PROFESORES DEL CURSO PROBABILIDAD Y ESTADISTICA

FUNDAMENTOS Y APLICACIONES 1980.

1.- DR. OCTAVIO A. RASCON CHAVEZ
Investigador del Instituto de Ingeniería
Profesor de la División de Estudios de
Posgrado de la Facultad de Ingeniería
UNAM
México 20, D.F.
Tel. 550.52.15 Ext.4475

Insurgentes Sur 1023 Edificio Detroit 6° Piso
México, D.F.
Tel. 598.04.89

2.- M. EN I. AUGUSTO VILLARREAL ARANDA
Profesor de la División de Estudios de
Posgrado
Facultad de Ingeniería
UNAM
México 20, D.F.
Tel. 550.52.15 Ext. 4475

CURSO: PROBABILIDAD Y ESTADISTICA. FUNDAMENTOS Y APLICACIONES.
FECHA: Del 17 de marzo al 30 de abril de 1980
HORARIO: lunes, miércoles y viernes de 18 a 21 h.
COORDINADOR: Dr. Octavio A. Rascón Chávez

| T E M A | PROFESOR | FECHA | HORARIO |
|---|---|---|---|
| Introducción<br>Muestreo Aleatorio simple<br>Procesamiento de la información | Octavio A. Rascón Chávez | marzo 17 | 18.00 a 21.00 h. |
| Medidas de tendencia central. Medidas de dispersión. Distribución conjunta de frecuencias | " | marzo 19 | 18.00 a 21.00 h. |
| Regresión Lineal | " | marzo 24 | " |
| Correlación Lineal | " | marzo 26 | " |
| Análisis de series en el tiempo.<br>Predicción | " | marzo 28 | " |
| Teoría de conjuntos<br>Eventos y espacio de eventos | " | marzo 31 | " |
| Probabilidad de eventos | " | abril 2 | " |
| Variables aleatorias. Momentos y esperanzas.<br>Distribución de Probabilidades | " | abril 7 | " |
| Distribuciones de Bernoulli, hipergeométrica, binominal y de Poisson. Proceso de Poisson. | " | abril 9 y 11 | " |
| Distribuciones uniforme, exponencial y normal | " | abril 14 | " |
| Estimación puntual. Distribuciones muestrales | M. en I. Augusto Villarreal | abril 16 y 18 | " |
| Estimación por Intervalos | " | abril 21 y 23 | " |
| Prueba de hipótesis de medios, variaciones y proporciones | " | abril 25 y 28 | " |

| T E M A | PROFESOR | FECHA | HORARIO |
|---|---|---|---|
| Prueba de bondad de ajuste en regresión lineal y en distribuciones de probabilidades | Octavio A. Rascon Chávez. | abril 30 | 18.00 a 21.00 h. |

# EVALUACION DEL PERSONAL DOCENTE

CURSO: PROBABILIDAD Y ESTADISTICA. FUNDAMENTOS Y APLICACIONES

FECHA: del 17 de marzo al 30 de abril, 1980

| CONFERENCISTA | DOMINIO DEL TEMA | EFICIENCIA EN EL USO DE AYUDAS AUDIOVISUALES | MANTENIMIENTO DEL INTERES. (COMUNICACION CON LOS ASISTENTES, AMENIDAD, FACILIDAD DE EXPRESION). | PUNTUALIDAD | |
|---|---|---|---|---|---|
| 1. DR. OCTAVIO A. RASCON CHAVEZ | | | | | |
| 2. M. en I. AUGUSTO VILLAREAL ARANDA | | | | | |
| 3. | | | | | |
| 4. | | | | | |
| 5. | | | | | |
| 6. | | | | | |
| 7. | | | | | |
| 8. | | | | | |
| 9. | | | | | |
| ESCALA DE EVALUACION: 1 a 10 | | | | | |

*JDV

# EVALUACION DE LA ENSEÑANZA



SU EVALUACION SINCERA NOS AYUDARA A MEJORAR LOS PROGRAMAS POSTERIORES QUE DISEÑAREMOS PARA USTED.

| TEMA | ORGANIZACION Y DESARROLLO DEL TEMA | GRADO DE PROFUNDIDAD LOGRADO EN EL TEMA | GRADO DE ACTUALIZACION LOGRADO EN EL TEMA | UTILIDAD PRACTICA DEL TEMA | |
|---|---|---|---|---|---|
| INTRODUCCION | | | | | |
| ESTADISTICA DESCRIPTIVA | | | | | |
| PROBABILIDAD | | | | | |
| INFERENCIA ESTADISTICA | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| ESCALA DE EVALUACION: 1 a 10 | | | | | |

Lidy

3

# EVALUACION DEL CURSO

| | CONCEPTO | EVALUACION |
|---|---|---|
| 1. | APLICACION INMEDIATA DE LOS CONCEPTOS EXPUESTOS | |
| 2. | CLARIDAD CON QUE SE EXPUSIERON LOS TEMAS | |
| 3. | GRADO DE ACTUALIZACION LOGRADO CON EL CURSO | |
| 4. | CUMPLIMIENTO DE LOS OBJETIVOS DEL CURSO | |
| 5. | CONTINUIDAD EN LOS TEMAS DEL CURSO | |
| 6. | CALIDAD DE LAS NOTAS DEL CURSO | |
| 7. | GRADO DE MOTIVACION LOGRADO EN EL CURSO | |

ESCALA DE EVALUACION DE 1 A 10

1. ¿Qué le pareció el ambiente en la División de Educación Continua?

| MUY AGRADABLE | AGRADABLE | DESAGRADABLE |
|---|---|---|
| | | |

2. Medio de comunicación por el que se enteró del curso:

| PERIODICO EXCELSIOR ANUNCIO TITULADO DIVISION DE EDUCACION CONTINUA | PERIODICO NOVEDADES ANUNCIO TITULADO DIVISION DE EDUCACION CONTINUA | FOLLETO DEL CURSO |
|---|---|---|
| | | |

| CARTEL MENSUAL | RADIO UNIVERSIDAD | COMUNICACION CARTA, TELEFONO, VERBAL, ETC. |
|---|---|---|
| | | |

| REVISTAS TECNICAS | FOLLETO ANUAL | CARTELERA UNAM "LOS UNIVERSITARIOS HOY" | GACETA UNAM |
|---|---|---|---|
| | | | |

3. Medio de transporte utilizado para venir al Palacio de Minería:

| AUTOMOVIL PARTICULAR | METRO | OTRO MEDIO |
|---|---|---|
| | | |

4. ¿Qué cambios haría usted en el programa para tratar de perfeccionar el curso?

______________________________

______________________________

______________________________

5. ¿Recomendaría el curso a otras personas?

| SI | NO |
|---|---|
| | |

6. ¿Qué cursos le gustaría que ofreciera la División de Educación Continua?

7. La coordinación académica fue:

| EXCELENTE | BUENA | REGULAR | MALA |
|---|---|---|---|
| | | | |

8. Si está interesado en tomar algún curso intensivo ¿Cuál es el horario más conveniente para usted?

| LUNES A VIERNES DE 9 A 13 H. Y DE 14 A 18 H. (CON COMIDAS) | LUNES A VIERNES DE 17 A 21 H. | LUNES, MIERCOLES Y VIERNES DE 18 A 21 H. | MARTES Y JUEVES DE 18 A 21 H. |
|---|---|---|---|
| | | | |

| VIERNES DE 17 A 21 H. SABADOS DE 9 A 14 H. | VIERNES DE 17 A 21 H. SABADOS DE 9 A 13 Y DE 14 a 18 H. | O T R O |
|---|---|---|
| | | |

9. ¿Qué servicios adicionales desearía que tuviese la División de Educación Continua, para los asistentes?

10. Otras sugerencias:

centro de educación continua
división de estudios de posgrado
facultad de ingeniería unam

# PROBABILIDAD Y ESTADISTICA FUNDAMENTOS Y APLICACIONES

## PROBLEMAS RESUELTOS

DR. OCTAVIO A. RASCON CHAVEZ

MARZO, 1 9 8 0

PROBLEMA 1. Las temperaturas registradas en un lugar a diversas horas son las indicadas en la siguiente tabla:

Tabla 1

| | | | |
|---|---|---|---|
| 39.0 | 32.0 | 11.9 | 38.8 |
| 33.8 | 22.9 | 24.7 | 30.7 |
| 16.2 | 10.3 | 31.8 | 35.7 |
| 28.1 | 3.5 | 33.8 | 32.1 |
| 5.3 | 16.2 | 7.4 | 33.3 |
| 18.2 | 36.0 | 26.6 | 10.4 |
| 10.4 | 16.8 | 16.1 | 3.3 |
| 31.9 | 6.0 | 7.7 | 20.8 |
| 19.5 | 4.5 | 25.4 | 3.1 |
| 21.2 | 28.2 | 39.0 | 23.5 |
| 35.1 | 3.3 | 24.2 | 5.6 |

a) Defina los intervalos de agrupamiento de los datos anteriores.

b) Con los eventos definidos en el inciso a) elabore: una tabla de datos agrupados;

c) Una tabla de distribución de frecuencias.

d) Dibuje el histograma, el polígono de frecuencias y la curva de frecuencias relativas acumuladas.

PROBLEMA 2: Se tienen los siguientes tiempos en segundos para diversas operaciones de una pala mecánica.

| CARGA | MOV. 1. | DESCARGA | MOV. 2. | TOTAL |
|---|---|---|---|---|
| 25 | 10 | 2 | 8 | 45 |
| 19 | 7 | 2 | 8 | 36 |
| 19 | 10 | 2 | 9 | 40 |
| 20 | 8 | 2 | 11 | 41 |
| 25 | 10 | 2 | 10 | 47 |
| 21 | 7 | 2 | 8 | 38 |
| 21 | 8 | 2 | 10 | 41 |
| 26 | 10 | 2 | 13 | 51 |
| 22 | 11 | 2 | 8 | 43 |
| 18 | 7 | 2 | 8 | 35 |

Se quiere reducir la variabilidad en tiempo total.

¿Cuál operación debe estudiarse como fuente principal de variabilidad?

Solución:

Obtener $\overline{X}$, $S^2$, $S$, $V$ para operaciones de carga, movimiento 1 y movimiento 2

Operación de carga

| X | X-$\overline{X}$ | (X-$\overline{X}$)$^2$ |
|---|---|---|
| 25 | 3.4 | 11.56 |
| 19 | -2.6 | 6.76 |
| 19 | -2.6 | 6.76 |
| 20 | -1.6 | 2.56 |
| 25 | 2.4 | 11.56 |
| 21 | -0.6 | 0.36 |
| 21 | -0.6 | 0.36 |
| 26 | 4.4 | 19.36 |
| 22 | 0.4 | 0.16 |
| 18 | -3.6 | 12.96 |
| 216 | 0.0 | 72.4 |

$$\overline{X} = \frac{216}{10} = 21.6 \text{ Seg.}$$

$$S^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N} = \frac{72.4}{10}$$

$$S^2 = 7.24 \text{ Seg}^2.$$

$$S = \sqrt{7.24 \text{ Seg}^2.} = 2.69 \text{ Seg.}$$

$$V = \frac{S}{\overline{X}} = \frac{2.69 \text{ seg.}}{21.6 \text{ seg.}} = 0.12$$

MOVIMIENTO 1

| $X$ | $X^2$ | $X - \overline{X}$ | $(X - \overline{X})^2$ |
|---|---|---|---|
| 10 | 100 | 1.2 | 1.44 |
| 7 | 49 | -1.8 | 3.24 |
| 10 | 100 | 1.2 | 1.44 |
| 8 | 64 | -0.8 | 0.64 |
| 10 | 100 | 1.2 | 1.44 |
| 7 | 49 | -1.8 | 3.24 |
| 8 | 64 | -0.8 | 0.64 |
| 10 | 100 | 1.2 | 1.44 |
| 11 | 121 | 2.2 | 4.84 |
| 7 | 49 | -1.8 | 3.24 |
| 88 | 796 | 0.0 | 21.6 |

$$\overline{X} = \frac{88}{10} = 8.8 \text{ seg.}$$

$$S^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N} = \frac{21.6}{10} = 2.16 \text{ seg}^2$$

$$S = \sqrt{2.16 \text{ seg}^2}. = 1.47 \text{ seg.}$$

$$V = \frac{1.47 \text{ seg.}}{8.8 \text{ seg.}} = 0.17$$

O bien usando

$$S^2 = \overline{X^2} - \overline{X}^2$$

$$\overline{X}^2 = \frac{796}{10} = 79.6 \text{ Seg}^2.$$

$$S^2 = 79.6 - (8.8)^2$$

$$S^2 = 79.6 - 77.44 = 2.16 \text{ seg}^2.$$

MOVIMIENTO 2

| $X$ | $X^2$ | $X - \overline{X}$ | $(X - \overline{X})^2$ |
|---|---|---|---|
| 8 | 64 | -1.3 | 1.69 |
| 8 | 64 | -1.3 | 1.69 |
| 9 | 81 | -0.3 | 0.09 |
| 11 | 121 | 1.7 | 2.89 |
| 10 | 100 | 0.7 | 0.49 |
| 8 | 64 | -1.3 | 1.69 |
| 10 | 100 | 0.7 | 0.49 |
| 13 | 169 | 3.7 | 13.69 |
| 8 | 64 | -1.3 | 1.69 |
| 8 | 64 | -1.3 | 1.69 |
| 93 | 891 | 0.0 | 26.10 |

$$\overline{X} = \frac{93}{10} = 9.3 \text{ seg.}$$

$$S^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N} = \frac{26.10}{10} = 2.61 \text{ seg}^2$$

$$S = \sqrt{2.61 \text{ Seg}^2.} = 1.62 \text{ Seg.}$$

$$V = \frac{S}{\overline{X}} = \frac{1.62}{9.3} \frac{\text{Seg.}}{\text{Seg.}} = 0.17$$

O bien usando

$$S^2 = \overline{X^2} - \overline{X}^2$$

$$\overline{X^2} = \frac{891}{10} = 89.1$$

$$S^2 = 89.1 - (9.3)^2$$

$$S^2 = 89.1 - 86.49$$

$$S^2 = 2.61 \text{ Seg}^2.$$

Por lo tanto, las principales fuentes de variabilidad son los movimientos 1 y 2, ya que tienen el mayor coeficiente de variación, de 0.17.

PROBLEMA 3.- En proceso de control de calidad en una fábrica de explosivos, se sacó la siguiente muestra aleatoria del peso de 40 cartuchos de dinamita.

| Peso en gramos | | | |
|---|---|---|---|
| 38.4 | 38.3 | 36.1 | 39.8 |
| 37.1 | 37.7 | 37.6 | 40.8 |
| 38.6 | 37.4 | 38.3 | 38.1 |
| 38.5 | 37.1 | 39.2 | 37.8 |
| 37.4 | 36.5 | 38.7 | 36.7 |
| 37.3 | 36.3 | 38.2 | 38.3 |
| 39.0 | 38.0 | 36.2 | 39.0 |
| 37.7 | 39.2 | 38.8 | 38.3 |
| 39.5 | 37.0 | 39.5 | 36.9 |
| 37.4 | 38.2 | 39.2 | 38.8 |

Usando los datos básicos determinar media, rango, variancia, desviación estándar y mediana.

Agrupar los datos y trazar el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas. Así mismo calcular la media, la variancia, la desviación estándar y la mediana.

a) MEDIA

$$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{N} = \frac{1522.90}{40} = 38.07$$

b) RANGO

$$R = 40.8 - 36.1 = 4.7$$

c) VARIANCIA

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{N} = \frac{44.05}{40} = 1.10$$

d) DESVIACION ESTANDAR

$$S = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (X_i - \overline{X})^2} = \sqrt{1.10} = 1.049$$

| X | X - $\overline{X}$ | (X - $\overline{X}$)$^2$ |
|---|---|---|
| 38.4 | 0.33 | 0.11 |
| 37.1 | -0.97 | 0.94 |
| 38.6 | 0.53 | 0.28 |
| 38.5 | 0.43 | 0.18 |
| 37.4 | -0.67 | 0.45 |
| 37.3 | -0.77 | 0.59 |
| 39.0 | 0.93 | 0.86 |
| 37.7 | -0.37 | 0.14 |
| 39.5 | 1.43 | 2.04 |
| 37.4 | -0.67 | 0.45 |
| 38.3 | 0.23 | 0.05 |
| 37.7 | -0.37 | 0.14 |
| 37.4 | -0.67 | 0.45 |
| 37.1 | -0.97 | 0.94 |
| 36.5 | -1.57 | 2.46 |
| 36.3 | -1.77 | 3.13 |
| 38.0 | -0.07 | 0.00 |
| 39.2 | 1.13 | 1.28 |
| 37.0 | -1.07 | 1.14 |
| 38.2 | 0.13 | 0.02 |

| X | X - $\overline{X}$ | (X - $\overline{X}$)$^2$ |
|---|---|---|
| 36.1 | -1.97 | 3.88 |
| 37.6 | -0.47 | 0.22 |
| 38.3 | 0.23 | 0.05 |
| 39.2 | 1.13 | 1.28 |
| 38.7 | 0.63 | 0.40 |
| 38.2 | 0.13 | 0.02 |
| 36.2 | -1.87 | 3.50 |
| 38.3 | 0.73 | 0.53 |
| 39.5 | 1.43 | 2.04 |
| 39.2 | 1.13 | 1.28 |
| 39.8 | 1.73 | 2.99 |
| 40.8 | 2.73 | 7.45 |
| 38.1 | 0.03 | 0.00 |
| 37.8 | -0.27 | 0.07 |
| 36.7 | -1.37 | 1.88 |
| 38.3 | 0.23 | 0.05 |
| 39.0 | 0.93 | 0.86 |
| 38.3 | 0.23 | 0.05 |
| 36.9 | -1.17 | 1.37 |
| 38.8 | 0.73 | 0.53 |
| | 1522.9 | 44.05 |

e) MEDIANA

| | | | |
|---|---|---|---|
| 36.1 | 37.4 | 38.2 | 38.8 |
| 36.2 | 37.4 | 38.3 | 39.0 |
| 36.3 | 37.4 | 38.3 | 39.0 |
| 36.5 | 37.6 | 38.3 | 39.2 |
| 36.7 | 37.7 | 38.3 | 39.2 |
| 36.9 | 37.7 | 38.4 | 39.2 |
| 37.0 | 37.8 | 38.5 | 39.5 |
| 37.1 | 38.0 | 38.6 | 39.5 |
| 37.1 | 38.1 | 38.7 | 39.8 |
| 37.3 | 38.2 | 39.8 | 40.8 |

M = 38.2

f) HISTOGRAMA

No. de Intervalos de clase $= \frac{40}{5} = 8$

Ancho: $\frac{4.7}{8} \doteq 0.6$

| CLASE LIMITE DE | FRECUENCIA | LIMITES REALES INF. | LIMITES REALES SUP. |
|---|---|---|---|
| 36.1 - 36.6 | 4 | 36.05 | 36.65 |
| 36.7 - 37.2 | 5 | 36.65 | 37.25 |
| 37.3 - 37.8 | 8 | 37.25 | 37.85 |
| 37.9 - 38.4 | 9 | 37.85 | 38.45 |
| 38.5 - 39.0 | 7 | 38.45 | 39.05 |
| 39.1 - 39.6 | 5 | 39.05 | 39.65 |
| 39.7 - 40.2 | 1 | 39.65 | 40.25 |
| 40.3 - 40.8 | 1 | 40.25 | 40.85 |

g) POLIGONO DE FRECUENCIAS

| MARCA DE CLASE | FRECUENCIA | FRECUENCIA RELATIVA |
|---|---|---|
| 36.35 | 4 | 0.100 |
| 36.95 | 5 | 0.125 |
| 37.55 | 8 | 0.200 |
| 38.15 | 9 | 0.225 |
| 38.75 | 7 | 0.175 |
| 39.35 | 5 | 0.125 |
| 39.95 | 1 | 0.025 |
| 40.55 | 1 | 0.025 |
| | | $\Sigma$ = 1.000 |

h) FRECUENCIAS ACUMULADAS

| LIMITE REAL SUPERIOR | FRECUENCIA ACUMULADA | FRECUENCIA RELATIVA ACUMULADA |
|---|---|---|
| 36.65 | 4 | 0.1 |
| 37.25 | 9 | 0.225 |
| 37.85 | 17 | 0.425 |
| 38.45 | 26 | 0.650 |
| 39.05 | 33 | 0.825 |
| 39.65 | 38 | 0.950 |
| 40.25 | 39 | 0.975 |
| 40.85 | 40 | 1.00 |

i) MEDIA

$$\bar{X} = \frac{\sum_{i=1}^{K} x_i}{N} = \frac{1522.20}{40} = 39.06$$

| MARCA DE CLASE X | FRECUENCIA f | Xf |
|---|---|---|
| 36.35 | 4 | 145.40 |
| 36.95 | 5 | 184.75 |
| 37.55 | 8 | 300.40 |
| 38.15 | 9 | 343.35 |
| 38.75 | 7 | 271.25 |
| 39.35 | 5 | 196.75 |
| 39.95 | 1 | 39.95 |
| 40.55 | 1 | 40.55 |
| | | 1522.20 |

j) VARIANCIA

$$S^2 = \frac{\sum_{i=1}^{K} f_i \, (X_i - \overline{X})^2}{N} = \frac{41.42}{40} = 1.036$$

| MARCA DE CLASE X | FRECUENCIA f | $X - \overline{X}$ | $(X - X^2)$ | $f(X - X)^2$ |
|---|---|---|---|---|
| 36.35 | 4 | -1.71 | 2.92 | 11.68 |
| 36.95 | 5 | -1.11 | 1.23 | 6.15 |
| 37.55 | 8 | -0.51 | 0.26 | 2.08 |
| 38.15 | 9 | 0.09 | 0.01 | 0.09 |
| 38.75 | 7 | 0.69 | 0.48 | 3.36 |
| 39.35 | 5 | 1.29 | 1.66 | 8.30 |
| 39.95 | 1 | 1.89 | 3.57 | 3.57 |
| 40.55 | 1 | 2.49 | 6.20 | 2.49 |
| | | | | 41.42 |

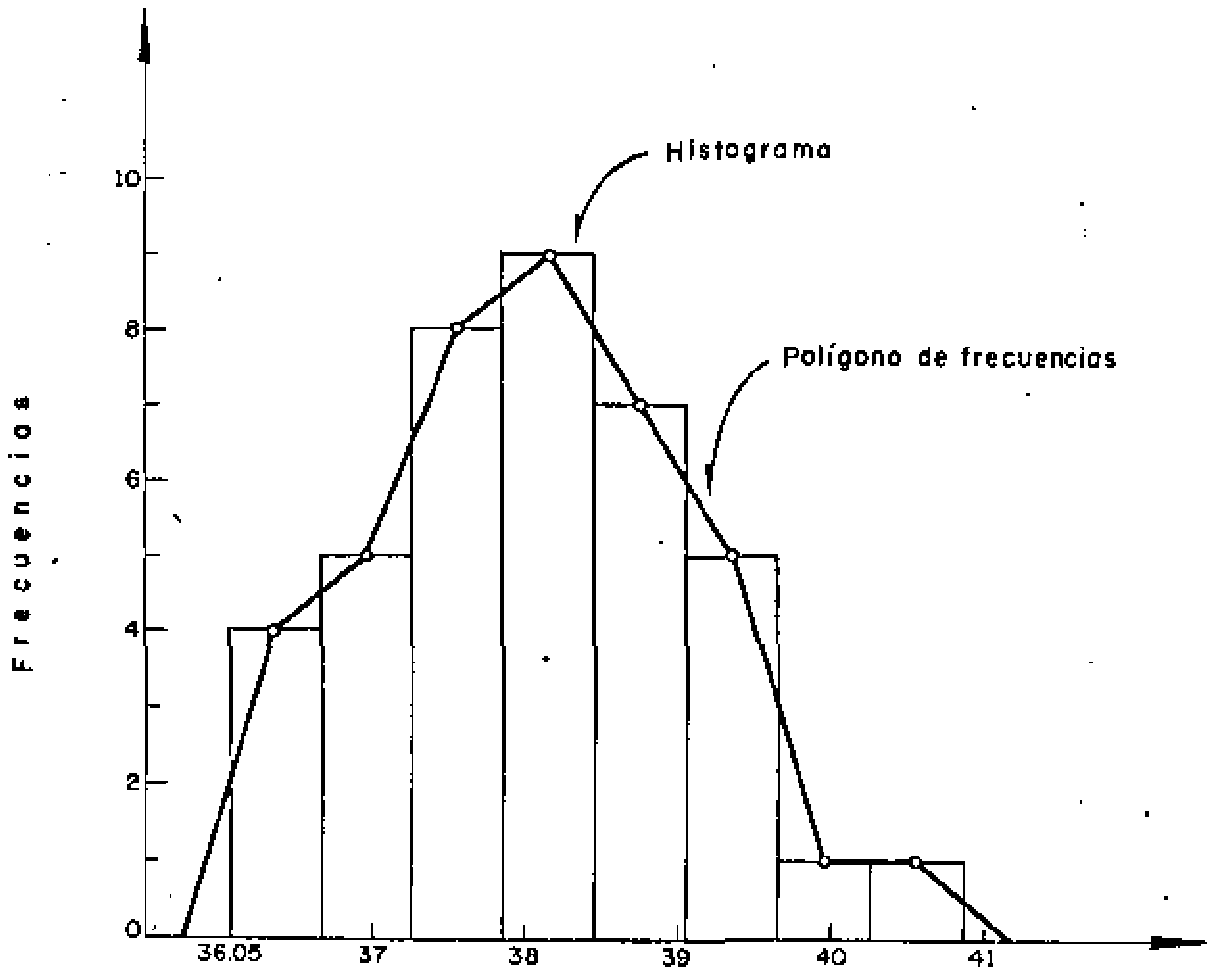
k) DESVIACION ESTANDAR

$$S = \sqrt{\frac{\sum_{i=1}^{K} f_i \; (X_i - \overline{X})^2}{N}} = \sqrt{1.035} = 1.017$$

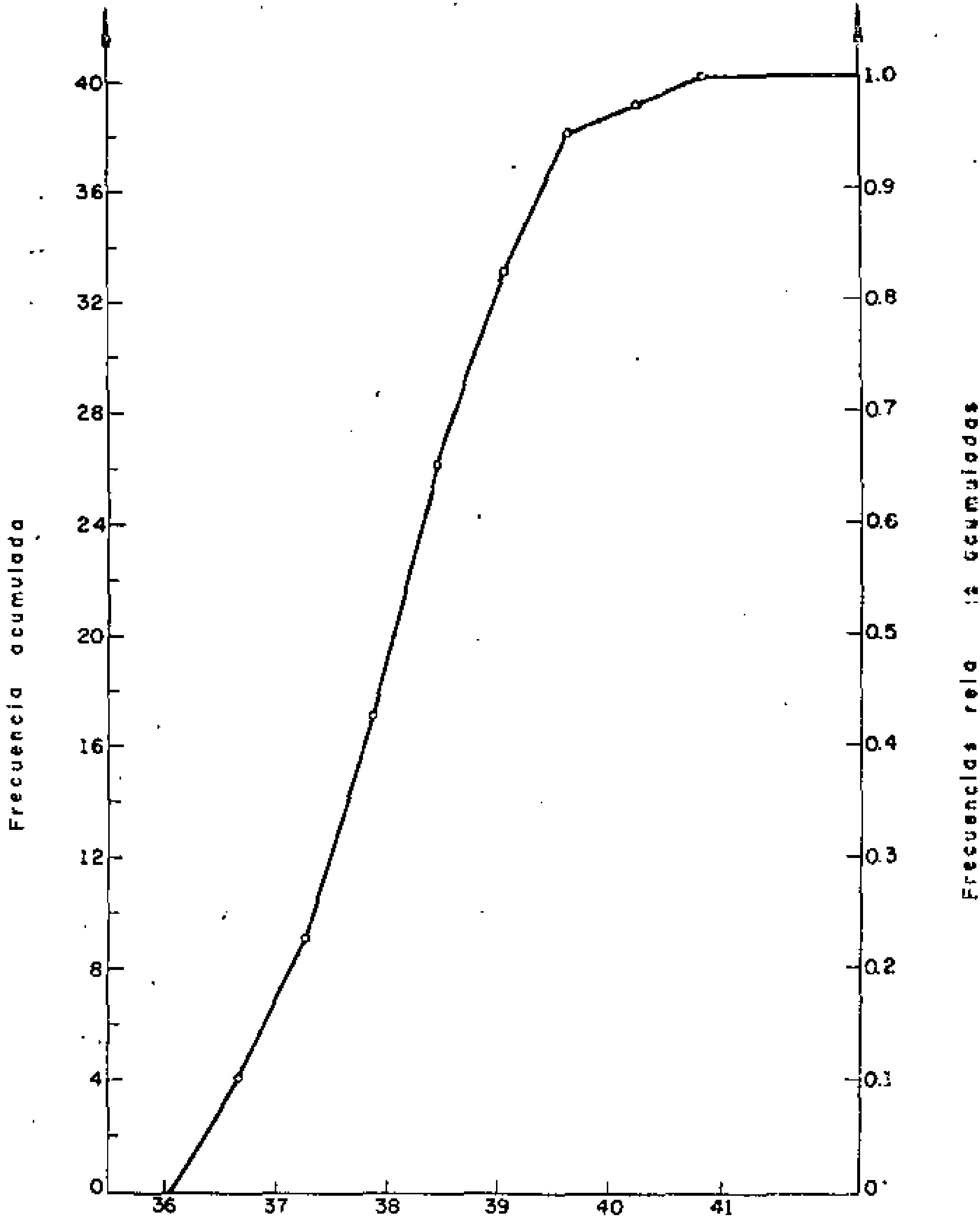
l) MEDIANA

$$Med = L_M + \frac{N/2 - F_M}{f_M} \; d_M$$

$$= 37.85 + \frac{(20 - 17)}{9} \times 0.6 = 38.05$$




Histograma y Polígono de frecuencias




Frecuencias acumuladas

PROBLEMA 4

Datos Básicos

| lbs. |
|---|
| 10,350 |
| 9,300 |
| 9,600 |
| 10,300 |
| 9,400 |
| 10,600 |
| 10,100 |
| 9,900 |
| 9,500 |
| 10,200 |
| 9,300 |
| 9,550 |
| 9,550 |
| 10,500 |
| 10,200 |

Datos Ordenados

| lbs. |
|---|
| 9,300 |
| 9,300 |
| 9,400 |
| 9,500 |
| 9,550 |
| 9,550 |
| 9,600 |
| 9,900 |
| 10,100 |
| 10,200 |
| 10,200 |
| 10,300 |
| 10,350 |
| 10,500 |
| 10,600 |

Agrupar, efectuar la transformación $Y = X - 10,000$ y calcular $\overline{X}$ a partir de $\overline{Y}$.

$$Y = X - 10,000.$$

Tabla 1.
(Agrupamiento arbitrario)

| | Límites | | M. C. | F | Y | Yf |
|---|---|---|---|---|---|---|
| | 9300 - 9500 | | 9,400 | 4 | -600 | -2400 |
| | 9500 - 9700 | | 9,600 | 3 | -400 | -1200 |
| | 9700 - 9900 | | 9,800 | 1 | -200 | - 200 |
| | 9900 - 10100 | | 10,000 | 1 | 0 | 0 |
| | 10100 - 10300 | | 10,200 | 3 | 200 | 600 |
| | 10300 - 10500 | | 10,400 | 2 | 400 | 800 |
| | 10500 - 10700 | | 10,600 | 1 | 600 | 600 |
| Σ | | | | 15 | | -1800 |

$$Y = -\frac{1800}{15} = -120 \quad ; \qquad \bar{X} = \bar{Y} + 10,000$$

$$\bar{X} = -120 + 10,000 = 9880 \text{ lb}$$

A continuación encuentre $\bar{X}$ para los mismos datos usando

$$Y = \frac{X - C_1}{C_2}$$

con $C_1 = 10,000$ (Marca de clase)

$C_2 = 200$ (Ancho de intervalo)

TABLA II

| Límites | M. C. | f | Y | Yf |
|---|---|---|---|---|
| 9300 - 9500 | 9,400 | 4 | -3 | -12 |
| 9500 - 9700 | 9,600 | 3 | -2 | - 6 |
| 9700 - 9900 | 9,800 | 1 | -1 | - 1 |
| 9900 -10100 | 10,000 | 1 | 0 | 0 |
| 10100 -10300 | 10,200 | 3 | 1 | 3 |
| 10300 -10500 | 10,400 | 2 | 2 | 4 |
| 10500 -10700 | 10,600 | 1 | 3 | 3 |
| | | 15 | | -9 |

$$Y = \frac{X - 10,000}{200} \qquad (C_1 = 10,000 \ , \ C_2 = 200)$$

$$\overline{Y} = \frac{-9}{15} = 0.6 \qquad \longleftarrow \overline{Y} = \frac{\sum_{i=1}^{N} f_i y_i}{N}$$

$$\overline{X} = C_2 \ \overline{Y} + C_1$$

$$\overline{X} = 200 \ (-0.6) + 10,000$$

$$\overline{X} = -120 \ + 10,000 \ = 9880 \text{ lb}$$

(Coincide con el resultado de la página anterior)

Usando el método corto obtener la variancia y la desviación estándar de los datos de X.

$$Y = \frac{X - 10000}{200}$$

| M.C. | f | Y | Yf | $Y^2$ | $Y^2f$ |
|---|---|---|---|---|---|
| 9400 | 4 | -3 | -12 | 9 | 36 |
| 9600 | 3 | -2 | - 6 | 4 | 12 |
| 9800 | 1 | -1 | - 1 | 1 | 1 |
| 10000 | 1 | 0 | 0 | 0 | 0 |
| 10200 | 3 | 1 | 3 | 1 | 3 |
| 10400 | 2 | 2 | 4 | 4 | 8 |
| 10600 | 1 | 3 | 3 | 9 | 9 |
| | 15 | | -9 | 28 | 69 |

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i f_i}{N} = \frac{-9}{15} = -.6$$

$$\overline{Y^2} = \frac{\sum_{i=1}^{N} Y_i^2 f_i}{N} = \frac{69}{15} = 4.6$$

$$S_y^2 = \overline{Y^2} - \overline{Y}^2 = 4.6 - (-.6)^2 = 4.6 - 0.36 = 4.24$$

como $\quad S_y^2 = \dfrac{S_x^2}{C_2^2} \quad \rightarrow \quad S_x^2 = C_2^2 \; S_y^2$

$$S_x^2 = (200)^2 (4.24) = 169,600 \quad Seg^2$$

$$\therefore \; S_x = \sqrt{169,600 \text{ seg}^2} = 411.83 \text{ lb}$$

$$V = \frac{Sx}{\bar{x}} = \frac{411.83}{9880} = 0.04168$$

| X | f | Y | fY | $Y^2$ | $Y^2 f$ |
|---|---|---|---|---|---|
| 9300 | 2 | -14 | -28 | 196 | 392 |
| 9400 | 1 | -12 | -24 | 144 | 144 |
| 9500 | 1 | -10 | -10 | 100 | 100 |
| 9550 | 2 | -9 | -18 | 81 | 162 |
| 9600 | 1 | -8 | -8 | 64 | 64 |
| 9900 | 1 | -2 | -2 | 4 | 4 |
| 10100 | 1 | 2 | 2 | 4 | 4 |
| 10200 | 2 | 4 | 8 | 16 | 32 |
| 10300 | 1 | 6 | 6 | 36 | 36 |
| 10350 | 1 | 7 | 7 | 49 | 49 |
| 10500 | 1 | 10 | 10 | 100 | 100 |
| 10600 | 1 | 12 | 12 | 144 | 144 |
| | 15 | | -45 | | 1231 |

$$Y = \frac{X - 10{,}000}{50}, \quad Y_1 = \frac{9300 - 10000}{50} = 14, \quad Y_2 = \frac{9400 - 10000}{50} = 12$$

$$\bar{Y} = \frac{-45}{15} = -3$$

$$\bar{Y} = \frac{\bar{X} - 10{,}000}{50} \Rightarrow \bar{X} = 50\bar{Y} + 10{,}000$$

$$\bar{X} = 50(-3) + 10{,}000 = 9850$$

$$\overline{Y^2} = \frac{1231}{15} = 82 \; ; \; S_Y^2 = \overline{Y^2} - \bar{Y}^2 = 82 - (-3)^2 = 73$$

$$S_Y^2 = \frac{S_X^2}{50} \Rightarrow S_X^2 = 50^2 S_Y^2 = 2500 \times 73 = 182{,}500$$

$$S_X = \sqrt{182{,}500} = 427.2$$

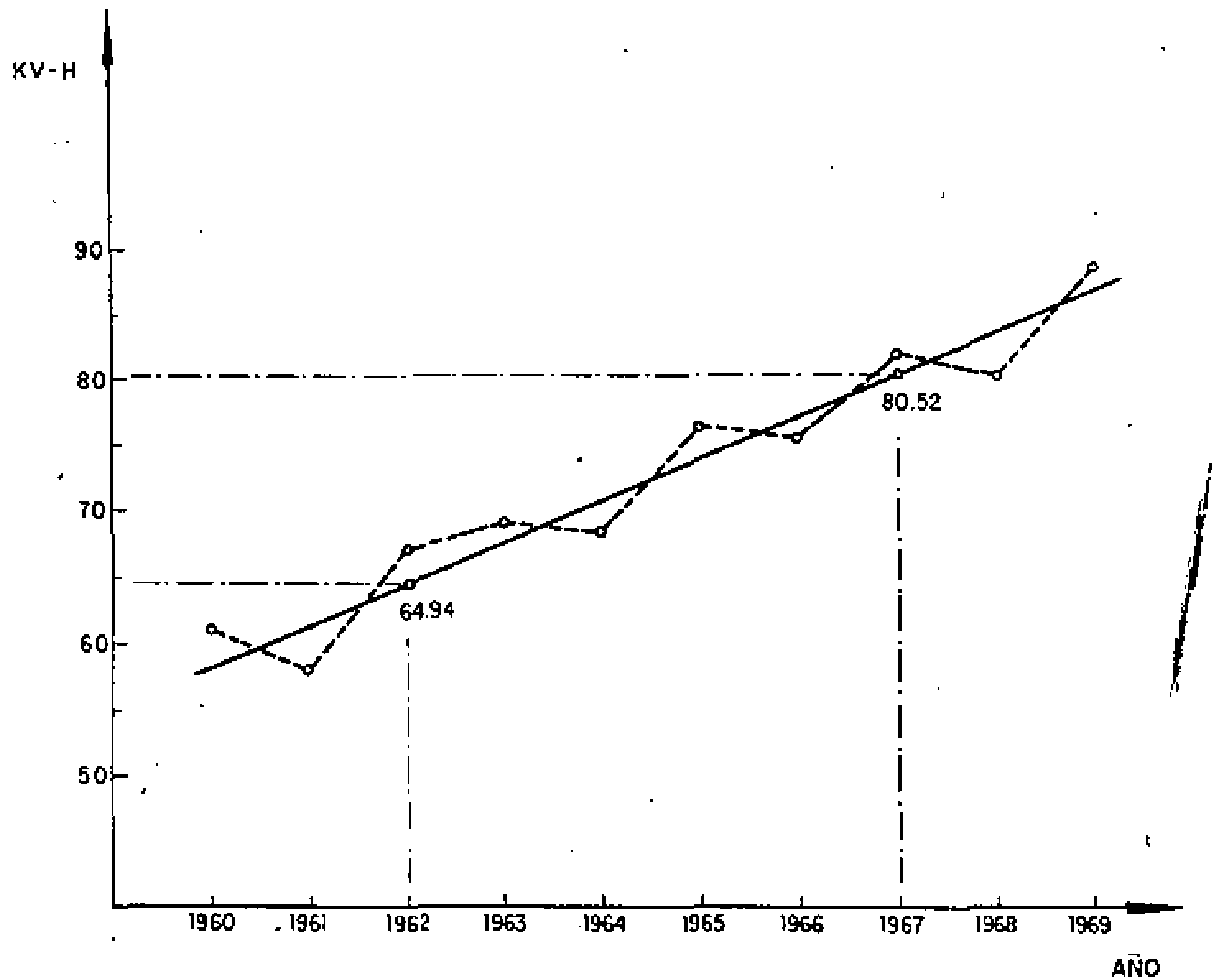
$$C.V_X = \frac{427.2}{9850} = 4.34\,\%$$

*PROBLEMA 5*. PARA LA SIGUIENTE SERIE DE TIEMPOS, DETERMINAR LA RECTA DE TENDENCIA GENERAL, MEDIANTE EL METODO DE DOS PROMEDIOS.

| AÑO | Y | |
|---|---|---|
| 1960 | 61.2 | |
| 61 | 58.3 | |
| 62 | 67.1 | |
| 63 | 69.2 | |
| 64 | 68.9 | $\Sigma = 324.7$ |
| 65 | 76.1 | |
| 66 | 75.9 | |
| 67 | 82.0 | |
| 68 | 80.1 | |
| 69 | 88.5 | $\Sigma = 402.6$ |

$\overline{Y}_1 = \frac{324.7}{5} = 64.94,$ $t_1 = 1962$

$\overline{Y}_2 = \frac{402.6}{5} = 80.52,$ $t_2 = 1967$



KV-H
90
80
70
60
50
80.52
64.94
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
AÑO

PROBLEMA 6. La tabla A contiene 20 parejas de datos donde:

$Y_i$ = Contenido de agua determinado por un método convencional.

$X_i$ = Contenido de agua por método I.

Mediante el método corto obtener la recta de regresión.

TABLA A

| Y | X |
|---|---|
| 62 | 58 |
| 47 | 50 |
| 19 | 17 |
| 15 | 17 |
| 65 | 68 |
| 28 | 24 |
| 63 | 60 |
| 37 | 35 |
| 20 | 21 |
| 9 | 9 |
| 64 | 69 |
| 56 | 57 |
| 53 | 54 |
| 2 | 5 |
| 49 | 48 |
| 53 | 50 |
| 15 | 14 |
| 22 | 20 |
| 13 | 13 |
| 35 | 34 |

TABLA DE CONTEO

| Y \ X | 0.5-10.5 | 10.5-20.5 | 20.5-30.5 | 30.5-40.5 | 40.5-50.5 | 50.5-60.5 | 60.5-70.5 |
|---|---|---|---|---|---|---|---|
| 0.5-10.5 | // 2 | | | | | | |
| 10.5-20.5 | | //// 4 | / 1 | | | | |
| 20.5-30.5 | | / 1 | / 1 | | | | |
| 30.5-40.5 | | | | / 2 | | | |
| 40.5-50.5 | | | | | / 2 | | |
| 50.5-60.5 | | | | | / 1 | / 2 | |
| 60.5-70.5 | | | | | | / 2 | / 2 |

| Y \ X | 5.5 | 15.5 | 25.5 | 35.5 | 45.5 | 55.5 | 65.5 | $f_Y$ | $Y'$ | $f_Y Y'$ | $Y'^2$ | $f_Y Y'^2$ | $\Sigma f_{ixy} X'Y'$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5.5 | 2 | | | | | | | 2 | -3 | -6 | | | |
| 15.5 | | 4 | 1 | | | | | 5 | -2 | -10 | | | |
| 25.5 | | 1 | 1 | | | | | 2 | -1 | - 2 | | | |
| 35.5 | | | | 2 | | | | 2 | 0 | 0 | | | |
| 45.5 | | | | | 2 | | | 2 | 1 | 2 | | | |
| 55.5 | | | | | 1 | 2 | | 3 | 2 | 6 | | | |
| 65.5 | | | | | | 2 | 2 | 4 | 3 | 12 | | | |
| $f_x$ | 2 | 5 | 2 | 2 | 3 | 4 | 2 | 20 | | 2 | | | |
| $X'$ | -3 | -2 | -1 | 0 | 1 | 2 | 3 | | | | | | |
| $f_x X'$ | -6 | -10 | -2 | 0 | 3 | 8 | 6 | -1 | | | | | |
| $X'^2$ | | | | | | | | | | | | | |
| $f_x X'^2$ | | | | | | | | | | | | | |
| $\Sigma f_{ixy} X' Y'$ | | | | | | | | | | | | | |

Usando

$$X' = \frac{X - C_1}{C_2} \quad ; \qquad Y' = \frac{Y - C_3}{C_4}$$

con $C_1 = C_3 = 35.5$

$$C_2 = C_4 = 10$$

$$X' = \frac{X - 35.5}{10}$$

$$Y' = \frac{Y - 35.5}{10}$$

$$\overline{X}' = \frac{\sum_{i=1}^{k} f_i X_i'}{N} = \frac{-1}{20} = -0.05$$

$$\overline{Y}' = \frac{\sum_{i=1}^{M} f_i Y_i'}{N} = \frac{2}{20} = 0.10$$

| Y \ X | 5.5 | 15.5 | 25.5 | 35.5 | 45.5 | 55.5 | 65.5 | $f_y$ | $Y'$ | $f_yY'$ | $Y'^2$ | $f_yY'^2$ | $\Sigma f_{ixy}X'Y'$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5.5 | 9 2 18 | | | | | | | 2 | -3 | -6 | 9 | 18 | 18 |
| 15.5 | | 4 4 16 | 2 1 2 | | | | | 5 | -2 | -10 | 4 | 20 | 18 |
| 25.5 | | 2 1 2 | 1 1 1 | | | | | 2 | -1 | - 2 | 1 | 2 | 3 |
| 35.5 | | | | 0 2 0 | | | | 2 | 0 | 0 | 0 | 0 | 0 |
| 45.5 | | | | | 1 2 2 | | | 2 | 1 | 2 | 1 | 2 | 2 |
| 55.5 | | | | | 2 1 1 | 4 2 8 | | 3 | 2 | 6 | 4 | 12 | 10 |
| 65.5 | | | | | | 6 2 12 | 9 2 18 | 4 | 3 | 12 | 9 | 36 | 30 |
| $f_x$ | 2 | 5 | 2 | 2 | 3 | 4 | 2 | 20 | | 2 | | 90 | 81 |
| $X'$ | -3 | -2 | -1 | 0 | 1 | 2 | 3 | | | | | | |
| $f_xX'$ | -6 | -10 | -2 | 0 | 3 | 8 | 6 | -1 | | | | | |
| $X'^2$ | 9 | 4 | 1 | 0 | 1 | 4 | 9 | | | | | | |
| $f_x X'^2$ | 18 | 20 | 2 | 0 | 3 | 16 | 18 | 77 | | | | | |
| $\Sigma f_{ixy}X'Y'$ | 18 | 18 | 3 | 0 | 4 | 20 | 18 | 81 | | | | | |

$\Sigma f_{ixy} X'Y' = 81$

$\Sigma f_{ix} X'^2 = 77$

$\Sigma f_{iy} Y'^2 = 90$

$\overline{X'^2} = \frac{77}{20} = 3.85$ $\qquad$ $\overline{Y'^2} = \frac{90}{20} = 4.5$

$S^2_{X'} = \overline{X'^2} - \overline{X}'^2 = 3.85-(-0.05)^2 = 3.847;$ $\quad$ $S^2_{Y'} = 4.5-(0.1)^2 = 4.49$

$$m = \frac{\frac{1}{N}\sum_{i=1}^{N} f_{ixy}\, X'_i Y'_i - \bar{X}'\, \bar{Y}'}{S^2_{X'}} \quad \frac{C_4}{C_2}$$

$C_2 = C_4 = 10$ ; $N = 20$

$\bar{X}' = -0.05$ ; $\bar{Y}' = 0.10$

$S^2_{X'} = 3.847$

$$\sum_{i=1}^{k} f_{ixy}\, X'_i Y'_i = 81$$

$$m = \frac{\left(\frac{81}{20} - (-0.05)(0.1)\right)10}{(3.847)(10)} = \frac{4.055}{3.847} = 1.05$$

Además $C_1 = C_3 = 35.5$

$\bar{X} = C_2\bar{X}' + C_1 = 10(-0.05) + 35.5 = 35$

$\bar{Y} = C_4\bar{Y}' + C_3 = 10(0.1) + 35.5 = 36.5$

$b = \bar{Y} - m\bar{X} = 36.5 - (1.05)(35.) = -0.25$

$\tilde{Y} = 1.05\ X - 0.25$

Para estos mismos datos obtener la ecuación de la recta de regresión sin agrupar los datos.

| X | Y | XY | $X^2$ |
|---|---|---|---|
| 58 | 62 | 3596 | 3364 |
| 50 | 47 | 2350 | 2500 |
| 17 | 19 | 323 | 289 |
| 17 | 15 | 255 | 289 |
| 68 | 65 | 4420 | 4624 |
| 24 | 28 | 672 | 576 |
| 60 | 63 | 3780 | 3600 |
| 35 | 27 | 945 | 1225 |
| 21 | 20 | 420 | 441 |
| 9 | 9 | 81 | 81 |
| 69 | 64 | 4416 | 4761 |
| 57 | 56 | 3192 | 3249 |
| 54 | 53 | 2862 | 2916 |
| 5 | 2 | 10 | 25 |
| 48 | 49 | 2352 | 2304 |
| 50 | 53 | 2650 | 2500 |
| 14 | 15 | 210 | 196 |
| 20 | 22 | 440 | 400 |
| 13 | 13 | 169 | 169 |
| 34 | 35 | 1190 | 1156 |
| 723 | 717 | 34,333 | 34,665 |

$$m = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i Y_i - \bar{X}\,\bar{Y}}{\overline{X^2} - \bar{X}^2}$$

$$N = 20$$

$$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{717}{20} = 35.85$$

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{723}{20} = 36.15$$

$$\frac{1}{N}\sum_{i=1}^{N} X_i Y_i = \frac{34333}{20} = 1716$$

$$\overline{X^2} = \frac{\sum_{i=1}^{N} X_i^2}{N} = \frac{34665}{20}$$

$$\overline{X^2} = 1733.25$$

$$m = \frac{1716 - 35.85(36.15)}{1733.25 - (36.15)^2}$$

$$m = \frac{1716 - 1295.98}{1733.25 - 1306.82} = \frac{420.02}{426.43} = 0.98$$

$$b = \bar{Y} - m\bar{X}$$

$$b = 35.85 - 0.98(36.15) = 0.42$$

$$\tilde{Y} = 0.98\,X + 0.42$$

PROBLEMA 7. PARA LOS SIGUIENTES DATOS, AJUSTAR UNA PARABOLA POR EL METODO DE MINIMOS CUADRADOS.

| AÑO | Y |
|---|---|
| 1850 | 23.2 |
| 1860 | 31.4 |
| 1870 | 39.8 |
| 1880 | 50.2 |
| 1890 | 62.9 |
| 1900 | 76.0 |
| 1910 | 92.0 |
| 1920 | 105.7 |
| 1930 | 122.8 |
| 1940 | 131.7 |
| 1950 | 151.1 |

USANDO $t' = \frac{t - 1900}{10}$ $(C_1 = 1900, \quad C_2 = 10)$

| AÑO | $t'$ | Y | $t'^2$ | $t'^3$ | $t'^4$ | $t'Y$ | $t'^2Y$ |
|---|---|---|---|---|---|---|---|
| 1850 | -5 | 23.2 | 25 | -125 | 625 | -116. | 580.0 |
| 1860 | -4 | 31.4 | 16 | - 64 | 256 | -125.6 | 502.4 |
| 1870 | -3 | 39.8 | 9 | - 27 | 81 | -119.4 | 358.2 |
| 1880 | -2 | 50.2 | 4 | - 8 | 16 | -100.4 | 200.8 |
| 1890 | -1 | 62.9 | 1 | - 1 | 1 | - 62.9 | 62.9 |
| 1900 | 0 | 76.0 | 0 | 0 | 0 | 0 | 0 |
| 1910 | 1 | 92.0 | 1 | 1 | 1 | 92.0 | 92. |
| 1920 | 2 | 105.7 | 4 | 8 | 16 | 211.4 | 422.8 |
| 1930 | 3 | 122.8 | 9 | 27 | 81 | 368.4 | 1105.2 |
| 1940 | 4 | 131.7 | 16 | 64 | 256 | 526.8 | 2107.2 |
| 1950 | 5 | 151.1 | 25 | 125 | 625 | 755.5 | 3777.5 |
| | 0 | 886.8 | 110 | 0 | 1958 | 1429.8 | 9209.0 |

$\Sigma t_i' = 0$

$\Sigma y_i = 886.8$

$\Sigma t_i'^2 = 110$

$\Sigma t_i'^3 = 0$

$\Sigma t_i'^4 = 1958$

$\Sigma t_i' y_i = 1429.8$

$\Sigma t'^2 Y = 9209.0$

$N = 11$

RECORDANDO QUE:

$$b_0N + b_1\Sigma t + b_2\Sigma t^2 = \Sigma Y$$

$$b_0\Sigma t + b_1\Sigma t^2 + b_2\Sigma t^3 = \Sigma ty$$

$$b_0\Sigma t^2 + b_1\Sigma t^3 + b_2\Sigma t^4 = \Sigma t^2y$$

SE TIENE QUE:

$$11b_0 + 0b_1 + 110b_2 = 886.8 \qquad (1)$$

$$0b_0 + 110b_1 + 0b_2 = 1429.8 \qquad (2)$$

$$110b_0 + 0b_1 + 1958b_2 = 9209.0 \qquad (3)$$

DE LA EC 2:

$$110b_1 = 1429.8$$

$$b_1 = 13.0$$

MULTIPLICANDO LA EC 1 POR 10 Y RESTANDOLE LA EC 3:

$$\begin{array}{rcr} 110b_0 + 1100b_2 & = & 8868.0 \\ -110b_0 - 1958b_2 & = & -9209.0 \\ \hline -858.b_2 & = & 341.0 \end{array}$$

$$b_2 = \frac{341}{858} = 0.39744$$

SUSTITUYENDO EN EC 1:

$$11b_0 + 110(0.39744) = 886.8$$

$$11b_0 = 886.8 - 43.72 = 843.08$$

$$b_0 = \frac{843.08}{11} = 76.6436$$

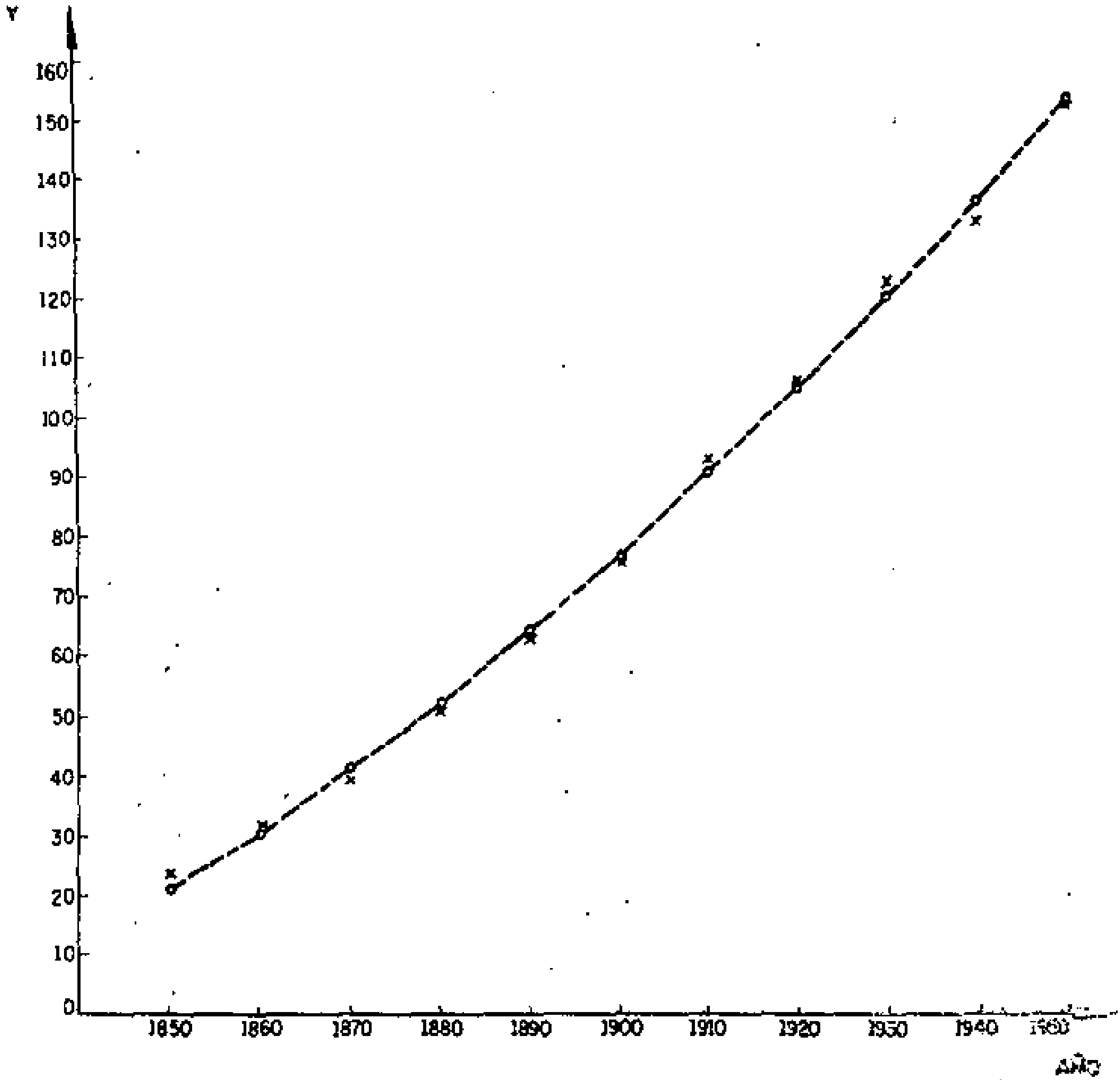
POR LO TANTO:

$$\tilde{Y} = 76.6436 + 13.0\ t' + 0.3974\ t'^2$$

DONDE EL ORIGEN $t' = 0$

CORRESPONDE A $t_i = 1900$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $t_i'$ | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
| X | 1850 | 1860 | 1870 | 1880 | 1890 | 1900 | 1910 | 1920 | 1930 | 1940 | 1950 |
| Y | 21.58 | 31.00 | 41.22 | 52.23 | 64.04 | 76.646 | 90.04 | 104.23 | 119.22 | 135.00 | 151.58 |


Y
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
AÑO

PROBLEMA 8. PARA LOS DATOS DE $X$ y $Y$ PRESENTADOS MAS ADELANTE, AJUSTAR MEDIANTE EL METODO DE MINIMOS CUADRADOS UNA CURVA DE LA FORMA $\bar{Y} = ab^X$, DONDE $a$ y $b$ SON CONSTANTES.

SACANDO LOGARITMOS BASE 10 A AMBOS MIEMBROS:

$$\log \tilde{Y} = \log ab^X$$

$$\log \bar{Y} = \log a + \log b^X$$

$$\log \tilde{Y} = \log a + X \log b$$

HACIENDO $\bar{Y}' = \log \tilde{Y}$

$$b' = \log a$$

$$m = \log b$$

SE TIENE: $\bar{Y}' = b' + m X$

RECORDANDO QUE PARA UNA RECTA AJUSTADA POR MINIMOS CUADRADOS SE TIENE:

$$M = \frac{\frac{1}{N} \Sigma X Y - \overline{X}\,\overline{Y}}{\overline{X^2} - \overline{X}^2} \quad ; \qquad b = \overline{Y} - m \overline{X}$$

| Y | Y' = log Y | X | $X^2$ | XY' |
|---|---|---|---|---|
| 32 | 1.5051 | 0 | 0 | 0 |
| 47 | 1.6721 | 1 | 1 | 1.6721 |
| 65 | 1.8129 | 2 | 4 | 3.6258 |
| 92. | 1.9638 | 3 | 9 | 5.8914 |
| 132 | 2.1206 | 4 | 16 | 8.4823 |
| 190 | 2.2788 | 5 | 25 | 11.3938 |
| 275 | 2.4393 | 6 | 36 | 14.6358 |
| Σ | 13.7926 | 21 | 91.0 | 45.7014 |

$$\overline{Y'} = \frac{13.7926}{7} = 1.9704$$

$$\overline{X} = \frac{21}{7} = 3$$

$$\overline{X^2} = \frac{91}{7} = 13.0$$

$$m = \frac{\frac{45.7014}{7} - 3(1.9704)}{13-(3)^2} = 0.1544$$

$$b' = \overline{Y'} - m\overline{X} = 1.9704 - 0.1544(3) = 1.5072$$

RECORDANDO QUE:

$$m = \log b, \quad b = 10^{0.1544} = 1.4269$$

$$b' = \log a, \quad a = 10^{1.5072} = 32.1514$$

$$\tilde{Y} = 32.1514(1.4269)^X$$

PROBLEMA 9

Calcular promedios móviles centrados de orden 4.

| AÑO | Y | Promedio móvil | Promedio móvil centrado |
|---|---|---|---|
| 1955 | 50 | | |
| 1956 | 36.5 | | |
| | | 43.5 | |
| 1957 | 43 | | 42.12 |
| | | 40.73 | |
| 1958 | 44.5 | | 40.93 |
| | | 41.13 | |
| 1959 | 38.9 | | 39.83 |
| | | 38.53 | |
| 1960 | 38.1 | | 37.80 |
| | | 37.08 | |
| 1961 | 32.6 | | 37.43 |
| | | 37.78 | |
| 1962 | 38.7 | | 38.16 |
| | | 38.53 | |
| 1963 | 41.7 | | 38.68 |
| | | 38.83 | |
| 1964 | 41.1 | | |
| 1965 | 33.8 | | |

PROBLEMA 10. Como resultado del censo habitacional en una ciudad, se determinó que el número anual de edificios construidos en ella de 1956 a 1965 fueron los indicados en la siguiente tabla. Determinar la tendencia de construcción mediante promedios móviles de orden 3, 4 y 8, y mediante los métodos de dos promedios y mínimos cuadrados (tendencia lineal en estos dos últimos casos).

PROMEDIOS MOVILES

| Año, T | Edificios x $10^3$, Y | Prom. móv. orden 3 | Suma orden 4 | Suma orden 8 | Prom. móv. orden 4 | Prom. móv. cent. orden 8 | Prom. móv. cent. orden 4 |
|---|---|---|---|---|---|---|---|
| 1956 | 5.3 | | | | | | |
| 1957 | 5.2 | 5.4 | | | 5.6 | | |
| 1958 | 5.7 | 5.7 | | | 5.7 | 5.65 | 5.65 |
| 1959 | 6.2 | 5.9 | 22.4 | 45.3 | 6.0 | 5.90 | 5.85 |
| 1960 | 5.8 | 6.3 | 22.9 | 47.2 | 6.3 | 6.20 | 6.15 |
| 1961 | 6.6 | 6.4 | 24.3 | 49.6 | 6.5 | 6.40 | 6.40 |
| 1962 | 6.7 | 6.7 | 25.3 | 51.3 | 7.0 | 6.75 | 6.75 |
| 1963 | 6.9 | 7.2 | 26.0 | 54.1 | 7.4 | 7.20 | 7.20 |
| 1964 | 7.9 | 7.4 | 28.1 | 57.7 | | | |
| 1965 | 8.1 | | 29.6 | | | | |

DOS PROMEDIOS

1a. parte $= \dfrac{28.2}{5} = 5.64$

2a. parte $= \dfrac{36.2}{5} = 7.24$

MINIMOS CUADRADOS

La recta de regresión se calculará con un cambio previo de variable:

$$X = T - 1956$$

| X | | Y | Xy | $X^2$ |
|---|---|---|---|---|
| 0 | (1956) | 5.3 | 0 | 0 |
| 1 | (1957) | 5.2 | 5.2 | 1 |
| 2 | (1958) | 5.7 | 11.4 | 4 |
| 3 | (1959) | 6.2 | 18.6 | 9 |
| 4 | (1960) | 5.8 | 23.2 | 16 |
| 5 | (1961) | 6.6 | 33.0 | 25 |
| 6 | (1962) | 6.7 | 40.2 | 36 |
| 7 | (1963) | 6.9 | 48.3 | 49 |
| 8. | (1964) | 7.9 | 63.2 | 64 |
| 9 | (1965) | 8.1 | 72.9 | 81 |
| 45 | | 64.4 | 316.0 | 285 |

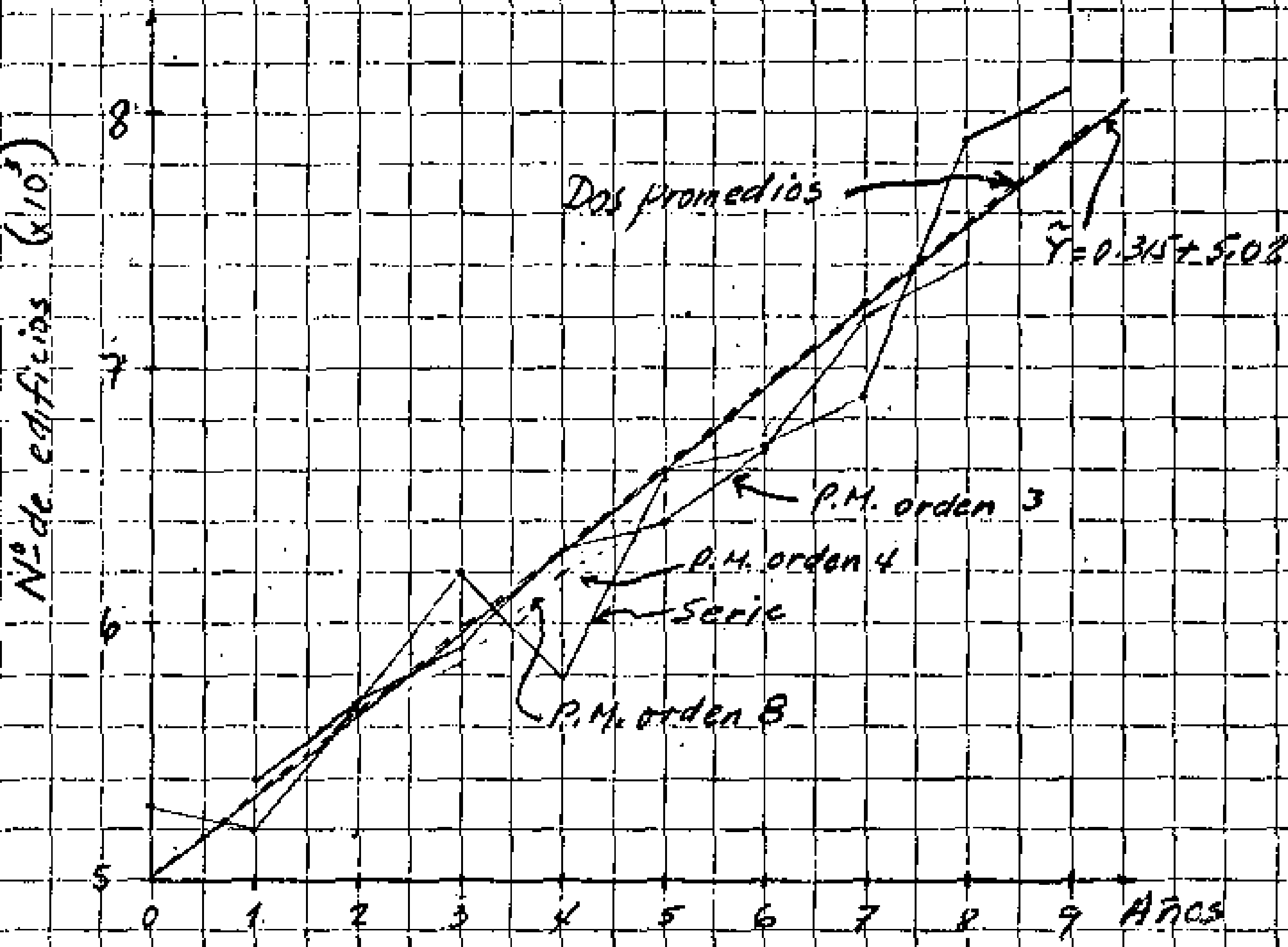
$\overline{X} = 4.5$ $\quad \overline{Y} = 6.44$ $\quad \frac{1}{N}\Sigma xy = 31.6$ $\quad \overline{X^2} = 28.5$

$\overline{X}^2 = 20.25$ $\quad \overline{X}\overline{Y} = 29.0$

$$m = \frac{31.6 - 29.0}{28.5 - 20.25} = 0.315$$

$$b = \overline{y} - m\overline{x} = 6.44 - 1.42 = 5.02$$

$$\boxed{\hat{y} = 0.315\ x + 5.02}$$


Nº de edificios ($\times 10^3$)
8
7
6
5
0
1
2
3
4
5
6
7
8
9
Años
Dos promedios
$\tilde{Y} = 0.315t + 5.02$
P.M. orden 3
P.M. orden 4
Serie
P.M. orden 8

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

PROBABILIDAD Y ESTADISTICA FUNDAMENTOS Y APLICACIONES

PROBLEMAS RESUELTOS
(Continuación)

DR. OCTAVIO A. RAGON CHAVEZ.
Marzo, 1980.

PROBLEMA 11. Con los siguientes datos, efectuar un análisis de correlación de X con $Y_\alpha$ y con $Y_\beta$, y comparar: $\rho_{XY\alpha}$ con $\rho_{XY\beta}$ para determinar cuál predomina.

| X | $Y_\alpha$ | $Y_\beta$ |
|---|---|---|
| 12 | 5.27 | 5.77 |
| 18 | 5.68 | 5.04 |
| 24 | 6.25 | 6.99 |
| 30 | 7.2 | 7.39 |
| 36 | 8.02 | 8.88 |
| 42 | 8.71 | 8.9 |
| 44 | 8.42 | 9.0 |

Solución:

$$S_x = \sqrt{\sum_{i=1}^{N} \frac{(X_i-\overline{X})^2}{N}} \quad ; \quad S_y = \sqrt{\sum_{i=1}^{N} \frac{(Y_i-\overline{Y})^2}{N}}$$

Covariancia: $$S_{xy} = \sum_{i=1}^{N} \frac{(X_i-\overline{X})(y_i-\overline{Y})}{N}$$

Coeficiente de correlación: $$\rho_{xy} = \frac{S_{xy}}{S_x S_y}$$

| X | Y | $(X_i - \bar{X})$ | $(Y_i - \bar{Y})$ | $(X_i - \bar{X})^2$ | $(Y_i - \bar{Y})^2$ | $(X_i - \bar{X})(Y_i - \bar{Y})$ |
|---|---|---|---|---|---|---|
| 12 | 5.27 | -17.43 | -1.81 | 303.8 | 3.28 | 31.55 |
| 18 | 5.68 | -11.43 | -1.4 | 130.64 | 1.96 | 16.0 |
| 24 | 6.25 | - 5.43 | -8.3 | 29.48 | 0.69 | 4.51 |
| 30 | 7.2 | 0.57 | 0.12 | 0.32 | 0.01 | 0.07 |
| 36 | 8.02 | 6.57 | 0.94 | 43.16 | 0.88 | 6.18 |
| 42 | 8.71 | 12.57 | 1.63 | 158.0 | 2.66 | 20.44 |
| 44 | 8.42 | 14.57 | 1.34 | 212.28 | 1.80 | 19.52 |
| 206 | 49.55 | | | 877.68 | 11.28 | 98.32 |

$\bar{X} = \frac{206}{7} = 29.43$ $\quad S_x^2 = \frac{877.68}{7} = 125.38$ $\quad S_y^2 = \frac{11.28}{7} = 1.61,$ $\quad S_{xy} = \frac{98.32}{7} = 14.05$

$\bar{Y} = \frac{49.55}{7} = 7.08$ $\quad S_x = \sqrt{125.38} = 11.2,$ $\quad S_y = \sqrt{1.61} = 1.27$

$$\rho_{xya} = \frac{14.05}{11.2 \times 1.27} = 0.98777$$

| $X$ | $Y$ | $(X_i - \overline{X})$ | $(Y_i - \overline{Y})$ | $(X_i - \overline{X})^2$ | $(Y_i - \overline{Y})^2$ | $(X_i - \overline{X})(Y_i - \overline{Y})$ |
|---|---|---|---|---|---|---|
| 12 | 5.77 | -17.43 | -1.65 | 303.80 | 2.72 | 28.76 |
| 8 | 5.04 | -11.43 | -2.38 | 130.64 | 5.66 | 27.20 |
| 24 | 6.99 | - 5.43 | -0.43 | 29.48 | -0.18 | 2.33 |
| 30 | 7.39 | 0.57 | -0.03 | 0.32 | 0.0009 | -0.02 |
| 36 | 8.88 | 6.57 | 1.46 | 43.16 | 2.13 | 9.59 |
| 42 | 8.90 | 12.57 | 1.48 | 158.00 | 2.19 | 18.60 |
| 44 | 9.00 | 14.57 | 1.58 | 212.28 | 2.50 | 23.02 |
| 206 | 51.97 | | | 877.68 | 15.38 | 109.48 |

$\overline{X} = \frac{206}{7}\ 29.43 \qquad S_x^2 = \frac{877.68}{7} = 125.38, \qquad S_Y^2 = \frac{15.38}{7} = 2.2, \qquad S_{xy} = \frac{109.48}{7} = 15.69$

$\overline{Y} = \frac{51.97}{7} = 7.42 \qquad S_x = \sqrt{125.38} = 11.2, \quad S_Y = \sqrt{2.2} = 1.48$

$$\rho_{xy_\beta} = \frac{15.64}{(11.2)(1.48)}$$

$$\rho_{xy\beta} = 0.94353 < \rho_{xy\alpha}$$

PROBLEMA 12. Los resultados de probar 10 vigas de madera son los siguientes:

Calcular $\rho_{xy}$, recta de regresión y variancias, explicada e inexplicada.

| X en Kg. | $\Delta$ en cm |
|---|---|
| 950 | 0.33 |
| 1050 | 0.37 |
| 750 | 0.28 |
| 900 | 0.30 |
| 700 | 0.27 |
| 650 | 0.28 |
| 950 | 0.35 |
| 850 | 0.40 |
| 600 | 0.26 |
| 700 | 0.29 |

| X | Y | X - $\bar{X}$ | Y - $\bar{Y}$ | (X - $\bar{X}$) (Y - $\bar{Y}$) | (X - $\bar{X}$)$^2$ | (Y-$\bar{Y}$)$^2$ |
|---|---|---|---|---|---|---|
| 950 | 0.33 | 140 | .017 | 2.38 | 19,600 | 0.000289 |
| 1050 | 0.37 | 240 | .057 | 13.68 | 57,600 | 0.003249 |
| 750 | 0.28 | - 60 | - .033 | 1.98 | 3,600 | 0.001089 |
| 900 | 0.30 | 90 | - .013 | - 1.17 | 8,100 | 0.000169 |
| 700 | 0.27 | -110 | - .043 | 4.73 | 12,100 | 0.001849 |
| 650 | 0.28 | -160 | - .033 | 5.28 | 25,600 | 0.001089 |
| 950 | 0.35 | 140 | .037 | 5.18 | 19,600 | 0.001369 |
| 850 | 0.40 | 40 | .087 | 3.48 | 1,600 | 0.007569 |
| 600 | 0.26 | -210 | - .053 | 11.13 | 44,100 | 0.002809 |
| 700 | 0.29 | -110 | - .023 | 2.53 | 12,100 | 0.000529 |
| 8100 | 3.13 | | | 49.20 | 204,000 | 0.02001 |

$$\therefore \quad \bar{X} = \frac{8100}{10} = 810 \text{ Kg.} \qquad S_{xy} = \frac{49.20}{10} = 4.92$$

$$\bar{Y} = \frac{3.13}{10} = 0.313 \text{ cm.} \qquad S_x^2 = \frac{204,000}{10} = 20400, \quad S_x = 142.83$$

$$S_y^2 = \frac{0.02001}{10} = 0.002001, \quad S_y = 0.04473$$

$$\rho_{xy} = \frac{4.92}{(142.83)(.04473)} = .7701$$

$$m = \rho_{xy}\frac{S_y}{S_x} = 0.7701\ \frac{0.04473}{142.83} = 0.0002411$$

$$b = \overline{Y} - m\,\overline{X}$$

$$b = 0.313 - 0.000241(810)$$

$$b = 0.11779 \doteq 0.118$$

$$\boxed{\hat{Y} = 0.000241\ X + 0.118}$$

La variancia de la estimación $S^2_{y|x}$ será

$$S^2_{y|x} = S^2_y(1-\rho^2_{xy})$$

$$S^2_{y|x} = 0.002001\ \left[1 - (0.7701)^2\right] = 0.000814$$

Pero $\quad S^2_y = S^2_{\hat{y}} + S^2_{y|x}$ ,

por lo que

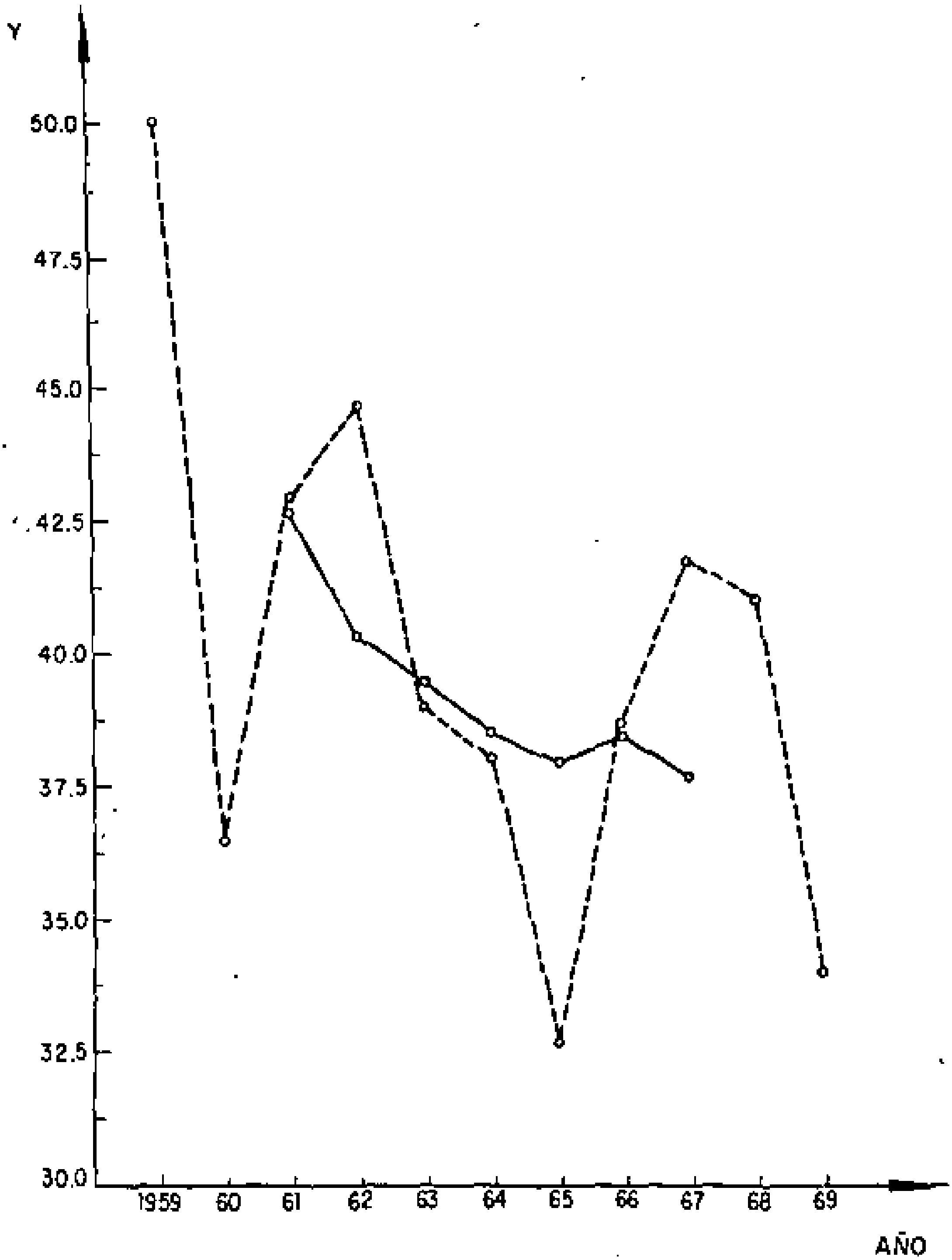
$$S^2_{\hat{y}} = S^2_y - S^2_{y|x} = 0.002001 - 0.000814 = 0.001187$$

PROBLEMA 13. OBTENGA LOS PROMEDIOS MOVILES DE ORDEN 5 DE LOS DATOS SIGUIENTES:

| AÑO | Y | SUMA 5 años | PROMEDIO MOVIL 5 años |
|---|---|---|---|
| 1959 | 50.0 | | |
| 1960 | 36.5 | | |
| 1961 | 43.0 | 212.9 | 42.58 |
| 1962 | 44.5 | 201.0 | 40.2 |
| 1963 | 38.9 | 197.0 | 39.4 |
| 1964 | 38.1 | 192.80 | 38.56 |
| 1965 | 32.6 | 190.0 | 38.0 |
| 1966 | 38.7 | 192.2 | 38.44 |
| 1967 | 41.7 | 187.9 | 37.58 |
| 1968 | 41.1 | | |
| 1969 | 33.8 | | |

LA REPRESENTACION SERA





PROBLEMA 14.-Los datos presentados en la tercer columna de la tabla siguiente corresponden al consumo mensual de energía eléctrica durante los años de 1960 a 1963, en cientos de miles de Kw/h. Calcular la componente estacional para cada mes, determinando a la tendencia, T, mediante el método de mínimos cuadrados, tomando como valores de Y a los promedios mensuales de consumo de cada año.

SOLUCION:Determinar la recta tomando como origen al tiempo medio de -- 1960, al cual correponde el promedio del consumo de ese año.

| Año,t | y | ty | $t^2$ |
|---|---|---|---|
| 0 | 4.88 | 0 | 0 |
| 1 | 5.88 | 5.88 | 1 |
| 2 | 6.48 | 12,96 | 4 |
| 3 | 7.08 | 21.24 | 9 |
| $\Sigma = 6$ | 24.32 | 40.08 | 14 |

$\bar{t} = 6/4 = 1.5$ , $\bar{Y} = 24.32/4 = 6.08$

$\frac{1}{N} \Sigma ty = 10,02$ , $\overline{t^2} = 14/4 = 3.5$ , $\sigma_t^2 = 3.5 - 1.5^2 = 1.25$

$$m = \frac{\frac{1}{N} \Sigma ty - \bar{t}\bar{Y}}{\sigma_t^2} = \frac{10.02 - 9.12}{1.25} = 0.72$$

$b = \bar{Y} - m\bar{t} = 6.08 - 0.72 \times 1.5 = 5$

$\tilde{Y} = 0.72\ t + 5$

Con origen en enero de 1960, y con el tiempo en meses la recta resulta ser:

$T = 0.72\ (t - 0.5) / 12 + 5 - 0.72 \times 0.5$

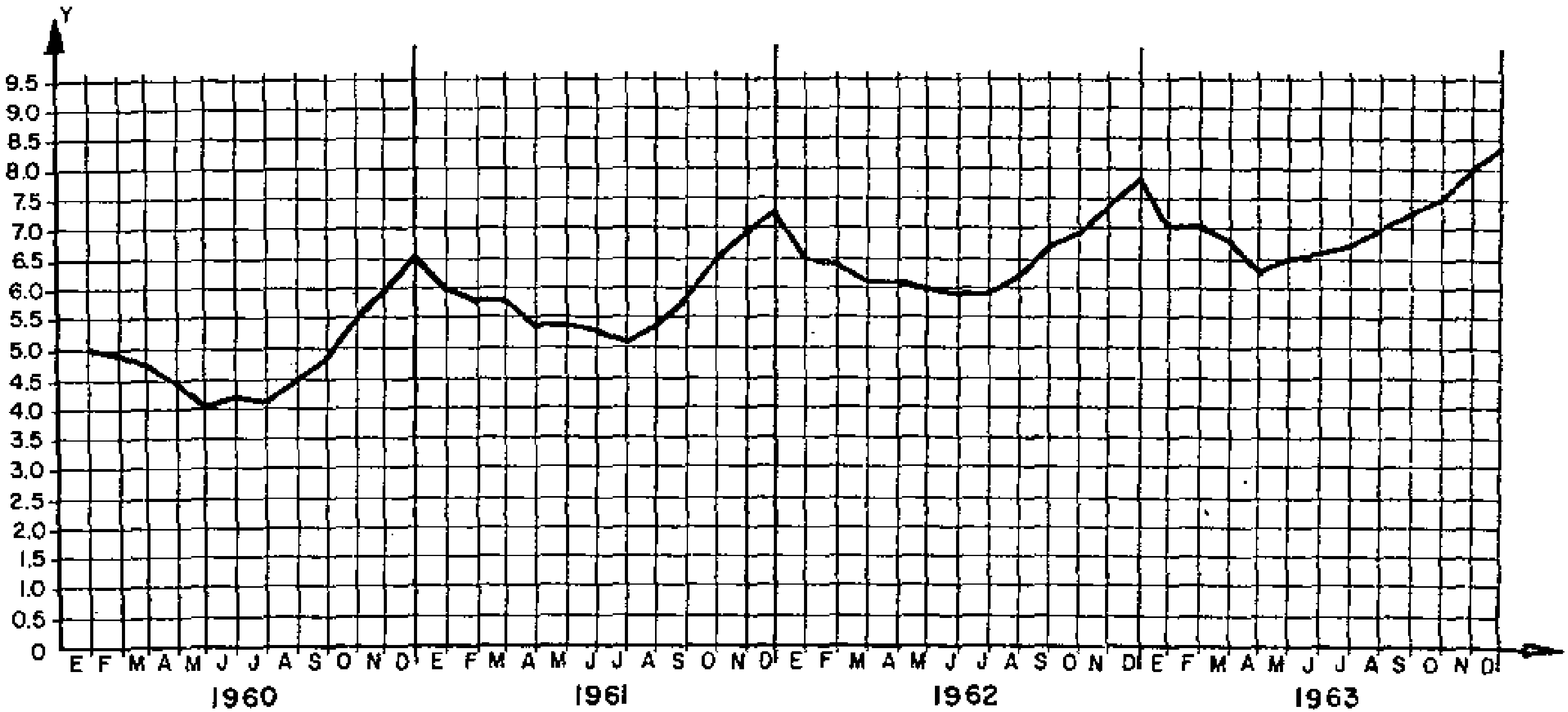
$T = 0.03\ t + 4.64$

| AÑO | MES | CONSUMO Y | SUMA | PROMEDIO MOVIL CENTRADO | PORCENTAJE DEL PROMEDIO MOVIL | E | T | ET | $\frac{Y}{ET}$ = CI |
|---|---|---|---|---|---|---|---|---|---|
| 1960 ( $\bar{Y}$ = 4.33 ) | E | 5.0 | | | | 105.6 | 4.67 | 493.2 | 101.4 |
| | F | 4.9 | | | | 102.4 | 4.73 | 484.4 | 101.2 |
| | M | 4.7 | | | | 99.4 | 4.79 | 476.1 | 98.7 |
| | A | 4.4 | | | | 95.5 | 4.85 | 463.2 | 95.0 |
| | M | 4.0 | | | | 93.9 | 4.91 | 461.0 | 86.8 |
| | J | 4.2 | | | | 91.5 | 4.97 | 454.8 | 92.3 |
| | J | 4.1 | | 4.9 | 83.7 | 87.1 | 5.03 | 438.1 | 93.6 |
| | A | 4.5 | | 5.0 | 90.0 | 90.7 | 5.09 | 461.7 | 97.5 |
| | S | 4.8 | | 5.1 | 94.1 | 98.4 | 5.15 | 506.8 | 94.7 |
| | O | 5.5 | | 5.2 | 105.8 | 106.7 | 5.21 | 555.9 | 98.9 |
| | N | 5.9 | | 5.3 | 111.3 | 112.2 | 5.27 | 591.3 | 99.8 |
| | D | 6.5 | 58.5 | 5.4 | 120.4 | 117.4 | 5.33 | 625.7 | 103.9 |
| 1961 ( $\bar{Y}$ = 5.88 ) | E | 6.0 | 59.5 | 5.5 | 109.1 | 105.6 | 5.39 | 569.2 | 105.4 |
| | F | 5.8 | 60.4 | 5.5 | 105.5 | 102.4 | 5.45 | 558.1 | 103.9 |
| | M | 5.8 | 61.5 | 5.6 | 103.6 | 99.4 | 5.51 | 547.7 | 105.9 |
| | A | 5.4 | 62.5 | 5.7 | 94.7 | 95.5 | 5.57 | 531.9 | 101.5 |
| | M | 5.4 | 63.9 | 5.8 | 93.1 | 93.9 | 5.63 | 528.7 | 102.1 |
| | J | 5.3 | 65.0 | 5.9 | 89.8 | 91.5 | 5.69 | 520.6 | 101.8 |
| | J | 5.1 | 66.0 | 5.9 | 86.4 | 87.1 | 5.75 | 500.8 | 101.8 |
| | A | 5.4 | 66.9 | 6.0 | 90.0 | 90.7 | 5.81 | 527.0 | 102.5 |
| | S | 5.8 | 67.9 | 6.0 | 96.7 | 98.4 | 5.87 | 577.6 | 100.4 |
| | O | 6.5 | 68.9 | 6.0 | 108.3 | 106.7 | 5.93 | 632.7 | 102.7 |
| | N | 6.9 | 69.9 | 6.1 | 113.1 | 112.2 | 5.99 | 672.1 | 102.7 |
| | D | 7.2 | 70.6 | 6.1 | 118.0 | 117.4 | 6.05 | 710.3 | 101.4 |
| 1962 ( $\bar{Y}$ = 6.38 ) | E | 6.5 | 71.1 | 6.2 | 104.8 | 105.6 | 6.11 | 645.2 | 100.7 |
| | F | 6.4 | 71.7 | 6.3 | 101.6 | 102.4 | 6.17 | 631.8 | 101.3 |
| | M | 6.1 | 72.0 | 6.3 | 96.8 | 99.4 | 6.23 | 619.3 | 98.5 |
| | A | 6.1 | 72.7 | 6.4 | 95.3 | 95.5 | 6.29 | 600.7 | 101.5 |
| | M | 6.0 | 73.3 | 6.4 | 93.8 | 93.9 | 6.35 | 596.3 | 100.6 |
| | J | 5.9 | 73.9 | 6.5 | 90.8 | 91.5 | 6.41 | 586.5 | 100.6 |
| | J | 5.9 | 74.7 | 6.5 | 90.8 | 87.1 | 6.47 | 563.5 | 104.7 |
| | A | 6.2 | 75.5 | 6.6 | 93.9 | 90.7 | 6.53 | 592.3 | 104.7 |
| | S | 6.7 | 76.4 | 6.6 | 101.5 | 98.4 | 6.59 | 648.5 | 103.3 |
| | O | 6.9 | 76.8 | 6.6 | 104.5 | 106.7 | 6.65 | 700.6 | 97.2 |
| | N | 7.3 | 77.2 | 6.7 | 109.0 | 112.2 | 6.71 | 752.9 | 97.0 |
| | D | 7.8 | 77.8 | 6.7 | 116.4 | 117.4 | 6.77 | 794.8 | 98.1 |

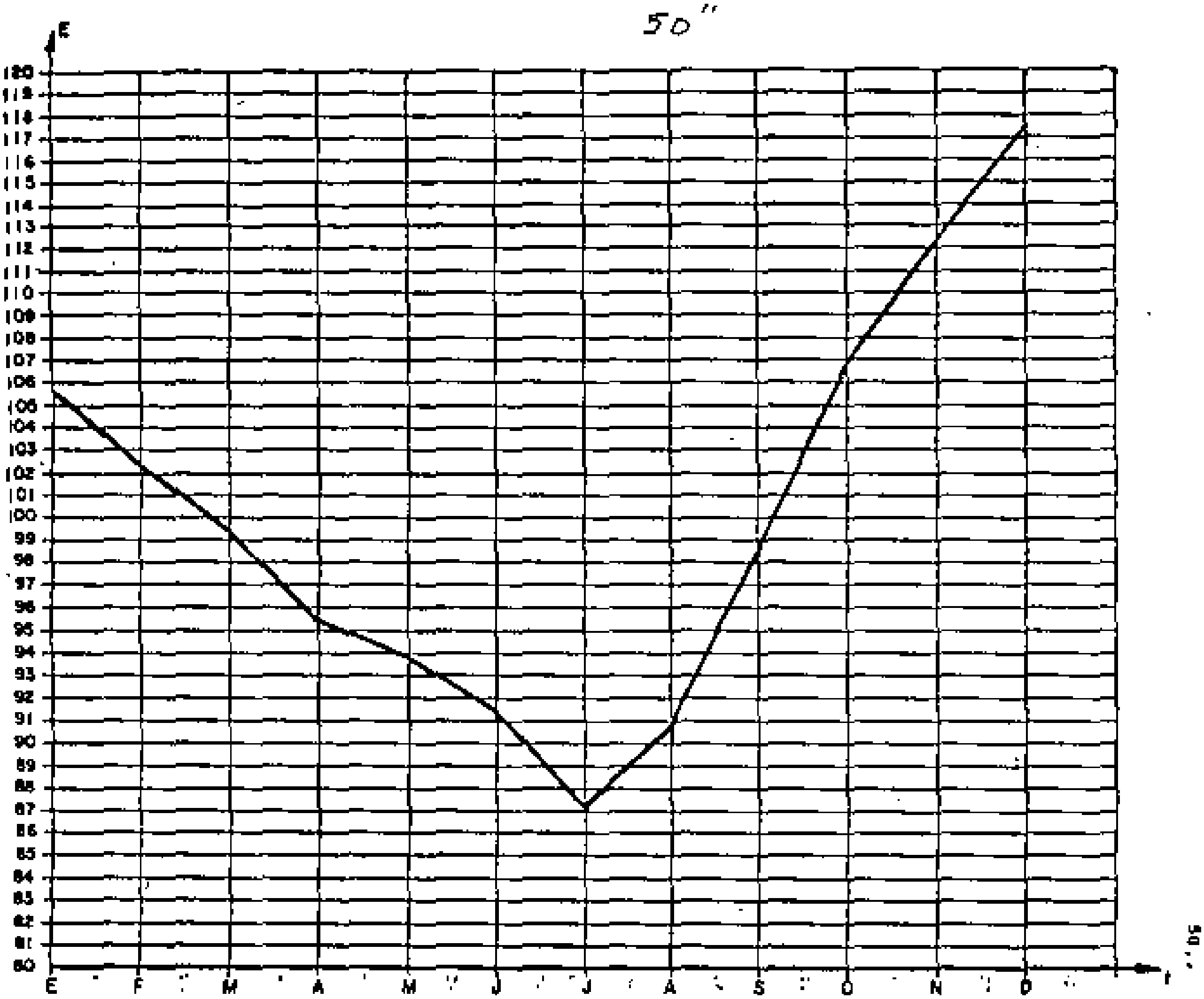
| AÑO | MES t | CONSUMO Y | SUMA | PROMEDIO MOVIL CENTRADO | PORCENTAJE DEL PROMEDIO MOVIL | E | T | ET | Y/ET = CI |
|---|---|---|---|---|---|---|---|---|---|
| 1963 ( $\bar{Y}$ = 7.03 ) | E | 7.0 | 78.3 | 6.8 | 102.9 | 105.6 | 6.83 | 721.2 | 97.1 |
| | F | 7.0 | 78.9 | 6.9 | 101.4 | 102.4 | 6.89 | 705.5 | 99.2 |
| | M | 6.8 | 79.6 | 6.9 | 98.6 | 99.4 | 6.95 | 690.8 | 98.4 |
| | A | 6.3 | 79.8 | 7.0 | 90.0 | 95.5 | 7.01 | 669.5 | 94.1 |
| | M | 6.5 | 80.3 | 7.0 | 92.9 | 93.9 | 7.07 | 663.9 | 97.9 |
| | J | 6.6 | 81.0 | 7.1 | 93.0 | 91.5 | 7.13 | 652.4 | 101.2 |
| | J | 6.7 | 81.8 | | | 87.1 | 7.19 | 626.2 | 107.0 |
| | A | 7.0 | 82.6 | | | 90.7 | 7.25 | 657.6 | 106.4 |
| | S | 7.2 | 83.1 | | | 98.4 | 7.31 | 719.3 | 100.1 |
| | O | 7.5 | 83.7 | | | 106.7 | 7.37 | 786.4 | 95.4 |
| | N | 8.0 | 84.4 | | | 112.2 | 7.43 | 833.6 | 96.0 |
| | D | 8.3 | 84.9 | | | 117.4 | 7.49 | 879.3 | 94.4 |

| MES | PORCENTAJES DEL PROMEDIO MOVIL | | | | MEDIANA | INDICE ESTACIONAL E, EN % |
|---|---|---|---|---|---|---|
| | 1960 | 1961 | 1962 | 1963 | | |
| ENE | | 109.1 | 104.8 | 102.9 | 104.8 | 105.6 |
| FEB | | 105.5 | 101.6 | 101.4 | 101.6 | 102.4 |
| MAR | | 103.6 | 96.8 | 98.6 | 98.6 | 99.4 |
| ABR | | 94.7 | 95.3 | 90.0 | 94.7 | 95.5 |
| MAY | | 93.1 | 93.8 | 92.9 | 93.1 | 93.9 |
| JUN | | 89.8 | 90.8 | 93.0 | 90.8 | 91.5 |
| JUL | 83.7 | 86.4 | 90.8 | | 86.4 | 87.1 |
| AGO | 90.0 | 90.0 | 93.9 | | 90.0 | 90.7 |
| SEP | 94.1 | 96.7 | 101.5 | | 96.7 | 99.4 |
| OCT | 105.8 | 108.3 | 104.5 | | 105.8 | 106.7 |
| NOV | 111.3 | 113.1 | 109.0 | | 111.3 | 112.2 |
| DIC | 120.4 | 118.0 | 116.4 | | 116.4 | 117.4 |
| | | | | | 1190.2 | 1200.0 |

$$\frac{1200}{1190.2} = 1.0032$$


Y
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D E F M A M J J A S O N D
1960
1961
1962
1963

50"



PROBLEMA 15.- En la tabla siguiente se presentan los valores del producto nacional bruto, de 1950 a 1960.

Calcular:

a) Los índices simples, tomando como año base a 1950 y 1955.

b) Los índices eslabón

c) Los índices simples a partir de los índices eslabón, para los años base 1950 y 1955.

Fuente: Cámara Nacional de la Industria de la Construcción. Ponencia No. 70. "La demanda y la planeación de la Industria de la Construcción en México".
Bustamante J. y Escobedo R.

| AÑO | Valor del Producto Nacional Bruto |
|---|---|
| 1950 | 41 |
| 1951 | 52 |
| 1952 | 59 |
| 1953 | 58 |
| 1954 | 72 |
| 1955 | 87 |
| 1956 | 99 |
| 1957 | 114 |
| 1958 | 127 |
| 1959 | 136 |
| 1960 | 154 |

a) Indices simples con año base 1950: 

| | | | | |
|---|---|---|---|---|
| V 1950 | = | | = | 100 % |
| V 1951/1950 | = | 52/41 | = | 126.8 % |
| V 1952/1950 | = | 59/41 | = | 143.9 % |
| V 1953/1950 | = | 58/41 | = | 141.5 % |
| V 1954/1950 | = | 72/41 | = | 175.6 % |
| V 1955/1950 | = | 87/41 | = | 212.2 % |
| V 1956/1950 | = | 99/41 | = | 241.5 % |
| V 1957/1950 | = | 114/41 | = | 278.0 % |
| V 1958/1950 | = | 127/41 | = | 309.8 % |
| V 1959/1950 | = | 136/41 | = | 331.7 % |
| V 1960/1950 | = | 154/41 | = | 375.6 % |

Indices simples con año base 1955:

| | | | | |
|---|---|---|---|---|
| V 1950/1955 | = | 41/87 | = | 47.1 % |
| V 1951/1955 | = | 52/87 | = | 59.7 % |
| V 1952/1955 | = | 59/87 | = | 67.8 % |
| V 1953/1955 | = | 58/87 | = | 66.6 % |
| V 1954/1955 | = | 72/87 | = | 82.6 % |
| V 1955 | = | | = | 100.0 % |
| V 1956/1955 | = | 99/87 | = | 113.7 % |
| V 1957/1955 | = | 114/87 | = | 131.0 % |
| V 1958/1955 | = | 127/87 | = | 145.9 % |
| V 1959/1955 | = | 136/87 | = | 156.3 % |
| V 1960/1955 | = | 154/87 | = | 176.8 % |

b) Indices eslabón:

| | | | | |
|---|---|---|---|---|
| V 1951/1950 | = | 52/41 | = | 126.8 % |
| V 1952/1951 | = | 59/52 | = | 113.4 % |
| V 1953/1952 | = | 58/59 | = | 98.3 % |
| V 1954/1953 | = | 72/58 | = | 124.1 % |
| V 1955/1954 | = | 87/72 | = | 120.8 % |
| V 1956/1955 | = | 99/87 | = | 113.7 % |
| V 1957/1956 | = | 114/99 | = | 115.1 % |
| V 1958/1957 | = | 127/114 | = | 111.4 % |
| V 1959/1958 | = | 136/127 | = | 107.0 % |
| V 1960/1959 | = | 154/136 | = | 113.2 % |

c) Indices simples con año base 1950:

1950=1.00=100%

1951=1.00x1.268=126.8%

1952=1.00x1.268x1.134=143.9%

1953=1.00x1.268x1.134x0.983=141.5%

1954=1.00x1.268x1.134x0.983x1.241=175.6%

1955=1.00x1.268x1.134x0.983x1.241x1.208=212.2%

1956=1.00x1.268x1.134x0.983x1.241x1.208x1.137=241.5%

1957=1.00x1,268x1.134x0.983x1.241x1.208x1.137x1.151=278.0%

1958=1.00x1.268x1.134x0.983x1.241x1.208x1.137x1.151x1.113=309.8%

1959=1.00x1.268x1.134x0.983x1.241x1.208x1.137x1.151x1.113x1.07=331.7%

1960=1.00x1.268x1.134x0.983x1.241x1.208x1.137x1.151x1.113x1.07x1.132=375.6%

Indices simples con año base 1955:

1950 = 1.00 X 1.208 X 1.241 X 98.3 X 1.134 X 1.268 = 47.1 %

1951 = 1.00 X 1.208 X 1.241 X 98.3 X 1.134 = 59.7 %

1952 = 1.00 X 1.208 X 1.241 X 98.3 = 67.8 %

1953 = 1.00 X 1.208 X 1.241 = 66.6 %

1954 = 1.00 X 1.208 = 82.6 %

1955 = 1.00 = 100 %

1956 = 1.00 x 1.137 = 1.137 %

1957 = 1.00 x 1.137 x 1.151 = 131.0 %

1958 = 1.00 x 1.137 x 1.151 x 1.113 = 145.9 %

1959 = 1.00 x 1.137 x 1.151 x 1.113 x 1.07 = 156.3 %

1960 = 1.00 x 1.137 x 1.151 x 1.113 x 1.07 x 1.132 = 176.8 %

$\overline{x'} = \frac{8}{22} = 0.3636;\ \overline{Y'} = \frac{3}{22} = 0.1364;\ \overline{x'^2} = \frac{42}{22} = 1.9091,\ \overline{y'^2} = \frac{63}{22} = 2.8636$

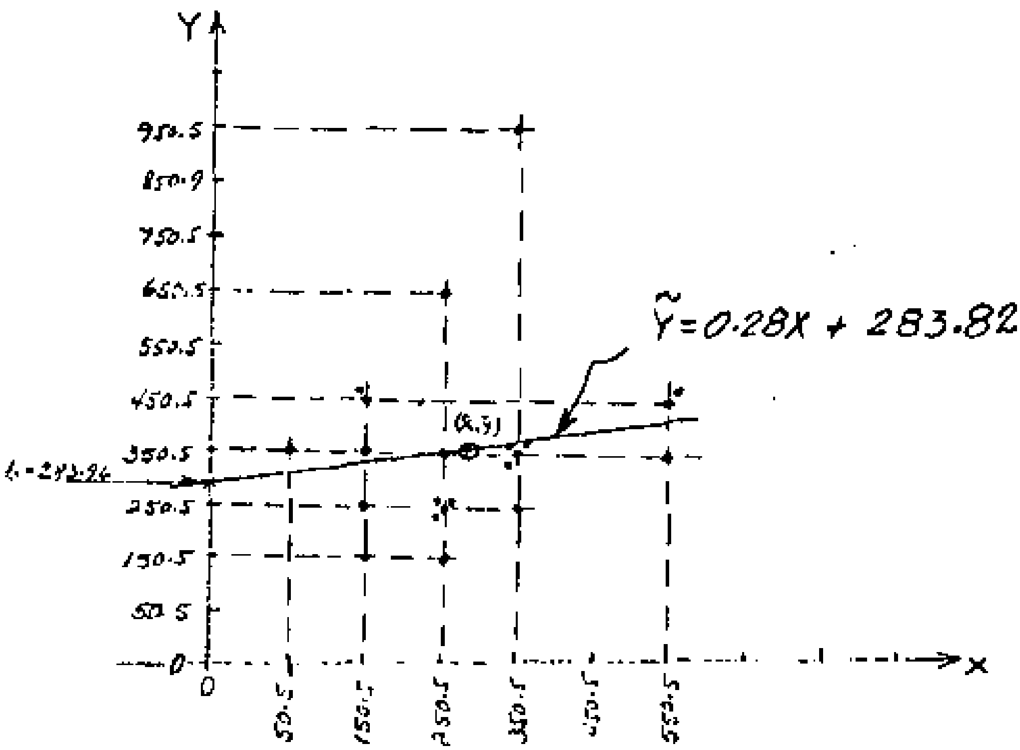
$S^2(x') = 1.9091-(0.3636)^2 = 1.7769;\ S^2(y') = 2.8636-(0.1364)^2 = 2.8450$

$\tilde{x} = C_2\overline{x'} + C_1 = 36.36 + 250.5 = 286.86$

$\tilde{y} = C_4\overline{Y'} + C_3 = 13.64 + 350.5 = 364.14$

$$m = \frac{100}{100}\ \frac{\frac{1}{22}\,12 - (0.3636)(0.1364)}{1.7769} = \frac{0.4959}{1.7769} = 0.28$$

$b = \tilde{y} - m\tilde{x} = 364.14 - 0.28 \times 286.86 = 283.82$




TAREA. CALCULAR LA RECTA DE REGRESION, Y TRAZAR LA GRAFICA CORRESPONDIENTE, DE LOS DATOS AGRUPADOS DE LA TABLA 3.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

# PROBABILIDAD Y ESTADISTICA FUNDAMENTOS Y APLICACIONES

## INFERENCIA ESTADISTICA

M.en I. AUGUSTO VILLARREAL ARANDA

Marzo, 1980.

Palacio de Minería   Calle de Tacuba 5,   primer piso   México 1, D. F.   Tels: 521-40-20   5 Líneas

# INFERENCIA ESTADISTICA

Augusto Villarreal Aranda

DIVISION DE ESTUDIOS DE POSGRADO
FACULTAD DE INGENIERIA

Noviembre , 1979

# 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 varianza, 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. Distribucion 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 $\mu$ y $\sigma$, 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 $\sigma^2$ 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

$$P[496 \leq X \leq 500] = P[-2.22 \leq Z \leq -0.74] =$$

$$= P[-2.22 \leq Z \leq 0] - P[-0.74 \leq Z \leq 0]$$

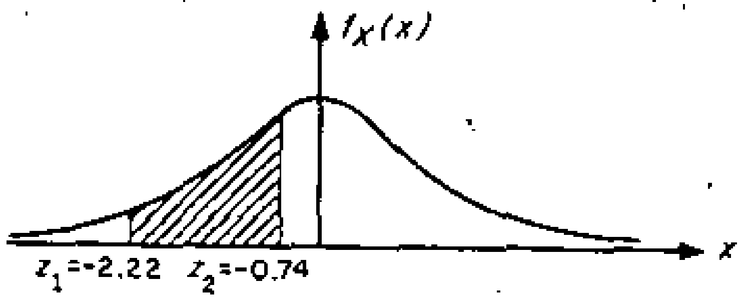



*Fig 4.1 Distribución normal correspondiente al ejemplo*

Recurriendo a la tabla de áreas bajo la curva normal estándar entre 0 y $Z$ queda finalmente .

$$P[496 \leq X \leq 500] = 0.4868 - 0.2704 = 0.2164$$

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

$$P[X \geq 510] = P[Z \geq 2.96] = P[Z \geq 0] - P[0 \leq Z \leq 2.96] =$$

$$= 0.5 - 0.4985 = 0.0015$$

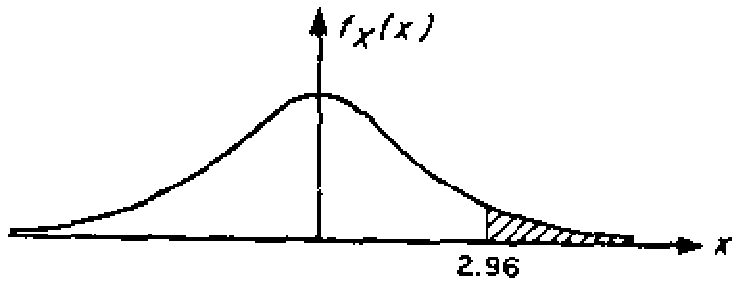



*Fig 4.2 Distribución normal correspondiente al ejemplo*

## 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 $\mu_X$ y $\mu_Y$ y desviaciones estándar $\sigma_X$ y $\sigma_Y$, tiene los siguientes parámetros:

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2} = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

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

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X}\,\frac{N_X - n_X}{N_X - 1} + \frac{\sigma^2_X}{n_Y}\,\frac{N_Y - n_Y}{N_Y - 1}}$$

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

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} \quad ; \quad \sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}$$

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 fabrica 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

$$P\left[\bar{X}_A > \bar{X}_B + 0.35\right] = P\left[Z > 3\right] = 0.500 - 0.4987 = 0.0013$$

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

$$P\left[\bar{X}_A > \bar{X}_B + 0.10\right] = P\left[Z > -2\right] = 0.5 + 0.4772 = 0.9772$$

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 $\mu_S$, y el parámetro correspondiente de la población es $\theta$, se dice que $S$ es un estimador insesgado de $\theta$ si

$$\mu_S = \theta$$

Por otra parte, si la estadística $S_n$ de la muestra tiende a ser igual al parámetro $\theta$ 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 média, 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.

## 8. 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 $\theta$, y sea $\sigma_S$ la desviación estándar (conocida o estimada) de su distribución muestral. La probabilidad, $1 - \alpha$, de que el valor de $\theta$ 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

$$P[S - z_c \sigma_S \leq \theta \leq S + z_c \sigma_S] = 1 - \alpha$$

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 $\theta$, $(S - z_c \sigma_S, S + z_c \sigma_S)$, 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 $\alpha$ se obtiene de la tabla de áreas bajo la curva normal o de la tabla 8.1 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 $\mu$ 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 $\mu$ 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.

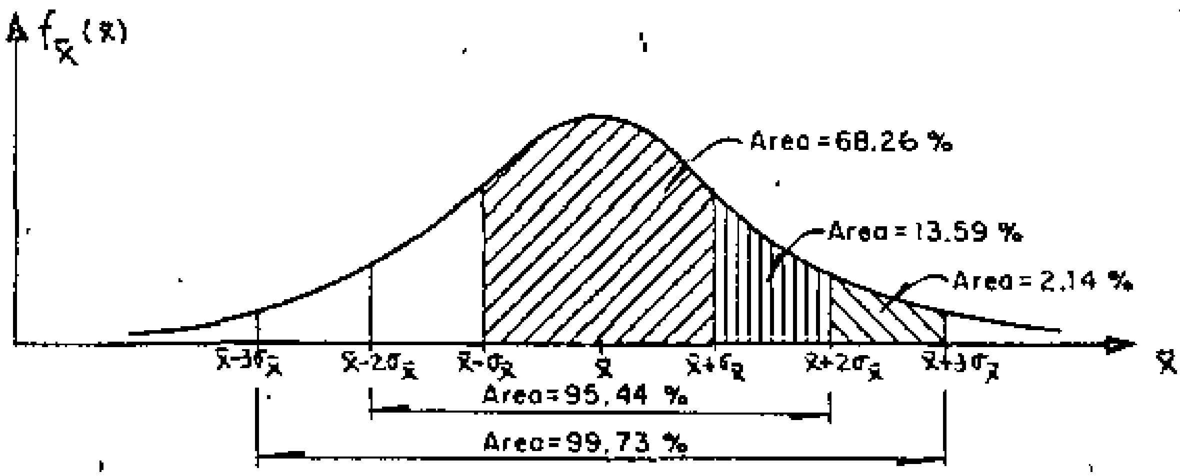



Fig 8.1

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, $\mu$, 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 $\sigma$ 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 $\mu_X$ 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 $72 \pm 1$ puntos.

a. Si se estima a $\sigma$ 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

$$(69.296,\quad 74.704)$$

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

$$\frac{384.16}{n} \quad \frac{1018 - n}{1017} < 4$$

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 $72 \pm 1$ 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.

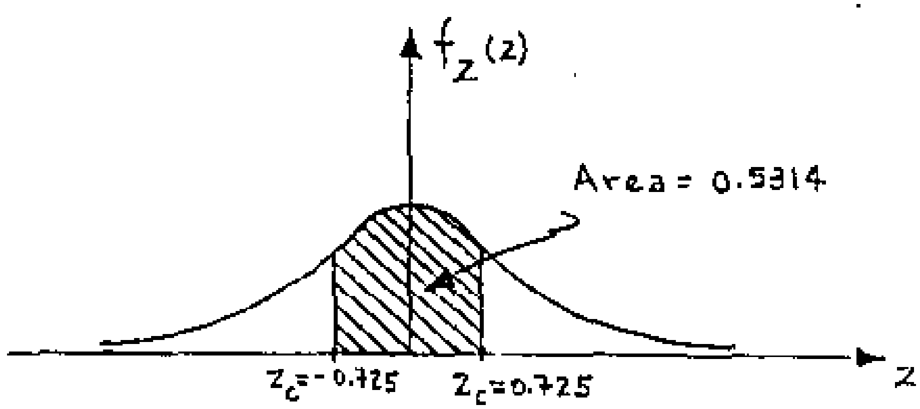



Fig 9.1

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 remplazo 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 $\sigma_X$ y $\sigma_Y$ 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 $\sigma_X$ y $\sigma_Y$ con $S_X$ y $S_Y$, respectivamente

$$1400 - 1200 \pm 1.96 \sqrt{\frac{(120)^2}{150} \quad \frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200} \quad \frac{5000 - 200}{5000 - 1}}$$

$$200 \pm 1.96\ (11.04)$$

$$200 \pm 21.638$$

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*, $\alpha$, 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, $1 - \alpha$, 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, $\mu_S$, de la distribución muestral de la estadística S es $\mu_1$, en contra de la hipótesis alternativa que establece que $\mu_S = \mu_2$, donde $\mu_2 > \mu_1$, es decir

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S = \mu_2$$

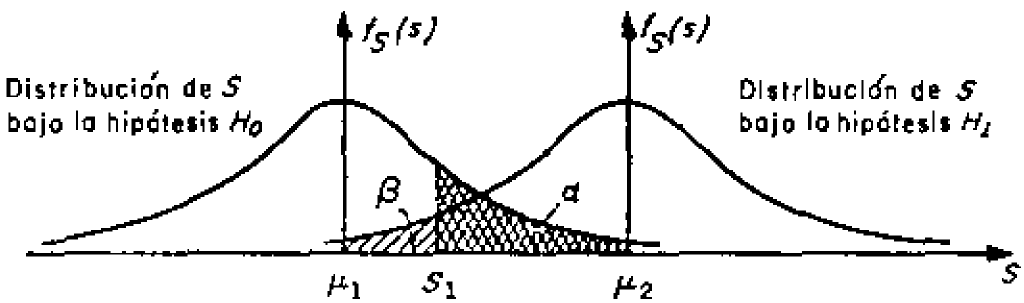
En la fig 13.1 se muestra en forma gráfica la relación entre los errores tipos I y II en el caso en el que la regla de decisión para aceptar o rechazar $H_0$ es la siguiente:

> *Si el valor de la estadística S obtenido de una muestra excede de cierto valor crítico $S_1$, rechácese $H_0$; en caso contrario, acéptese.*

Es evidente que si $H_0$ es verdadera, entonces $\alpha$ (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 $\beta$ (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 $\alpha$, pero se incrementa la $\beta$; 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.

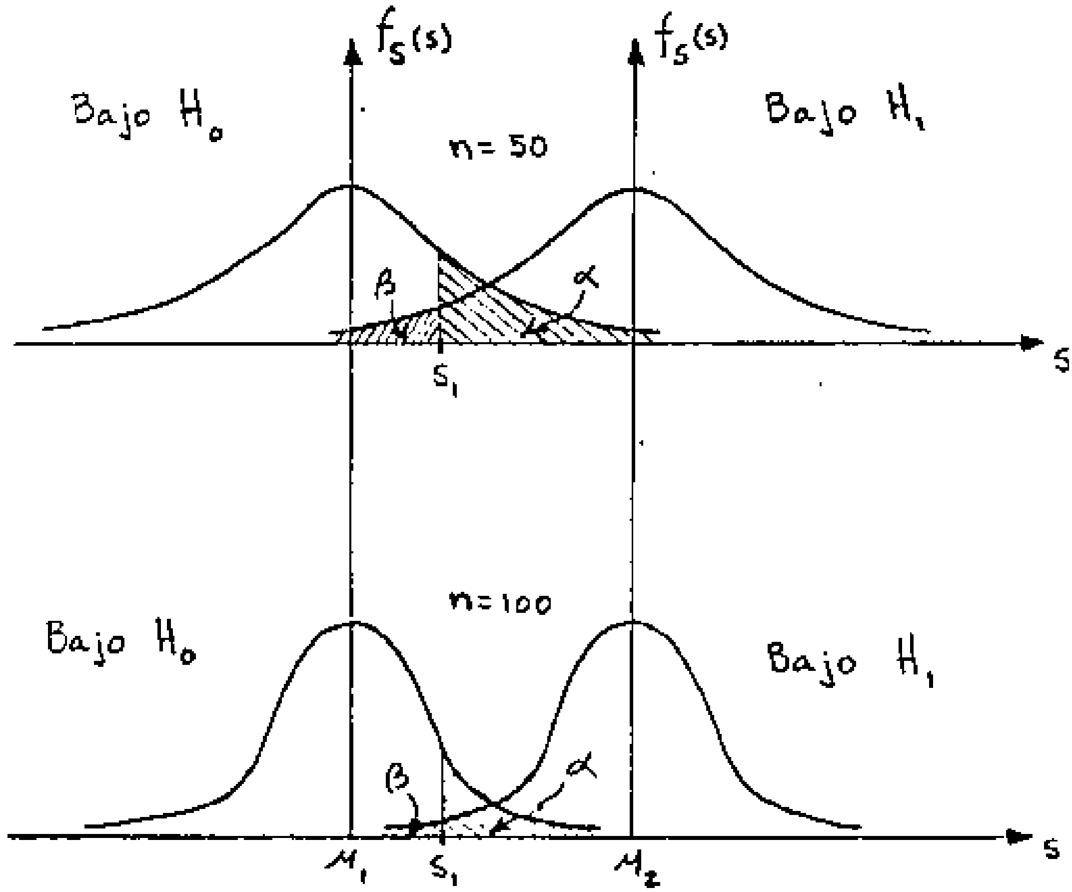



Fig 13.2

Sin embargo, esta técnica de reducción simultánea de ámbos tipos de errores no siempre puede ponerse en práctica, debido a razones de costo, tiempo, etc.

14. Regiones críticas, de rechazo o de significancia. Regiones de aceptación.

Cuando una hipótesis nula no se acepta se dice que se *rechaza a un nivel de significancia del $\alpha$ por ciento*, o que el valor estandarizado de la estadística involucrada es *significativo a un nivel de significancia $\alpha$*.

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 $\sigma_S$, que la variable $Z$ resulta de estandarizar a $S$, que la hipótesis nula, $H_0$, es que la media de $S$ vale $\mu_S$, y que la hipótesis alternativa $H_1$ es que dicha media es diferente de $\mu_S$, 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

$$-2.58 \leq Z \leq 2.58$$

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 de 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 $-2.58 \leq Z \leq 2.58$, y la de rechazo es $Z > 2.58$ y $Z < -2.58$.

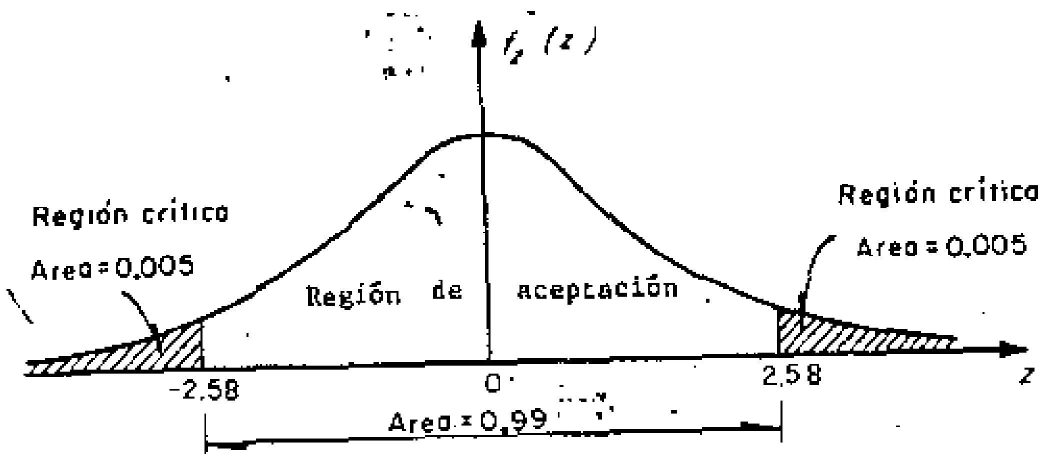



*Fig 14.1 Región de significancia*

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, $\alpha$ | 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 $\neq$ (diferente de), como en el siguiente caso

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S \neq \mu_1$$

en donde $\mu_S$ es la media de la estadística $S$, y $\mu_1$ es un valor fijo.

En los casos

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S < \mu_1$$

y

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S > \mu_1$$

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 $\sigma$ 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 $\mu$ y $\sigma$ 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 $\alpha$ seleccionado. En cualquiera de los casos anteriores, el valor o valores críticos se pueden obtener de la tabla 14.1, para valores comunes de $\alpha$.

*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 $\mu$ 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 $\mu$ y $\sigma$ 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 $\sigma$, 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 des-

viación 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 $\sigma$ de la población mediante $S_X$ de la muestra.

Considerando a la población infinita, y suponiendo como verdadera a $H_0$, se tiene que

$$\mu_{\bar{X}} \doteq \mu = 905 \text{ kg}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{50}} = 5.94$$

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 14.1); 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 $\mu_X$ y $\mu_Y$, y desviaciones estándar $\sigma_X$ y $\sigma_Y$. 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

$$Z = \frac{X - Y - \mu_{\bar{X}-\bar{Y}}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{X - Y - 0}{\sigma_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}}$$

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 $\mu_A$ y $\mu_B$ 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

$$H_0 : \mu_X = \mu_Y$$

$$H_1 \quad \mu_X > \mu_Y$$

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.

## 3.4 Muestras pequeñas

Como ya se indicó, para muestras grandes ($n \geq 30$) 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 ($\chi^2$)

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 estras aleatorias extraídas de poblaciones normales. Puesto que $S_X$ no puede ser negat..a, es de esperarse que su distribución muestral no sea una curva normal, ya que esta

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 $\sigma_X$ y si para cada muestra se calcula el valor de la estadística,

$$\chi^2 = \frac{n\ S_X^2}{\sigma^2} \tag{3.14}$$

donde $S_X^2$ es la variancia de la muestra.

El número de grados de libertad, $\nu$, 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 $\chi^2$ 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 $\nu$.

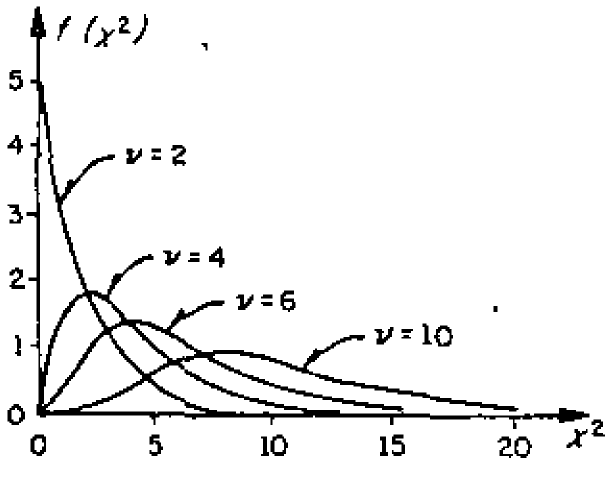
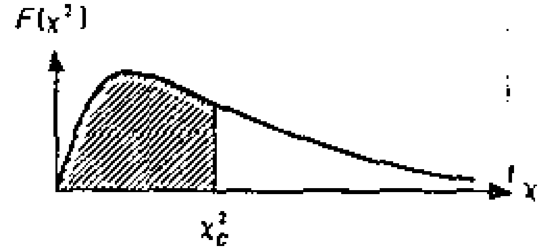



*Fig 21. Distribución Ji cuadrada para distintos valores de* $\nu$




## TABLA 8. VALORES CRITICOS $\chi^2_c$

| $\nu$ | $\chi^2_{.995}$ | $\chi^2_{.99}$ | $\chi^2_{.975}$ | $\chi^2_{.95}$ | $\chi^2_{.90}$ | $\chi^2_{.75}$ | $\chi^2_{.50}$ | $\chi^2_{.25}$ | $\chi^2_{.10}$ | $\chi^2_{.05}$ | $\chi^2_{.025}$ | $\chi^2_{.01}$ | $\chi^2_{.005}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | 7.76 | 5.39 | 3.36 | 1.92 | 1.06 | .711 | .483 | .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 | 23.2 | 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 | 11.73 | 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 | 90.12 | 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 variancia

Tal como se hizo para la distribución normal, se pueden establecer intervalos de confianza para la variancia de la población en términos de la variancia de una muestra extraída de ella, a un nivel de confianza dado $1 - \alpha$, si se hace uso de los valores críticos $\chi_c^2$ de la tabla 8. Por lo tanto, un intervalo de confianza para la estadística $\chi^2$, estaría dado por

$$\chi_{c'}^2 < \frac{n S_X^2}{\sigma^2} < \chi_c^2$$

donde $\chi_{c'}^2$ y $\chi_c^2$ 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 $\sigma^2$ a un nivel de confianza $1 - \alpha$.

3.4.1.2 Prueba de hipótesis para la variancia

La prueba de hipótesis para la variancia de una población normal se efectúa calculando el valor de la estadística $\chi^2$ 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 variancia 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 variancia 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 $\chi^2$ 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 $\chi^2$ fuera mayor que el de $\chi^2$ 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 $\chi^2$ 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

$$F = \frac{S_X^2}{S_Y^2} \tag{3.15}$$

# TABLA 9. VALORES $F_c$ PARA $\alpha = 0.01$

| $r_2$ = Grados de libertad del denominador | $r_1$ = 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 | 15.50 | 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 | 9.77 | 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 | 6.87 |
| 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 | 4.83 |
| 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 | 9.66 | 7.22 | 6.22 | 5.68 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 | 4.40 | 4.25 | 4.10 | 4.03 | 3.93 | 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 | 2.75 |
| 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 | 3.08 | 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 | 8.03 | 5.79 | 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 | 4.83 | 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 | 4.14 | 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 | 6.65 | 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, $\nu_X$ y $\nu_Y$, 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 $\nu_X$ y $\nu_Y$ y un nivel de significancia de 0.01. Cuando los grados de libertad $\nu_X$ o $\nu_Y$ 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).

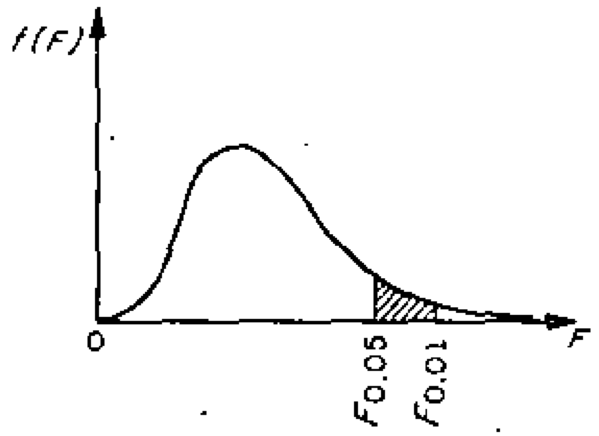



*Fig 22. Distribución F.*

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-

lan 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 prensadora 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 $\mu$ y variancia desconocida, para cada muestra se puede calcular la estadística $T$ definida mediante la fórmula

$$T = \frac{\overline{X} - \mu}{S_X}\sqrt{n - 1} \qquad (3.16)$$

donde $X$ es el promedio y $S_X$ la desviación estándar de la muestra.

La distribución muestral de $T$ (fig 23) está dada por la ecuación

$$f(t) = \frac{U}{(1 + \frac{t^2}{\nu})^{(\nu + 1)/2}}$$

Ojo: 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.

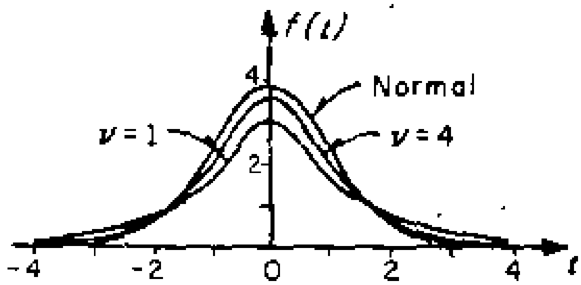



*Fig 23. Distribución t de Student para distintos valores de ν*

En la fig 23 se aprecia que conforme $\nu$ (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, $\mu$, 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,

$$-t_c < \frac{\overline{X} - \mu}{S_X} \sqrt{n-1} < t_c$$

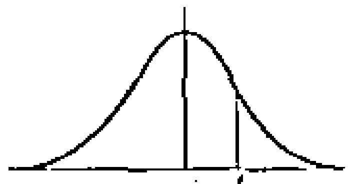
representa un intervalo de confianza para $t$, a partir del cual se puede estimar que $\mu$ se encuentra dentro del intervalo

$$\overline{X} - t_c \frac{\sigma_X}{\sqrt{n-1}} < \mu < \overline{X} + t_c \frac{\sigma_X}{\sqrt{n-1}}$$

En términos generales, los límites de confianza para la media de la población se representan como

$$\overline{X} \pm t_c \frac{\sigma_X}{\sqrt{n-1}}$$

## TABLA 10. VALORES $t_c$ PARA LA DISTRIBUCION $t$ DE STUDENT




| $\nu$ | $t_{.995}$ | $t_{.99}$ | $t_{.975}$ | $t_{.95}$ | $t_{.90}$ | $t_{.80}$ | $t_{.75}$ | $t_{.70}$ | $t_{.60}$ | $t_{.55}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 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 | .136 |
| 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 | 2.06 | 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 |
| $\infty$ | 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

$$T = \frac{\bar{X} - \bar{Y}}{\epsilon \sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \qquad (3.17)$$

donde

$$\epsilon = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \qquad (3.18)$$

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 $\sigma_X^2$ y $\sigma_Y^2$ 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

$$c = \sqrt{\frac{n_X\ S_X^2 + n_Y\ S_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{12(0.40)^2 + 12(0.36)^2}{12 + 12 - 2}} = 0.397$$

por lo cual

$$t = \frac{5.3 - 5.0}{0.397\sqrt{\frac{1}{12} + \frac{1}{12}}} = 1.85$$

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.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

PROBABILIDAD Y ESTADISTICA: FUNDAMENTOS Y APLICACIONES

EJEMPLO Y EJERCICIOS

DR. OCTAVIO RASCON CHAVEZ

ABRIL 1980

EJEMPLO

CON OBJETO DE VERIFICAR LA CONSISTENCIA INTERNA DE UNA PRUEBA PSICOLOGICA, ESTA SE APLICO DOS VECES A CADA UNA DE DOS MUESTRAS ALEATORIAS. ESTAS MUESTRAS SE EXTRAJERON DE NIÑOS DEL CUARTO GRADO DE DOS ESCUELAS DISTINTAS, "A" y "B". LAS CALIFICACIONES DE LA PRIMERA APLICACION CORRESPONDEN A LA VARIABLE X; LAS DE LA SEGUNDA APLICACION (15 DIAS DESPUES DE LA PRIMERA), CORRESPONDEN A LA VARIABLE Y.

a. CALCULAR EL COEFICIENTE DE CORRELACION DE X y Y PARA CADA ESCUELA, Y PARA LAS DOS ESCUELAS JUNTAS, Y PROBAR LA HIPOTESIS DE QUE $\rho_{XY}>0$ EN CADA CASO.

b. PROBAR LA HIPOTESIS DE QUE $\mu_X = \mu_Y$ PARA AMBAS ESCUELAS JUNTAS, Y PARA CADA ESCUELA POR SEPARADO.

c. PROBAR LA HIPOTESIS DE QUE

1. $\mu_{X_A} = \mu_{X_B}$

2. $\mu_{Y_A} = \mu_{Y_B}$

3. $\sigma^2(X_A) = \sigma^2(X_B)$

4. $\sigma^2(Y_A) = \sigma^2(Y_B)$

FORMULAS

$\bar{X} = \Sigma X_i/n$, $\bar{Y} = \Sigma Y_i/n$, $S^2(X) = \Sigma X^2/n - \bar{X}^2$, $S^2(Y) = \Sigma Y_i^2/n - \bar{Y}^2$,

$S^2(d) = \Sigma d_i^2/n - \bar{d}^2$, $t_d = \dfrac{(\bar{X}-\bar{Y})\sqrt{n-1}}{S_d}$, $t_\rho = \rho_{XY}\sqrt{\dfrac{n-2}{1-\rho^2}}$

$$t_{\bar{X}-\bar{Y}} = \frac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{n_X S^2(X) + n_Y S^2(Y)}{n_X+n_Y-2}\left(\dfrac{1}{n_X}+\dfrac{1}{n_Y}\right)}}, \quad F = \frac{S_M}{S_m}$$

DONDE $S_M^2$ y $S_m^2$ SON ESTIMACIONES INSESGADAS DE LAS VARIANCIAS MAYOR Y MENOR, RESPECTIVAMENTE, DE LAS DOS QUE SE ESTAN COMPARANDO.

RESPUESTAS A LOS INCISOS a y b

ESCUELA A

| X | Y | $X^2$ | $Y^2$ | XY | d=x-y | $d^2$ |
|---|---|---|---|---|---|---|
| 34 | 35 | 1156 | 1225 | 1190 | -1 | 1 |
| 39 | 36 | 1521 | 1296 | 1404 | 3 | 9 |
| 40 | 40 | 1600 | 1600 | 1600 | 0 | 0 |
| 35 | 38 | 1225 | 1444 | 1330 | -3 | 9 |
| 30 | 29 | 900 | 841 | 870 | 1 | 1 |
| 28 | 26 | 784 | 676 | 728 | 2 | 4 |
| 33 | 34 | 1089 | 1156 | 1122 | -1 | 1 |
| 38 | 40 | 1444 | 1600 | 1520 | -2 | 4 |
| 32 | 39 | 1024 | 1521 | 1248 | -7 | 49 |
| 37 | 35 | 1369 | 1225 | 1295 | 2 | 4 |
| 26 | 26 | 676 | 676 | 676 | 0 | 0 |
| 40 | 39 | 1600 | 1521 | 1560 | 1 | 1 |
| 32 | 30 | 1024 | 900 | 960 | 2 | 4 |
| 33 | 34 | 1089 | 1156 | 1122 | -1 | 1 |
| 38 | 33 | 1444 | 1089 | 1254 | 5 | 25 |
| 34 | 39 | 1156 | 1521 | 1326 | -5 | 25 |
| 35 | 37 | 1225 | 1369 | 1295 | -2 | 4 |
| 584 | 590 | 20326 | 20816 | 20500 | -6 | 142 |

$\bar{X} = \frac{584}{17} = 34.352941$ ; $\bar{X}^2 = 1180.1245$

$\bar{Y} = \frac{590}{17} = 34.705882$ ; $\bar{Y}^2 = 1204.4982$

$\bar{d} = -6/17 = -0.3529411$ ; $\bar{d}^2 = 0.1245674$

$S^2(X) = \frac{20326}{17} - 1180.1245 = 15.5225$ ; $S(X) = 3.9398604$

$S^2(Y) = \frac{20816}{17} - 1204.4982 = 19.9723$ ; $S(Y) = 4.4690379$

$S_d{}^2 = \frac{142}{17} - 0.1245674 = 8.2283737$ ; $S_d = 2.8685141$

$$\rho_{xy} = \frac{(20500/17)-(34.352941)(34.705882)}{(3.9398604)(4.4690379)} = \underline{\underline{0.7742887}}$$

$$t_\rho = 0.774 \sqrt{\frac{17-2}{1-0.744^2}} = 0.774 \times 6.116 = 4.73 > 2.13$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE QUE $\rho_{xy}=0$ CON UN NIVEL DE CONFIANZA DEL 95%.

$$t_d = \left| \frac{(34.353-34.706)\sqrt{16}}{2.869} \right| = 0.492 < 2.12$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_X = \mu_Y$ CON UN 95% DE NIVEL DE CONFIANZA.

ESCUELA B

| X | Y | $X^2$ | $Y^2$ | XY | d=X-Y | $d^2$ |
|---|---|---|---|---|---|---|
| 39 | 41 | 1521 | 1681 | 1599 | -2 | 4 |
| 27 | 36 | 729 | 1296 | 972 | -9 | 81 |
| 33 | 31 | 1089 | 961 | 1023 | 2 | 4 |
| 37 | 36 | 1369 | 1296 | 1332 | 1 | 4 |
| 35 | 36 | 1225 | 1296 | 1260 | -1 | 1 |
| 31 | 33 | 961 | 1089 | 1023 | -2 | 4 |
| 33 | 32 | 1089 | 1024 | 1056 | 1 | 1 |
| 39 | 40 | 1521 | 1600 | 1560 | -1 | 1 |
| 39 | 35 | 1521 | 1225 | 1365 | 4 | 16 |
| 27 | 29 | 891 | 841 | 783 | -2 | 4 |
| 32 | 36 | 1024 | 1296 | 1152 | -4 | 16 |
| 34 | 35 | 1156 | 1225 | 1190 | -1 | 1 |
| 35 | 34 | 1225 | 1156 | 1190 | 1 | 1 |
| 36 | 42 | 1296 | 1764 | 1512 | -6 | 36 |
| 34 | 34 | 1156 | 1156 | 1156 | 0 | 0 |
| 29 | 31 | 841 | 961 | 899 | -2 | 4 |
| 540 | 561 | 18614 | 19867 | 19072 | -21 | 175 |

$\bar{X} = \frac{545}{16} = 33.75$ ; $\bar{X}^2 = 1139.0625$ ; $\bar{d} = -1.3125$ ; $\bar{d}^2 = 1.7226562$

$\bar{Y} = \frac{561}{16} = 35.0625$ ; $\bar{Y}^2 = 1229.3789$

$S_X^2(X) = \frac{18614}{16} - 1139.0625 = 24.3125$ ; $S(X) = 4.9307707$

$S^2(Y) = \frac{19867}{16} - 1229.3789 = 12.3086$ ; $S(Y) = 3.5083614$

$S_d^2 = \frac{175}{16} - 1.7226562 = 9.214844$ $\qquad S_d = 3.0355961$

$$\rho_{xy} = \frac{(19072/16)-(33.75)(35.0625)}{(4.9307707)(3.5083614)} = \underline{0.4994934}$$

$$t_\rho = 0.499\sqrt{\frac{14}{0.751}} = 2.154 < 2.15$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE QUE $\rho_{xy}=0$, CON UN 95% DE NIVEL DE CONFIANZA.

$$t_d = \left|\frac{(33.75-35.063)\sqrt{15}}{3.036}\right| = 1.67 < 2.13$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_X = \mu_Y$, CON UN 95% DE NIVEL DE CONFIANZA.

AMBAS ESCUELAS JUNTAS

$\Sigma x_i = 1124$, $\Sigma y_i = 1151$, $\Sigma x_i^2 = 38940$, $\Sigma y_i^2 = 39572$, $\Sigma d_i = -27$, $\Sigma d_i^2 = 317$

$\bar{X} = \frac{1124}{33} = 34.060606$ ; $\bar{X}^2 = 1160.1248$ ; $\bar{d} = \frac{-27}{33} = -0.8181818$ ; $\bar{d}^2 = 0.669421$

$\bar{Y} = \frac{1151}{33} = 34.878787$ ; $\bar{Y}^2 = 1216.5297$

$S^2(X) = \frac{38940}{33} - 1160.1248 = 19.8752$ ; $S(X) = 4.458161$

$S^2(Y) = \frac{40683}{33} - 1216.5297 = 16.2884$ ; $S(Y) = 4.0358889$

$S_d^2 = \frac{317}{33} - 0.6694214 = 8.9366392$ ; $S_d = 2.9894212$

$$\rho_{xy} = \frac{(39572/33)-(34.060606)(34.878787)}{(4.458161)(4.0358889)} = \underline{0.6201924}$$

$$t_d = \left| \frac{(34.061 - 34.879)\sqrt{32}}{2.989} \right| = 1.548 < 2.04$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_X = \mu_Y$, CON UN 95% DE NIVEL DE CONFIANZA.

$$t_\rho = 0.620\sqrt{\frac{31}{0.616}} = 4.398 > 2.04$$

POR LO QUE SE RECHAZA LA HIPOTESIS DE QUE $\rho_{xy}=0$, CON UN NIVEL DE CONFIANZA DEL 95%.

RESPUESTAS AL INCISO c

$$t_{\bar{X}_A-\bar{X}_B} = \frac{34.35-33.75}{\sqrt{\frac{17x15.52+16x24.31}{31}\left(\frac{1}{17}+\frac{1}{15}\right)}} = \frac{0.6}{\sqrt{\frac{263.84+388.96}{31}(0.126)}} =$$

$$= 0.368 < 2.04$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_{X_A} = \mu_{X_B}$, CON UN 95% DE NIVEL DE CONFIANZA.

$$t_{\bar{Y}_A-\bar{Y}_B} = \frac{|34.71-35.06|}{\sqrt{\frac{17x19.97+16x12.31}{31}(0.126)}} = \frac{|-0.35|}{\sqrt{\frac{339.49+196.96}{31}(0.126)}}$$

$$= 0.24 < 2.04$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\mu_{Y_A} = \mu_{Y_B}$, CON UN 95% DE NIVEL DE CONFIANZA.

PARA LA PRUEBA DE IGUALDAD DE VARIANCIAS USAREMOS

$$F = \frac{24.31\sqrt{\frac{16}{15}}}{15.52\sqrt{\frac{17}{16}}} = 1.57 < 3.41 = F_{0.01}(15,16)$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\sigma^2_{X_A} = \sigma^2_{X_B}$ CON UN 98% DE NIVEL DE CONFIANZA.

$$F = \frac{19.97\sqrt{\frac{17}{16}}}{12.31\sqrt{\frac{16}{15}}} = \frac{20.58}{12.71} = 1.62 < 3.41$$

POR LO QUE SE ACEPTA LA HIPOTESIS DE QUE $\sigma^2_{Y_A} = \sigma^2_{Y_B}$ CON UN 98% DE NIVEL DE CONFIANZA.

EJEMPLO

EN UN ESTUDIO CON FINES ANTROPOLOGICOS SE OBTUVO UNA MUESTRA ALEATORIA DEL TAMAÑO DE LA CABEZA DE LOS INDIGENAS ORIGINARIOS DE CIERTA REGION TROPICAL. LOS DATOS AGRUPADOS SE PRESENTAN EN LA SIGUIENTE TABLA. PROBAR LA HIPOTESIS DE QUE ESTOS DATOS CORRESPONDEN A UNA VARIABLE CON DISTRIBUCION NORMAL.

| INTERVALO DE VALORES, mm | FRECUENCIA OBSERVADA, $f_i$ | FRECUENCIA ESPERADA, $e_i$ | $f_i-e_i$ | $(f_i-e_i)^2$ | $\frac{(f_i-e_i)^2}{e_i}$ |
|---|---|---|---|---|---|
| 171.5-175.5 | 3 | 2.4 | 0.6 | 0.36 | 0.15 |
| 175.5-179.5 | 9 | 10.5 | -1.5 | 1.25 | 0.12 |
| 179.5-183.5 | 29 | 33.1 | -4.1 | 16.81 | 0.51 |
| 183.5-187.5 | 76 | 71.3 | 4.7 | 22.09 | 0.31 |
| 187.5-191.5 | 104 | 104.2 | -0.2 | 0.04 | 0.00 |
| 191.5-195.5 | 110 | 108.8 | 1.8 | 3.24 | 0.03 |
| 195.5-199.5 | 88 | 77.3 | 10.7 | 114.49 | 1.48 |
| 199.5-203.5 | 30 | 37.5 | -7.5 | 56.25 | 1.50 |
| 203.5-207.5 | 6 | 13.0 | -7.0 | 49.00 | 3.77 |
| 07.5-211.5 | 4 | 3.0 | 1.0 | 1.00 | 0.33 |
| 211.5-215.5 | 2 | 0.5082 | 1.4918 | 2.23 | 4.58 |
| 215.5-219.5 | 1 | 0.0462 | 0.9538 | 0.910 | 19.69 |
| | | | | TOTAL: | 32.27 |

$$\chi^2 = 32.27 < 124 = \chi^2_{0.95}$$

POR LO QUE LA HIPOTESIS NULA NO PUEDE RECHAZARSE CON UN 95% DE NIVEL DE CONFIANZA.

EJERCICIO

Se piensa que la emisión de partículas radioactivas de cierta fuente ocurre según una distribución de probabilidades de Poisson. El número de partículas emitidas en 100 intervalos consecutivos de 10 seg quedó distribuido de la siguiente manera

| Nº de partículas | 0 | 1 | 2 | 3 | 4 | ≥5 |
|---|---|---|---|---|---|---|
| Nº de intervalos (frecuencia) | 11 | 30 | 25 | 20 | 10 | 4 |

Probar la hipótesis de que efectivamente se trata de una distribución de Poisson. Usar $\alpha = 0.01$.

Solución.

Puesto que no nos indican un valor del parámetro de la distribución necesitamos estimarlo a partir de la información dada arriba:

$$\hat{\lambda} = \{(0\times 11)+(1\times 30)+(2\times 25)+(3\times 20)+(4\times 10)+(5\times 4)\}/100$$

$= 2.00$ partículas/intervalo

La distribución de Poisson es entonces:

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{2^x e^{-2}}{x!} = P(X=x)$$

$p_1 = f_X(0) = 2^0 e^{-2}/0! = 0.135$ ; $np_1 = 100\times 0.135 = 13.5$

$p_2 = f_X(1) = 2^1 e^{-2}/1! = 0.270$ ; $np_2 = 100\times 0.270 = 27.0$

$p_3 = f_X(2) = 2^2 e^{-2}/2! = 0.270$ ; $np_2 = 27.0$

$p_4 = f_X(3) = 2^3 e^{-2}/3! = 0.180$ ; $np_3 = 18.0$

$p_5 = f_X(4) = 2^4 e^{-2}/4! = 0.090$ ; $np_4 = 9.0$

$p_6 = P(X\geq 5) = 1 - F_X(5) + f_X(5) = 0.055$ ; $np_6 = 5.5$

$$\chi^2 = \frac{(11-13.5)^2}{13.5} + \frac{(30-27.0)^2}{27.0} + \frac{(25-27.0)^2}{27.0} + \frac{(20-18)^2}{18} + \frac{(10-9.0)^2}{9.0} + \frac{(4-5.5)^2}{5.5}$$

$= \underline{1.687}$. $\qquad \nu = 6-1-1 = 4$ ($\nu = N - r - 1$ ; $r$ = Nº de estimaciones hechas con los datos)

$\chi^2_{0.99,4} = 13.277 > 1.687$ ∴ se acepta la hipótesis nula

EJERCICIO

La duración de los transformadores producidos en una fábrica fue medida en una muestra de 50 elementos tomados al azar, obteniéndose la siguiente distribución de frecuencias:

| Intervalo N° | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Intervalo de tiempo | $0 \le t < 1$ | $1 \le t < 2$ | $2 \le t < 3$ | $t \ge 3$ |
| Frecuencia | 21 | 16 | 9 | 4 |

Probar la hipótesis de que la distribución de probabilidades de la variable aleatoria "duración de los transformadores" es exponencial con parámetro $\lambda = 0.45$ años$^{-1}$ usar $\alpha = 0.05$.

Solución

Las frecuencias esperadas son: $nP(x_1 \le X < x_2)$

donde $n$ = tamaño de la muestra

$$p_1 = P(0 \le X < 1) = \int_0^1 (0.45) e^{-0.45t}\, dt = 0.362 \; ; \; 50p_1 = 18.10$$

$$p_2 = P(1 \le X < 2) = \int_1^2 (0.45) e^{-0.45t}\, dt = 0.232 \; ; \; 50p_2 = 11.60$$

$$p_3 = P(2 \le X < 3) = \int_2^3 (0.45) e^{-0.45t}\, dt = 0.145 \; ; \; 50p_3 = 7.25$$

$$p_4 = P(X > 4) = \int_3^\infty (0.45) e^{-0.45t}\, dt = \underline{0.259} \; ; \; 50p_4 = \underline{12.95}$$

$$0.998 \doteq 1 \qquad 49.90 \doteq 50$$

$$\chi^2 = \frac{(21-18.10)^2}{18.10} + \frac{(16-11.6)^2}{11.6} + \frac{(9-7.25)^2}{7.25} + \frac{(4-12.95)^2}{12.95} = \underline{\underline{8.74}}$$

$\chi^2_{0.95,3} = 7.81 < 8.74$ por lo que se rechaza la hipótesis nula con un 5% de nivel de significancia

TAREA: Sacar una muestra de 50 números de una tabla de números aleatorios y probar la hipótesis de que provienen de una distribución uniforme de 0 a 1, previa reducción a decimales.

# PRUEBAS DE HIPOTESIS E INTERVALOS DE CONFIANZA EN REGRESION LINEAL

Si el modelo que relaciona a Y con X es lineal, entonces

$$Y = MX + B$$

Si no se conocen M y B, 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 varianza de la estimación de Y con base en X.

Se puede demostrar que, si se conoce $\sigma^2_{y|x}$, entonces:

$$Var(m) = \sigma^2_m = \sigma^2_{y|x} \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$Var(b) = \sigma^2_b = \sigma^2_{y|x}/n + \frac{\bar{x}^2 \sigma^2_{y|x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$Var(mX + b) = \sigma^2_{y|x}/n + \frac{\sigma^2_{y|x}(x_* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sigma_{\tilde{y}}$$

Si $\sigma^2_{y|x}$ no se conoce, se puede obtener una estimación insesgada de ella mediante la ecuación

$$\boxed{s^2_{y|x} = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2}$$

Intervalos de confianza: $\sigma_{y|x}$ CONOCIDA

a. Para la ordenada en el origen, B,

$$\boxed{b \pm z_c \sigma_b}$$

donde. $z_c = P(Z < z_c) = 1 - \alpha/2$ ; $\alpha$ = nivel de significancia

b. Para la pendiente, M:

$$\boxed{m \pm z_c \sigma_m}$$

c. Para la predicción, $Y_i$:

$$\boxed{\tilde{Y}_i \pm z_c \sigma_{\tilde{y}}}$$

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, $B$: $b \pm t_c \sigma_b$

$$b \pm t_c S_{y|x}\left\{\sqrt{\left(\frac{\overline{x^2}}{n s_x^2}\right)+\frac{1}{n}}\right\} \text{ o } b \pm t_c \frac{S_{y|x}}{[illegible]}\sqrt{\frac{\overline{x^2}}{n}\frac{1}{s_x}+\frac{1}{n}}$$

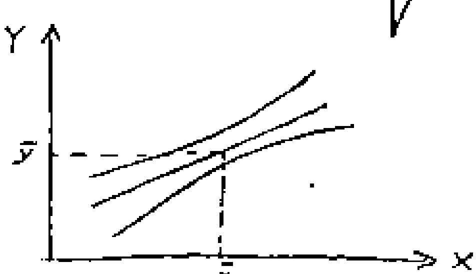
donde $t_c$ es el valor crítico de un nivel de significancia $\alpha$ correspondiente a una distribución $t$ de Student con $\nu = n-2$ grados de libertad, y $s_x^2$ es la variancia (sesgada) de la muestra de $x$.

b. Para la pendiente, $M$: $m \pm t_c \sigma_m$

$$m \pm t_c\, s_{y|x}/\sqrt{n s_x^2} \quad \text{o} \quad m \pm t_c \frac{S_{y|x}}{s_x\sqrt{n}}$$

c. Para la predicción, $Y_i$: $\tilde{Y}_i \pm t_c \sigma_{\tilde{y}}$

$$\tilde{Y}_i \pm t_c S_{y|x}\sqrt{\frac{1}{n}+\frac{(x_i-\bar{x})^2}{n s_x^2}}$$



EJEMPLO

La formación de alcohol en un proceso de fermentación 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 mínimos cuadrados se obtiene

$$\tilde{Y} = 0.225X + 13.01$$

$(\bar{x} = 47.5,\ \bar{y} = 23.7)$

1. Intervalos de confianza con $\sigma_{y/x} = 0.8$ (conocida). $\alpha = 0.05$.

$$\sigma_b = \frac{0.8}{\sqrt{6}} + \sqrt{\frac{0.8 \times 47.5^2}{437.5}} = 0.661$$

donde $\sum_{i=1}^{6} (x_i - \bar{x})^2 = 437.5$

$$b \pm z_c \sigma_b = 13.01 \pm 1.96 \times 0.661 = (11.71, 14.31)$$

$$\sigma_m = \frac{0.8}{\sqrt{437.5}} = \frac{0.8}{20.92} =$$

$$m \pm z_c \sigma_m = 0.225 \pm 1.96 \times 0.0382 = 0.225 \pm 0.075 =$$

$$= (0.150, 0.300)$$

2. Intervalos de confianza con $\sigma_{y/x}$ desconocida.

En este caso $S_{y/x} = \frac{1}{n-2}\sum_{i=1}^{6} (y_i - 0.225x_i - 13.01)^2 = \sum_{i=1}^{6} (y_i - \tilde{y}_i)^2/(n-$

<u>TAREA</u>: CALCULAR $S_{y/x}$ y los intervalos de confianza correspondientes a $B$ y $M$, con $\alpha = 0.95$ $(t_c = t_{0.975,4} = 2.776)$, y para $\tilde{Y}$ correspondiente a $X=50$.

TAREA: Hacer estimaciones de intervalos de confianza para $\alpha = 0.05$ y $\alpha = 0.01$ de b, m y $Y_i$, este último para un $X = x_i$ que seleccione cada quien. Utilizar uno de los problemas de regresión dejados como tarea anteriormente.

Pruebas de hipótesis

a. Para la ordenada en el origen

Se demuestra que
$$\frac{B - b_0}{S_{y/x}\sqrt{\frac{\overline{x^2}}{n S_x^2}}} = \frac{B - b_0}{\frac{S_{y/x}}{S_x}\sqrt{\frac{\overline{x^2}}{n}}} = T$$

tiene distribución t de Student con $D = n-2$ g. de l.

Si se desea probar la hipótesis ~~de~~

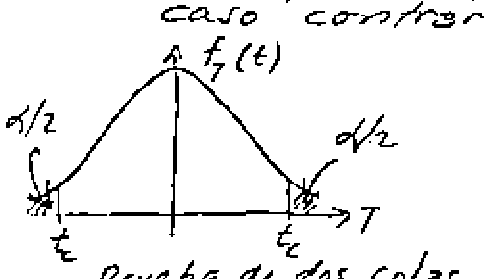
$H_0: B = b_0$

$H_1: B \neq b_0$

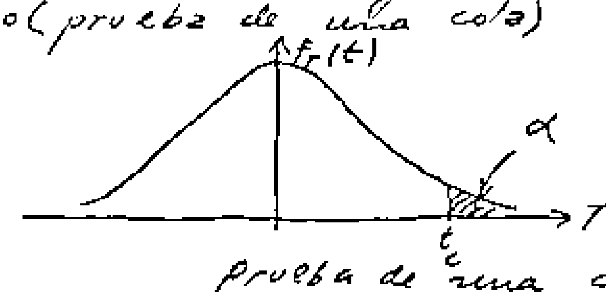
basta sustituir a $B = b$ en la ecuación anterior y evaluar $T = t$, es decir,

$$t = \frac{b - b_0}{\frac{S_{y/x}}{S_x}\sqrt{\frac{\overline{x^2}}{n}}}$$

Se aceptará $H_0$ si $|t| < |t_c|$; en caso contrario se rechazará (prueba de dos colas). Si $H_1$ fuera $B > b_0$, se aceptará si $t < t_c$, y se rechazará en caso contrario (prueba de una cola)



prueba de dos colas



prueba de una cola

b. Para la pendiente, M.

Análogamente, para M:

$$\frac{M - m_0}{S_{y/x}/\sqrt{n S_x^2}} = \frac{M - m_0}{\frac{S_{y/x}}{S_x \sqrt{n}}} = T$$

también tiene distribución t de student con $D = n-2$ grados de libertad

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 |

$$m = 0.093, \quad b = 0.032, \quad S_{y/x}^2 = 0.01258$$

$$S_x^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 82.50/10 = 8.25; \quad \Sigma x_i^2 = 285, \quad \overline{x^2} = \frac{285}{10} = 28.5$$

a. Probar la hipótesis de que $B = 0$

b. Probar la hipótesis de que $M = 0.1$

Con $\alpha = 0.01$ y $S_{y/x}$ desconocida

a. $H_0: B = 0$; $H_1: B \neq 0$

$$t = \frac{b - b_0}{S_{y/x}\sqrt{\frac{\overline{x^2}}{n S_x^2}}} = \frac{0.032 - 0}{\sqrt{0.01258}\sqrt{\frac{28.5}{10 \times 8.25}}} = 0.486$$

$$t_c = t_{0.995, 8} = 3.355 > 0.486 \therefore \text{se acepta } H_0$$

b. $H_0: M = 0.1$; $H_1: M \neq 0.1$

$$\cancel{t = \frac{b - b_0}{\frac{S_{y/x}}{S_x}}}$$

$$t = \frac{m - m_0}{\frac{S_{y/x}}{S_x \sqrt{n}}} = \frac{0.093 - 0.1}{\frac{\sqrt{0.01258}}{\sqrt{8.25 \times 10}}} = 0.567$$

$$< 3.355$$

$\therefore$ se acepta $H_0$

Prueba de hipótesis para el coeficiente de correlación, $\rho_{xy}$

Prueba

$H_0: \rho_{xy} = 0 \quad ; \quad H_1: \rho_{xy} \neq 0$

Se demuestra que en este caso de que $X$ y $Y$ son independientes $(\rho = 0)$, la estadística

$$T = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}}$$

tiene distribución $t$ 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 correlación $r_{xy} = 0.931$.

Probar la hipótesis de que $\rho_{xy} = 0$. Usar $\alpha = 0.05$

$H_0: \rho_{xy} = 0 \quad ; \quad H_1: \rho_{xy} \neq 0$

$$t = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0.931\sqrt{\frac{30-2}{1-0.931^2}} = 13.448$$

$$t_{0.975, 28} = 2.048 < 13.448$$

$\therefore$ se rechaza $H_0$ a un nivel de confianza del 95%.

DIRECTORIO DE ASISTENTES AL CURSO PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES ( DEL 17 DE MARZO AL 30 DE ABRIL DE 1980 )

| NOMBRE Y DIRECCION | EMPRESA Y DIRECCION |
|---|---|
| 1. RAUL CASTAÑEDA TRUJILLO<br>Medellín 355-3<br>Col. Roma<br>México 7, D. F. | INSTITUTO MEXICANO DEL SEGURO SOCIAL<br>Cuauhtémoc No. 330<br>Col. Roma<br>México 7, D. F.<br>Tel: 7-61-11-22 |
| 2. RAUL CORREA M.<br>Fray S. T. de Mier 906-13-302<br>Col. J. Balbuena<br>México 9, D. F.<br>Tel; 5-53-35-99 | AMERICAN EXPRESS<br>Amburgo 64-1er. Piso<br>Col. Juárez<br>México 6, D. F.<br>Tel: 5-53-35-99 |
| 3. FRANCISCO CASTRO JUAREZ<br>Oriente 180-A No. 5-A<br>Col. Prado Ermita<br>México 13, D. F.<br>Tel: 5-39-67-16 | INGENIERIA Y PROCESAMIENTO ELECTRONICO,S.A.<br>Cda. de San Francisco No. 6-3er. Piso<br>Col. del Valle<br>México 12, D. F.<br>Tel: 5-59-87-14 Ext. 74 |
| 4. GONZALO N. CRUZ BERISTAIN<br>Andador No. 51 de Acoxpa No. 20<br>Villa Coapa<br>México 22, D. F.<br>Tel: 5-94-33-90 | FACULTAD DE INGENIERIA, UNAM<br>Ciudad Universitaria<br>Col. Narvarte<br>México 12, D. F.<br>Tel: 5-19-94-05 |
| 5. TOMAS E. CHAVEZ RAMIREZ<br>Eten No. 619 Depto. 2<br>Col. Lindavista<br>México 14, D. F. | COMISION FEDERAL DE ELECTRICIDAD<br>Oklahoma No. 85-3er. Piso<br>Col. Nápoles<br>México 18, D. F.<br>Tel: 5-43-44-51 |
| 6. EDGAR C. DE LA GARZA MORATO<br>Antonio León y Gama No. 9-A<br>Col. Obrera<br>México 8, D. F.<br>Tel: 7-61-17-05 | CENTRO NACIONAL DE INFORMACION Y ESTADISTICAS DEL TRABAJO<br>Patriotismo No. 98<br>Col. Escandón<br>México 18, D. F.<br>Tel: 2-77-47-22 |
| 7. LUIS FELIPE EGUIA RAMOS<br>12 de octubre No. 49<br>Jardines del Sur<br>México 23, D. F.<br>Tel: 6-76-06-04 | PETROLEOS MEXICANOS<br>Marina Nacional No. 329<br>Col. Anáhuac<br>México 17, D. F.<br>Tel: 5-31-61-49 |

DIRECTORIO DE ASISTENTES AL CURSO PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES ( DEL 17 DE MARZO AL 30 DE ABRIL DE 1980 )

| NOMBRE Y DIRECCION | EMPRESA Y DIRECCION |
|---|---|
| 8. PABLO FERNANDEZ FERNANDEZ<br>Paseo de la Reforma 51-15o. Piso<br>México 4, D. F.<br>Tel: 5-35-54-83 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Paseo de la Reforma 51 15o. Piso<br>México 4, D. F.<br>Tel: 5-35-54-83 |
| 9. CARLOS RODRIGO GARCIA ONTIVEROS<br>Carolina No. 80-DEPTO. 1<br>Col. Nápoles<br>México 18, D. F.<br>Tel: 5-63-50-15 | INGENIERIA, EQUIPOS Y CONSTRUCCIONES, S.A.<br>Detroit No. 9 Dep. 1102<br>Col. Nápoles<br>México 18, D. F.<br>Tel: 5-98-48-33 |
| 10. ROSÁRIO AMILCAR GOMEZ LUGO<br>Manuel González No. 184-B<br>Tlaltelolco<br>México 3, D. F. | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Paseo de la Reforma No. 51-15o. Piso<br>Col. Tabacalera<br>México 4, D. F.<br>Tel: 5-66-92-83 |
| 11. IRMA LETICIA GONZALEZ S. | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS |
| 12. ALFONSO HERNANDEZ ISLAS<br>Av. San Jerónimo No. 9<br>El Rosal Contreras<br>México 20, D. F. | INSTITUTO MEXICANO DEL SEGURO SOCIAL<br>Cuauhtémoc No. 330<br>Col. Roma<br>México 7, D. F.<br>Tel: 7-61-11-22 Ext. 21-27 |
| 13. MANUEL A. HUIDOBRO GARCIA<br>Astorga No. 88<br>Lomas Estrella<br>México 13, D. F. | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>San Bernabé No. 549<br>San Jerónimo Lídice<br>México 20, D. F.<br>Tel: 5-95-39-50 |
| 14. SERGIO HURTADO PEDRAZA<br>Albino García No. 270-6<br>Viaducto Piedad<br>México 13, D. F.<br>Tel: 5-38-88-31 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Ignacio Ramírez No. 20<br>Col. San Rafael<br>México 4, D. F.<br>Tel: 5-66-38-48 |

DIRECTORIO DE ASISTENTES AL CURSO PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES ( DEL 17 DE MARZO AL 30 DE ABRIL DE 1980 )

| NOMBRE Y DIRECCION | EMPRESA Y DIRECCION |
|---|---|
| 15. ARMANDO JAIME RAMOS<br>Retorno 203-20<br>Unidad Modelo<br>México 13, D. F.<br>Tel: 5-92-11-44 | DIRECCION GENERAL DE R. D. D. F.<br>Edisón 170-3er. Piso<br>Col. San Rafael<br>México 13, D. F.<br>Tel: 5-81-38-36 |
| 16. LEON GALDINO JUAREZ MARTINEZ<br>Baja California No. 80-6<br>Col. Roma Sur<br>México 7, D. F.<br>Tel: 5-84-86-82 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Ignacio Ramírez No. 20<br>Col. San Rafael<br>México 4, D. F.<br>Tel: 5-66-36-25 |
| 17. MANUEL ENRIQUE LOPEZ MUÑOZ<br>Mitla 361-B<br>Col. Narvarte<br>México 12, D. F.<br>Tel: 5-79-31-42 | ACUEDUCTOS Y ALCANTARILLADOS-COSTA RICA<br>Costa Rica |
| 18. LUIS LORENZO LUNA ITURBE<br>Miguel Schultz No. 96-2<br>Col. San Rafael<br>México 4, D. F.<br>Tel: 5-66-37-22 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Ignacio Ramírez No. 20<br>Col. San Rafael<br>México 4, D. F.<br>Tel: 5-66-39-11 |
| 19. CARLOS MAROTO CABRERA<br>Zacatecas 143-14<br>Col. Roma<br>México 7, D. F.<br>Tel: 5-84-16-39 | FACULTAD DE QUIMICA, UNAM<br>Ciudad Universitaria<br>México 20, D. F.<br>Tel: 5-48-65-00 Ext. 274 |
| 20. RENE MARTINEZ ALVARADO<br>Edificio 9o. A-304<br>Unidad Cuitlahuac<br>México, D. F.<br>Tel: 3-55-39-68 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Paseo de la Reforma No. 51-15<br>Col. Tabacalera<br>Tel: 5-66-97-83 |
| 21. MA. TERESA MEDINA V.<br>Tagle 202<br>Pachuca, Hgo. | UNIVERSIDAD AUTONOMA DE HIDALGO<br>Abasolo 600<br>Pachuca, Hgo.<br>Tel: 2-65-35 Ext. 41. |

DIRECTORIO DE ASISTENTES AL CURSO PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES ( DEL 17 DE MARZO AL 30 DE ABRIL DE 1980 )

| NOMBRE Y DIRECCION | EMPRESA Y DIRECCION |
|---|---|
| 22. JUANA MA. MENDOZA ANDRADE<br>Retorno 47 No. 35<br>Col. Avante<br>México 21, D. F.<br>Tel: 6-77-62-54 | SECRETARIA DE ASENTAMIENTOS HUMANOS Y OBRAS PUBLICAS<br>Lago Poniente No. 16-1er. Piso<br>Col. Natívitas<br>México D. F.<br>Tel: 6-74-17-39 |
| 23. MIGUEL T. MORENO SERRANO<br>José T. Cuellar 116 Depto. 1<br>Col. Obrera<br>México 8, D. F.<br>Tel: 5-88-82-65 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Ignacio Ramírez No. 20-4o. Piso<br>Col. San Rafael<br>México 4, D. F.<br>Tel: 5-66-38-48 |
| 24. MARGARITA MUÑOZ FUENTES<br>Chac - Mool 6-3<br>Unidad Independencia<br>México 20, D. F.<br>Tel: 5-95-68-14 | DIRECCION GENERAL DE PLANIFICACION<br>Pino Suárez No. 15<br>México 1, D. F.<br>Tel: 5-22-64-38 |
| 25. NELIDA R. PARRA MALDONADO<br>Cda. Plan de Ayala 10-4<br>Nextitla<br>México 17, D. F.<br>Tel: 5-47-52-21 | LAB. SALUD PUBLICA I.M.S.S.<br>Hosp. Red. C.M.N.<br>Col. Doctores<br>México 7, D. F.<br>Tel: 5-47-52-21 |
| 26. JOSE MANUEL PERDOMO RANGEL<br>Sur 1-9 No. 624<br>Sector Popular<br>México 13, D. F.<br>Tel: 5-94-61-01 | CALCULOS Y PROYECTOS RODRIGUEZ, S.A.<br>C.I. No. 31 Edif. K.19<br>Pantitlan<br>México 9, D. F.<br>Tel: 5-58-54-24 |
| 27. CESAR AUGUSTO PUGA CEH<br>Horacio Nelson No. 65-101<br>Col. Moderna<br>México 13, D. F.<br>Tel: 6-96-01-03 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Paseo de la Reforma No. 45-10o. Piso<br>México 1, D. F.<br>Tel: 5-92-01-08 |
| 28. MANUEL E. RODRIGUEZ GARNICA<br>José S. Trujillo No. 226-4<br>San Alvaro<br>México 17, D. F.<br>Tel: 5-27-93-10 | SECRETARIA DE ASENTAMIENTOS HUMANOS Y OBRAS PUBLICAS<br>Xola y Universidad<br>Col. Narvarte<br>México 17, D. F.<br>Tel: 5-19-76-60 |

DIRECTORIO DE ASISTENTES AL CURSO PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES ( DEL 17 DE MARZO AL 30 DE ABRIL DE 1980 )

| NOMBRE Y DIRECCION | EMPRESA Y DIRECCION |
|---|---|
| 29. JOSE LUIS RUEDA PALOMEQUE<br>Pino No. 522-2<br>Col. Arenal<br>México 15, D. F.<br>Tel: 3-55-80-07 | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Paseo de la Reforma No. 51-17o. Piso<br>México 1, D. F.<br>Tel: 5-66-97-69 |
| 30. JUAN JAVIER SILLER LEYVA<br>Sacramento 219<br>Col. del Valle<br>México 12, D. F.<br>Tel: 5-23-67-00 | CONSEJO NACIONAL DE CIENCIA Y TECNOLOGIA<br>Insurgentes Sur 1814<br>México, D. F. |
| 31. VICTOR A. SOTELO CORNEJO<br>Circunvalación No. 57<br>Unidad Barrientos<br>Tlalnepantla, Edo. de México<br>Tel: 5-65-28-30 | SECRETARIA DE ASENTAMIENTOS HUMANOS Y OBRAS PUBLICAS<br>Xola y Universidad<br>Col. Narvarte<br>México 12, D. F. |
| 32. JUAN VALLE FRAGOSO<br>Aztecas No. 176<br>Col. La Romana<br>Tlalnepantla<br>México, D. F. | LUGATOM, S. A.<br>Km. 24.5 Carr. Puente de Vigas<br>Tepotzotlan,<br>Cuautitlan, México |
| 33. GABRIEL VENTURA SUAREZ<br>Fray Juan de Torquemada No. 9-10<br>Col. Obrera<br>México 8, D. F.<br>Tel: 5-30-47-04 | |
| 34. OSCAR VERA SMITH<br>Torquemada No. 139-3<br>Col. Obrera<br>México 8, D. F. | SECRETARIA DE AGRICULTURA Y RECURSOS HIDRAULICOS<br>Reforma 51-15o. Piso<br>Col. Tabacalera<br>México 4, D. F.<br>Tel: 5-66-97-83 |
| 35. MA. DORALISA VILLARROEL MELO<br>Av. Universidad 2042-1206<br>Copilco-Universidad<br>México 20, D. F.<br>Tel: 5-48-33-35 | INSTITUTO DE INGENIERIA, UNAM<br>Ciudad Universitaria<br>México 20, D. F.<br>Tel: 5-50-52-15 Ext. 3604 |

DIRECTORIO DE ASISTENTES AL CURSO PROBABILIDAD Y ESTADISTICA, FUNDAMENTOS Y APLICACIONES ( DEL 17 DE MARZO AL 30 DE ABRIL DE 1980 )

| NOMBRE Y DIRECCION | EMPRESA Y DIRECCION |
|---|---|
| 36. EFREN ZARAGOZA RAMOS<br>Amores 403-2<br>Col. del Valle<br>México 12, D. F.<br>Tel: 5-23-23-95 | DIRECCION GENERAL DE PROMOCION DEPORTIVA<br>Av. División del Nte. 2333<br>México 13, D. F.<br>Tel: 5-34-13-78 |