ESTADÍSTICA

CAPÍTULO 4

MEDIDAS DE VARIABILIDAD Y DE FORMA

T. Teddy Canelas Verduguez

U.M.S.A.

UNIVERSIDAD MAYOR DE SAN ANDRES
FACULTAD DE CIENCIAS ECONÓMICAS Y FINANCIERAS
CARRERAS: AUDITORÍA Y ECONOMÍA

MATERIA:

"ESTADÍSTICA"

CAPÍTULO 4. MEDIDAS DE VARIABILIDAD Y DE FORMA

CATEDRÁTICO: T.Teddy Canelas Verduguez
Master en Estadística Matemática

COLABORACIÓN: Marianela Martinez
Ingeniero y Doctor

La Paz–Bolivia

Nº Depósito Legal: 4-1-482-08
Título: ESTADÍSTICA
Edición: Cuarta
Impreso en: Imp. "EL CLON" 2200660

DEDICATORIA:

Con amor y gratitud.

A mi madre:

Julia Verduguez Larraín

A mi esposa:

Elsa Rivero Aparicio

PRESENTACÍON

*Con el marcado propósito de que el presente texto de "**ESTADÍSTICA**", sea accesible, particularmente, por los estudiantes se presenta por Capítulos en hojas de tamaño medio oficio y con la apariencia de un cuaderno corriente.*

Se ha escrito también con la esperanza de que el potencial lector, encuentre novedad y sencillez, y principalmente empiece a adquirir una cultura estadística para comprender mejor lo que se dice, de los hechos o acontecimientos, del mundo real en el que vivimos, después de que estos fueron investigados "científicamente".

El profesional de cualquier área, llámese Auditor, Economista, Administrador de Empresas, médico, etc., que tenga una formación adecuada de la Estadística, será sin duda alguna, un mejor profesional en su campo.

La Paz, Abril de 2008

LOS EDITORES

CAPÍTULO 4

MEDIDAS DE VARIABILIDAD Y DE FORMA

CAPÍTULO 4. MEDIDAS DE VARIABILIDAD Y DE FORMA

4.1 INTRODUCCION.

Continuando el análisis de un conjunto de datos, en este capítulo se estudian las medidas descriptivas de:

- variabilidad y
- de forma

Las medidas de variabilidad son indicadores estadísticos del grado de dispersión o variación de los datos principalmente respecto al promedio.

Las medidas de forma son indicadores del grado de relación con una curva simétrica denominada normal. Estos indicadores se agrupan en medidas de:

1) Asimetría y
2) Curtosis o de alargamiento.

4.2 RECORRIDO O RANGO, R.

El recorrido o rango es la variación total del conjunto de datos, entre el menor valor y el mayor. Entonces:

DEFINICIÓN ***RECORRIDO.***

Es la diferencia del máximo valor menos el mínimo. En símbolos:

$$R = Máx - Mín\ .$$

EJEMPLO. 4.1 Recorrido o Rango de Datos no Agrupados.

En la ciudad de Santa Cruz, durante un mes de verano, los 7 vendedores de una empresa de equipos de calefacción y aire acondicionado (ECAA) vendieron las siguientes cantidades de unidades de aire acondicionado: 7, 10, 2, 15, 6, 8 y 5. ¿Cuál es el recorrido de esta variable, unidades de aire acondicionado?

SOLUCIÓN.

FORMULA: $R = Máx - Mín$

Datos:

$$7,\ 10,\ \underset{Mín}{2},\ \overset{Máx}{15},\ 6,\ 8,\ 5.$$

Reemplazando: $R = 15 - 2 = 13$ ▲

RESPUESTA.- El recorrido o rango es 13 y significa que comparando las ventas de los 7 vendedores existe una variación de 13 unidades de aire acondicionado entre el que vendió más y el que vendió menos

El "**recorrido**", de un conjunto de datos nos indica la variación total por diferencia de dos datos. Este indicador puede ser muy engañoso ya que no informa absolutamente nada del grado de variabilidad del resto de datos y nosotros particularmente estamos interesados en el grado de concentración de la totalidad de los datos alrededor de la media.

Un promedio es representativo de un conjunto de datos, cuando la mayoría de los datos están cerca de dicho promedio, en otras palabras diremos que un promedio es más representativo o tiene mayor representatividad, cuando la variabilidad de los datos alrededor de dicho promedio es más pequeña.

4.3 RANGOS MODIFICADOS, [$v.g. R_{mod}$]

Son indicadores que tienden a corregir la diferencia anotada, eliminando datos en ambos extremos del recorrido o rango y midiendo la variación de un porcentaje de los datos que se ubican al centro. Entonces se puede considerar "**Rangos Modificados**" de un, digamos 70% central, esto significa, que se elimina el mismo porcentaje en ambos extremos, concretamente, en este caso se elimina el 15% de los valores más pequeños y el 15% de los más grandes. De este modo, por ejemplo, el "**Rango del 50% Central**", resulta ser un rango modificado, llamado, rango intercuartil, (RIC). Así

$$RIC = Q_3 - Q_1$$

Otros rangos modificados de uso común son los rangos centrales del 80, 90 y 95 por ciento.

EJEMPLO. 4.2 Cálculo de Rangos Modificados.

Una vez ordenados de menor a mayor los datos del ejemplo 4.1 se pide:

a) Calcular el Rango intercuartil, RIC.
b) Calcular el Rango modificado del 80% central.

SOLUCIÓN:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Datos ordenados, con n = 7	2,	5,	6,	7,	8,	10,	15.

a) Cálculo, Rango Intercuartil, RIC.

FORMULA: $RIC = Q_3 - Q_1$ (1)

Previo:

$$Q_3 = X_{\frac{3n}{4}+\frac{1}{4}} = X_{\frac{21}{4}+\frac{1}{4}} = X_{5,75} = 8 + (0,75)(10-8)$$

$$= 8 + 1,5 = 9,5 \blacktriangle$$

$$Q_1 = X_{\frac{n}{4}+\frac{1}{2}} = X_{\frac{7}{4}+\frac{1}{2}} = X_{2,25} = 5 + (0,25)(6-5)$$

$$= 5 + 0,25 \doteq 5,2 \blacktriangle$$

Reemplazando en (1): $RIC = 9,5 - 5,2 = 4,3 \therefore$ $\boxed{RIC = 4,3\ u_s\ ECAA}$

b) Cálculo, Rango Modificado del 80% central.

FORMULA: $R_{80\%C} = P_{90} - P_{10}$ (2)

Previo:

$$P_{90} = X_{\frac{90n}{100}+\frac{1}{2}} = X_{\frac{630}{100}+\frac{1}{2}} = X_{6,8} = 10 - 0.8(15-10)$$

$$= 10 + 4 = 14,0 <$$

$$P_{10} = X_{\frac{10n}{100}+\frac{1}{2}} = X_{\frac{70}{100}+\frac{1}{2}} = X_{1,2} = 2 + 0,2(5-2)$$

$$= 2 + 1,5 = 3,5 \triangleleft$$

Reemplazando en (2): $R_{80\%C} = 14,0 - 3,5 = 10,5$

Luego: $\boxed{R_{80\%C} = 10,5\ u.\ ECAA}$

RESPUESTA: Se lee el rango modificado del 80% central es 10,4 unidades de Equipos de Calefacción y aire acondicionado.

4.4 DESVIACIÓN MEDIA, D.M.:

DEFINICIÓN **DESVIACIÓN MEDIA.**

"Es la media de los desvíos absolutos de los valores de la variable respecto a su media".
Notación.- *Se utiliza:* $D.M._x$.

Por definición:

a) **POBLACIÓN.** (Parámetro) : $D.M._x = \dfrac{\sum_{i=1}^{N} |x_i - \mu|}{N}$

b) **MUESTRA,** (Estadígrafo): $D.M._x = \dfrac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$

OBSERVAR QUE:

1) $|x_i - \bar{x}|$ se llama desvío o desviación absoluta de x_i respecto de la media, $\bar{x}$.

2) Para datos agrupados, los desvíos se multiplica por su frecuencia n_i , donde x_i es el punto medio de la clase. Así: $D.M._x = \dfrac{\sum |x_i - \bar{x}| n_i}{n}$

¡HOLA!: ¿Por qué piensa Usted que en la definición se tomó los valores absolutos de los desvíos, $(x_i - \bar{x})$?.

FORMULAS **DESVIACIÓN MEDIA.**

Las fórmulas de la desviación media, para datos (a) no agrupados y (b) agrupados de una muestra, (Estadígrafos), son :

a) DATOS NO AGRUPADOS:

$$D.M. = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

b) DATOS AGRUPADOS. *(k clases, con,* $n_1 + n_2 + \cdots + n_k = n$ *)*

$$D.M. = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| n_i}{n}$$

Donde: (1) x_i *= punto medio de la clase* i

(2) $|x_i - \bar{x}|$ *= Valor absoluto de los desvíos* $(x_i - \bar{x})$

EJEMPLO. 4.3 Cálculo de la Desviación Media, D.M.(Con datos no agrupados)

SOLUCIÓN:(datos del ejemplo 4.1).
DATOS: 2, 5, 6, 7, 8, 10, 15; n = 7.

Formulas: $D.M._x = \dfrac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$ (1) ;

Previo calcular la media: $\bar{x} = \frac{\sum_{i=1}^{7} x_i}{n} = \frac{2+5+6+7+8+10+15}{7} = \frac{53}{7} = 7,571428571$

Luego: $\boxed{\bar{x} = 7,6 u.ECAA.}$

Reemplazando en (1):

$$D.M_x = \frac{|2-\bar{x}|+|5-\bar{x}|+|6-\bar{x}|+|7-\bar{x}|+|8-\bar{x}|+|10-\bar{x}|+|15-\bar{x}|}{7}$$

$$D.M_x = \frac{5,6+2,6+1,6+0,6+0,4+2,4+7,4}{7} = \frac{20,6}{7} = 2,9388$$

Luego: $\boxed{D.M_x = 2.94 \quad \textit{unidades de ECAA}}$

EJEMPLO. 4.4 Cálculo de la Desviación Media, D.M. (Datos agrupados)

SOLUCIÓN:(datos del ejemplo 2.1, son los sueldos mensuales de una muestra de 50 empleados de la empresa "ACE" del país. Año 20XX).

HOJA DE TRABAJO Nº .2

(000 Bs.) SALARIOS (Miles Bs.)	PM x_i	n_i	$x_i n_i$	$\bar{x}$ = 2,94 $\lvert x_i-\bar{x}\rvert$	$\lvert x_i-\bar{x}\rvert n_i$	$(x_i-\bar{x})n_i$
1,1 – 1,7	1,4	6	8,4	1,54	9,24	-9,24
1,8 – 2,4	2,1	9	18,9	0,84	7,56	-7,56
2,5 – 3,1	2,8	14	39,2	0,14	1,96	-1,96
3,2 – 3,8	3,5	13	45,5	0,56	7,28	7,28
3,9 – 4,5	4,2	7	29,4	1,26	8,82	8,82
4,6 – 5,2	4,9	0	0,0	1,96	0,00	0,00
5,3 – 5,9	5,6	1	5,6	2,66	2,66	2,66
TOTAL	-.-	**n = 50**	**147,0**	-.-	**37,52**	**0,00**

NOTA.- Ultima columna suma cero ¿Por qué?

Reemplazando: $\bar{x} = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \frac{147,0}{50} = 2,94$

$$D.M_x = \frac{\sum_{i=1}^{7} |x_i - \bar{x}| n_i}{n} = \frac{37,52}{50} = 0,7504$$

Luego: $\boxed{D.M_x = Bs.750,40}$

RESPUESTA.- Los datos se apartan en promedio de la media aritmética ($\bar{x}$ = Bs. 2 940) en Bs. 750.

4.5 VARIANZA Y DESVIACIÓN ESTANDAR:

Una medida importante de la estadística es la **"Varianza"**.

DEFINICIÓN ***VARIANZA.***

"Es la media de los cuadrados de las desviaciones de los valores de la variable respecto a su media".

Notación.- *Para una muestra, la varianza se indica por:* S^2 *y en el caso de la población por:* σ^2

La **varianza** se indica y define por:

a) **Población**. (Parámetro): $V[X] = \sigma^2 = \dfrac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$

b) **Muestra**. (Estadígrafo): $V[X] = S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$

DEFINICIÓN **_DESVIACIÓN ESTÁNDAR._**

"Es la raíz cuadrada positiva de la varianza" y se indica por: S cuando es una muestra, y por σ cuando es una población.

La desviación estándar se indica y define por.

a) **Población:** $\sigma = +\sqrt{\sigma^2}$

b) **Muestra.** $S = -\sqrt{S^2}$

OBSERVAR QUE:

1) Si la variable, X peso de un estudiante se mide en kilogramos, ¿Cómo interpretaría S^2 expresada en kilogramos al cuadrado? Un intento. "la variabilidad de los pesos de los estudiantes en promedio es de 50 kilogramos al cuadrado" ¿Verdad que es risible?

2) La desviación estándar se expresa en la misma unidad de medida que la de la variable, entonces la interpretación no presentará la dificultad anotada para la varianza. Una buena razón para preferir la ¿Verdad?

3) Notación: De la varianza, $V[x]$ ó σ^2 ó S^2 y de la desviación estándar: $DS[x]$ ó σ ó S

4) El cálculo manual de la varianza aplicando la definición en forma directa es las más de las veces molestoso o moroso y con el propósito de disminuir la molestia se transforma a otra, llamada formula de cálculo de la varianza.

Desarrollo de:

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = ?$$

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}\left(x_i^2 - 2\overline{x}x_i + \overline{x}^2\right) = \sum_{i=1}^{n}x_i^2 - 2\overline{x}\sum_{i=1}^{n}x_i + n\overline{x}^2$$

$$= \sum_{i=1}^{n}x_i^2 - 2\overline{x}\cdot n\overline{x} - n\overline{x}^2 = \sum_{i=1}^{n}x_i^2 - n\overline{x}^2$$

Reemplazando :

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \frac{\sum_{i=1}^{n}x_i^2 - n\overline{x}^2}{n} = \frac{\sum_{i=1}^{n}x_i^2}{n} - \overline{x}^2$$

Cálculo con datos no agrupados

FÓRMULA: $S^2 = \dfrac{\sum x_i^2 - n\overline{x}^2}{n}$ ó $S^2 = \dfrac{\sum_{i=1}^{n}x_i^2}{n} - \overline{x}^2$

Similarmente para datos agrupados:

FÓRMULA: $V[x] = S^2 = \dfrac{\sum x_i^2 n_i - n\overline{x}^2}{n}$ ó $S^2 = \dfrac{\sum_{i=1}^{k} x_i^2 n_i}{n} - \overline{x}^2$

5) En capítulos posteriores se explica que ciertos estadígrafos son buenos estimadores de los parámetros. Así $\overline{x}$ es el mejor estimador de μ, pero no así S^2 para σ^2

4.6 CUASI-VIARIANZA Y DESVIACIÓN ESTANDAR-C:

***Cuasi-varianza o varianza-c muestral.- Este estadígrafo**, se indica y define por,*

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

La **"Cuasi-varianza"** o **"Varianza-C"**, es mejor estimador que la varianza muestral S^2. Por ello se usa, S^2 como estimador de σ^2 La mayoría, de los libros de texto de la materia de Estadística, o si se prefiere **casi todos**, denominan varianza muestral a la "Cuasi-varianza" muestral. Suponemos que es para evitar ciertas molestias y facilitar su manejo en su aplicación.

El término **"Cuasi-varianza"** o **"varianza-c"** sólo difiere del término varianza en el denominador que en lugar de N ó n, es (N – 1) ó (n – 1); es decir, que el denominador disminuye en una unidad.

La desviación estándar-c, es la raíz cuadrada positiva de la Cuasi-varianza.

FORMULAS: VARIANZA, Cuasi-Varianza, Desviación Estándar

Se presentan las fórmulas para una muestra. Para las fórmulas de la población sólo cambiar los símbolos que corresponden ya que las definiciones no cambian:

***1) VARIANZA**. Notación:* $V[x]$ ó S^2

a) DATOS NO AGRUPADOS: $(x_1, x_2, \ldots, x_n)$

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}$$

b) DATOS AGRUPADOS, clases: $(n_1 + n_2 + \ldots + n_k = n)$

$$S^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 n_i}{n}$$

***2) CUASI-VARIANZA.** Notación:* $V_c[x]$ ó S^2

Sólo cambiar en la fórmula el denominador n de la varianza por (n – 1) para la cuasi-varianza.

Así, para datos agrupados y no agrupados.

a) $S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2 n_i}{n-1}$; b) $S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$

3) ***DESVIACIÓN ESTANDAR.*** *Notación:* S^2 ó $DS[x]$

Es la raíz cuadrada positiva de la varianza

4) ***DESVIACIÓN ESTANDAR-C.*** *Notación:* S^2 ó $DS_c[x]$ *Es la raíz cuadrada positiva de la cuasi-varianza.*

5) ***FORMULAS DE CÁCULO.*** *- Para datos agrupados*

a) $S^2 = \dfrac{\sum_{i=1}^{n} x_i^2 n_i}{n} - \overline{x}^2$ b) $S^2 = \dfrac{\sum_{i=1}^{n} x_i^2 n_i - n\overline{x}^2}{n-1}$

c) $S = +\sqrt{S^2}$ d) $S = +\sqrt{S^2}$

SUGERENCIA. *- Primero calcular* S^2 ó S^2 *para después obtener* S ó S

NOTA.- Nuestra posición es que las definiciones de los términos estadísticos sea única o la misma se trate de la población o de una muestra. (La definición debe ser la misma sean los cálculos con todos los datos de la población o de una parte de la misma)

EJEMPLO. 4.5 Cálculo de la Varianza, Cuasi-Varianza, Desviación Estándar y desviación Estandar-C.(datos del ejemplo 4.1, no agrupados)

SOLUCIÓN:

DATOS: n = 7, Son 2, 5, 6, 7, 8, 10, 15 ventas unidades de E.C.A.A.

1) **Calcular la Varianza y Cuasi-Varianza.**

(Esta HOJA DE TRABAJO puede omitirse, depende de Ud.)

Símbolo x_i	**Valores de X**	$\overline{x}$ **=7,6** $x_i - \overline{x}$	$(x_i - \overline{x})^2$	x_i^2
x_1	2	-5,6	31,36	4
x_2	5	-2,6	6,76	25
x_3	6	-1,6	2,56	36
x_4	7	-0,6	0,36	49
x_5	8	0,4	0,16	64
x_6	10	2,4	5,76	100
x_7	15	7,4	54,76	225
Total	**53**	-	**101,72**	**503**

FORMULAS:

a) **VARIANZA:** $S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \dfrac{\sum x_i^2 - n\overline{x}^2}{n} = \dfrac{\sum x_i^2}{n} - \overline{x}^2$

b) CUASI-VARIANZA, $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$

Cálculos:

Media: $\bar{x} = \frac{\sum_{i=1}^{n} x}{n} = \frac{53}{73} = 7.5714; \quad \bar{x} = 7,6uE.C.A.A.$

a) VARIANZA:

- Por definición la Varianza:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{101,72}{7} = 14,5314$$

- Formula de cálculo:

$$S^2 == \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{503}{7} - (7,5714)^2 = 14,5310$$

Luego: $\boxed{S^2 = 14.53(uE.C.A.A.)^2}$

OBSERVAR QUE:

En la formula de cálculo como $\bar{x}$ **se debe tener el cuidado de tomar con muchos decimales** (Aquí son 4 decimales) para obtener resultados de mayor exactitud de S^2. Con $\bar{x} = 7,6$ es:

$\therefore \boxed{S^2 = 14.10(uE.C.A.A.)^2}$

Hay una diferencia de 43 centésimos de $(uE.C.A.A.)^2$

b) CUASI-VARIANZA

- Por definición de Cuasi-Varianza:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{101,72}{6} = 16,953 \therefore S^2 = 16,95$$

- Formula de cálculo:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{503 - 7(7,5714)^2}{6} = \frac{101,7173}{6} = 16,9529$$

Luego: $\boxed{S^2 = 16.95(uE.C.A.A.)^2}$

2) DESVIACIÓN ESTÁNDAR Y DESVIACIÓN ESTANDAR-C

a) DESVIACIÓN ESTÁNDAR

$S=\sqrt{S^2}=\sqrt{147,5314}=3,81 \quad \therefore \boxed{S=3,81uE.C.A.A.}$ ♣

b) DESVIACIÓN ESTÁNDAR-C

$\boxed{S=+\sqrt{S^2}=\sqrt{3,81} \quad uE.C.A.A. \therefore S=4,12uE.C.A.A.}$ ♣

EJEMPLO. 4.6 Cálculo de la Varianza, Cuasi-Varianza, Desviación Estándar y desviación Estándar-C.(datos agrupados)

HOJA DE TRABAJO

SALARIOS (000 Bs.)	PM x_i	n_i	$\bar{x}=2,94$ $(x_i-\bar{x})^2 n_i$	$\bar{x}=2,94$ $(x_i-\bar{x})n_i$	$x_i n_i$	$x_i^2 n_i$
1,1 – 1,7	1,4	6	14,2296	- 9,24	8,4	11,76
1,8 – 2,4	2,1	9	6,3504	- 7,56	18,9	39,69
2,5 – 3,1	2,8	14	0,2744	- 1,96	39,2	109,76
3,2 – 3,8	3,5	13	4,0768	7,28	45,5	159,25
3,9 – 4,5	4,2	7	11,1132	8,82	29,4	123,48
4,6 – 5,2	4,9	0	0,0000	0,00	0,0	0,00
5,3 – 5,9	5,6	1	7,0756	2,66	5,6	31,36
Total	- , -	**N = 50**	**43,1200**	**0,00**	**147,0**	**475,30**

DATOS: n = 50;

FORMULAS: VARIANZA: $S^2=\dfrac{\sum_{i=1}^{k}(x_i-\bar{x})^2 n_i}{n}$

CUASI-VARIANZA, $S^2=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}=\dfrac{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}{n-1}$

CÁLCULOS:

a) VARIANZA Y DESVIACIÓN ESTÁNDAR:

- Por definición la Varianza:

$$S^2=\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2}{n}=\frac{43,12}{50}=0,8624 ◁$$

- Formula de cálculo:

$$S^2=\frac{\sum x_i^2}{n}-\bar{x}^2=\frac{475,3}{50}-(2,94)^2=9,506-8,6436=0,8624$$

Y $S=\sqrt{0,8694}=0,92866$

Luego: $\boxed{S^2=0,8624(Bs.)^2 \quad y \quad S=0,92866Bs.}$

b) CUASI-VARIANZA Y DESVIACIÓN ESTÁNDAR-C .

- Por definición de Cuasi-Varianza:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{43,12}{50-1} = \frac{43,12}{49} = 0,88$$

- Formula de cálculo:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{475,3 - 50(2,94)^2}{50-1} = \frac{43,12}{49} = 0,88$$

Y $\boxed{S^2 = -\sqrt{0,88} = 0,938}$

Luego: $\boxed{S^2 = 0,88(Bs.)^2 \text{ y } S^2 = 0,938\,Bs.}$

4.7 COEFICIENTE DE VARIABILIDAD:

Es una medida relativa de variabilidad al evaluar la variación absoluta (desviación estándar - c) en la relación con la medida de los datos. Es útil, cuando se compara la variabilidad de dos conjuntos de datos.

El coeficiente de variabilidad se indica y define por,

a) POBLACIÓN: $C.V = \dfrac{\sigma}{\mu}$

b) MUESTRA: $C.V = \dfrac{S}{\bar{x}}$ ó $\dfrac{S}{\bar{x}}$

EJEMPLO. 4.7 Cálculo del Coeficiente de variación

Los siguientes tiempos en minutos fueron registrados por corredores de 400 y 1600 metros de un equipo universitario de pista.

Tiempos\Univ.	1	2	3	4	5
a) 400 metros	0,92	0,98	1,04	0,90	0,99
b) 1600 metros	4,52	4,35	4,60	4,70	4,50

Después de ver esta muestra de tiempos, uno de los entrenadores comentó que los corredores de 400 metros corrían con más consistencia. Emplee la desviación estándar-c y el coeficiente de variación para resumir la variabilidad en los datos. ¿El coeficiente de variación indica que es cierta la afirmación del entrenador?

SOLUCIÓN:

a. Para 400 metros: De los cálculos $\sum x = 4,83$ $\bar{x} = 0,966$

$S = 0,0504$ $S = 0,0564$

$$C.V. = \frac{S}{\bar{x}} = \frac{0,05639}{0,966} = 0,058 \Rightarrow \boxed{C.V. = 5,8\%}$$

b. Para 1600 metros. De los cálculos $\sum x = 22,67$ $\bar{x} = 4,534$

$S = 0,11586$ $S = 0,12954$

$$C.V. = \frac{S}{\bar{x}} = \frac{0,129538}{4,534} = 0,029 \Rightarrow \boxed{C.V. = 2,9\%}$$

c. La variabilidad relativa es menor para los 1600 metros, ya que $C.V. = 2{,}9\%$

4.8 PROPIEDADES DE LA VARIANZA

$\boxed{P.1}$.- *Sea $y = ax + b$, una función lineal de x, entonces:*

$$V[y] = V[ax + b] = a^2 V[x].$$

Casos: a) $a = 0 \Rightarrow V[b] = 0$ (la varianza de una constante es cero)

b) $b = 0 \Rightarrow V[ax] = a^2 V[x]$

c) $a = 1 \Rightarrow V[x + b] = V[x]$

$\boxed{P.2}$.- *La suma de cuadrados de las desviaciones de los valores de la variable respecto a un punto (número) a es mínimo si y sólo si a es la media $\overline{x}$.*

Observar que $\sum_{i=1}^{n}(x_i - \overline{x})^2$ es el numerador de la varianza.

DEMOSTRACIÓN:

Sea $Z = \sum_{i=1}^{n}(x_i - a)^2 = \text{mín}$

Para un mínimo la derivada de Z respecto a x debe ser cero. Así,

$$\frac{dZ}{dx} = 0 \Rightarrow 2\sum_{i=1}^{n}(x_i - a) = 0$$

$$2\sum_{i=1}^{n}(x_i - a) = 0 \Rightarrow \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}a = 0$$

$$\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}a$$

$$\sum_{i=1}^{n}x_i = na$$

$$a = \frac{\sum_{i=1}^{n}x_i}{n} \quad \text{por definición de media} \quad \boxed{a = \overline{x}}$$

$\boxed{P.3}$.- *Sea A un conjunto de datos de tamaño n_A y varianza S_A^2 y B otro conjunto de datos de tamaño n_B y varianza S_B^2. Si ambos tienen la misma media $\overline{x}$, entonces la varianza combinada (agrupando ambos conjuntos A y B) es:*

$$S_{A+B}^2 = \frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B}$$

Dice: Que S_{A+B}^2 es la media ponderada de las dos varianzas.

Esta propiedad es válida para más de dos grupos de datos.

4.9 USO DE LA DESVIACIÓN ESTÁNDAR. (EN LA DESCRIPCIÓN DE DATOS).

El teorema de Chebyshev determina el porcentaje de datos que se encuentran dentro el intervalo definido por $\overline{X}$ más y menos k desviaciones estándar, $(\overline{x} \pm kS)$. Especificamente,

A. TEOREMA DE CHEBYSHEV.

"Para un conjunto cualquiera de datos, sin importar cómo estén distribuidos, la proporción (fracción) de ellos que se encuentra dentro de k desviaciones estándar de la media de los datos", (k > 1), es "al menos".

$$P\left[\left|x-\overline{x}\right| \leq k\sigma\right] \leq 1-\frac{1}{k^2}$$

$$\left(1-\frac{1}{k^2}\right)$$

OBSERVAR QUE.

1. k es mayor que 1.

2. $(\overline{x} \pm kS)$ = Intervalo definido por $\overline{x}$ más y menos k desviaciones estándar, S

3. "al menos" significa que la proporción (fracción) podría ser mayor.

4. Este teorema es verdadero para todo conjunto de datos, sin ninguna restricción sobre la forma de la distribución de los datos.

EJEMPLO. 4.8 Uso de la Desviación Estándar (Teorema de Chebyshev)

En un conjunto de 36 datos, (a) ¿cuál es el porcentaje de datos que están dentro de 2 desviaciones estándar de la media? y (b) ¿mínimamente cuantos y cuales son, para los siguientes datos, ordenados de menor a mayor, $a_1, a_2, a_3, \ldots, a_{36}$, si la media es: $\overline{x}$ y la desviación estándar S

SOLUCIÓN.

a) Si $k = 2$, el porcentaje es: $1-\frac{1}{k^2} = 1-\frac{1}{2^2} = \frac{3}{4} = 0,75 = 75\%$

RESPUESTA.- Al menos 75% de los datos se encuentran en el intervalo $\overline{x} \pm 2S$ es decir, están dentro de 2 desviaciones estándar, S, de la media, $\overline{x}$.

b) Son:

1. Cantidad de datos: 75% de 36 datos = 0,75 x 36 = 27 datos

2. Minimamente, los 27 datos del total 36, son los a_i que satisfacen la relación,

$$(\overline{x}-2S) \leq a_i \leq (\overline{x}+2S)$$

O equivalentemente $\left|\overline{x}-a_i\right| \leq 2S$

Donde:

$\overline{x}$ es la medida y s la desviación estándar, de los datos.

También se podría escribir para todo x_i, tal que x_i está en el intervalo $(\overline{x} \pm 2S)$.

Brevemente: Para todo $x_i \in (\overline{x} \pm 2S)$

B. REGLA EMPIRICA:

Cuando la muestra es grande y tiene aproximadamente forma de campana (curva normal), entonces, basados, precisamente en el modelo de distribución normal, se aplica la siguiente regla empírica.

Aproximadamente:

a) El 68% de los datos están a menos de una desviación estándar de la media. Están en el intervalo $(\overline{x} \pm S)$..

b) El 95%, están en el intervalo $(\overline{x} \pm 2S)$. A dos desviaciones estándar de $\overline{X}$

c) El 99%, están en el intervalo $(\overline{x} \pm 3S)$ A tres desviaciones estándar de $\overline{X}$

Curva Normal

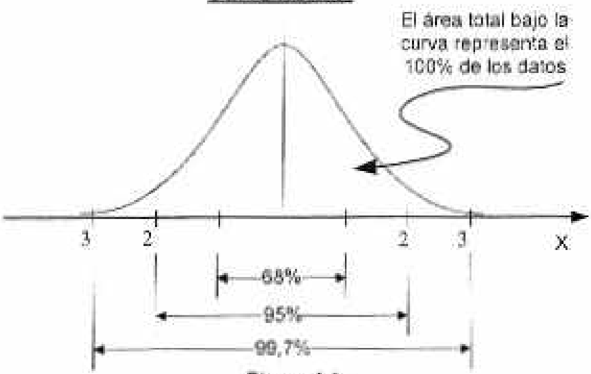


Figura 4.1.

C. VALORES ESTANDAR, Z .

Los valores de una variable X, podemos transformarlos en "valores estandarizados" ó "valores estándar Z", al usar la media y desviación estándar de x.

Los valores estándar o estandarizados se indica y define por.

$$Z = \frac{x_i - \overline{x}}{S}; \quad i = 1, 2, \ldots, n$$

Se indica con la letra Z y es igual a la diferencia de X menos su media y dividida entre la desviación estándar.

Donde:

X_i = Valor de i- ésimo de la variable x de una muestra de tamaño n.

$\overline{X}$ = la media muestral

S = Desviación estándar-c de la muestra

Z_i = "valor estándar" ó "valor Z" del valor $i-ésimo$ de la variable x.

OBSERVAR QUE.

1) El valor Z, se puede interpretar como la cantidad (el número de veces) de desviaciones estándar que dista x_i de la media, $\bar{x}$

2) También se puede expresar la interpretación anterior, como medir las distancias $(x_i - \bar{x})$ en la unidad de medida S ;es decir, medir en unidades de desviación estándar.

3) Si $x_i = 19$, $\bar{x} = 10$, $S = 6$, entonces. $Z = \frac{x-\bar{x}}{S} = \frac{19-10}{8} = \frac{9}{6} = 1{,}5$

significa que x_i es 1,5 desviaciones estándar mayor a la media

D. CONTROL ESTADÍSTICO DE PROCESOS.

Además del control de promedios $(\bar{x})$ de procesos, interesa el control de la variabilidad de los procesos.

Para vigilar y controlar de que el proceso de producción se mantenga estable, se determinan las desviaciones estándar-c de los subgrupos racionales; es decir, calcular S para muestras secuénciales del proceso y mostrar en forma idéntica a la explicada anteriormente en la grafica de corridas, ahora bajo el nombre de "Grafica S".

4.10 MOMENTOS

El momento r-ésimo de una variable X, respecto a un punto a, es la media de $(x-a)^r$; es decir,

$$\boxed{M\left[(x-a)^r\right]}$$

Donde:

M = es el operador media (media aritmética).

Entonces se lee. "Media de la potencia $r-ésima$ de, $(x-a)$".

CASOS:

1. $a = 0 \Rightarrow m_r = M\left[x^r\right]$ Se llama **Momento** respecto al origen ó sólo momento.

2. $a = \bar{x} \Rightarrow M_r = M\left[(x-\bar{x})^r\right]$ Se llama Momento respecto a la media o momento central o **centrado.**

3. $a = T \Rightarrow M\left[(x-T)^r\right]$ Se llama Momento respecto a un punto convencional T ó momento **reducido.**

FORMULAS. *Notaciones:*

MOMENTOS — **DATOS NO AGRUPADOS** — **DATOS AGRUPADOS**

1. $m_r = M\left[x^r\right] = \dfrac{\sum_{i=1}^{n} x_i^r}{n}$ ó $\dfrac{\sum x^r}{n}$; $m_r = \dfrac{\sum^{n} x^r n_i}{n}$

2. $M_r = M\left[(x-\bar{x})^r\right] = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^r}{n}$ ó $\dfrac{\sum(x-\bar{x})^2}{n}$; $m_r = \dfrac{\sum_{i=1}^{n}(x-\bar{x})^r n_i}{n}$

Donde:

1.- m_r = *es el momento r-ésimo de x*

2.- M_r = *es el momento centrado de orden r de x*

EJEMPLO. 4.9 Cálculo de Momentos de Orden r; r = 1, 2, 3 y 4

Para los datos de la Tabla que se encuentra en el ejemplo 4.4. sueldos de los trabajadores de la empresa ACE S.A. a) Calcular los momentos de primer y segundo orden y b) Calcular los momentos centrados de primer, segundo, tercer y cuarto orden.

SOLUCIÓN.

HOJA DE TRABAJO

PM (000 Bs.) x_i	Nro. de Trabajo n_i	$x_i n_i$	$(x_i-\bar{x})n_i$	$\bar{x}$ =2,94 $(x_i-\bar{x})^2 n_i$	$(x_i-\bar{x})^3 n_i$	$(x_i-\bar{x})^4 n_i$	$x_i^2 n_i$
1,4	6	8,4	-9,24	14,2296	-21,913584	33,74691936	11,76
2,1	9	18,9	-7,56	6,3504	-5,334336	4,48084224	39,69
2,8	14	39,2	-1,96	0,2744	-0,038416	0,00537824	109,76
3,5	13	45,5	7,28	4,0768	2,283008	1,27848448	159,25
4,2	7	29,4	8,82	11,1132	14,002632	17,64331632	123,48
4,9	0	0,0	0,00	0,0000	0,000000	0,00000000	0,00
5,6	1	5,6	2,66	7,0756	18,821096	50,06411536	31,36
Sumas	50	147,0	0,00	43,1200	7,820400	107,21905600	475,30
			$(\sum \div 50) =$	0,8624	0,156408	2,14438112	
			$\sqrt{(*)} =$	0,9287	0,395485	1,46437055	

1. $m_1 = M[x] = \dfrac{\sum_{i=1}^{n} x_i n_i}{n} = \dfrac{147,0}{50} = 2,94 \quad Luego: \; m_1 = \bar{x} = Bs.\,2940$

2. $m_2 = M\left[x^2\right] = \dfrac{\sum x_i^2 n_i}{n} = \dfrac{475,30}{50} = 9,506 \quad Luego: m_2 = 9'506.000\,(Bs)^2$

b. CALCULOS DE MOMENTOS CENTRADOS

1. $M_1 = M\left[(x-\bar{x})^1\right] = 0 \quad Luego: M_1 = 0$

2. $M_2 = M\left[(x-\bar{x})^2\right] = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2 n_i}{n} = \dfrac{43,12}{50} = 0,8624 <$

$$Luego: M_2 = V[x] = 862.400\,(Bs.)^2$$

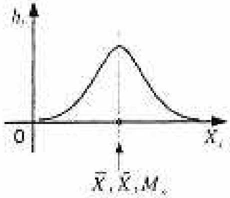
3. $$M_3 = M\left[(x-\overline{x})^3\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^3 n_i}{n} = \frac{7,8204}{50} = 0,156408 \quad \triangleleft$$

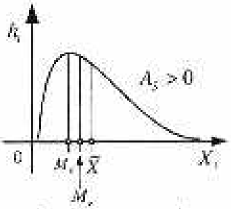
$$Luego: M_3 = 156'408.000\,(Bs.)^3$$

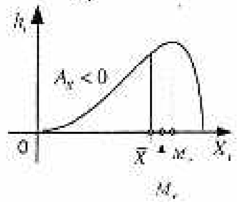
4. $$M_4 = M\left[(x-\overline{x})^4\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^4 n_i}{n} = \frac{107,219056}{50} = 2,14438112 \quad \triangleleft$$

$$Luego: M_4 = 2'144.381'120.000\,(Bs.)^4$$

4.11 MEDIDAS DE FORMA.







Las medidas de forma son unidades estadísticas que buscan medir dos aspectos de una distribución.

a. La simetría alrededor de un promedio. (como **medidas de Asimetría**) y

b. El alargamiento de la curva de frecuencias alrededor de la posición central o la frecuencia relativa de los valores muy alejados de la posición central. (como medidas de curtosis, alargamiento o de apuntamiento)

a) ASIMETRIA.-

Si se tiene una curva de frecuencias perfectamente simétricas, en forma de campana, entonces las medidas de tendencia central, los promedios, coinciden; es decir, la media, la mediana y la moda tienen el mismo valor.

Si la curva no es simétrica y muestra una cola a la derecha, tal que la moda es menor que la mediana y esta menor que la media, $(M_O < M_e < \overline{x})$, se dice que la distribución es asimétrica positiva $(A_S > 0)$ o una distribución con cola a la derecha.

Si la curva tiene una cola a la izquierda, tal que $(M_O > M_e > \overline{x})$ entonces se dice que es asimétrica negativa $(A_S < 0)$

DEFINICIÓN — ***COEFICIENTE DE ASIMETRIA DE MOMENTOS.***

El coeficiente de Asimetría, de momentos, se indica y define por,

$$A_s = \frac{M_3}{(M_2)^{\frac{3}{2}}} = \frac{Momento\ centrado\ de\ orden\ 3}{(Momento\ centrado\ de\ orden\ 2\ )^{\frac{3}{2}}} = \frac{M_3}{\sigma^3}$$

Casos:

a) $A_s = 0$, *Simétrica*

b) $A_s > 0$, *Asimétrica positiva (cola a la derecha)*

c) $A_s < 0$, *Asimétrica negativa (cola a la izquierda)*

Las definiciones de ***"momentos"*** *es la misma tanto para la población como para la muestra. Desarrollemos para la muestra, (momentos centrados para datos no agrupados).*

$$M_3 = M\left[(x-\overline{x})^3\right] = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^3}{n}$$

$$M_2 = M\left[(x-\overline{x})^2\right] = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n} = V[x]$$

EJEMPLO. 4.10 Cálculo del Coeficiente de Asimetría

Del ejemplo 4.4 se tiene: el momento centrado de segundo y tercer orden, $M_2 = 0,8624$ y $M_3 = 0.156408$

De donde, para la distribución de la variable X, "sueldo de los trabajadores" de la empresa ACE S.A., el coeficiente de asimetría es:

$$A_s = \frac{M_3}{M_2^{3/2}} = \frac{0.1564}{\sqrt{(0.8624)^3}} \doteq 0,1953$$

RESPUESTA: La distribución de sueldos tiene una ligera asimetría positiva o asimetría hacia la derecha.

DEFINICIÓN ***COEFICIENTE DE ASIMETRIA DE PEARSON.***

FORMULAS:

a). POBLACIÓN: $A_{sp} = \dfrac{3(\mu - Med)}{\sigma}$

b). MUESTRA: $A_{sm} = \dfrac{3(\overline{x} - Med)}{S}$

b) CURTOSIS O ALARGAMIENTO.-

La curva de frecuencias puede ser alargada o aplastada o como la curva normal.

Si la curva es más alargada o puntiaguda, la distribución se denomina leptocúrtica; si es aplanada, es platicúrtica y si es como la normal se dice que es mesocúrtica.

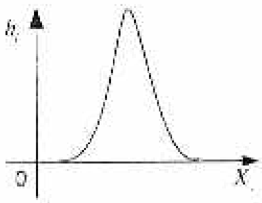


Figura 4.2.Leptocúrtica
(Alargada)

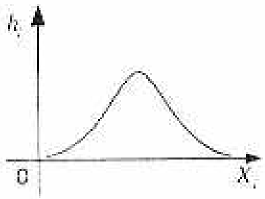


Figura 4.3.Mesocúrtica
(Normal)

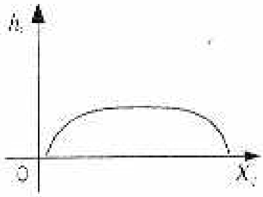


Figura 4.4. Planticúrtica
(Aplastada)

DEFINICIÓN: ***COEFICIENTE DE CURTOSIS O APUNTAMIENTO***

Por definición:

$$k = \frac{M_4}{M_2^2} = \frac{\text{momento centrado de orden 4}}{(\text{momento centrado de orden 2})^2} = \frac{M_4}{\sigma^4}$$

Casos:

1) *k = 3 ⇒ Mesocúrtica (Como la normal)*

2) *k < 3 ⇒ Platicúrtica o aplastada*

3) *k > 3 ⇒ Leptocúrtica o apuntada*

*La definición del **"momento centrado"** sea la población o la muestra, es misma. En la fórmula algunos símbolos cambian, como $\overline{x}$ y μ. Para la muestra:*

MUESTRA: Fórmulas del 4º y 2º momentos

$$M_4 \overset{Def}{=} M\left[(x-\overline{x})^4\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^4}{n}$$

$$M_2 \overset{Def}{=} M\left[(x-\overline{x})^2\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^2}{n} = V[x]$$

Nota.-1) Para la población se reemplazará $\overline{x}$ por μ

*2) Algunas veces **b = (k - 3)** se usa como medida de curtosis.*

EJEMPLO. 4.11 Cálculo del Coeficiente de Curtosis.

Del ejemplo 4.4 se tiene: el momento centrado de segundo y cuarto orden, $M_2 = 0{,}8624$ y $M_4 = 2{,}1444$.

El coeficiente de curtosis para la distribución de la variable X, "sueldos de los trabajadores" de la empresa ACE. S.A., es

$$k = \frac{M_4}{M_2^2} = \frac{2{,}1444}{\left(\sqrt{0{,}8624}\right)^4} = \frac{2{,}1444}{0{,}8624^2} = 2{,}88$$

RESPUESTA: como $k < 3$, decimos que es ligeramente platicúrtica, Prácticamente es mesocúrtica.

4.12 ANALISIS EXPLORATORIO DE DATOS:

El análisis exploratorio de datos comprende una serie de técnicas sencillas, sean numéricas o graficas, de la estadística descriptiva en el análisis de los de una muestra con la finalidad de,

- detectar la presencia de valores anómalos o atípicos y

- fundamentalmente resumir los datos y descubrir la naturaleza y estructura de la población madre.

Conocimiento necesario para la adecuada o correcta aplicación de los modelos y técnicas de la inferencia estadística.

Dos de las principales técnicas son,

- El diagrama de tallo y hojas (ver Cap. 2)
- El diagrama de caja, que describe el conjunto de datos, mediante el uso de lo que se ha venido en llamar, "resumen de los 5 números". Estos 5 números son:

1. Valor mínimo.
2. Primer cuartil (Q_1)
3. Mediana $(Me = Q_2)$
4. Tercer Cuartil (Q_3)
5. Valor Máximo

Considerando a las cuartiles (Q_1), $(Me = Q_2)$ y (Q_3) como los "estadígrafos resistentes" a pequeñas alteraciones de los datos.

*El análisis exploratorio, aplica las **técnicas gráficas** disponibles como **estrategia básica** para revelar la estructura de los datos.*

EL DIAGRAMA DE TALLO Y HOJAS.-

Es prácticamente un histograma con la gran ventaja de mostrar los valores (datos) individualmente sin pérdida de información, ya que en el histograma se pierden los valores individuales al agrupar los datos en clases.

EL DIAGRAMA DE CAJA.-

El diagrama de caja, como técnica gráfica del análisis exploratorio de datos, busca visualizar, especialmente, la asimetría de la distribución de los valores de la variable utilizando las tres cuartilas.

Además el diagrama muestra los valores extremos o atípicos y/o los anómalos.

En esta gráfica la "CAJA" o "RECTÁNGULO", limitado por Q_1 y Q_3, la primera y tercera cuartila, representa el 50% central de los datos. La línea vertical dentro de la caja indica la posición de la mediana (2º cuartil) y divide la caja en dos regiones y cada una representa el 25% de los datos.

Empezando a media altura de los lados del rectángulo (caja) se trazan a izquierda y derecha líneas horizontales punteadas (brazos) y se señalan en cada una las "barreras internas" y "externas", ubicada a 1,5 y 3,0 unidades del rango intercuartil, RIC, respectivamente. Concretamente:

BARRERAS INTERNAS:

Brazo izquierdo: $Q_1 - 1{,}5(RIC)$

Brazo derecho: $Q_3 + 1{,}5(RIC)$

BARRERAS EXTERNAS:

Brazo izquierdo: $Q_1 - 3{,}0(RIC)$

Brazo derecho: $Q_3 + 3{,}0(RIC)$

Donde.

El rango intercuartil: $RIC = (Q_3 - Q_1)$

Los datos ubicados entre las barreras se denominan "datos anómalos moderados" ó sólo "alejados."

Los datos más allá de la barrera externa se denomina datos "anómalos extremos" o sólo "muy alejados".

Se debe indicar con *, Asterisco, un dato "atípico"

En el análisis exploratorio los datos atípicos (alejados) deben ser investigados sobre su validez y confiabilidad. Si algún dato con error de medición u otro, no puede ser corregido debe rechazarse.

La figura 4.5 nos muestra las características de un diagrama de caja.

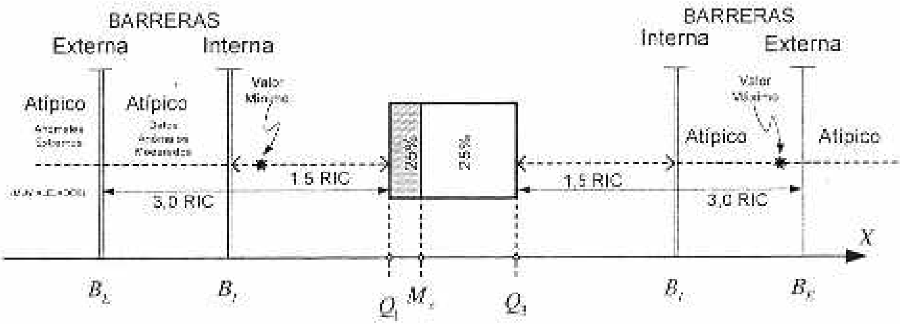


Figura 4.5. DIAGRAMA DE CAJA

4.13 ANALISIS ESTADISTICO DE ASOCIACIÓN ENTRE DOS VARIABLES

Con frecuencia el interés de un estudio, es la relación o grado de asociación entre dos variables.

Medidas estadísticas básicas de la relación entre dos variables son.

- La covarianza y
- El coeficiente correlación

Supongamos una muestra de n pares de valores observados (x_i, y_i); $i = 1, 2, \ldots, n$ de las variables X y Y, y su diagrama de dispersión, figura 4.6.

Al trazar dos perpendiculares a los ejes del plano cartesiano, en $\overline{x}$ y $\overline{y}$, el diagrama se divide en 4 cuadrantes.

Para cualquier punto P, de coordenadas (x_i, y_i), definamos las desviaciones.

$$a_i = (x_i - \overline{x}) \quad \text{y} \quad b_i = (y_i - \overline{y})$$

Por un examen rápido de los puntos del diagrama en los 4 cuadrantes.

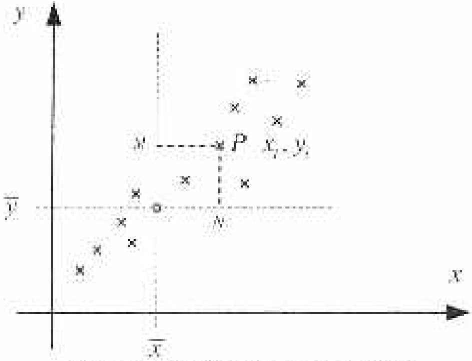


Figura 4.6. DIAGRAMA DE DISPERSIÓN

1) Si están en el cuadrante I, el producto $a_i b_i$ es positivo.

2) Si están en el cuadrante II, el producto $a_i b_i$ es negativo.

3) Si están en el cuadrante III, el producto $a_i b_i$ es positivo.

4) Si están en el cuadrante IV, el producto $a_i b_i$ es negativo.

De este modo $\sum a_i b_i$; es decir,

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \quad ; \quad (\alpha)$$

sirve como una medida de asociación entre las variables X y Y. Veamos una explicación:

a) Si la relación es positiva, la mayoría de los puntos estarán en los cuadrantes I y III, entonces (α) tiende a ser positiva.

b) Pero si la relación es negativa, la mayoría de los puntos estarán en II y IV, entonces (α) tiende a ser negativa.

c) Si no hubiera relación entre X y Y, los puntos se distribuirían en los 4 cuadrantes por igual entonces (α) tiende a ser muy pequeña.

d) Con base a la argumentación anterior, se define un estadígrafo útil, **la COVARIANZA para cuantificar el grado de asociación lineal entre dos variables.**

La medida de la covarianza adolece de dos defectos: La unidad de medida y el tamaño de la muestra, n

a) Si n aumenta, entonces crece la covarianza.

b) Si X se mide en metros primero, pero después se decide usar la unidad centímetros, sólo debido a este cambio, la covarianza crece.

Sin embargo, lo deseable es que la medida de la covarianza como medida del grado de relación entre dos variables determinadas no debería cambiar, (variar).

DEFINICIÓN: **COVARIANZA**

La covarianza entre dos variables X y Y con n pares de datos (x_i, y_i) dispuestos en una tabla de distribución conjunta, con valores distintos, x_i; $(i = 1, 2, \ldots, k)$ y y_j; $(j = 1, 2, \ldots, h)$, se indica y define por,

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{h}(x_i - \overline{x})(y_j - \overline{y})n_{ij} \quad donde: \quad n = \sum_{i=1}^{k}\sum_{j=1}^{h}n_{ij}$$

Donde:

1) n_{ij}, la frecuencia conjunta, denota el número de veces que se ha observado simultáneamente el par (x_i, y_j)

2) Si $n_{ij} = 1$, entonces la fórmula es;

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{h}(x_i - \overline{x})(y_j - \overline{y})$$

3) Si los datos se presentan en un listado de n pares (x_i, y_i), entonces la fórmula es como la (2); o sea,

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

$$Desarrollando: \ S_{xy} = \frac{\sum_{i=1}^{n}x_i y_i - n\overline{x}\,\overline{y}}{n} = \frac{\sum_{i=1}^{n}x_i y_i}{n} - \overline{x}\,\overline{y}$$

4) OBSERVACIÓN: En la aplicación práctica de estas fórmulas:

a) Población?.- La n se reemplaza por N

b) Muestra?.- La n se reemplaza por (n-1)

Para resolver estos defectos, se define otra medida, denominada coeficiente de correlación de Pearson.

DEFINICIÓN: **COEFICIENTE DE CORRELACIÓN**

Para una muestra, el coeficiente de correlación momento-producto de Pearson, se indica y define por,

$$r_{xy} \equiv r = \frac{S_{xy}}{S_x S_y}$$

$$donde:$$

$$S_{xy} = Covarianza\ entre\ x\ y\ y$$

$$S_x = Desviación\ estandar\ de\ x$$

$$S_y = Desviación\ estandar\ de\ y$$

Variación del valor del coeficiente de correlación.

$$\boxed{-1 \le r \le 1}$$

El valor del coeficiente de correlación varía entre -1 *y* $+1$

Para valores próximos a -1 *ó* -1 *, indica una fuerte asociación.*

Para $r = \pm 1$ *, hay una relación lineal perfecta. Si* $r - 1$ *relación negativa y si* $r - 1$ *relación positiva (directa).*

Para valores próximos a cero, se dice que no existe relación entre ambas variables.

EJEMPLO. 4.12 Cálculo de la Covarianza y el Coeficiente de Correlación.

En la tabla 4.1 Se ve el valor en libros por acción y el dividendo anual de 15 empresas de servicios en los Estados Unidos

a) **Calcular la covarianza y las varianzas. Suponer:**

1) unidad: dólares
2) unidades: bolivianos (Se deja como ejercicio)

b) **Calcular e interpretar el coeficiente de correlación muestral.**

TABLA No 4.1

Valor en libros Sus: x_i	22,44	20,89	22,09	14,48	20,73	19,25	20,37
Dividendo anual: y_i	2,40	2,98	2,06	1,09	1,96	1,55	2,16

Valor en libros: x_i	23,43	12,14	23,31	16,23	0,56	0,84	18,05	12,45
Dividendo: y_i	1,60	0,80	1,94	3,00	0,28	0,84	1,80	1,21

SOLUCIÓN:

a) **Calcular la covarianza y las varianzas. Suponer:**

Varianza de X:

$$V[x] \equiv S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2 - n\bar{x}^2}{n} = \frac{4.987,3342 - 15\left(\frac{250,26}{15}\right)^2}{15} =$$

$$= \frac{811.99636}{15} = 54,13309067 \qquad \therefore \boxed{S_x \doteq 7.357519}$$

Varianza de Y:

$$V[y] \equiv S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{\sum y_i^2 - n\bar{y}^2}{n} = \frac{52,4335 - 15\left(\frac{25,67}{15}\right)^2}{15} =$$

$$= \frac{8,503573}{15} = 0,566905 \qquad \therefore \boxed{S_y \doteq 0,752931}$$

HOJA DE TRABAJO

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
22,44	2,40	53,8560	503,5536	5,7600
20,89	2,98	62,2522	436,3921	8,8804
22,09	2,06	45,5054	487,9681	4,2436
14,48	1,09	15,7832	209,6704	1,1881
20,73	1,96	40,6308	429,7329	3,8416
19,25	1,55	29,8375	370,5625	2,4025
20,37	2,16	43,9992	414,9369	4,6656
26,43	1,60	42,2880	698,5449	2,5600
12,14	0,80	9,7120	147,3796	0,6400
23,31	1,94	45,2214	543,3561	3,7636
16,23	3,00	48,6900	263,4129	9,0000
0,56	0,28	0,1568	0,3136	0,0784
0,84	0,84	0,7056	0,7056	0,7056
18,05	1,80	32,4900	325,8025	3,2400
12,45	1,21	15,0645	155,0025	1,4641
250,26	**25,67**	**486,1926**	**4987,3342**	**52,4335**

Covarianza de X y Y, :

$$Cov(x,y) \equiv S_{xy} = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{n} = \frac{\sum x_i y_i - n\overline{x}\,\overline{y}}{n}$$

$$= \frac{486{,}1926 - 15\left(\frac{250{,}26}{15}\right)\left(\frac{25{,}67}{15}\right)}{15} = \frac{57{,}91432}{15} = 3{,}860954667$$

$$\therefore \boxed{S_{xy} \doteq 3{,}860955}$$

b) **Calcular coeficiente de correlación muestral.**

$$r = \frac{S_{xy}}{S_x S_y}$$

$$r = \frac{3{,}860955}{(7.357519)(0.752931)} \doteq 0{,}69696 \qquad \therefore \boxed{r = 0{,}6970}$$

4.14 ESTADISTICA DESCRIPTIVA: UN COMENTARIO FINAL.

Para reforzar y consolidar ideas: "Así como usted puede describir un determinado paisaje, también puede describir un determinado conjunto de datos. Precisamente esta tarea corresponde a la "Estadística Descriptiva", respondiendo a preguntas como ser:

a) ¿Cómo son los datos?

b) ¿Cuál el comportamiento general de los datos o de su distribución?

c) ¿Esta distribución de datos que propiedades tiene o mejor cuales son las particularidades o características propias en cuanto a su estructura y aspectos esenciales de (a) posición, (b) variabilidad y (c) forma?

d) ¿Cuál es el promedio de mayor representatividad de los datos?

e) ¿Cuál es el grado de variabilidad de los datos?

f) ¿En cuanto a la forma de la distribución cuál la medida de asimetría y de curtosis?

g) Finalmente, ¿Cuál la interpretación que se puede formular sobre las medidas descriptivas resumen, calculadas en el proceso del análisis de datos?.

h) Como comentario final, diremos que todo el trabajo de recolección, presentación, análisis e interpretación de los datos, tenga valor, debe realizarse en un marco estricto de objetividad, confiabilidad, veracidad y responsabilidad, no hacerlo así, es actuar en un marco no ético ó nada ético.

CAPÍTULO 4 PROBLEMAS RESUELTOS

MEDIDAS DE VARIABILIDAD Y DE FORMA

PROBLEMA 1.- Diagrama de Tallo y Hojas. Tabla de Distribución de Frecuencias. Desviación Estándar y Diagrama de Caja

Para los datos redondeados al boliviano más próximo, correspondiente a la variable de la tabla 4.2 "salarios semanales",x una muestra de 40 obreros, se pide:

a. Construir un diagrama de "tallo y Hoja".

b. Elaborar dos distribuciones de frecuencias de clase constante, de siete y seis clases, empleando la información de (a).

c. Elaborar una Tabla de distribuciones de frecuencias iniciando la primera clase con un límite inferior de Bs. 113 y aplicando un intervalo de clase constante de Bs. 9.

d. Calcular la desviación estándar y

e. Construir un "diagrama de caja" de (c).

TABLA 4.2 SALARIOS SEMANALES (En Bs.)

125	129	137	131
134	153	145	115
132	122	151	117
137	138	129	136
128	127	143	127
123	132	134	128
128	139	124	138
133	146	126	131
144	134	141	147
123	144	127	122

SOLUCIÓN.

a. Diagrama de "tallo y Hoja".

a.1. DIAGRAMA DE TALLO Y HOJAS

TALLO	HOJAS
11	5 7
12	3 3 2 4 2
	5 8 8 9 7 9 6 7 7 8
13	4 2 3 2 4 4 1 1
	7 8 9 7 6 8
14	4 4 3 1
	6 5 7
15	3 1

a.2. DIAGRAMA DE TALLO Y HOJAS AJUSTADO

TALLO	HOJAS	n_i
11	5 7	2
12	2 2 3 3 4	5
	5 6 7 7 7 8 8 8 9 9	10
13	1 1 2 2 3 4 4 4	8
	6 7 7 8 8 9	6
14	1 3 4 4	4
	5 6 7	3
15	1 3	2
	Total	40

b. **DISTRIBUCIOPNES DE FRECUENCIAS** Elaborar dos distribuciones de frecuencias de clase constante, de siete y seis clases, empleando la información de (a) (b.1). Siete clases: $\frac{R}{\#C}=\frac{40}{7}\doteq 5.7 \therefore \Delta = 6$ (b.2).- Seis clases: $\frac{R}{\#C}=\frac{40}{6}\doteq 6.6 \therefore \Delta = 7$

(b.1). Siete clases	
SALARIOS	**Nro. DE OBREROS**
113 – 118	2
119 – 124	5
125 – 130	10
131 – 136	9
137 – 142	6
143 – 148	6
149 – 154	2
TOTAL	40

(b.2).- Seis clases	
SALARIOS	**Nro. DE OBREROS**
113 – 119	2
120 – 126	7
127 – 133	13
134 – 140	9
141 – 147	7
148 – 154	2
TOTAL	40

OBSERVAR QUE: Son respuestas posibles.

c. DISTRIBUCIOPNES DE FRECUENCIAS. Con Δ = 9. y origen.

La respuesta se elabora con las columnas [1] y [2] de la Hoja de trabajo.

HOJA DE TRABAJO

SALARIO	Nro. OBREROS n_i	FRONTERAS $x'_{i-1} - x'_i$	PM. x_i	$x_i n_i$	$x_i^2 n_i$	N_i
[1]	[2]	[3]	[4]	[5]	[6]	[7]
113 – 121	2	112,50 – 121,50	117	234	27,378	2
122 – 130	15	121,50 – 130,50	126	1890	236,140	17
131 – 139	14	130,50 – 139,50	135	1890	255,150	31
140 – 148	7	139,50 – 148,50	144	1008	145,152	38
149 – 157	2	148,50 – 157,50	153	306	46,818	40
Total	**40**			**5,328**	**712,638**	-.-

d. DESVIACIÓN ESTÁNDAR

$$S^2=\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2 n_i}{n}=\frac{\sum_{i=1}^{k}x_i^2 n_i - n\bar{x}^2}{n}=$$

$$=\frac{712{,}638-40\left(5{,}328/40\right)^2}{40}=\frac{2{,}948.4}{40}=$$

$$=73.71 \Rightarrow S=\sqrt{73.71}=8.5854528 \quad \therefore \boxed{S=Bs.8{,}59}$$

e. DIAGRAMA DE CAJA

CALCULO DE LAS CUARTILAS

FORMA 1.- DATOS NO AGRUPADOS.

$$Q_1=X_{\frac{n}{4}+\frac{1}{2}}=X_{\frac{40}{4}+\frac{1}{2}}=x_{10.5}=127$$

$$Q_2=Me=X_{\frac{n}{2}+\frac{1}{2}}=x_{20.5}=132$$

$$Q_3=X_{\frac{3n}{4}+\frac{1}{2}}=x_{30.5}=138+0{,}5(139-138)=138+{,}05=138{,}5$$

$$RIC=Q_3-Q_1=138.5-127=11.5$$

$$Min=115 \quad Bs.$$

$$Max=153 \quad Bs.$$

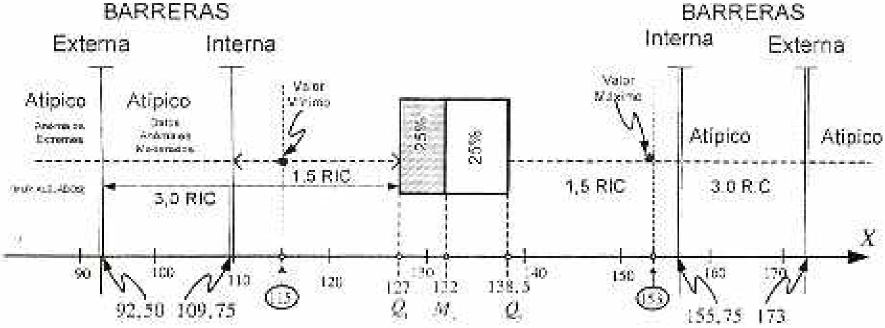


Figura 4. DIAGRAMA DE CAJA

BARRERAS:

A la derecha: *Interna*: $Q_3 + 1{,}5RIC = 155{,}75$

Externa: $Q_3 - 3.0RIC = 173{,}75$

A la izquierda: *Interna*: $Q_1 - 1{,}5RIC = 127 - 1{,}5(11{,}5) = 109.75$

Externa: $Q_1 - 3{,}0RIC = 127 - 3{,}0(11{,}5) = 92{,}50$

OBSERVAR QUE.

- No hay valores atípicos
- Poca sesgada a la derecha (Asimetría positiva)

FORMA 2.- DATOS AGRUPADOS.

FORMULA DE LA MEDIANA:

$$\boxed{M_e = X_{i=1} + \frac{\left(\frac{n}{2} - N_{i=1}\right)\Delta}{n_i}}$$

1) $\frac{n}{2} = \frac{40}{2} = 20; Med = 130{,}5 + \frac{(20-17)9}{14} = 130.5 - 1.93 = 132{,}43$

2) $\frac{n}{4} = 10; \quad Q_1 = 121{,}5 + \frac{(10-2)9}{15} = 121.5 + 4.8 = 126.3$

3) $3\frac{n}{4} = 30; \quad Q_3 = 130{,}5 + \frac{(30-17)9}{14} = 130.5 + 8.36 = 138.86$

$RIC = Q_3 - Q_1 = 138{,}86 - 126{,}30 = 12{,}56$

$Max = 153Bs.$

$Min = 115Bs.$

EJERCICIO: Construir con estos datos como en la forma 1.

OBSERVAR QUE:

En las tres cuartilas, calculadas en ambas formas, hay poca diferencia.

Veamos: "(en Bs.)"

NO AGRUPADOS	AGRUPADOS
$Q_1 =$ 127	$Q_1 =$ 126,3
$Q_2 =$ 132	$Q_2 =$ 132,4
$Q_3 =$ 138,5	$Q_3 =$ 138,9

PROBLEMA 2. Medidas de variabilidad.

En una empresa, la distribución de salarios tienen una media aritmética de Bs. 1 500 y una desviación estándar de Bs. 200. En la mesa de negociaciones se plantea las siguientes dos alternativas de solución al actual conflicto:

Alternativa A: un aumento general del 40%, de los salarios;

Alternativa B: un aumento general del 30% de los salarios y un bono adicional de Bs. 150; a cada obrero. ¿Cuál de las alternativas propuestas le conviene aceptar al sindicato de trabajadores de la empresa?. Fundamente su respuesta.

SOLUCIÓN:

Datos:

$x = "Salario"$

$\overline{x} = Bs.1.500 \quad ; \qquad S = Bs.200$

Alternativas:

A: 40% de incremento a los salarios.

B: 30% + Bs. 150 (bono) constante incremento.

Alternativa A. ; Y = 1,4 x

$$\overline{y} = M[Y] = M[1.4x] = 1,4\overline{x} = 1.4\times(1.500) = Bs.2100$$

$$S^2 = V[Y] = V[1.4x] = 1,4^2V[x] = 1.4^2\times S^2 = 1,4^2\times 200^2 = 280^2$$

$$\therefore \boxed{S = 280}$$

Alternativa B. Y = 1,3 x + 150

$$\overline{y} = M[Y] = M[1.3x+150] = 1.3\overline{x}+150 = Bs.2100$$

$$S^2 = V[Y] = V[1.3x-150] = 1.3^2V[x] = 1.3^2\times 200^2$$

$$\therefore \boxed{S = 260}$$

RESPUESTA:

$A: \quad \overline{x}_A = Bs.2100; \quad S = Bs.280$

$B: \quad \overline{x}_B = Bs.2100; \quad S = Bs.260$

En ambas alternativas la media es la misma, Bs. 2100, entonces deberá elegirse la alternativa de menor variabilidad, respecto a la, media. Por ello, se elige la alternativa B, tiene desviación estándar menor (Bs. 260 < Bs. 280). Relativamente es más beneficiosa para los trabajadores de salarios bajos.

PROBLEMA 3. Medidas de dos muestras

Una muestra de 60 datos y otra de 40 datos de la misma población, (variable x). la primera muestra tiene media de 120 y desviación estándar de 6. Y la segunda una media de 130 y una desviación estándar de 5. Reunir ambas muestras formando una sola muestra y calcular la varianza de los 100 datos

SOLUCIÓN:

Datos:

Muestra 1, $\quad n_1 = 60 \quad ; \qquad \overline{x}_1 = Bs.120 \quad ; \qquad S_1^2 = Bs.^2\ 36$

Muestra 2, $\quad n_2 = 40 \quad ; \qquad \overline{x}_2 = Bs.130; \qquad S_2 = Bs.^2\ 25$

MEDIA: De la unión de ambas muestras.

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1+n_2} \quad \text{Reemplazando:}$$

$$\overline{x} = \frac{60(120)+40(130)}{60+40} = \frac{12.400}{100} = 124 \therefore \overline{x} = 124$$

Por definición de varianza:
Muestra 1:

$$V[X_1] = \frac{\sum X_1^2}{n_1} - \bar{X}^2$$

$$36 = \frac{\sum X_1^2}{60} - 120^2 \Rightarrow \sum X_1^2 = 60(36 - 120^2) \therefore \sum X_1^2 = 866160$$

Muestra 2:

$$V[X_2] = \frac{\sum X_2^2}{n_2} - \bar{X}_2^2$$

$$25 = \frac{\sum X_2^2}{40} - 130^2 \Rightarrow \sum X_2^2 = 40(25 - 130^2) \therefore \sum X_2^2 = 677000$$

La suma de los cuadrados de los 100 datos (unión de ambas muestras) es

$$\text{Es } \sum X^2 = \sum X_1^2 - \sum X_2^2 = 866.160 - 677.000 = 1.543.160 \quad +$$

Varianza de la unión $\therefore [X_1 \oplus X_2]$

$$V[X] = V[X_1 \oplus X_2] = \frac{\sum X^2}{n} - \bar{X}^2$$

$$= \frac{1.543.160}{100} - 124^2$$

$$= 55.6$$

Luego: $\boxed{V[x] = Bs.^2\ 55.6}$

PROBLEMA 4. Coeficiente de variación. Planilla mensual de 130 empleados.

El coeficiente de variación de los ingresos de 130 empleados de una empresa es 65%. Después de ajustar todos los sueldos en Bs. 70, el coeficiente de variación pasa a 60%. La dirección ejecutiva fija después un sueldo mínimo de Bs. 470, lo que beneficia a 30 personas que antes de reajuste ganaban menos de Bs. 400, con un sueldo medio de Bs. 350. ¿A cuánto asciende el pago mensual total de la empresa, después del doble reajuste?

DATOS:

$n = 130$ — $S_{min} = Bs.\,470$

$a = Bs.\,70\ (ajuste)$ — $n_1 = 30\ personas\ menos\ de\ Bs.\,400$

$CV_1 = 65\%$ — $\bar{x}_1 = Bs.\,350$

$CV_2 = 60\%$

SOLUCIÓN:

$$CV_1 = \frac{S}{\bar{X}} = 0{,}65 \Rightarrow S = 0{,}65\bar{X}$$

$$CV_2 = \frac{S}{\bar{X}+70} = 0{,}60 \Rightarrow S = 0{,}60(\bar{X} - 70)$$

$$0{,}65\bar{X} = 0{,}60\bar{X} + (0{,}6)(70)$$

$$0{,}05\bar{X} = 42$$

$$\therefore \bar{X} = 840$$

ANTES DEL REAJUSTE:

Planilla de 130 empleados: Total	X	$n\bar{x}$	130(840)	109.200( )
Planilla de 30 empleados: Subtotal	X	$n_1\bar{x}_1$	30(350)	10.500( )
Planilla de 100 empleados: ganancia subtotal				98.700

DESPUÉS DEL REAJUSTE:

Planilla de (130-30)=100 personas,subtotal	98.700	(100)(70)	105.700
Planilla de 30 personas subtotal	(30)(470)		14.100
130 personas	TOTAL DESPUES DEL DOBLE REAJUSTE		Bs.119.800

PROBLEMA 5. Asimetría, curtosis. Estaturas de 100 estudiantes.

Utilice la distribución de frecuencias de estaturas de la tabla 4.3. Para encontrar la estatura media de los 100 estudiantes hombres de la universidad XYZ.

TABLA 4.3. Estatura de 100 estudiantes varones

Estatura (pulg)	Marcas de clase	Frecuencia		u_i	$u_i n_i$	u_i'	$u_i' n_i$	$x_i^2 n_i$	$u_i'^2 n_i$
60 - 62	61	5	305	-6	-30	-2	-10	18605	20
63 - 65	64	18	1152	-3	-54	-1	-18	73728	18
66 - 68	67	42	2814	0	0	0	0	188538	0
69 - 71	70	27	1890	3	81	1	27	132300	27
72 - 74	73	8	584	6	48	2	16	42632	32
		100	**6745**	**0**	**45**	**0**	**15**	**455803**	**97**

$$media: \quad \bar{x} = \frac{\sum_{i=1}^{n} n_i \bar{x}_i}{n} = \frac{6.745}{100} \Rightarrow \boxed{\bar{x} = 67,45 \ pulg}$$

Hacer: $u_i = x_i - T$ con $T = 67$

$$\left.\begin{array}{l} \therefore x = T - u \\ \bar{x} = T - \bar{u} \end{array}\right\} \bar{u} = \frac{\sum u_i n_i}{n} = \frac{45}{100} = 0,45$$

$$\bar{x} = 67 - 0,45 \quad Luego: \quad \boxed{\bar{x} = 67,45 \ pulg}$$

Hacer: $u_i' = \dfrac{x_i - T}{\Delta}$

$$\left.\begin{array}{l} \therefore x = T + \Delta u' \\ \bar{x} = T + \Delta \bar{u}' \end{array}\right\} \bar{u}' = \frac{\sum u_i' n_i}{n} = \frac{15}{100} = 0,15$$

$$\bar{x} = 67 + 3(0,15) = 67 - 0,45 \quad Luego: \quad \boxed{\bar{x} = 67,45 \ pulg}$$

$$V[x] = S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i}{n}$$

$$= \frac{\sum_i x_i^2 n_i - \bar{x}^2 n}{n} = \frac{\sum x_i^2 n_i}{n} - \bar{x}^2 = \frac{455.803}{100} - (67,45)^2$$

$$Luego: \quad \boxed{V[x] = 8,5275}$$

$$DS[x] = S = \sqrt{S^2} = \sqrt{8,5275} = 2,9202$$

$$Luego: \quad \boxed{DS[x] = S = 2,92 \ pulg}$$

$$u' = \frac{x_i - T}{\Delta}$$

$$V[x] = \Delta^2 V[u'] \left\{ \begin{array}{l} V[u'] = \dfrac{\sum u_i'^2 n_i}{n} - (\bar{u}')^2 \\ = \dfrac{97}{100} - (0,15)^2 = 0,97 - (0,15)^2 \\ = 0,9475 \end{array} \right.$$

$$V[x] = 3^2 (0,9475)$$

$$S^2 = V[x] = 8,5275 \ pulg^2$$

$$S = DS[x] = 2,92 \ pulg$$

Momentos centrados

$$M_1 \overset{Def}{=} M\left[(x-\overline{x})^1\right] = 0 \quad \boxed{Luego: M_1 = 0}$$

$$M_2 \overset{Def}{=} M\left[(x-\overline{x})^2\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^2 n_i}{n} = \frac{852,75}{100} = 8,5275 \triangleleft \boxed{M_2 = 8,53}$$

$$M_3 \overset{Def}{=} M\left[(x-\overline{x})^3\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^3 n_i}{n} = \frac{-269,325}{100} = -2,69325 \triangleleft \boxed{M_3 = -2,69}$$

$$M_4 \overset{Def}{=} M\left[(x-\overline{x})^4\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^4 n_i}{n} = \frac{19937,59313}{100} = 199,3759313 \triangleleft \boxed{M_4 = 199,38}$$

$$M_5 \overset{Def}{=} M\left[(x-\overline{x})^5\right] = \frac{\sum_{i=1}^{k}(x_i-\overline{x})^5 n_i}{n} = \frac{19578,11513}{100} = -195,7811513 \triangleleft \boxed{M_4 = 195,78}$$

Asimetría: $\alpha_3 = A_s = \frac{M_3}{S^3} = \frac{-2,69325}{(2,9202)^3} = -10,8153\%$

Curtosis: $\alpha_4 = \frac{M_4}{S^4} = \frac{199,3759}{(2,9202)^4} = 274,17\%$

CALCULOS CASIO : CFX-9850G PLUS

$\overline{x} = 67,45$

$$M_2 = \frac{1}{100}\left[(61-67,45)^2 * 5 + (64-67,45)^2 \ldots\right] \boxed{M_2 = 8,53}$$

PROBLEMA 6.	**Asociación entre dos variables**

Como medidas del movimiento general de precios de las acciones en el mercado de valores se usan el Promedio Industrial Don Jones (PIDJ) y el indice S&P 500 de Standard and Poor.

El PIDJ se basa en los movimientos de precio para 30 empresas grandes; el S&P 500 esta formado por las acciones de 500 empresas. Algunos dicen que el S&P 500 es una mejor medida del desempeño del mercado accionario, porque su base es más amplia. La Tabla 4.4 muestra los precios al cierre para el PIDJ y el S&P 500, durante las 10 últimas semanas de 2007.

TABLA 4.4. Precios al cierre para el Promedio Industrial Dow Jones y el indice S&P 500 durante las 10 últimas semanas de 2007

Fecha	Dow Jones	S&P 500
Octubre 26	8715	952
Noviembre 2	8442	925
Noviembre 9	8581	938
Noviembre 16	8572	938
Noviembre 23	8881	973
Noviembre 30	8823	965
Diciembre 7	9149	994
Diciembre 14	8838	963
Diciembre 21	8755	957
Diciembre 28	8678	946

Fuente: Datos hipotéticos

a. Construir un diagrama de dispersión de los datos

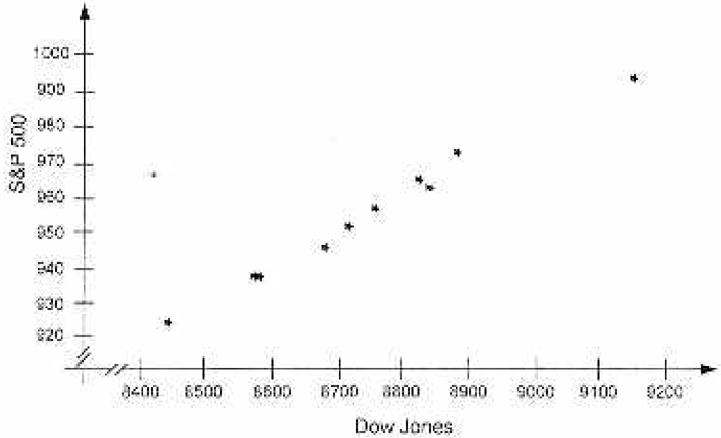
b. Calcular la covarianza

c. Calcular el coeficiente de correlación

d. Estos datos, ¿tienen mala correlación o su asociación es muy estrecha?

SOLUCIÓN.

a. Construir un diagrama de dispersión de los datos



b. Calcular la covarianza

Se realiza una hoja de trabajo:

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
8715	952	8296680	75951225	906304
8442	925	7808850	71267364	855625
8581	938	8048978	73633561	879844
8572	938	8040536	73479184	879844
8881	973	8641213	78872161	946729
8823	965	8514195	77845329	931225
9149	994	9094106	83704201	988036
8838	963	8510994	78110244	927369
8756	957	8379492	76667536	915849
8679	946	8210334	75325041	894916
87436	**9551**	**83545378**	**764855846**	**9125741**

La media es:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{87436}{10} = 8743,6$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{9551}{10} = 955,1$$

$$Cov(x,y) \equiv S_{xy} = \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{n} = \frac{\sum x_i y_i - n\overline{x}\,\overline{y}}{n}$$

$$= \frac{83545378 - 10\left(\frac{87436}{10}\right)\left(\frac{9551}{10}\right)}{10} = \frac{35254,4}{10} = 3525,44$$

$$\therefore \boxed{S_{xy} \doteq 3525.44}$$

c. Calcular el coeficiente de correlación

Varianza de X:

$$V[x] \equiv S_x^2 = \frac{\sum(x_i-\overline{x})^2}{n} = \frac{\sum x_i^2 - n\overline{x}^2}{n} = \frac{764855846 - 10\left(\frac{87436}{10}\right)^2}{10} =$$

$$= 35043.64 \qquad \therefore \boxed{S_x \doteq 187.1994658}$$

Varianza de Y:

$$V[y] \equiv S_y^2 = \frac{\sum(y_i-\overline{y})^2}{n} = \frac{\sum y_i^2 - n\overline{y}^2}{n} = \frac{9125741 - 10\left(\frac{9551}{10}\right)^2}{10} =$$

$$= 358.09 \qquad \therefore \boxed{S_y \doteq 18,9232661}$$

$$r = \frac{S_{xy}}{S_x S_y}$$

$$r = \frac{3525,44}{(187,1994658)(18,9232661)} \doteq 0,995205176 \qquad \therefore \boxed{r = 0,9952}$$

d. Estos datos,¿tienen mala correlación o su asociación es muy estrecha?

Como se denota r es muy próximo a 1 , lo cual indica una estrecha correlación entre ambos índices; además es positiva por tanto es una correlación directa, significa que el aumento de uno de los índices provoca el aumento del otro ó en caso contrario la disminución de uno de ellos provoca la disminución del otro.

CAPITULO. 4 PROBLEMAS COMPLEMENTARIOS

GRUPO A. MEDIDAS DE VARIABILIDAD Y DE FORMA

MÉTODOS:

1. Se tiene una muestra de tamaño 5, con valores de 10, 20, 12, 17 y 16. Calcule el rango y el rango intercuartil.

2. Se tiene una muestra de tamaño 5, con valores de 10, 20, 12, 17 y 16. Calcule la varianza y la desviación estándar.

3. Se tiene una muestra de tamaño 8, con valores de 27, 25,20, 15, 30, 34, 28 y 25. Calcule el recorrido intercuartil, la varianza y la desviación estándar.

APLICACIONES:

4. La oficina de visitantes de Hawai reúne datos sobre la cantidad de personas que visitan las islas. Los datos siguientes son una muestra representativa de visitantes (en miles) durante varios días de noviembre de 1994 (The Honolulu Advertiser, 28 de diciembre de 1994).

Del resto el Continente Americano, Canadá y Europa						
108.70	112.25	94.01	144.03	162.44	161.61	76.20
102.11	110.87	79.36	129.04	95.16	114.16	121.88
De Asia y el Pacífico:						
29.89	41.13	40.67	40.41	43.07	24.86	
31.81	21.60	27.34	64.57	32.98	41.31	

   a) Calcule la media y la mediana de la cantidad de visitantes de ambas fuentes de procedencias.

   b) ¿Qué comparaciones pueden hacerse entre las cantidades de visitantes de las dos fuentes de procedencia?

5. El diario Los Ángeles Times informa con regularidad el índice de calidad de aire de varias zonas del sur de California. Una muestra de valores del índice de la calidad de aire para Pomona dio los siguientes datos: 28, 42, 58, 48, 45, 55, 60, 49 y 50.

   a) Calcule el rango y el rango intercuartil

   b) Calcula la varianza del muestra y la desviación estándar de la muestra.

   c) Una muestra de índices de calidad de aire para Anaheim dio un promedio de 48.5, una varianza de 136 y una desviación estándar de 11.66. ¿Qué comparaciones pueden establecerse entre las calidades de aire en Pomona y en Anaheim, con base en estas medidas estadísticas descriptivas.?

6. Suponga que con los datos siguientes se trazan los histogramas de la cantidad de días que Dawson Supply , Inc ., y J.C. Clark Distributors necesitan para surtir pedidos (Véase la figura 3.2).

Días para entrega de Dawson Supply	11	10	9	10	11	11	10	11	10	10
Días para entrega de Clark Distributors	8	10	13	7	10	11	10	7	15	12

   Emplee el rango y la desviación estándar para sustentar la observación anterior, referente a que Dawson Supply tienen tiempos de entrega más consistentes y confiables

7. Un departamento de producción aplica un procedimiento de muestreo para verificar la calidad de artículos recién elaborados. Para ello, recurre a la siguiente regla de decisión en una estación de inspección: si muestra de 14 artículos tiene varianza mayor que 0.005 , se

debe parar la línea de producción y efectuar las representaciones necesarias. Suponga que acaban de reunirse los siguientes datos:

3.43	3.45	3.43	3.48	3.52	3.50	3.39
3.48	3.41	3.38	3.49	3.45	3.51	3.50

¿Debe detenerse la producción? ¿por que?

8. Se tiene una muestra con media de 30 y desviación estándar de 5. aplique el teorema de Chebyshev para determinar la proporción, o porcentaje, de los datos dentor de cada uno de los siguientes intervalos.

a) de 20 a 40.
b) de 15 a 45.
c) de 22 a 38.
d) de 18 a 42.
e) de 12 a 48.

9. Unos datos cuya distribución tiene forma de campana presentan una media de 30 y 5 desviación estándar. Aplique la regla empírica para determinar la proporción, o porcentaje, de los datos dentro de cada uno de los siguientes intervalos.

a) de 20 a 40.
b) de 15 a 445.
c) de 25 a 35.

10. El costo promedio de mano de obra en las reparaciones de televisores, en Chicago, es $ 90.06 (The Wall Street Journal, de 2 de enero de 1998). Suponga que a desviación estándar es $20,00.

a) ¿Cuál es el valor z para una reparación cuyo costo de mano de obra fue $ 17.00?.
b) ¿Cuál es el valor z para una reparación cuyo costo de mano de obra fue $ 18.00?
c) Interprete los valores z obtenidos en los incisos a y b. comente si se deberían considerar valores atípicos.

11. Según la agencia Roth Young Personnel Service, los salarios de gerente de tiendas departamentales van de $ 30.000 a $ 26.000 dólares (Nacional Business Employment Weeklly, 16 a 22 de octubre de 1994). Suponga que los datos siguientes son salarios anuales para una muestra de gerentes de tienda. Los datos están en miles de dólares.

33.7	45.4	44.0	47.5	59.6
45.1	37.7	43.9	48.3	53.0
39.5	42.9	51.0	35.6	41.5
49.5	45.4	58.2	55.4	62.3
32:2	45.9	47.6	56.2	56.8
48.8	31.3	51.2	43.2	54.4

a) Calcula la media y desviación estándar.

b) Un gerente de tienda n Memphis, Tennessee, gana $28,500 dólares al año. Calcule el valor z para este gerente y diga si cree que ese salario debe considerarse como un valor atípico.

c) Calcule los valores z para salarios de $30,000, $45.000, $60,000 y $75,000 dólares. ¿Debe considerarse atípico alguno de ellos?

12. Considere los datos de salarios del ejercicio 33. y emplee el teorema de Chebyshev para determinar el porcentaje de gerentes de tienda cuyos salarios estén entre los siguientes límites.

a) De $ 30,700 a $ 63,100.
b) De $ 28.200 a $ 65. 600.

13. En un articulo de la revista Atlantic Monthly (mayo de 1989) se describen los coeficientes de inteligencia y las tasas de natalicad. Las calificaciones de inteligencia tienen una distribución en forma de campana, con media 100 y desviación estándar15.

a) ¿Qué porcentaje de la población debe tener un coeficiente de inteligencia entre 85 y 115?
b) ¿Qué porcentaje de la población debe tener un coeficiente de inteligencia entre 70 y 130?
c) ¿Qué porcentaje de la población debe tener un coeficiente de inteligencia mayor que 130?
d) Una persona cuyo coeficiente de inteligencia sea mayor que 145 debe considerarse un genio. ¿Respalda la regla empirica esta afirmación'? Explique sus razones.

14. El estadounidense promedio pasó 4.4 horas diarias viendo televisión durante 1996 (Statistical Abstracto f the United Status, 1996). Suponga que la desviación estándar es 2.1 horas.

a) Aplique el teorema de Chebyshev para calcular el porcentaje de individuos que ven de 0.2 a 8.6 horas por dia.
b) Haga lo anterior para calcular el porcentaje de individuos que ve TV. De 0.0 a 8.8 horas por dia.
c) Aplique la regla empirica para calcular el porcentaje de individuos que ven TV. De 0.2 a 8.6 horas por dia. ¿Cree usted que es adecuada la aplicación de la regla empirica en este caso?

MÉTODOS:

15. Se tiene una muestra de tamaño 8 con los valores 27, 25, 20, 15, 30, 34, 28 y 25. elabore el resumen de cinco números para esos datos.

16. Trace el diagrama de caja para los datos del ejercicio 35.

17. Elabore el resumen de cinco números y trace el diagrama e caja para los siguientes datos: 5, 15, 18, 10, 8, 12, 16, 10 y 6.

18. Un conjunto de datos tiene un primer cuartil de 42 y un tercero de 50. calcule los límites inferior y superior. ¿Se debe considerar atípico un valor de datos de 65

APLICACIONES:

19. Una de las metas de toda administración es ganar lo más posible en relación con el capital invertido en la empresa. Una medida de éxito en alcanzarla es el retorno sobre la aportación, que es la relación de las ganancias netas entre el valor de las acciones para 25 empresas (Standard & Poor's Stock Reports, noviembre 1997)

9.0	19.6	22.9	41.6	11.4
15.8	52.7	17.3	12.3	5.1
17.3	31.1	9.6	8.6	11.2
12.8	12.2	14.5	9.2	16.6
5.0	30.3	14.7	19.2	6.2

a) Forme el resumen de cinco números.
b) Determine los límites inferior y superior.
c) ¿Parece haber valores atípicos? ¿Cómo podría un analista financiero usar esta información?
d) Trace un diagrama de caja

20. A continuación aparece una muestra de ventas anuales, en millones de dólares, de 21 empresas farmacéuticas (Business Week, 25 de abril de 1994).

8408	1374	1872	8879	2459	11413
608	14138	6452	1850	2818	1356
10498	7478	4019	4341	739	2127
3653	5794	8305			

a) Forme el resumen de cinco números.

b) Calcule los limites inferior y superior.
c) ¿Parece haber valore atípicos?

21. Las ventas de Jonson & Jonson publica calificaciones de funcionamiento y de calidad para muchos productos de consumo. Se publicaron calificaciones generales de una muestra de 16 Videocaseteras de precio intermedio en Consumer Reports 1992 buying Guide. Las marcas y las calificaciones aparecen en la siguiente tabla.

Fabricante	Calificación
Fisher	77
General Electric	81
Hitachi	89
J.C. Penny	78
JVC.	79
Magnavox	80
Montgomery Ward	78
Mitsubishi	90
Panasonic	77
Philips	73
Quasar	72
Radio Shack	76
RCA	79
Sanyo	75
Sony	86
Toshiba	79

a) Determine la calificación promedio y la mediana general.
b) Determine el primer y el tercer cuartil.
c) Elabore el resumen de cinco números.
d) Una evaluación similar de cámaras de video proporcionó calificaciones que tuvieron una media de 82.56, una desviación estándar 6.39 y un resumen de cinco números 75, 77, 82, 86, 93. comparé los datos de calificaciones de Consumer Reports de Videocaseteras con los de cámara de video. Trace los diagramas de caja de ambos.
e) ¿Hay valores atípicos en los datos de Videocaseteras? Explique su respuesta.

22. El informe del Instituto de Pérdidas en Accidentes y Choques, sobre Datos de pérdidas en Carreteras (septiembre de 1988) califica los modelos de automóviles con base en la cantidad de pólizas reclamadas a causa de accidentes. Un índice de calificaciones de 100 se considera normal. Un menor valor de índice de calificación se considera mejor, porque indica un modelo de automóvil más seguro. A continuación se observan las calificaciones de 20 automóviles medianos y 20 compactos.

Automóviles medianos	81	91	93	127	68	81	60	51	58	75
	100	103	119	82	128	76	68	81	91	82
Automóviles compactos	73	100	127	100	124	103	119	108	109	113
	108	118	103	120	102	122	96	133	80	140

a) Forme un resumen de cinco números para los coches medianos y otro para los compactos.

b) Trace los diagramas de caja.

c) Con base en los resúmenes, llega a una conclusión acerca de la seguridad de lso automóviles medianos a comparación con los compactos.

MÉTODOS:

23. A continuación, cinco observaciones de dos variables.

x_i	4	6	11	3	16
y_i	50	50	40	60	30

a) Trace un diagrama de dispersión con x en el eje horizontal.

b) ¿Qué indica el diagrama de dispersión que trazó en el inciso a cerca de la relación entre dos variables?

c) Calcule e interprete la covarianza de la muestra de los datos.

d) Calcule e interprete el coeficiente de correlación de la muestra de los datos.

24. En seguida aparecen cinco observaciones de dos variables.

x_i	6	11	15	21	27
y_i	6	9	6	17	12

a) Trace un diagrama de correlación para estos datos.

b) ¿Qué indica el diagrama de correlación sobre una posible relación entre x y y?

c) Calcule e interprete la covarianza de la muestra de los datos.

d) Calcule e interprete el coeficiente de correlación de la muestra de los datos.

25. Un estudio del departamento de tránsito acerca de la velocidad y la distancia recorrida para automóviles medianos arrojó los datos siguientes..

x_i	30	50	40	55	30	25	60	25	50	55
y_i	28	25	25	23	30	32	21	35	26	25

26. Calcule e interprete el coeficiente de correlación de la muestra para estos datos

Tabla 4.5 PRECIOS AL CIERRE PARA EL PROMEDIO INDUSTRIAL DOW JONES Y EL INDICE & P 500 DURANTE LOS ULTIMOS MESES DE 1997

Fecha		Dow Jones	S & P 500
Octubre	24	7715	942
Octubre	31	7442	915
Noviembre	7	7581	928
Noviembre	14	7572	928
Noviembre	21	7881	963
Noviembre	28	7823	955
Diciembre	5	8149	984
Diciembre	12	7838	953
Diciembre	19	7758	947
Diciembre	26	7679	936

27. El Promedio Industrial Dow Jones (PIDJ) y el índice S&P 500 de Standard and Poor se usan como medidas del movimiento general en el mercado accionario. El PIDJ se basa en los movimientos de precios de 30 empresas grandes; el S&P 500 está formado por las acciones de 500 empresas. Algunos dicen que el S&P 500 es una mejor medida del desempeño del mercado accionario, porque su base es más amplia. La Tabla 4.5 muestra los precios al cierre para PIDJ y el S&P 500, durante las 10 últimas semanas de 1997.

a) Calcule el coeficiente de correlación de la muestra para los precios al cierre.

b) Estos datos, ¿tienen mala correlación o su asociación es muy estrecha?.

28. En la tabla 4.6 se ve el valor en libros por acción y el dividendo anual de 15 empresas de servicios (Barron´s 2 de enero 1995).

a) Trace un diagrama de dispersión con el valor en libros en el eje horizontal.

b) Calcule e interprete el coeficiente de correlación muestral.

TABLA 4. 6 VALORES EN LIBROS POR ACCION PARA 15 EMPRESAS ELECTRICAS

Compañía	Valor en Libros ($)	dividendo Anual ($)	Compañía	Valor en Libros ($)	dividendo Anual ($)
Am Elec.	22.44	2.40	Centerior	12.14	0.80
Con Ed.	20.89	2.98	Cons N Gas	23.31	1.94
Detrit Ed.	22.09	2.06	Houston Ind.	16.23	3.00
Niag Mohn	14.48	1.09	NorAm Enrgy	0.56	0.28
Pac G & E	20.73	1.96	Panh East	0.84	0.84
Peco	19.25	1.55	Peoples En	18.05	1.80
Pub Sv Ent	20.37	2.16	SCEsopr	12.45	1.21
Unicomp Co	26.43	1.60			

29. Vea los datos de la muestra en la siguiente distribución de frecuencias

Clase	Promedio	Frecuencia
3-7	5	4
8-12	10	7
13-17	15	9
18-22	20	5

a) Calcula la media y la muestra.

b) Calcula la varianza y la desviación estándar de la muestra

APLICACIONES:

30. El promedio de calificaciones para los alumnos de una preparatoria se basa en un cálculo de media ponderada. En la mayor parte de las preparatorias (en estados Unidos). Se asignan los siguientes valores a las clasificaciones: A (4), B (3), C (2), D (1) y F (0). Después de acreditar 60 horas en cursos, un alumno ha obtenido 9 horas de A, 15 de b, 33 de C y 3 de D.

a) Calcule la calificación promedio del alumno.

b) Los alumnos de esta preparatoria deben tener un promedio de 2.5 en sus primeras 60 horas de cursos para ingresar a la carrera de Administración. ¿Será admitido este alumno?

31. En una gasolinera se formó la siguiente distribución de frecuencias de galones de gasolina vendidos por automóvil, en una muestra de 680 vehículos.

Clase	Frecuencia
0-4	74
5-9	192
10-14	280
15-19	105
20-24	23
25-29	6
	Total 680

32. Calcule la media, la varianza y la desviación estándar para estos datos agrupados. Si la gasolinera espera atender a unos 120 automóviles en cierto día, ¿Cuánto puede ser la cantidad vendida de galones de gasolina?.

33. En una encuesta de suscriptores de la reviste Fortune, se preguntó lo siguiente: "¿cuántos de los últimos cuatro números ha leído usted?" La siguiente distribución de frecuencias resume 500 respuestas (perfil Nacional de Suscriptores de Fortune, 1994).

Cantidad leída	Frecuencia
0	15
1	10
2	40
3	85
4	350
	Total 500

a) ¿Cuál es la media de la cantidad de revista que lee un suscriptor de Fortune.?
b) ¿Cuál es la desviación estándar de la cantidad de revistas leídas?

MISCELANEOS:

34. El estadounidense promedio gasta $ 65,88 dólares mensuales saliendo a cenar (The Des Moines Register, 5 de diciembre de 1997). En una muestra de adultos jóvenes se obtuvieron los siguientes gastos, en dólares, durante el mes anterior.

253	101	245	467	131	0	225
80	113	69	198	95	129	124
11	178	104	161	0	118	151
55	152	134	169			

a) Calcule la media, la mediana y la moda.
b) En vista de los resultados del inciso a, ¿Aparentemente estas personas gastan más o menos lo mismo que un estadounidense promedio?
c) Calcule el primer y el tercer cuartil.
d) Calcule el rango y el rango intercuartil.
e) Calcule la varianza y la desviación estándar.¿Hay algunos valores atípicos?

35. A continuación un amuestra de vencimiento de 10 bonos, negociados en la bolsa de valores de Nueva York (Barron´s 5 de enero de 1998)

Emisor	Rendimiento (%)	Emisor	Rendimiento (%)
Argosy	12.6	Caterpillar	6.3
Chase Maniatan	6.7	Dow	6.8
IBM	7.0	Lucent	6.7
Mobil	7.3	Pacifific Bell	6.7
RJR Nabisco	8.1	Service Mdse.	8.6

a) Calcule las siguientes medidas descriptivas:
b) Media, median y moda.
c) Primer y tercer cuartiles.
d) Rango y rango intercuartil.
e) Varianza y desviación estándar.
f) Coeficiente de varianza.

36. Se efectuó una encuesta acerca de la capacidad de los fabricantes de computadoras para resolver con rapidez sus problemas (PC Computing, noviembre de 1997). Se obtuvieron los siguientes resultados:

Empresa	**Días para resolver el problema**	**Empresa**	**Días para resolver el problema**
Compaq	13	Gateway	21
Packard	27	Digital	27
Quartex	11	IBM	12
Dell	14	Hewlett – Packard	14
NEC	14	AT & AT	20
AST	17	Toshiba	37
acer	16	Micron	17

a) ¿Cuáles son la media y la mediana de la cantidad de días necesarios para resolver problemas?

b) ¿Cuál fue la varianza y la desviación estándar?

c) ¿Qué fabricante tiene el mejor registro?

d) ¿Cuál es el valor de z para Packard Bell?

e) ¿Hay algunos valores atípicos?.

37. Una muestra de 10 acciones en la bolsa valores de nueva York. (The Wall Street Journal, 26 de enerpo 1998) tiene las siguientes relacione de precio a rendimiento: 8, 4, 6, 7, 3, 11, 4, 6, 4 7. Con estos datos calculo la media, la mediana, la moda, el rango, la varianza y la desviación estándar.

38. Dos modos que usan los empleados para ir a trabajar diariamente son el transporte público y el automóvil. A continuación vamos una muestra de tiempos de cada modo. Las cifras son minutos.

Transporte público	28	29	32	37	33	25	29	32	41	34
Automóvil	29	31	33	32	34	30	31	32	35	33

a) Calcule la media de la muestra del tiempo que se lleva en cada modo de trasporte.

b) Calcule la desviación estándar de la muestra para cada modo de trasporte.

c) Con base a los resultados de los incisos a y b,

d) ¿Qué modo de trasporte debe preferirse? Explique sus razones.

e) Trace un diagrama de caja para cada modo.

f) Al comparar los diagramas de caja, ¿se respalda la conclusión sus del inciso?

39. Las calificaciones del examen final de 25 alumnos de estadística son las siguientes

56	77	84	84	42	61	44	95	98	84
93	62	96	78	88	58	62	79	85	89
89	97	53	76	75					

a) Determine el resumen de cinco números.

b) Trace un diagrama de caja

40. En la revista Road & Track de octubre de 1994 apareció la siguiente muestra de las evaluaciones de neumáticos y la capacidad de carga de los neumáticos automotrices.

Calificación del neumático	Capacidad de Carga
75	853
82	1047
85	1135
87	1201
88	1235
91	1356
92	1389
93	1433
105	2039

a) Trace un diagrama de dispersion para los datos con la evaluación de los neumáticos en el eje horizontal.

b) ¿Cuál es el coeficiente de correlación de la muestra, y que indica ésta acerca de la relación entre calificación de neumático y capacidad de carga?

41. En una muestra de cinco fondos en el mercado de dinero, que se presenta a continuación, se anotan los plazos (en días y la cantidad depositada. Calcule la media ponderada para determinar el plazo promedio (en días) de los depósitos en esos cinco fondos.

Plazo	Valor (millones de dólares)
20	20
12	30
7	10
5	15
6	10

42. En una técnica de pronósticos, que se llama promedios móviles, se emplea la media de los n periodos más recientes para pronosticar el valor siguiente en una serie temporal de datos. Con un promedio móvil de tres periodos, se usan los tres más recientes para calcular el pronóstico se tiene un producto cuya demanda, durante los tres primeros meses de este año fue: enero, 800 unidades, febrero, 70 unidades y marzo 900 unidades.

a) Cuál es el pronóstico para abril, con promedio móvil de tres meses.

b) Una variación de esta técnica se llama promedios móviles ponderados. La ponderación permite que los datos de la serie más reciente reciban más peso o más importancia en el cálculo del pronóstico. Por ejemplo, en un móvil promedio ponderado de tres meses podría asignarse un peso de tres a los datos de un mes de antigüedad, de 2 para los de dos meses y de 1 para los tres meses de antigüedad. Con los datos anteriores calcule un promedio móvil ponderado de tres meses, que será el pronóstico para abril

43. A continuación se presenta una distribución de frecuencias de la duración de 20 llamadas telefónicas de larga distancia, en minutos. Calcule la media, la varianza y la desviación estándar de los datos.

Duración de la llamada en minutos	Frecuencia
4-7	4
8-11	5
12-15	7
16-19	2
20-23	1
24-27	1
	Total 20

44. Las cuentas por comidas en el restaurante francés La Maison tiene la distribución de frecuencias de la tabla siguiente. Càlcule la media, la varianza y la desviación estándar de esos datos.

Duración de la llamada en minutos	Frecuencia
25-34	2
35-44	6
45-54	4
55-64	4
65-74	2
75-84	2
	Total 20

GRUPO B. MEDIDAS DE VARIABILIDAD Y DE FORMA Y GRAFICOS

Estadística General y aplicada Capítulos 1,2,3 y4

1) Las cuentas por comidas en el restaurante francés La Maison tiene la distribución de frecuencias de la tabla siguiente. Calcule la media, la varianza y la desviación estándar de esos datos.

34	58	87	120	33	25	78	105	27	43
65	78	36	57	74	130	123	96	38	25

2) El director de un almacén está interesado en el número de quejas recibidas por el departamento de atención al cliente acerca de la calidad de los productos eléctricos vendidos. Los datos correspondientes a un periódico de diez semanas aparecen en la siguiente tabla.

Semana	Número de quejas	Semana	Número de quejas
1	13	6	4
2	15	7	21
3	8	8	11
4	16	9	3
5	8	10	15

a) Hallar el número medio de quejas semanales de esta población.
b) Hallar la mediana del número de quejas semanales de esta población.

3) Una compañía posee 12 parcelas de terreno edificable. Los tipos impositivos a efectos de tasación durante 1994 fueron para estas parcelas (en tanto por ciento).

21	22	27	36	22	29
22	23	22	28	36	33

Para esta población:

a) Hallar la media poblacional.
b) Hallar la mediana.
c) Hallar la moda.

4) Se toma una muestra de diez economistas y se les pide una predicción acerca del porcentaje de crecimiento del IPC para el próximo año. Las predicciones fueron .

3,6	3,1	3,9	3,7	3,5
3,7	3,4	3,0	3,6	3,4

a) Hallar la media muestral de las predicciones.
b) Hallar la mediana muestral

5) Se toma una de ocho estudiantes que viven en la residencia de cierto campus, y se les pide que evalúen la calidad de la comida en el comedor de la residencia en una escala que va de 1 (mala) a 7 (excelente). Las puntuaciones fueron.

2 4 2 3 5 4 3 2

a) Hallar la media muestral.
b) Hallar la mediana muestral

6) Una empresa está interesada en encontrar economistas. Para evaluar el nivel de los candidatos, se elabora un examen de 50 preguntas. En un estudio piloto, se somete a este examen a una muestra de diez economistas. los números de respuestas acertadas por los individuos de esta muestra fueron:

42	29	21	37	40
33	38	26	39	47

a) Hallar la media muestral del número de respuestas correctas.
b) Hallar la mediana de esta muestral

7) Una cadena de grandes almacenes tiene, diez establecimientos. Se analiza el volumen de ventas durante el periodo de Navidad y se comparan con las obtenidas en el mismo periodo del año anterior. Los porcentajes de incremento de ventas en dólares de los diez establecimientos fueron:

10,2	3,1	5,9	7,0	3,7
2,9	6,8	7,3	8,2	4,3

Para esta población:

a) Hallar la media del porcentaje de incremento de ventas en dólares.
b) Hallar la mediana.

8) Da un ejemplo económico real para el que la medida de centralización más adecuada sea.

Para esta población:

a) La media
b) La mediana.
c) La moda.

9) Retomemos los datos del Ejercicio 2, acerca del número de quejas semanales recibidas por el departamento de atención al cliente en un periodo de diez semanas.

a) Hallar la varianza poblacional y la desviación típica.
b) Hallar la media de las desviaciones absolutas.
c) Hallar el rango o recorrido.
d) Hallar el rango intercuartílico.

10) Situémonos en el contexto del Ejercicio 3, en el que se recogían los tipos impositivos de 12 parcelas.

a) Hallar la varianza muestral y la desviación típica.
b) Hallar la media de las desviaciones absolutas.
c) Hallar el recorrido.
d) Hallar el rango intercuartílico.

11) Retomemos los datos del Ejercicio 4, que se referían a predicciones acerca del porcentaje de incremento del IPC.

a) Hallar la varianza muestral y la desviación típica.
b) Hallar el rango intercuartílico

12) Para los datos del Ejercicio 5, que recogían las calificaciones otorgadas por una muestra de ocho estudiantes al servicio de comedor de la residencia de campus:

a) Hallar la varianza muestral y la desviación típica.
b) Hallar el rango intercuartílico.

13) Para los datos del Ejercicio 7, que se referían a los incrementos de ventas en diez establecimientos:

a) Hallar la varianza muestral y la desviación típica.
b) Hallar el recorrido.
c) Hallar el rango intercuartílico.

14) Se toma una muestra de 12 estudiantes matriculados en estadística y se les pregunta por el número de horas que emplearon en estudiar la asignatura en la semana anterior al examen final:

12	7	4	16	21	5
9	3	11	14	10	6

a) Hallar la media muestral.
b) Hallar la mediana muestral
c) Hallar la varianza muestral y la desviación típica.

d) Hallar el rango intercuartílico

15) Un auditor ha comprobado que el valor de las facturas pagadas por cierta empresa norteamericana tiene una media de 295 dólares, y una desviación típica de 63 dólares.

a) Hallar un intervalo en el cual se pueda garantizar que se encuentran el 60% de estos valores.
b) Hallar un intervalo en el cual se pueda garantizar que se encuentra el 80% de estos valores.

16) Los neumáticos de cierta marca tienen una duración de vida media de 29,000 kilómetros y desviación típica de 3.000 kilómetros.

a) Encontrar un intervalo en el que se pueda garantizar que se encuentran el 75% de los tiempos de vida de los neumáticos de esta marca.
b) Usando la regla empírica, encontrar un intervalo en el cual se estime que se encuentran aproximadamente el 95% de los tiempos de vida de los neumáticos de esta marca.

17) La tabla que aparece a continuación recoge los porcentajes de disminución de las acciones de los 25 mayores fondos de pensiones el viernes 13 de noviembre de 1989. Construir un histograma que sintetice estos datos de forma adecuada.

4,7	4,7	4,	4,7	3,0
4,4	5,0	3,3	3,8	6,4
3,3	3,6	4,7	4,4	5,4
3,0	4,9	5,2	4,2	3,3
4,1	6,0	5,8	4,9	3,8

18) La tabla que aparece a continuación recoge los porcentajes de trabajadores sindicados en cada uno de los 50 estados norteamericanos. Construir un histograma que sintetice estos datos de forma adecuada.

Estado	Trabajadores Sindicados (%)	Estado	Trabajadores Sindicados (%)
Alabama	19,2	Montana	24,1
Alaska	26,2	Nebraska	15,3
Arizona	13,8	Nevada	22,9
Arkansas	15,0	Nuevo Hampshire	13 ,3
California	23,7	Nueva Jersey	23,0
Colorado	15,2	Nuevo Mexico	12,1
Connecticut	21,9	Nueva York	39,2
Delaware	21,7	California del Norte	6,5
Florida	11,7	Dakota del Norte	14,7
Georgia	13,6	Ohio	29,5
Hawai	32,1	Oklahoma	13,5
Idaho	14,3	Oregon	23,1
Illinois	31,5	Pennsylvania	34,2
Indiana	29,3	Rhode Island	27,1
Iowa	19,2	California del Sur	6,7
Kansas	12,8	Dakota del Sur	10,3
Kentucky	22,4	Tennessee	17,7
Louisiana	16,0	Texas	11,0
Maine	18,3	Utah	13,0
Maryland	21,0	Veermont	17,5
Massachussets	24,4	Virginia	12,7
Michigan	34,6	Washington	33,1
Minnesota	24,4	West Virginia	36,8
Mississippi	12,4	Wisconsin	27,8
Missouri	30,0	Wyoming	14,9

19) Se somete a los 40 estudiantes de una clase a una encuesta para evaluar a su profesor, según una escala que va de 1 (malo) hasta (excelente) . Los resultados se recogen en la siguiente tabla.

PUNTUACIÓN	NÚMERO DE ESTUDIANTES
1	1
2	7
3	15
4	10
5	7

a) Hallar la media.
b) Hallar la mediana de estas puntuaciones.
c) ¿Cuál es la puntuación modal?
d) Hallar la varianza y la desviación típica de esta población de puntuaciones

20) La tabla que aparece a continuación recoge las puntuaciones obtenidas en un examen por los 40 estudiantes de una clase. Construir un histograma que sintetice estos datos de forma adecuada.

54	56	56	59	60
62	62	66	67	68
68	70	70	73	73
73	75	77	78	79
79	81	81	82	83
83	85	86	86	88
89	89	90	90	91
93	93	94	95	98

21) Retomemos los datos del Ejercicio 19, que recogían las puntuaciones obtenidas en un examen por cada uno de los cuarenta alumnos de una clase.

a) A partir de la agrupación de datos usada para construir el histograma, estimar.

1) La media poblacional.
2) La desviación típica poblacional.
3) La mediana de esta población.
4) El rango intercuartílico de esta población.

b) Usando los datos del Ejercicio 19, calcular directamente los cuatro estadísticos poblacionales estimados en el apartado (a) y disminuir la calidad de dichas estimaciones.

22) Retomemos los datos del ejercicio 17, acerca de los porcentajes de trabajadores sindicados en cada uno de los 50 estados norteamericanos:

a) A partir de la agrupación de datos empleada para construir el histograma:

1) Estimar la media de los porcentajes de sindicalización.
2) Estimar la mediana.
3) Estimar la desviación típica
4) Estimar el rango intercuartílico.

b) Calcular ahora directamente la media y la desviación típica de las 50 observaciones y compararlas con las obtenidas en el apartado (a).

23) Se toma una muestra de 25 estudiantes. La tabla siguiente recoge la cantidad de tiempo empleado por cada uno de los miembros de dicha muestra en preparar un examen.

TIEMPO DE ESTUDIO (HORAS)	0 – 4	4 – 8	8 – 12	12 – 16	16 -20
NÚMERO DE ESTUDIANTES	3	7	8	5	2

a) Dibujar el histograma.
b) Hallar las frecuencias relativas.
c) Hallar las frecuencias relativas acumuladas y dibujar el histograma correspondiente.
d) Estimar la media muestral del tiempo del tiempo muestral.
e) Estimar la desviación típica muestral.
f) ¿En qué clase está la media muestral?
g) ¿Cuál es la clase modal?

24) Se toma una muestra de 20 analistas financieros y se les pide que hagan una predicción sobre las ganancias por acción de cierta empresa norteamericana para el próximo año. Los resultados aparecen resumidos en la tabla siguiente:

PREDICCIÓN (Dólares por acción)	NÚMERO DE ANALISTAS
9,95 – 10,45	2
10,45 – 10,95	8
10,95 – 11,45	6
11,45 – 11,95	3
11,95 – 12,45	1

a) Dibujar el histograma.
b) Hallar las frecuencias relativas de la muestra.
c) Hallar la frecuencia acumuladas de la muestra.
d) Hallar e interpretar las frecuencias relativas acumuladas de la muestra.
e) Estimar la media muestral de la predicción.
f) Estimar la varianza muestral y la desviación típica de la predicción.
g) Estimar la mediana muestral de la predicción.
h) Estimar el rango intercuartílico muestral
i) ¿Cuál es la clase modal?

25) Durante una epidemia de gripe, los tiempos de espera en cierto centro de salud fueron más largos de lo habitual. La siguiente tabla resume la distribución de los tiempos de espera para una muestra de 20 pacientes que visitan en centro de salud durante este periodo.

TIEMPO DE ESPERA	0 – 1	1 – 2	2 – 3	3 – 4
NÚMER DE PACIENTES	6	9	4	1

a) Dibujar el histograma.
b) Hallar las frecuencias relativas de la muestra.
c) Hallar e interpretar las frecuencias relativas acumuladas de la muestra.
d) Estimar la media muestral del tiempo de espera.
e) Estimar la varianza muestral y la desviación típica.
f) Estimar la mediana muestral.
g) Estimar el rango intercuartílico muestral.
h) ¿Cuál es la clase modal para esta muestra?

26) Se dispone de la siguiente información acerca de las rentas familiares en los hogares de cierta ciudad americana.

Rentas Familiares (en dólares)	**Frecuencia Relativa**
10,000 – 15,000	0,20
15,000 – 20,000	0,18
20,000 – 30,000	0,14
25,000 – 30,000	0,12
30,000 – 40,000	0,14
40,000 – 50,000	0,14
50,000 – 60,000	0,08

a) dibujar el histograma.

b) Estimar la media poblacional de la renta familiar.
c) Estimar la desviación típica poblacional de la renta familiar.
d) Estimar la mediana poblacional de la renta.
e) Comparar los estimadores obtenidos en los apartados (b) y (d) comentar las diferencias.

27) Se ha estimado, que la media de la cantidad de dinero que gastan en ropa las mujeres españolas es de 50,700 pesetas, mientras que para los hombres, la media es de 35,000 pesetas. Dibujar un diagrama de barras que represente esta información.

28) Según un estudio reciente, en Estados Unidos, mueren cada año, 43,000 mujeres a causa del cáncer de mama y 90,000 a causa de la diabetes. Dibujar un diagrama de barras que represente esta información.

29) En 1986 se produjeron 50,2 nacimientos por cada mil mujeres con una edad entre 15 y 19 años. En 1992, el número de nacimientos fue de 62,1 por cada mil mujeres de la misma edad. Dibujar un diagrama de barras que represente esta información.

30) En 1987 había 19,152 hombres y 602 mujeres cumpliendo condena en las cárceles de Illinois. En 1992 eran 29,089 hombres y 1,226 mujeres. Dibujar un diagrama de barras por componentes que represente esta información.

31) La siguiente tabla recoge los porcentajes de incremento del índice de precios al Consumo en los Estados Unidos a lo largo de un periodo de diez años. Dibujar un gráfico temporal para estos datos e interpretarlo verbalmente.

Año	1983	1984	1985	1986	1987
% INCREMENTO IPC	3,8	3,9	3,8	1,1	4,4
Año	1988	1989	1990	1991	1992
% INCREMENTO IPC	4,4	4,6	6,1	3,1	2,9

32) De las películas que están en cartelera en una gran ciudad, el 35% son dramas, el 30% comedias, un 10% son películas de acción, otros 10% de ciencia ficción, el 6% son películas, y el 2% son de terror. Construir un pictograma que represente esta información.

33) De todos los anuncios de bebidas alcohólicas en vallas publicitarias, el 75% son cerveza, el 20% de licores con alta graduación, el 7% sobre vino, y el 2% restante de bebidas con baja graduación alcohólica, construir un pictograma que represente esta información.

34) Dibujar un diagrama de caja para los datos del Ejercicio 22, que recogían los porcentajes de disminución de las acciones de los 25 mayores fondos de pensiones el viernes 13 de noviembre de 1989.

35) Tomar una muestra de 20 hombres y otra de 20 mujeres que tengan aproximadamente la misma edad y anotar sus alturas. Construir un diagrama de CIA. para cada muestra y comparar verbalmente los resultados.

36) La tabla 4.7 recoge, para cada uno de los 50 estados norteamericanos, los porcentajes del gasto total dedicados a educación y a bienestar público. Realizar un análisis de estos datos. Emplea para ello las técnicas numéricas y gráficas que te parezcan adecuadas para extraer la esencia de la información contenida en los datos.

37) Consigue la memoria anual de alguna gran empresa, describe las técnicas gráficas empleadas para presentar los datos en le informe y sugiere alguna mejora que puede hacerse.

38) Explica qué tipo de información proporcionan acerca de una población las siguientes medidas.

a) La media.
b) La mediana.
c) La desviación típica.
d) El rango intercuartílico.

TABLA 4.7 GASTO TOTAL DE CADA ESTADO PARA LA EDUCACIÓN Y BIENESTAR PÚBLICO

Estado	Educación	Bienestar Público	Estado	Educación	Bienestar Público
Alabama	15,4	13,5	Montana	9,7	9,3
Alaska	8,5	5,6	Nebraska	17,0	12,4
Arizona	16,7	5,6	Nevada	10,3	6,4
Arkansas	12,0	14,8	Nuevo Hampshire	13,4	14,9
California	9,2	11,0	Nueva Jersey	6,6	10,1
Colorado	20,0	7,5	Nuevo Mexico	14,8	9,3
Connecticut	8,7	18,1	Nueva York	5,1	6,4
Delaware	17,3	11,1	California del Norte	13,6	11,0
Florida	8,8	9,5	Dakota del Norte	16,2	8,8
Georgia	12,0	14,2	Ohio	11,8	11,8
Hawai	12,7	13,8	Oklahoma	16,1	16,8
Idaho	12,7	10,8	Oregon	11,8	13,1
Illinois	8,3	19,8	Pennsylvania	5,5	19,2
Indiana	17,2	8,8	Rhode Island	9,8	20,8
Iowa	13,9	14,4	California del Sur	15,1	10,7
Kansas	14,7	16,2	Dakota del Sur	15,5	13,7
Kentucky	12,0	13,9	Tennessee	14,6	14,7
Louisiana	11,3	13,1	Texas	17,0	12,7
Maine	9,1	19,5	Utah	19,6	10,3
Maryland	8,8	13,8	Vermont	15,7	14,1
Massachussets	5,9	23,9	Virginia	15,4	10,3
Michigan	11,3	19,8	Washington	15,1	13,7
Minnesota	12,7	10,3	West Virginia	9,7	9,3
Mississippi	11,3	14,3	Wisconsin	13,2	12,4
Missouri	12,2	16,2	Wyoming	11,4	5,3

39) Dos poblaciones constan de dos individuos cada una. La media de estas dos poblaciones es la misma, y también lo son sus desviaciones típicas. ¿Son necesariamente iguales los valores numéricos de los individuos de las dos poblaciones?

40) Los porcentajes de rentabilidad de los fondos de inversión de diez grandes empresas fueron.

27,9	11,6	17,6	26,6	15,6
12,4	22,4	18,5	22,9	25,0

Para esta población:

a) Hallar la media.
b) Hallar la mediana.
c) Hallar la varianza.
d) Hallar la desviación típica.
e) Hallar el recorrido.
f) Hallar el rango intercuartílico.

41) En la tabla siguiente aparecen los años de servicio acumulados por 355 trabajadores de una gran empresa antes de su retiro voluntario

Años de Experiencia	Número de empleados	Años de Experiencia	Número de empleados
0 – 1	4	8 – 9	11
1 – 2	41	9 –10	7
2 – 3	67	10 – 11	14
3 – 4	82	11 – 12	6
4 – 5	28	12 – 13	14
5 – 6	43	13 – 14	5
6 – 7	14	14 – 15	2
7 – 8	17		

Para esta población:

a) Hallar la media.
b) Hallar la mediana.
c) Hallar la varianza.
d) Hallar la desviación típica.
e) Hallar el recorrido.
f) Hallar el rango intercuartílico.

42) Sean $X_1, X_2, \ldots, X_N$ las N observaciones de una población con media μ. Sea K cualquier número. Probar que:

$$\sum_{i=1}^{N}(x_i - K)^2 = \sum_{i=1}^{N}(x_i - \mu)^2 + N(K-\mu)^2$$

A partir de esto, demostrar que el valor de K que minimiza: $\sum_{i=1}^{N}(x_i - K)^2$ es $K = \mu$.

43) En la tabla se recogen los porcentajes de rentabilidad a 5 años de vencimiento bonos de empresas privadas.

a) Representar estos datos en un histograma.
b) Representar estos datos en un diagrama de tallo y hojas.
c) Construir un diagrama de caja para estos.

97,9	91,3	69,0	83,6	63,0	86,3	121,3	73,8	90,4	76,6
99,7	91,4	82,7	94,3	45,9	86,5	90,3	85,6	83,6	81,7
83,5	93,7	91,3	83,1	79,6	106,3	92,4	77,4	79,2	85,3
96,6	94,5	88,3	74,2	77,5	71,5	82,8	81,5	92,1	94,7
62,9	74,6	83,0	77,6	87,3	82,1	62,6	84,2	69,5	75,1
83,1	77,3	79,2	98,1	57,4					

44) La siguiente tabla se refiere a los usos más comunes citados en una encuesta realizada a usuarios de ordenadores de pequeñas y medianas empresas. Construir un pictograma para representar esta información.

ÁREA	RESPUESTA (%)
Contabilidad	32
Procesos de texto	16
Hojas de cálculo	13
Bases de datos	12
Puntos de venta	4
Telecomunicaciones	1
Otros	22

En Estados Unidos, el 63% de los médicos tienen alguna especialidad; este porcentaje es del 48% en Alemania y del 37% en el Reino Unido. Construir un diagrama de barras que ilustre esta información

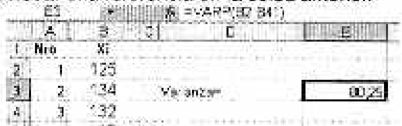
CAPÍTULO No. 4 APÉNDICE.

APÉNDICE. Nº 1 ANALISIS CON EXCEL

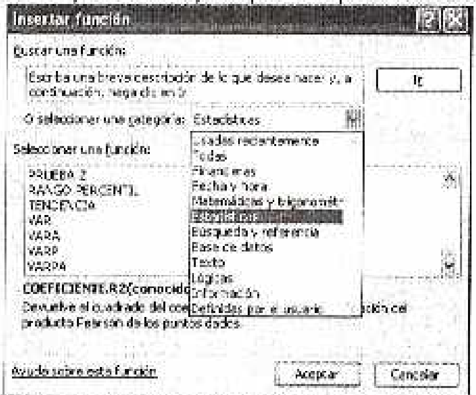
Excel como todas las otras hojas de cálculo, nos permite hacer un análisis de los datos, utilizaremos el problema 1 de los resueltos con lo cual realizamos diferentes cálculos, para lo cual se seguirá los siguientes pasos, después de tener los datos:

PASO 1.- Celda de salida de resultado

- Posicione en una celda donde desea obtener el resultado, opcionalmente puede colocar una referencia en la celda anterior.



- Pulse sobre la figura: *fx* de la barra superior para obtener la lista de funciones que desea realizar, Luego selecciona la categoría que en este caso es: "**estadísticas**" y de la lista elija la operación que desea realizar



- Luego le presenta el cuadro para elegir el rango de datos de entrada que para nuestro caso será el ejercicio indicado, con las funciones de:

VARIANZA.- Se utiliza la **VARP** calcula la varianza (la diferencia con **VARPA** la cual incluye en el cálculo texto y valores lógicos como VERDADERO o FALSO)

DESVIACIÓN ESTÁNDAR.- La función se denomina como **DESVESTP** y se obtiene:

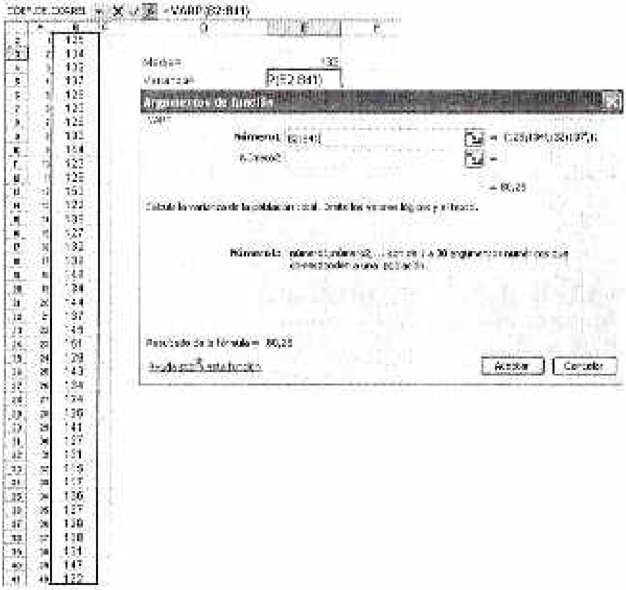
CUASI-VARIANZA.-El cálculo es operado con la función: **VARA,**

DESVIACIÓN ESTÁNDAR-C.- Se realiza por medio de la función **DESVEST**

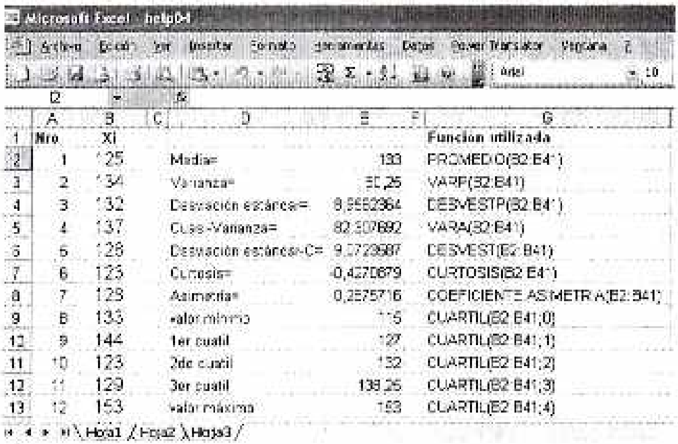
CUARTILES.- Para el cálculo de cuarteles se tiene la función **CUARTIL** con dos parámetros el primero es el rango de datos y la segúnda el número de cuartil como ser: 0 (valor mínimo), 1 (1er cuartil), 2 (segundo cuartil), 3 (3er cuartil) y 4 (el máximo valor), como ser:

CURTOSISI O ALARGAMIENTO.- Se utiliza la función de **CUROSIS()**

ASIMETRÍA(SESGO).-Se utiliza la función de **COEFICIENTE.ASIMETRIA()**



Por tanto se obtiene la siguiente tabla de resultados:



	A	B	C	D	E	F	G
1	Nro	Xi					Función utilizada
2	1	125		Media=	133		PROMEDIO(B2:B41)
3	2	134		Varianza=	80,25		VARP(B2:B41)
4	3	132		Desviación estándar=	8,9582364		DESVESTP(B2:B41)
5	4	137		Cuasi-Varianza=	82,307692		VARA(B2:B41)
6	5	128		Desviación estándar-C=	9,0723587		DESVEST(B2:B41)
7	6	123		Curtosis=	-0,4270879		CURTOSIS(B2:B41)
8	7	128		Asimetría=	0,2575716		COEFICIENTE.ASIMETRIA(B2:B41)
9	8	133		valor mínimo	115		CUARTIL(B2:B41;0)
10	9	144		1er cuatil	127		CUARTIL(B2:B41;1)
11	10	123		2do cuatil	132		CUARTIL(B2:B41;2)
12	11	129		3er cuatil	138,25		CUARTIL(B2:B41;3)
13	12	153		valor máximo	153		CUARTIL(B2:B41;4)

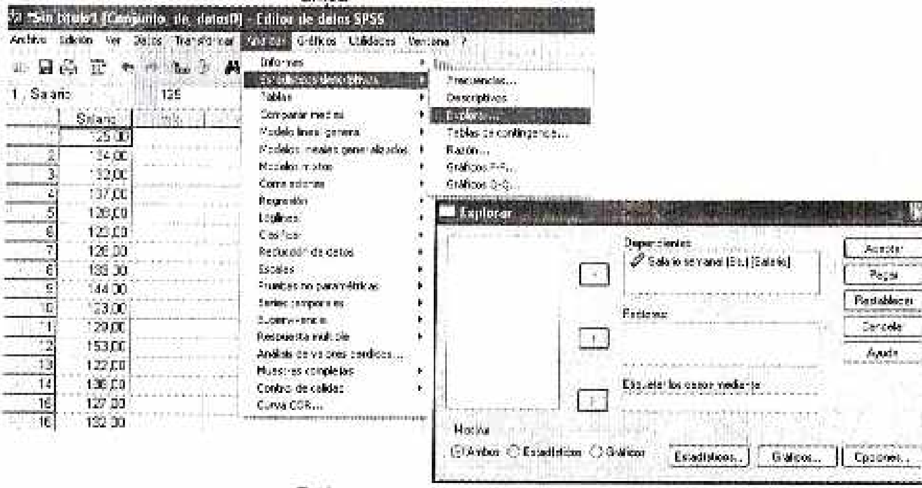
NOTA: Existe una diferencia entre los resultados del problema 1 y los obtenidos aquí por que del problema 1 están agrupados en clases.

APENDICE. Nº 2 ANALISIS CON SPSS.

De igual forma que la exposición anterior se utilizará el problema 1 de los problemas resueltos para la resolución por medio del SSPS.

PASO 1.- Teclear los datos en columna en el **editor de datos**

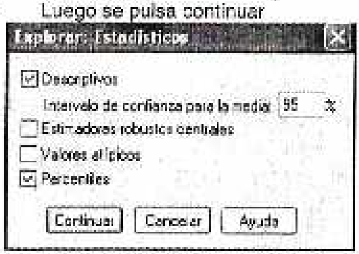
- En vista de variables: identificar el tipo de variable
- Para obtener el diagrama de tallos y hojas se procede a:
  - Barra de herramientas **Analizar**
  - **Estadísticos**
  - **Explorar**
- Luego se elige la variable apara el análisis que en este caso es la única



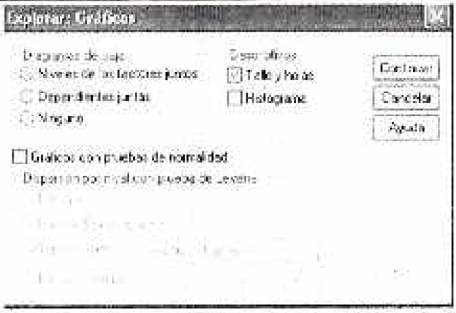
- En la parte de mostrar se escoge: **Ambos**: para que muestre los Estadísticos y los gráficos(donde se encuentra el diagrama de tallos y hojas)

PASO 2.- Opciones de los submenus

- En la pantalla de **Estadísticos** se elige: **Descriptivos** ; y **Percentiles.**
- Luego se pulsa continuar



- Luego en el submenu **Graficos**



- Se elige el diagrama de cajas **Nivles de los factores juntos** (genera una presentación para cada variable dependiente. En cada una se muestran diagramas de caja para cada uno de los grupos definidos por una variable de factor, que en el presente caso no influye)
- Luego en Descriptivos **Tallos y hojas** si se desea también el histograma.
- Luego se procede ha aceptar en la pantalla de **Explorar** y se obtendrá:

Diagrama de tallos y hojas

Salario semanal (Bs.)

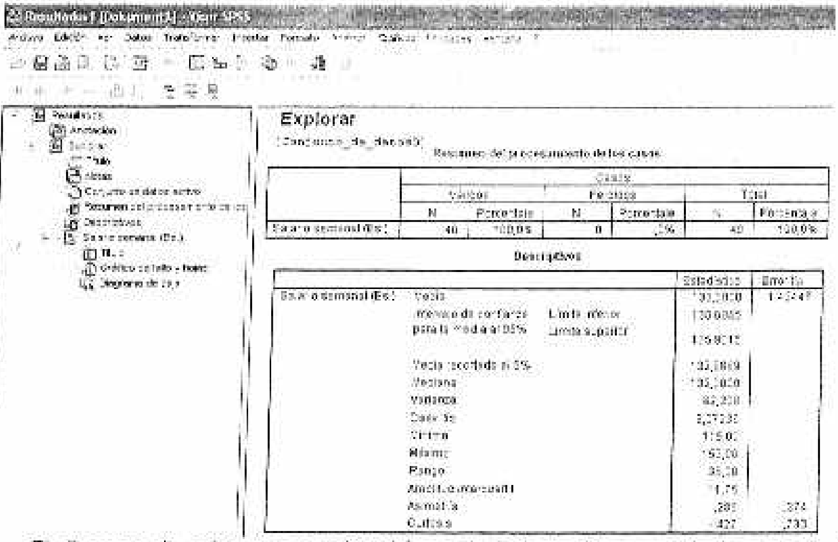
```
Salario semanal (Bs.) Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

      ,00       11 .
     2,00       11 .  57
     5,00       12 .  22334
    10,00       12 .  5677788899
     8,00       13 .  11223444
     6,00       13 .  577889
     4,00       14 .  1344
     3,00       14 .  567
     2,00       15 .  00

 Stem width:      10,00
 Each leaf:        1 case(s)
```

Percentiles

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Promedio ponderado(definición 1)	Salario semanal (Bs.)	117,2500	122,1000	127,0000	132,0000	136,7500	145,0000	150,0000
Bisagras de Tukey	Salario semanal (Bs.)			127,0000	132,0000	136,5000		

Los datos descriptivos son:



Explorar

[Conjunto_de_datos0]

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Salario semanal (Bs.)	40	100,0%	0	,0%	40	100,0%

Descriptivos

			Estadístico	Error típ.
Salario semanal (Bs.)	Media		133,2000	[illegible]
	Intervalo de confianza para la media al 95%	Límite inferior	130,0085	
		Límite superior	135,9015	
	Media recortada al 5%		133,0889	
	Mediana		132,0000	
	Varianza		82,200	
	Desv. típ.		9,07230	
	Mínimo		115,00	
	Máximo		153,00	
	Rango		38,00	
	Amplitud intercuartil		11,75	
	Asimetría		,285	,374
	Curtosis		-,427	,733

El diagrama de cajas, que se obtendrá es el siguiente diagrama más los valores de referencia: