

No. de	EDAD		Tipo de establecimiento. Nivel Diversificado		Preparación que recibí en el diversificado				Los conocimientos y habilidades adquiridas en el diversificado le han permitido comprensión de los cursos del 1er semestre, satisfactoriamente	
	18-22	23-27	Publico	Privado	Mala	Regular	Buena	Muy buena	Si	No
BOLETA										
1	1			1		1				1
2	1			1			1			1
3	1		1					1		1
4	1			1				1		1
5	1			1				1	1	1
30		1	1			1				1
31	1			1			1		1	1
32	1			1				1		1
33	1			1			1		1	1
34		1	1			1			1	1
35	1			1			1		1	1
36		1	1			1			1	1
SUMA	33	3	20	16	0	10	17	9	17	19
PORCENTAJE	92%	8%	56%	44%	0%	28%	47%	25%	47%	53%

Uniones de Estadística 20200328

ORGANIZAR Y RESUMIR LOS DATOS (ORDENAR LAS EVIDENCIAS)

2. CLASIFICAR LOS DATOS

I. ORDENAR LOS DATOS

No. de	EDAD		Tipo de establecimiento. Nivel Diversificado		Preparación que recibí en el diversificado				Los conocimientos y habilidades adquiridas en el diversificado le han permitido comprensión de los cursos del 1er semestre, satisfactoriamente	
	18-22	23-27	Publico	Privado	Mala	Regular	Buena	Muy buena	Si	No
BOLETA										
1	1			1		1				1
2	1			1			1			1
3	1		1					1		1
4	1			1				1		1
5	1			1				1	1	1
30		1	1			1				1
31	1			1			1		1	1
32	1			1				1		1
33	1			1			1		1	1
34		1	1			1			1	1
35	1			1			1		1	1
36		1	1			1			1	1
SUMA	33	3	20	16	0	10	17	9	17	19
PORCENTAJE	92%	8%	56%	44%	0%	28%	47%	25%	47%	53%

3. TABULAR LOS DATOS

4. RESUMIR LOS DATOS

EJEMPLO: MEDIDAS DE VARIABILIDAD O DISPERSIÓN

No.	CARNET	GENERO		PUNTEO	$x_i - \mu$	$ x_i - \mu $	$(x_i - \mu)^2$
		MASCULINO	FEMENINO				
1	18996	1	0	30	-27	27	729
2	15852	1	0	35	-22	22	484
3	13849	1	0	35	-22	22	484
4	14999	1	0	40	-17	17	289
5	17005	1	0	40	-17	17	289
6	15327	1	0	40	-17	17	289
7	20013	0	1	45	-12	12	144
8	18952	0	1	45	-12	12	144
9	16874	0	1	50	-7	7	49
10	19969	1	0	50	-7	7	49
11	22002	0	1	60	3	3	9
12	15108	1	0	60	3	3	9
13	14806	0	1	60	3	3	9
14	14973	1	0	60	3	3	9
15	21565	0	1	60	3	3	9
16	15494	1	0	60	3	3	9
17	15661	0	1	60	3	3	9
18	16383	0	1	65	8	8	64
19	18119	0	1	65	8	8	64
20	21110	1	0	65	8	8	64
21	13743	0	1	75	18	18	324
22	36989	1	0	75	18	18	324
23	15262	1	0	80	23	23	529
24	15485	0	1	80	23	23	529
25	15626	1	0	90	33	33	1,089
				1425	320	6000	

$\mu = \frac{1,425}{25} = 57$

- Rango (Oscilación, amplitud o recorrido)**
 $Rango = Dato\ mayor - Dato\ menor$
 $R = X_{mayor} - X_{menor} = 90 - 30 = 60$
- Desviación Media, Promedio de Desviación o Variación de la Media**
 $DM = \frac{\sum |x_i - \mu|}{N}$ (sólo para este cálculo todos los datos producto de la diferencia de $x_i - \mu$ se consideran como positivos y se representan entre dos líneas paralelas)
 $DM = \frac{320}{25} = 12.8$
- Varianza**
 $S^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{6000}{25} = 240$
- Desviación Típica o Desviación Estándar**
 $S = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = \sqrt{240} = 15.49$

$$\text{Flexyl} = \text{C.O.} = \frac{3(50.38 - 50.0)}{5.55} = 0.21$$

Resumen

medidas	Plastex	Syntel	Flexyl
\bar{x}	52.88	51.25	50.38
m_d	55.50	49.50	50.00
m_o	Amodal	49.00	Amodal
R	26.00	17.00	17.00
s^2	90.98	30.78	30.83
s	9.54	5.55	5.55
C.V.	18.04%	10.83%	11.02%
C.O.	-0.82	0.95	0.21

Análisis e interpretación.

Como puede notarse la marca Plastex, presenta una mejor situación a nivel de las medidas de tendencia central, (media y mediana) que las otras dos marcas, sin embargo cuando se examinan las medidas de dispersión, es evidente que resulta siendo la que tiene un comportamiento menos uniforme, y por lo tanto la media aritmética de 52.88 miles de bolsas por rollo, es menos segura que las de las otras marcas.

En cuanto a Syntel y Flexyl, puede decirse que la primera tiene una media de rendimiento mayor, y además un comportamiento más uniforme que la marca Flexyl, esto último se evidencia en el valor del coeficiente de variación. puesto que ambas tienen la misma desviación estándar.

Conclusión.

Con base en lo anterior es factible concluir que si se tiene la oportunidad de vender, es la marca Plastex la que debe ofrecerse.

En cuanto a la utilización en la producción, se recomienda preferentemente la marca Syntel, puesto que por la uniformidad en su rendimiento, es la que puede garantizar de mejor manera que los presupuestos o estimaciones de producción y/o de materia prima sean más confiables.

En cuanto a la marca Flexyl, se aconseja que sea utilizada, para cubrir contingencias como segunda opción, puesto que es de un rendimiento y uniformidad menor que Syntel.

Syntel

Desviación Estándar

$$s = \sqrt{30.78} = 5.55 \text{ miles de bolsas por rollo}$$

Flexyl

Varianza

$$S^2 = \frac{215.84}{7} = 30.83 \text{ miles de bolsas por rollo}$$

Flexyl

Desviación Estándar

$$S = \sqrt{30.83} = 5.55 \text{ miles de bolsas por rollo.}$$

Coefficiente de Variación

La Fórmula para calcular el Coeficiente de Variación es la siguiente.

$$C.V. = \frac{s}{\bar{X}} \times 100$$

Entonces.

$$9.54 = \frac{(0.02 - 8E.0E) \epsilon}{\dots} = 0.0 = \dots$$

$$\text{Plastex} = C.V. = \frac{9.54}{52.88} \times 100 = 18.04 \%$$

$$\text{Syntel} = C.V. = \frac{5.55}{51.25} \times 100 = 10.83 \%$$

$$\text{Flexyl} = C.V. = \frac{5.55}{50.38} \times 100 = 11.02 \%$$

Coefficiente de Oblicuidad

La Fórmula para calcular el Coeficiente de Oblicuidad es como sigue.

$$C_o = \frac{3(\bar{X} - m_d)}{S}$$

Entonces:

$$\text{Plastex} = C.O = \frac{3(52.88 - 55.5)}{9.54} = -0.82$$

$$\text{Syntel} = C.O = \frac{3(51.25 - 49.5)}{5.55} = 0.95$$

Plastex

n	x	(x - \bar{X}) ²
1	38	221.41
2	43	97.61
3	46	47.33
4	54	1.25
5	57	16.97
6	58	26.21
7	63	102.41
8	64	123.65
		$\Sigma = 636.84$

$$x - \bar{X} = 38 - 52.88 = 14.88$$

$$\leftarrow (14.88)^2 = 221.41$$

de igual manera se procede para todos los demás cálculos

Sustituyendo valores en la fórmula tenemos.

$$s^2 = \frac{636.84}{7} = 90.977 \approx 90.98$$

$$s = \sqrt{90.98} \approx 9.54 \text{ miles de bolsas por rollo}$$

Cálculo de la Desviación Estándar.

La fórmula empleada para series simples es la siguiente.

$$s = \sqrt{\frac{\Sigma(x - \bar{X})^2}{n - 1}}$$

Como podrá notarse, la desviación estándar es igual a la raíz cuadrada de la Varianza.

Entonces al sustituir los valores en la fórmula tenemos.

$$s = \sqrt{90.98} = 9.538 \approx 9.54 \text{ miles de bolsas por rollo}$$

Syntel			Flexyl		
n	x	(x - \bar{x}) ²	n	x	(x - \bar{x}) ²
1	45	39.06	1	44	40.07
2	46	27.56	2	44	40.07
3	49	5.06	3	48	5.66
4	49	5.06	4	49	1.90
5	50	1.56	5	51	0.38
6	54	7.56	6	53	6.86
7	55	14.06	7	53	6.86
8	62	115.56	8	61	112.78
		$\Sigma = 215.48$			$\Sigma = 215.84$

Syntel

Varianza

$$s^2 = \frac{215.48}{7} = 30.78 \text{ miles de bolsas por rollo}$$

$$s = \sqrt{30.78} = 5.55 \text{ miles de bolsas por rollo}$$

Plastex		Syntel		Flexyl	
n	x	n	x	n	x
1	38	1	45	1	44←
2	43	2	46	2	44←
3	46	3	49←	3	48
4	54	4	49←	4	49
5	57	5	50	5	51
6	58	6	54	6	53←
7	63	7	55	7	53←
8	64	8	62	8	61

De Acuerdo con lo anterior se puede calificar las series de la manera siguiente.

Plastex = amodal. (ningún valor se repite más de una vez)

Syntel = moda 49 miles de bolsas por rollo.

Flexyl = amodal (tiene dos modas)

Medidas de dispersión

Cálculo del Rango o Recorrido.

Para calcular el rango en todas las series de datos, se hace uso de la fórmula siguiente:

$$R = L_2 - L_1$$

Donde: L_2 = Valor más alto de la serie
 L_1 = Valor más bajo de la serie

Sustituyendo valores tenemos:

Plastex $R = 64 - 38 = 26$ miles de bolsas por rollo.

Syntel $R = 62 - 45 = 17$ miles de bolsas por rollo.

Flexyl $R = 61 - 44 = 17$ miles de bolsas por rollo.

Cálculo de la varianza y la desviación estándar

El cálculo de la varianza para series simples, se efectúa a través de la fórmula siguiente. (como es una muestra se utilizará el signo s)

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

Para poder utilizar la fórmula debemos construir en cada caso una columna de equis menos equis barra al cuadrado, así, El procedimiento empleado es como sigue:

Sustituyendo valores tenemos

$$\text{posición} = \frac{8 + 1}{2} = 4.5$$

En vista que las tres series están integradas por 8 elementos, la posición (4.5) será la misma para todas.

El paso siguiente es encontrar el valor en x_i , cuando la serie es par, como en éste caso, es necesario interpolar entre los dos valores (anterior y posterior a la posición) así:

	Plastex	Syntel	Flexyl
n	x	n	x
1	38	1	45
2	43	2	46
3	46	3	49
4	54	4	49
4.5 →		4.5 →	
5	57	5	50
6	58	6	54
7	63	7	55
8	64	8	62

Por lo tanto tenemos.

Plastex	Syntel	Flexyl
$m_d = \frac{54 + 57}{2}$	$m_d = \frac{49 + 50}{2}$	$m_d = \frac{49 + 51}{2}$
$m_d = 55.5$ miles de bolsas por rollo	$m_d = 49.5$ miles de bolsas por rollo	$m_d = 50.0$ miles de bolsas por rollo

Lo anterior significa que el valor que está a la mitad de la serie cuando ésta ha sido ordenada del menor al mayor valor, y por lo tanto sobre y bajo de él se encuentra el 50% de los valores, en cada uno de los casos estudiados es como sigue.

Plastex $m_d = 55.5$ miles de bolsas por rollo

Syntel $m_d = 49.5$ miles de bolsas por rollo

Flexyl $m_d = 50.0$ miles de bolsas por rollo

Cálculo de la moda

Para determinar el valor de la moda en series simples, se recurre a la simple observación, o bien practicando tarjas o marcas.

El primer paso que debe darse para proceder a efectuar el análisis e interpretación estadística de series simples, es ordenar las series de datos de acuerdo con la magnitud de los mismos así:

	Plastex		Syntel		Flexyl	
	n	x	n	x	n	x
1		38	1	45	1	44
2		43	2	46	2	44
3		46	3	49	3	48
4		54	4	49	4	49
5		57	5	50	5	51
6		58	6	54	6	53
7		63	7	55	7	53
8		64	8	62	8	61
Σ		423		410		403

Medidas de tendencia central

Cálculo de la Media Aritmética

Para series simples se utiliza la fórmula siguiente

$$\bar{X} = \frac{\Sigma X}{n} \quad \text{ó} \quad \mu = \frac{\Sigma X}{N}$$

En este caso como se trata de muestras utilizaremos el estadístico equis barra así:

Media Aritmética de Plastex	Media Aritmética de Syntel	Media Aritmética de Flexyl
$\bar{X} = \frac{423}{8}$	$\bar{X} = \frac{410}{8}$	$\bar{X} = \frac{403}{8}$
$\bar{X} = 52.85$	$\bar{X} = 51.25$	$\bar{X} = 50.375$

Es decir que Plastex tiene una media de rendimiento de 52.88 miles de bolsas por rollo.

Syntel tiene una media de rendimiento de 51.25 miles de bolsas por rollo y

Flexyl tiene una media de rendimiento de 50.38 miles de bolsas por rollo.

Cálculo de la mediana m_d

Para calcular la mediana de series simples, primero se calcula la posición en "n" así:

$$\text{posición} = \frac{n + 1}{2}$$

Varianza

Es la medida de dispersión que representa el promedio al cuadrado de las variaciones de los valores de los datos, con respecto al valor de la media, cuando pertenece a una muestra se identifica como s^2 y cuando pertenece a una población se identifica como σ^2

Desviación Estándar

Es la medida de dispersión que en unidades lineales, proporciona al investigador la información sobre el grado de variación existente entre los valores de los datos, sobre y bajo la media aritmética, cuando pertenece a una muestra se identifica como s y cuando pertenece a una población se identifica como σ

Coefficiente de Variación

Es la medida que proporciona al investigador información acerca de la participación de la desviación estándar en la conformación del valor de la media aritmética, se identifica como **C.V.**

Coefficiente de Oblicuidad

Es la medida que proporciona al investigador, información sobre la acumulación de los datos hacia uno u otro lado de la mediana, es decir, el grado de oblicuidad o sesgo en la distribución de los valores de los datos, se identifica como **C.O.**

CAPITULO II

Análisis e Interpretación de Series Simples, o Muestras Pequeñas

Los conjuntos de datos integrados por treinta (30) o menos elementos, se conocen como muestras pequeñas o series simples, para llevar a cabo el análisis e interpretación de ellas, se utilizarán las medidas de Tendencia Central y de Dispersión, definidas con anterioridad, es necesario hacer notar que el significado de cada una de ellas, es el mismo para series simples que para las muestras grandes, la diferencia es solamente la fórmula de cálculo para algunas.

Ejemplo: Problema No. 1

Una empresa, dedicada a la producción de bolsas de plástico, tiene en sus bodegas tres marcas distintas de esa materia prima y desea que usted como asesor administrativo le aconseje como utilizarlas, y que además le dé su opinión, si se tiene la posibilidad de vender una de ellas.

Según los datos estadísticos, los rendimientos son como sigue.

Marca	Rendimiento en miles de bolsas por rollo
plastex	63 - 58 - 43 - 46 - 38 - 54 - 57 - 64
Syntel	45 - 46 - 49 - 49 - 50 - 54 - 55 - 62
Flexyl	44 - 44 - 48 - 49 - 51 - 53 - 53 - 61

Interpretación

Es el sentido personal y objetivo, que el investigador debe darle tanto a los resultados numéricos obtenidos, como a las relaciones existentes entre las partes integrantes de la serie de datos y el suceso objeto de estudio.

Medidas De Tendencia Central

Son los índices numéricos que proporcionan al investigador información sobre la tendencia de los datos a agruparse generalmente hacia una posición central, las medidas de tendencia central, que se estudiarán en este trabajo son la media aritmética, la mediana y la moda.

Media Aritmética

Es el valor que mejor representa a todos los datos de la serie estudiada. Cuando pertenece a una muestra se identifica como **X** (**equis barra**) y cuando pertenece a una población se identifica como μ (**mu**)

Mediana

Es el valor que se encuentra exactamente a la mitad de la serie de datos, cuando éstos han sido ordenados de acuerdo a su magnitud, del menor al mayor o viceversa, normalmente se ordenan de menor a mayor, se identifica como m_d para muestras o M_d para poblaciones.

Moda

Es el valor que se repite con mayor frecuencia, entre todos los valores de la serie, se identifica como m_o para muestras o M_o para poblaciones.

De acuerdo con lo anterior, las series de datos pueden tener dos o más modas, en cuyo caso serán bimodales (dos modas) Trimodales (tres modas) Tétramodales (cuatro modas) o polimodales (más de cuatro modas). Sin embargo en el presente trabajo, se considerará como amodales (sin moda) a aquellas que tengan más de una moda, de igual manera se calificará a las que no tengan ninguna moda.

Medidas de Dispersión

Son los índices o indicadores numéricos, que proporcionan al investigador, información sobre el grado de variación existente entre los valores de los datos de la serie objeto de estudio.

En este documento abordaremos las siguientes: El Rango, la Varianza, la Desviación Estándar, el Coeficiente de Variación y el Coeficiente de Oblicuidad.

Rango o Recorrido

Es la medida de dispersión que representa el valor total de la variación en los valores de los datos, se obtiene de la diferencia entre el mayor valor y el menor valor de la serie de datos, se identifica

como R

44 - 44 = 48 - 49 = 51 - 53 = 61

Terminología Básica

Con el propósito de iniciar el estudio de la estadística, se abordan a continuación, algunos de los conceptos que se utilizarán con frecuencia en el presente trabajo, más adelante se definen otros que de acuerdo con el avance en el contenido son necesarios.

Estadística

Es la rama del conocimiento humano que trata de la recolección, organización, clasificación, presentación, análisis e interpretación de datos numéricos.

Campos de la Estadística

La estadística tiene dos campos de aplicación específicos así: estadística descriptiva o deductiva y estadística inductiva o inferencial.

Estadística Descriptiva o Deductiva

Es la aplicación de las técnicas de recolección, organización, clasificación, presentación, análisis e interpretación de datos estadísticos, a un determinado conjunto de ellos, ya sea una muestra o una población, cuyo resultado es válido únicamente para esos conjuntos de datos.

Estadística Inductiva o Inferencial

Es la técnica que con base en la probabilidad y a partir de la deducción o descripción, (estadística descriptiva o deductiva) es útil para elaborar interpretaciones, inferencias o inducciones válidas para un conjunto mayor de datos.

Datos Estadísticos

Son los conjuntos o colecciones de datos numéricos, que presentan características comunes, que les permiten ser comparados entre sí, y por ende analizados e interpretados.

Poblaciones

Son los conjuntos o colecciones completas de elementos, individuos, cosas, objetos o valores, que contienen la totalidad de observaciones posibles o valores de interés en un proceso de investigación.

Muestras

Son los subconjuntos o colecciones parciales de los elementos de una población.

Estadísticos o estadígrafos

Son las medidas numéricas, de tendencia central o de dispersión que se refieren a las características inherentes a las muestras.

Parámetros

Son las medidas numéricas, de tendencia central o de dispersión que se refieren a características inherentes a las poblaciones.

Análisis

Es la descomposición de una serie de datos en sus partes integrantes, para establecer la relación existente entre ellas y el suceso, objeto o variable de estudio.

CAPITULO I

ANTECEDENTES Y GENERALIDADES DE LA ESTADISTICA

Breve Historia de la Estadística

Como todas las ramas del conocimiento humano, la estadística ha pasado por diferentes etapas en su desarrollo, las que de manera sucesiva son abordadas hasta llegar a la estadística inferencial.

Los datos que pueden merecer el calificativo de estadísticos, se generaron hasta cuando las sociedades humanas empezaron a utilizar los diversos sistemas de escritura; sin embargo, cabe suponer que antes de ello tuvieron necesidad de practicar recuentos de población y de otros recursos de la sociedad o del territorio que ocupaban.

La estadística en su etapa inicial, se refería principalmente a la realización de conteos y a la presentación de la información en cuadros y gráficas.

La aplicación de la probabilidad, surgió de la necesidad de utilizar los datos para predecir con alguna precisión, los hechos futuros, lo que conllevó un gran impulso al desarrollo de la estadística, convirtiéndola en una herramienta muy útil para describir, manipular, analizar e interpretar la más diversa información relacionada con la evolución y desarrollo de las sociedades humanas y, por ende, para comprender, evaluar y controlar su funcionamiento, tanto a nivel general como en aspectos específicos.

probabilidad, la media aritmética y la desviación estándar de una variable aleatoria discreta. INDICE

página

Lo que se refiere a la distribución de la probabilidad de una variable aleatoria continua, se trata en el capítulo de la distribución normal.

11

Después de dejar constancia de mi reconocimiento a la Universidad de la Sorbona, Leonel Enrique Casoso Chinchipe, a la colaboración al proporcionarme en forma de estudiante, las tomadas sobre el Curso Métodos Cuantitativos, que he tomado del Profesor Thular. Las fórmulas y procedimientos estadísticos fueron tomados de la bibliografía que se cita en el epílogo. En lo que respecta a los ejemplos, ejemplos de adecuación y presentación del contenido, son reservados al autor de esta obra. La bibliografía y procedimientos estadísticos fueron tomados de la bibliografía que se cita en el epílogo.

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

MATERIAL DE APOYO PARA EL CURSO DE MÉTODOS CUANTITATIVOS I DE LA ESCUELA DE ADMINISTRACIÓN

Nueva Edición

Febrero del 2001. 500 ejemplares

EDICIÓN NO COMERCIAL

PROHIBIDA SU REPRODUCCIÓN

Tiraje Limitado

COLECCIÓN: "Textos Estadísticos"
DEPARTAMENTO DE PUBLICACIONES
FACULTAD DE CIENCIAS ECONÓMICAS
UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

Guatemala, Febrero del 2001

Presentación

Estos apuntes tienen como objetivo proporcionar a los estudiantes de la Escuela de Administración de la Facultad de Ciencias Económicas, un apoyo para el estudio del curso de Métodos Cuantitativos I.

Es necesario resaltar que la mejor forma de asimilar el conocimiento se logrará a través de la investigación y la práctica del estudiante, tanto en las aulas universitarias, como en su vida laboral, puesto que con el avance tecnológico y el desarrollo constante de la ciencia, es factible que se pongan en práctica nuevos y mejores métodos y sistemas estadísticos, que indudablemente fortalezcan lo que de manera sencilla se aporta en este trabajo.

El contenido se ha dividido en seis capítulos en la forma siguiente, en el primero se presenta una breve historia de la Estadística, así como alguna terminología básica.

El análisis e interpretación de series simples y muestras grandes, a través de las medidas de tendencia central y de dispersión, es objeto de tratamiento en los capítulos II y III.

En el capítulo IV se aborda la naturaleza e importancia de la probabilidad, algunos conceptos básicos, probabilidad de eventos simples y compuestos, el diagrama de árbol y el árbol de probabilidad.

El capítulo V contiene lo relacionado con la distribución de