta la hipótesis nula

$F_{0.05,2,32} = 3.30$, $F_{0.01,2,32} = 5.34$

$F_{0.05,1,32} = 4.15$, $F_{0.01,1,32} = 7.50$

d) Modelo I

Fuente	G. de l.	SS	MS	F_0
Grupo (ajustada)	$k-1$ 3	$SW_{A}+SA = 1650.0866$	610.695	14.6528
Residual	$N-k-1 = 35$	$SW_A\,SE = 1480.2165$	42.292	

$SE = 1240.222$

$SWA = 158.2303$

$SA = 1660.8563$

$SW = 230.5545$

$F_{0.05,3,35} \doteq 2.875 < F_0 = 14.6528$

Se rechaza la prueba de hipótesis nula

Directorio de Asistentes al curso: Diseño Estadístico de Experimentos 1982

1.	Eloy Altamirano Conde I M S S Luna 235 Guerrero Cuauhtémoc 06300 México, D.F. 529 95 09 Ext. 261	Portal No. 26 Pastores Naucalpan de Juárez, Edo. de México 373 49 68
2.	Eduardo Beltrán Juárez	
3.	Roberto Bocanegra Guzmán POLICYD, S.A. Av. la Presa s/n San Juan Ixhuatepec, Edo. de Méx. 596 08 88	Edif. Antonio Rosales C 32 Tlatelolco Cuauhtémoc México, D.F. 583 41 63
4.	Leonardo M. Díaz Angeles POLICYD, S.A. Av. la Presa s/n San Juan Ixhuatepec Tlanepantla, Edo. de Méx. 586 08 88 (24)	Lino Merino 847 Juan Escutia Iztapalapa 09100 México, D.F. 797 76 97
5.	Ma. del Carmen Díez Gallard Instituto Nacional de Investigaciones Pecuarias Km. 15.5 Carr. Méx. Toluca Palo Alto Cuajimalpa, Edo. de Méx. 570 31 00 Ext.140	Vía Láctea 64 Prado Churubusco Iztapalapa 04230 México, D.F. 582 57 09
6.	Roberto Enciso Rodríguez Unidad de Investigación Biomédica en Medicina Tradicional y Herbolaria Luz Saviñón 214 Del Valle B.Juárez 03100 México, D.F. 543 26 97	Guayaquil 73 Américas Naucalpan , Edo. de Méx. 560 36 64
7.	José Luis Galaviz Medina Secretaría de Agricultura y Recursos Hidráulicos Reforma 107-10° Piso Tabacalera México, D.F. 546 48 22	Mezquite 6-002 U. INFONAVIT Iztacalco México,D.F. 535 15 77
8.	Antonio Huelaz Lesbros U A M Calz del Hueso 1100 Col. Villa Quietud Coyoacán México, D.F. 594 70 02	Tecahuitaco 116-402 Ma. del Carmen B.Juárez México, D.F. 554 23 78
9.	Miguel Angel López López Cía. Minera Autlán Dom. Conocido Ayotla, Hgo. 43130 México	Dos Arbolitos 171 Col. B.Juárez Cd. Netzahualcoyotl, Edo. de México 57000 México
10.	Francisco López Paredes Colgate Palmolive, S.A. de C.V. Presa de la Angostura 225 Irrigación M. Hidalgo México, D.F. 557 00 22	Rinc. Monedas Edif. Soles 201 Pedregal de Carrasco Coyoacán 04700 México, D.F.
11.	Ricardo Moreno Chan Facultad de Medicina Veterinaria UNAM México, D.F. 550 52 15 Ext. 4965	Av. La Hacienda No. 92 Residencial Villa Coapa Tlalpán México, D.F. 594 37 36
12.	Carlos G. Pacheco Alarcón S E P Coordinador de Inv. Educativa Cerrada de Netzahualcoyotl No. 1 México, D.F. 522 31 64	Manta 653 Lindavista México 14, D.F. 586 31 60
13.	Mario Pilatasig Montaluisa División de Ciencias Básicas Ingeniería UNAM México, D.F.	Calle Madroño Lote 9 Agrarista Iztapalapa México, D.F.
14.	Héctor Quiroz Vilchis	
15.	Carlos Rangel Nafaile Instituto de Investigaciones en Materiales Circ. Ext. UNAM México, D.F. 550 52 15 Ext. 4744	Blvd. A. L. Mateos 1762-304 B.Juárez 01020 México,D.F. 651 39 63
16.	Jaime Rangel Mondragón Centro de Investigación y Estudios Avanzados del I PN Av. IPN NO. 2508 Zacatenco México, D.F. 754 02 00	Tacuba 86-4 Centro Cuauhtémoc 06000 México, D.F. 518 66 62
17.	Jesús Torres Calderón Colgate Palmolive, S.A. Presa la Angostura 225 Irrigación M. Hidalgo México, D.F. 557 00 22	Lag. S..Cristobal 126-3 Anáhuac M.Hidalgo 11320 México, D.F. 250 04 03

18. Pedro Urrua Rodríguez

19. Román Vázquez Barber
COVITUR,
D.D.F.
Av. Universidad 800-3° Piso
Sta. Cruz Atoyac
B.Juárez
México, D.F.
575 97 47

Vía Lactea No. 12 Depto. 6
Prado Churubusco
Coyoacán
México, D.F.