

Probabilidades y Estadística Ingeniería en Sistemas de Información UNIDADES 1, 2 y 3

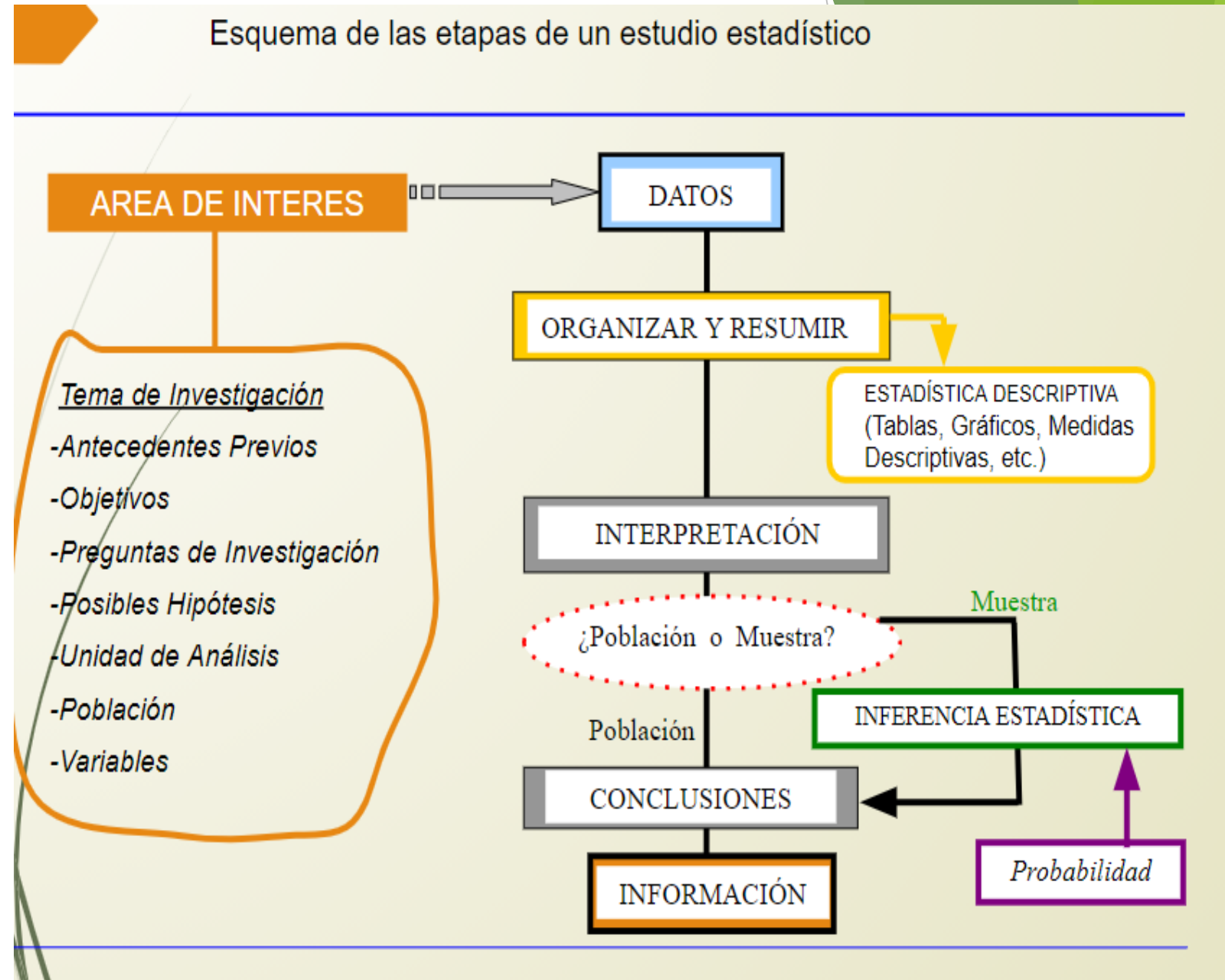
PARTE PRÁCTICA

Prof. Ana Strub - Prof. Clarisa Stefanich

Conceptos importantes. Unidad 1: Datos estadísticos y etapas para su análisis.

1. Estadística. Estadística descriptiva y Estadística inferencial.
2. El método científico de la Estadística.
3. Población. Muestra.
4. Unidad estadística o de observación. Unidad de relevamiento.
5. Dato estadístico. Clasificación

Etapas del estudio estadístico



Dato Estadístico

TIPOS DE DATOS:

Categóricos/CUALITATIVOS:

dicotómicos: (dos categorías)

mas categorías:

nominales: No existe orden obvio entre las categorías.

(país de origen, estado civil, diagnóstico.)

ordinales: Existe un orden natural entre las categorías.

(Tabaquismo: No fuma / ex-fumador / fuma ≤ 10
cigarrillos diarios / fuma > 10 cigarrillos diarios)

(Severidad de la patología: Ausente/leve/moderado/severo)

Dato estadístico

► **NUMÉRICOS** : el resultado de la observación o medición es un número

► **Discretos** : La variable sólo puede tomar un cierto conjunto de valores posibles. En general, aparecen por conteo. (número de miembros del hogar, número de intervenciones quirúrgicas, número de casos notificados de una cierta patología.)

► **Continuos** : Generalmente son el resultado de una medición que se expresa en unidades. Las mediciones pueden tomar teóricamente un conjunto infinito de valores posibles dentro de un rango. (altura, peso, pH, nivel de colesterol en sangre.)

Unidad 1. Ejercicio N°2:

En cada uno de los siguientes casos indique:

- a) ¿Cuál es la población objeto de estudio?
- b) ¿Cuál es la muestra?
- c) ¿Cuál es la unidad estadística y cuál la de relevamiento?
- d) ¿Cuál es la característica analizada y de qué tipo es?

2.1 El productor de un programa de FM, que se transmite por cierta radio cordobesa, desea conocer el grado de aceptación del programa, medido a través de si el mismo es escuchado o no, en comparación con otros que se emiten a la misma hora. Para ello, una consultora especializada se compromete a realizar una encuesta de 500 radioescuchas de la Ciudad de Córdoba a la hora que se transmite el programa y así poder evaluar la proporción de individuos que lo prefieren. Suponga que el alcance de la transmisión es hasta el fin del ejido municipal.

2.3 Cierta organismo del Gobierno Provincial desea conocer la evolución que experimenta la industria en toda la Provincia de Córdoba. Para ello se seleccionaron 400 empresas industriales de los distintos Departamentos de la Provincia, y entre otras preguntas figuran:

Número de personas empleadas en cada empresa

Porcentaje de utilización de la capacidad instalada.

2.1

Población: Todos los individuos que en la Ciudad de Córdoba, hasta el fin del Ejido Municipal, escuchan radio a la hora que se emite el programa.

Muestra: Los 500 radioescuchas que serán entrevistados.

Unidad estadística: Cada uno de los radioescuchas de la Ciudad de Córdoba

Unidad de relevamiento: Coincide con la unidad estadística.

Característica: la respuesta de cada radioescucha sobre si escucha o no ese programa. Estamos frente a una variable categórica o atributo.

2.3

Población: Todas las empresas Industriales de la provincia de Córdoba, en espacio y tiempo determinados.

Muestra: 400 empresas industriales, seleccionadas en los distintos Departamentos.

Unidad Estadística: Cada una de las empresas.

Unidad de Relevamiento: Registros de cada empresa.

Característica: Número de personas empleadas. Variable discreta.

Porcentaje de utilización de la capacidad instalada. Variable continúa.

2.8 Encuesta de Ocupación Hotelera

Programa organizado por: Secretaría de Turismo de Provincia de Córdoba

Objetivo general: medir el impacto del turismo internacional y del turismo interno sobre el sector hotelero de la Ciudad de Villa Carlos Paz a los fines de elaborar indicadores de la actividad hotelera tanto desde la perspectiva de la oferta como desde el punto de vista de la demanda. Se relevó a los Establecimientos hoteleros. Área de cobertura: Ciudad de Villa Carlos Paz. Período que abarca el operativo: enero de 2017

ENCUESTA	% OCUP	CATEG HOTEL	TURISTA NACIONAL
1	90	3	NO
2	70	1	SI
3	50	4	NO
4	77	4	NO
5	87	2	NO
6	45	3	SI
7	67	3	SI
8	78	5	NO
9	60	1	NO

NRO. ENCUESTA	Número de encuesta: 1,...,10
% OCUP	OCUPACIÒN DEL HOTEL EN %
CAT HOTEL	CATEGORIZACIÒN HOTEL. ESTRELLAS 1-5
TURISTA NACIONAL	SI - NO

2.8

Población: Todos los Hoteles de la ciudad de Carlos Paz activos durante enero de 2017.

Muestra: No se tomó muestra.

Unidad Estadística: Cada uno de los Hoteles

Unidad de Relevamiento: Registros en los hoteles de los Pasajeros de los hoteles, durante Enero de 2017.

Características: % de Ocupación: Variable cuantitativa continua. **Categoría del Hotel:** Variable cualitativa ordinal. **Turista Nacional:** variable cualitativa nominal.

Conceptos importantes. Unidad 2: Organización y Presentación de Datos Estadísticos

1. Formas de presentación de los datos
2. Tablas de distribución de frecuencias. Tipos
3. Gráficos. Tipos de acuerdo al tipo de variable y la forma de agrupación de los datos

PRESENTACIÓN DE DATOS ESTADÍSTICOS.

La presentación de datos estadísticos constituye en sus diferentes modalidades uno de los aspectos de más uso en la estadística descriptiva.

Presentación escrita

Presentación tabular

Presentación gráfica

Presentación escrita.
Esta forma de presentación de informaciones se usa cuando una serie de datos incluye pocos valores, por lo cual resulta más apropiada la palabra escrita como forma de escribir el comportamiento de los datos.

Presentación de todos y cada uno de los datos obtenidos

Presentación de datos en forma resumida mediante porcentajes o promedios

Cualitativos

- Diagrama de pastel
- Diagrama de barras
- Diagrama de barras compuestas
- Diagrama de Barras agrupadas
- Pictograma

Cuantitativos

- Histograma
- Polígono de Frecuencias
- Ojiva Porcentual
- Diagrama de dispersión
- Cuadros

CUADRO No. 1

NUMERO DE CASOS NUEVOS DIAGNOSTICADOS CON TUBERCULOSIS POR GRUPO DE EDAD
EUA 1973

GRUPOS DE EDAD (AÑOS)	NUMERO DE CASOS
0-4	1242
5-14	1081
15-24	2482
25-44	8153
45-64	10916
65 Y MAS	7124
TOTAL	30998

FUENTE: HEW, PHS, CDC, Reported Morbidity and Mortality in the United States, 1973.

Nombre de cuadro

Nombre de columnas

Nombre de Renglones

TOTAL

Separación de las columnas

Fuente

Ejercicio integrador - Unidad 1-2

Fueron encuestados los 50 primeros empleados de una industria alimenticia de la provincia de Córdoba. En la BD que se adjunta se detallan los datos relevados a cada empleado.

1) Haciendo uso del vocabulario técnico, defina cada concepto, siempre en relación al caso propuesto, y detalle el tipo de dato estadístico que se presenta.

a. Población b. Muestra c. Unidad estadística d. Unidad de relevamiento e. Características de las variables bajo estudio.

2) A partir de la BD suministrada y mediante el empleo del InfoStat construya la/s tabla/s de Distribución de Frecuencias con agrupación en lista o por intervalos, según corresponda. Calcule todas las frecuencias que conozca de las siguientes variables:

a. Salario (tabla con 5 intervalos de amplitud 8) b. Cantidad de hijos c. Área

Interprete, en cada tabla, las frecuencias para un valor o un intervalo de valores de la variable.

3) Con los datos tabulados o mediante el uso de funciones en InfoStat, responda en forma completa y clara, utilizando vocabulario técnico adecuado en la explicación.

4) Mediante el empleo del InfoStat grafique e interprete las variables analizadas en los puntos anteriores.

Resolución ejercicio integrador

Punto 1

- a) **Población:** Todos los empleados de una industria alimenticia de la provincia de Córdoba.
- b) **Muestra:** 50 empleados seleccionados.
- c) **Unidad estadística:** Cada empleado.
- d) **Unidad de relevamiento:** Cada empleado.
- e) **Características de las variables bajo estudio:**
 - Salario:** variable cuantitativa continua.
 - Cantidad de hijos:** variable cuantitativa discreta.
 - Área:** dato cualitativo nominal.

Resolución ejercicio integrador.

Punto 2

a.

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Salarios	1	[115.00	123.00)	119.00	2	0.04	2	0.04
Salarios	2	[123.00	131.00)	127.00	10	0.20	12	0.24
Salarios	3	[131.00	139.00)	135.00	19	0.38	31	0.62
Salarios	4	[139.00	147.00)	143.00	12	0.24	43	0.86
Salarios	5	[147.00	155.00]	151.00	7	0.14	50	1.00

Fuente: BD- clase práctica PyE

Tabla de distribución de frecuencias para datos agrupados en intervalos

b.

Variable	Clase	MC	FA	FR	FAA	FRA
Hijos	1	8	0.16	8	0.16	
Hijos	2	10	0.20	18	0.36	
Hijos	3	15	0.30	33	0.66	
Hijos	4	13	0.26	46	0.92	
Hijos	5	4	0.08	50	1.00	

Fuente: BD-clase PyE

Tabla de distribución de frecuencias para datos agrupados en lista o con agrupación simple

c.

Variable	Clase	Categorías	FA	FR
Área	1	Administración	21	0.42
Área	2	Comunicación	5	0.10
Área	3	Diseño	13	0.26
Área	4	Ingeniería	9	0.18
Área	5	RRHH	2	0.04

Fuente: BD- clase práctica PyE

Tabla de distribución de frecuencias para datos cualitativos

FA= ni Frecuencia absoluta simple
FR= hi Frecuencia relativa simple
FAA= Ni Frecuencia absoluta acumulada
FRA= Hi Frecuencia relativa acumulada

Ejemplos

El Rango de variabilidad, se obtiene como la diferencia del dato con el máximo valor y el dato con el mínimo valor: $R = X_{\max} - X_{\min}$

Tipo I.
Pocos datos y poca variabilidad.

Edad
5
6
7
7
8
10
14
17
18
18

Las frecuencias de los datos es la misma (uno) para cada dato de la lista)

$$R = 18 - 5 = 13$$

Tipo II.
Muchos datos y poca variabilidad.

$$R = 18 - 5 = 13$$

Edad	N°	%
5	3	3.9
6	7	9.2
7	15	19.7
8	16	21.1
10	8	10.5
14	12	15.8
17	9	11.8
18	6	7.9
Total	76	100

Cuando los datos son prácticamente los mismos, pero se repiten muchas veces, conviene ordenarlos y colocar su frecuencia.

Tipo III.
Muchos datos y Mucha variabilidad.

$$R = 39 - 0 = 39$$

Edad	N°	%
0-4	18	7.8
5-9	25	10.9
10-14	30	13.0
15-19	34	14.8
20-24	37	16.1
25-29	29	12.6
30-34	36	15.7
35-39	21	9.1
Total	230	100

Si son muchos datos, y son diferentes muchos de ellos (variabilidad), entonces, conviene agrupar los datos y colocar la frecuencia para cada grupo de datos.

Construcción de las tablas. Punto 2

Salario (tabla con 5 intervalos de amplitud 8)

Tabla de datos agrupados en intervalos:

Resulta adecuado agrupar los datos de esta manera cuando existen muchos valores observados diferentes y dichos valores se repiten.

Tablas de frecuencias

Opciones de la tabla de frecuencias | Ajustes

Variable: Salarios (Real)

☒ LI ☒ LS ☒ MC
☒ FA ☒ FR ☒ FAA ☒ FRA
☐ E(FA) ☐ E(FR) ☐ E(FAA) ☐ E(FRA)

☐ Usar cálculo automático en todas las variables

Número de clases

Calcular

☐ Automáticamente
☒ Personalizado 5

Valor mínimo 115 Intervalo de clase
Valor máximo 155 8

☒ Tratar a las variables enteras como conteos
☐ Intervalos cerrados por derecha

La opción intervalos cerrados por derecha debe dejarse destildada

Construcción de las tablas. Punto 2

Tablas de frecuencias

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Salarios	1	[115.00	123.00)	119.00	2	0.04	2	0.04
Salarios	2	[123.00	131.00)	127.00	10	0.20	12	0.24
Salarios	3	[131.00	139.00)	135.00	19	0.38	31	0.62
Salarios	4	[139.00	147.00)	143.00	12	0.24	43	0.86
Salarios	5	[147.00	155.00]	151.00	7	0.14	50	1.00

Infostat efectúa automáticamente los siguientes cálculos:

i	yi-1	yi	FA (ni) frecuencia absoluta simple	FR (hi)=ni/n frecuencia relativa simple	FAA (Ni) frecuencia absoluta acumulada	FRA (Hi)= Ni / n frecuencia relativa acumulada
1	115	123	2	2/50=0,04	2	2/50=0,04
2	123	131	10	10/50=0,20	2+10=12	12/50=0,24
3	131	139	19	19/50=0,38	2+10+19= 31	31/50=0,62
4	139	147	12	12/50=0,24	2+10+19+12= 43	43/50=0,86
5	147	155	7	7/50=0,14	50	1

n=50

Para calcular la amplitud de un intervalo:

Recorrido = R= Valor máx. - valor min. = 153-117= 36

Amplitud = m = R/c(cantidad de intervalos) = 36/5 = 7,2, se redondea a 8

Recorrido Ampliado = R´ = m*c= 8*5= 40

R´ - R = 40-36= 4 unidades que se reparten al comienzo y al final del primer y último intervalo:

valor mín = 115 (117-2)

Valor máx.= 155 (153+2)

i	yi-1	yi	FA (ni) Frecuencia absoluta simple	FD (N´ i) frecuencia absoluta desacumulada	FR (hi)=ni/n frecuencia relativa simple	FD (H´ i) frecuencia relativa desacumulada
1	115	123	2	50	2/50=0,04	50/50= 1
2	123	131	10	50 - 2= 48	10/50=0,20	48/50= 0,96
3	131	139	19	48 - 10= 38	19/50=0,38	38/50= 0,76
4	139	147	12	38 - 19= 19	12/50=0,24	19/50= 0,38
5	147	155	7	19- 12= 7	7/50=0,14	7/50= 0,14

Interpretación de frecuencias para cada tipo de agrupación.

Punto 2

Seleccione un renglón de la tabla de distribución de frecuencia e **interprete en términos del problema planteado**, para cada variable del punto 2), las frecuencias obtenidas (para un valor o un intervalo de la variable).

Tabla de frecuencias para datos agrupados en intervalos:

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Salarios	1	[115.00	123.00)	119.00	2	0.04	2	0.04
Salarios	2	[123.00	131.00)	127.00	10	0.20	12	0.24
Salarios	3	[131.00	139.00)	135.00	19	0.38	31	0.62
Salarios	4	[139.00	147.00)	143.00	12	0.24	43	0.86
Salarios	5	[147.00	155.00]	151.00	7	0.14	50	1.00

Interpretación correspondiente a las frecuencias del segundo intervalo:

$n_2 = 10$ Frecuencia absoluta simple del segundo intervalo. Es la cantidad de empleados que tienen salarios mayores o iguales que 123 y menores que 131 pesos.

$h_2 = 0,20$ Frecuencia relativa simple del segundo intervalo. Es la proporción de empleados que tienen salarios mayores o iguales que 123 y menores que 131 pesos.

$N_2 = 12$ Frecuencia absoluta acumulada del segundo intervalo. Es la cantidad de empleados que tienen salarios menores que 131 pesos.

$H_2 = 0,24$ Frecuencia relativa acumulada del segundo intervalo. Es la proporción de empleados que tienen salarios menores que 131

Interpretación de frecuencias para cada tipo de agrupación.

Punto 2

Tabla de frecuencias para datos agrupados en lista:

Resulta adecuado agrupar los datos de esta manera cuando existen pocos valores observados diferentes y dichos valores se repiten.

Variable	Clase	MC	FA	FR	FAA	FRA
Hijos	1	1	8	0.16	8	0.16
Hijos	2	2	10	0.20	18	0.36
Hijos	3	3	15	0.30	33	0.66
Hijos	4	4	13	0.26	46	0.92
Hijos	5	5	4	0.08	50	1.00

Interpretación correspondiente a las frecuencias de la tercera clase:

$n_3 = 15$ Frecuencia absoluta simple de la tercera clase. Es la cantidad de empleados que tienen 3 hijos.

$h_3 = 0,30$ Frecuencia relativa simple de la tercera clase. Es la proporción de empleados que tienen 3 hijos.

$N_3 = 33$ Frecuencia absoluta acumulada de la tercera clase. Es la cantidad de empleados que tienen 3 o menos hijos.

$H_3 = 0,66$ Frecuencia relativa acumulada de la de la tercera clase. Es la proporción de empleados que tienen 3 o menos hijos.

Interpretación de frecuencias para datos cualitativos.

Punto 2

Tabla de frecuencias para datos cualitativos:

Para un dato cualitativo nominal solo deben calcularse las frecuencias simples:

Variable	Clase	Categorías	FA	FR
Área	1	Administración	21	0.42
Área	2	Comunicación	5	0.10
Área	3	Diseño	13	0.26
Área	4	Ingeniería	9	0.18
Área	5	RRHH	2	0.04

Interpretación correspondiente a las frecuencias de la cuarta clase

$n_4 = 9$ Frecuencia absoluta simple de la cuarta clase. Es la cantidad de empleados que pertenecen al área Ingeniería.

$h_4 = 0,18$ Frecuencia relativa simple de la cuarta clase. Es la proporción de empleados que pertenecen al área Ingeniería.

Interpretación de frecuencias para las distintas variables analizadas. Punto 3

3) Con los datos tabulados o mediante el uso de funciones en InfoStat, responda en forma completa y clara, utilizando vocabulario técnico adecuado en la explicación.

a. ¿Qué cantidad de empleados tienen 3 o menos hijos?

Variable	Clase	MC	FA	FR	FAA	FRA
Hijos	1	1	8	0.16	8	0.16
Hijos	2	2	10	0.20	18	0.36
Hijos	3	3	15	0.30	33	0.66
Hijos	4	4	13	0.26	46	0.92
Hijos	5	5	4	0.08	50	1.00

Hay **33** empleados que tienen 3 o menos hijos.

b. ¿En qué proporción de empleados tiene un salario comprendido entre 123 y menos de 131 pesos? Equivale a decir 123 o más y menos de 131.

0,20

c. ¿Qué proporción de empleados tienen más de 2 hijos?

Variable	Clase	MC	FA	FR	FAA	FRA
Hijos	1	1	8	0.16	8	0.16
Hijos	2	2	10	0.20	18	0.36
Hijos	3	3	15	0.30	33	0.66
Hijos	4	4	13	0.26	46	0.92
Hijos	5	5	4	0.08	50	1.00

$$(15+13+4)/50 = 32/50$$

$$= 0,64$$

También lo podemos resolver de la siguiente manera:

0,30 + 0,26 + 0,08 = 0,64 Es la proporción de empleados que tienen más de 2 hijos.

d. ¿Cuántos empleados tienen un salario menor que 139 pesos?

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Salarios	1	[115.00	123.00)	119.00	2	0.04	2	0.04
Salarios	2	[123.00	131.00)	127.00	10	0.20	12	0.24
Salarios	3	[131.00	139.00)	135.00	19	0.38	31	0.62
Salarios	4	[139.00	147.00)	143.00	12	0.24	43	0.86
Salarios	5	[147.00	155.00]	151.00	7	0.14	50	1.00

31 es la cantidad de empleados que cobran menos de 139

e. ¿Cuántos obreros perciben semanalmente menos de 131 pesos?

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Salarios	1	[115.00	123.00)	119.00	2	0.04	2	0.04
Salarios	2	[123.00	131.00)	127.00	10	0.20	12	0.24
Salarios	3	[131.00	139.00)	135.00	19	0.38	31	0.62
Salarios	4	[139.00	147.00)	143.00	12	0.24	43	0.86
Salarios	5	[147.00	155.00]	151.00	7	0.14	50	1.00

Hay **12** obreros que perciben semanalmente menos de 131 pesos.

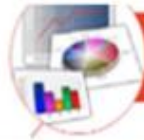
Interpretación de frecuencias para las distintas variables analizadas. Punto 3

f. ¿Qué proporción de empleados pertenecen al área Ingeniería?

Variable	Clase	Categorías	FA	FR
Área	1	Administración	21	0.42
Área	2	Comunicación	5	0.10
Área	3	Diseño	13	0.26
Área	4	Ingeniería	9	0.18
Área	5	RRHH	2	0.04

0,18 Es la proporción de empleados que pertenecen al área Ingeniería.

PRESENTACIÓN DE DATOS ESTADÍSTICOS.



Presentación gráfica



Gráficos Lineales, para agrupamientos en lista

- Gráfico de Bastones (Frecuencias absolutas o relativas)
- Gráfico Acumulativo de Frecuencias (Frecuencias acumuladas, sean absolutas o relativas)

Gráficos de Superficie, para agrupamiento en intervalos

- Histograma
- Polígono de Frecuencias
- Curva Suave

Frecuencias absolutas
Frecuencias relativas
Densidades de Frecuencias

Gráficos Especiales

Para variables categóricas

- Diagrama de Barras Horizontales
- Gráfica de Barras de Componentes de Porcentajes o Barra Porcentual
- Diagrama de Pastel o Círculo Radiado

Para otros fenómenos

- Gráfico de Zonas
- Diagrama de Pareto
- Diagrama de Tallo y Hoja

- Diagrama Escalonado
- Ojiva menor que y Ojiva mayor que
- Curva Acumulativa Suavizada

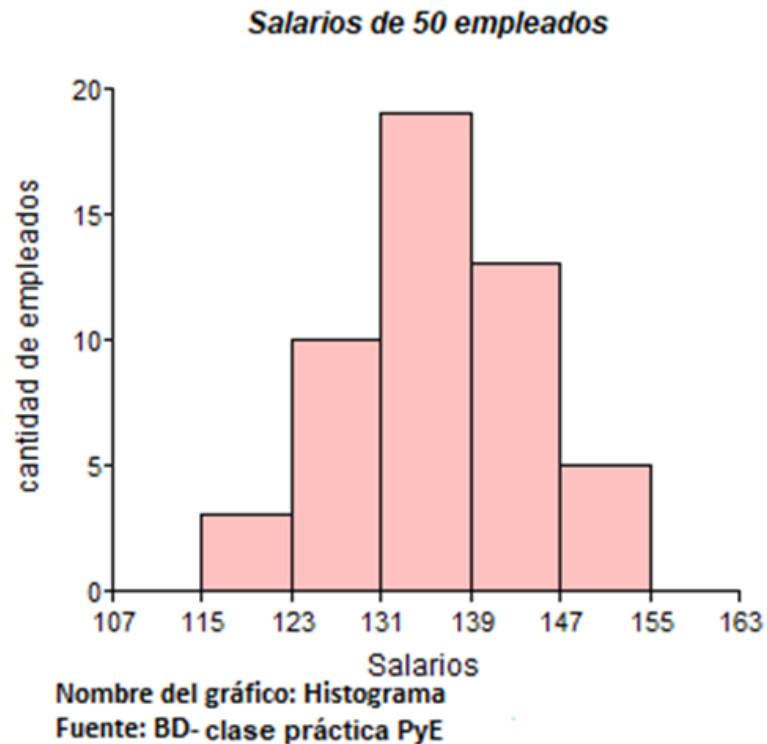
Frecuencias Acumuladas
Absolutas
Relativas

Tipos de gráficos

Gráficos. Punto 4

4) Mediante el empleo del InfoStat grafique e interprete las variables analizadas en los puntos anteriores.

El Histograma y la Ojiva solo se utilizan para representar datos agrupados en intervalos. No se utilizan para representar datos agrupados en lista.



Interpretación: Hasta el tercer intervalo a medida que se incrementan los salarios aumenta la cantidad de empleados. A partir de dicho intervalo a medida que se incrementan los salarios disminuye la cantidad de empleados. La mayor cantidad de empleados tiene salarios comprendidos entre 131 y 139 pesos, mientras que la menor cantidad de empleados se registra para el primero y el último intervalos.

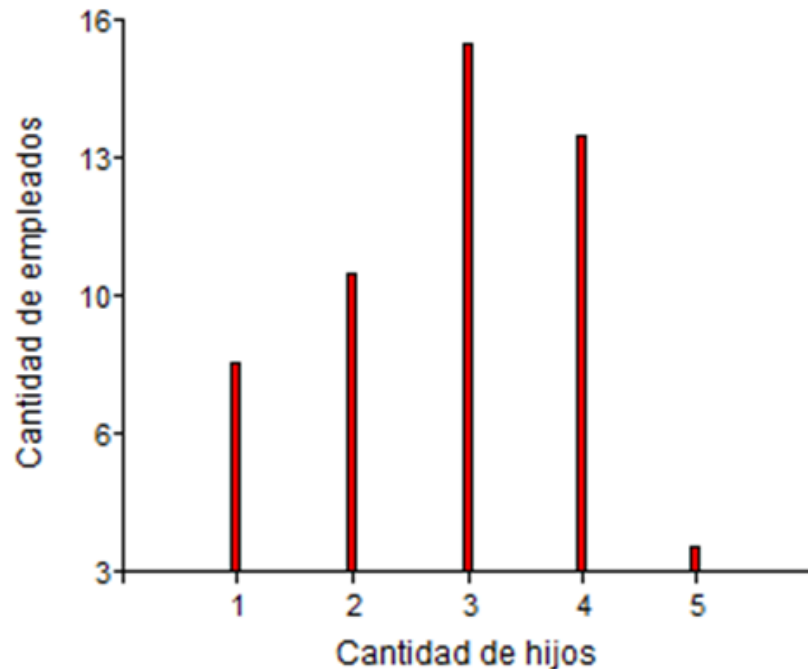


Interpretación: En el segundo intervalo el incremento en la cantidad de empleados se vuelve más acentuado, es más abrupto aún en los tres siguientes intervalos y se vuelve más suave en el último intervalo.

Gráficos. Punto 4

El Diagrama de Bastones y el Gráfico acumulativo de frecuencias se utilizan para representar datos agrupados en lista.

Cantidad de hijos de 50 empleados

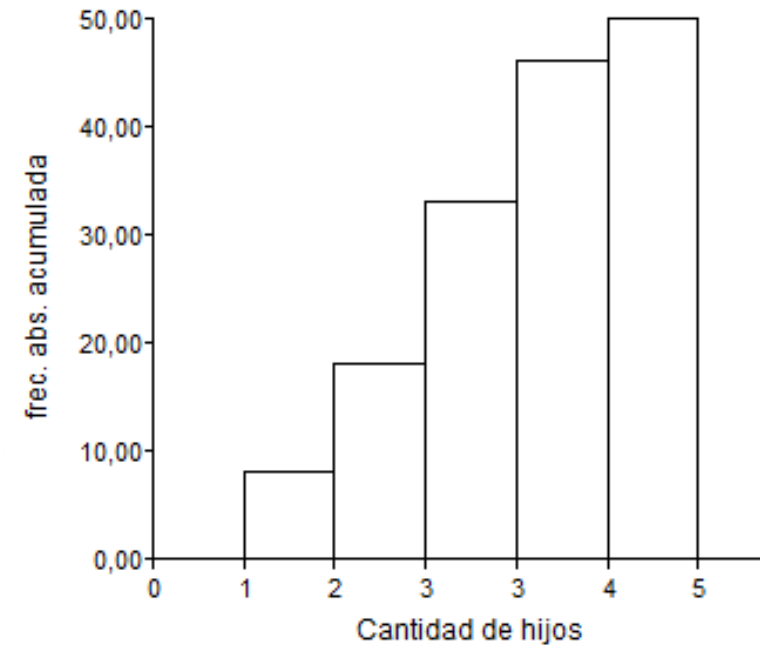


Nombre del gráfico: Diagrama de Bastones

Fuente: BD- clase práctica PyE

Interpretación: Hasta la tercera clase, a medida que se incrementa la cantidad de hijos se incrementa la cantidad de empleados. A partir de dicha clase a medida que aumenta la cantidad de hijos, disminuye la cantidad de empleados. La mayor cantidad de empleados tiene 3 hijos. La menor cantidad de empleados tiene 5 hijos.

Cantidad de hijos de 50 empleados

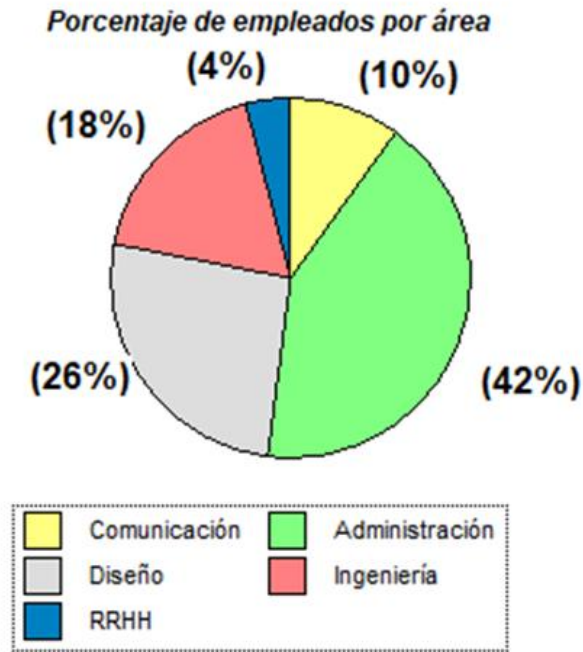


Nombre del gráfico: Gráfico acumulativo de frecuencias

Fuente: BD- clase práctica PyE

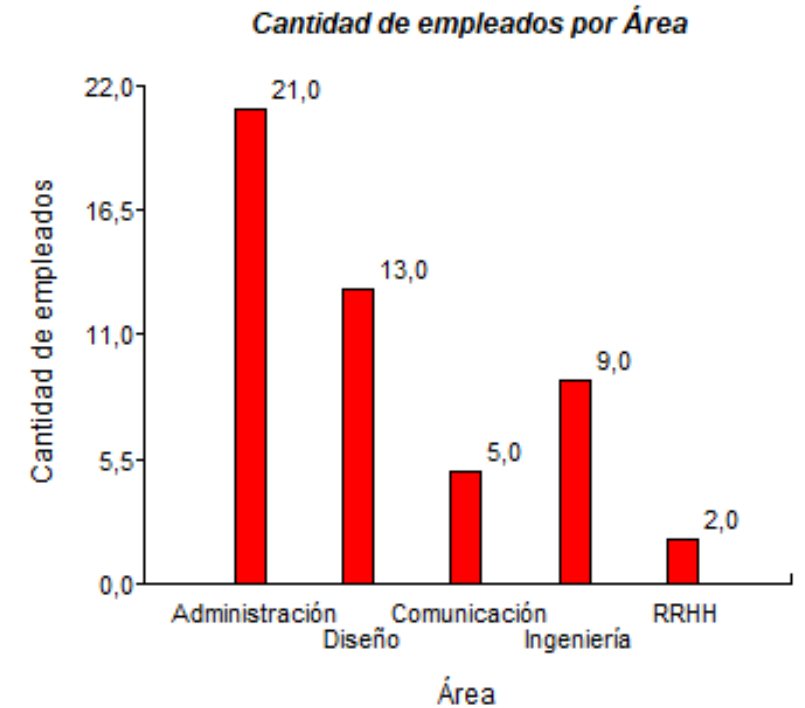
Interpretación: El incremento en la cantidad de empleados se produce de manera muy pareja hasta la cuarta clase, y se vuelve más leve desde la cuarta clase a la siguiente.

Gráficos. Punto 4



Nombre del gráfico: Diagrama de sectores o torta
Fuente: BD_Clase práctica PyE

Interpretación: El mayor porcentaje de empleados corresponde al área Administración y disminuyen para las áreas de Ingeniería y Diseño, en tanto que los menores porcentajes se registran para las áreas de Comunicación y RRHH.



Nombre del gráfico: Diagrama de barras verticales
Fuente: BD_Clase práctica PyE


Interpretación: La mayor cantidad de empleados corresponde al área Administración y disminuyen para las áreas de Ingeniería y Diseño, en tanto que las menores frecuencias absolutas se registran para las áreas de Comunicación y RRHH.

Conceptos importantes. Unidad 3: Descripción de Datos Estadísticos

- Medidas de posición: media, mediana, moda, cuartiles.
- Medidas de dispersión: rango o recorrido, varianza, desv. estándar, coeficiente de variación.
- Medidas de asimetría.
- Medidas de curtosis.

- Central (Media, Mediana y Moda)
- No central (Cuantiles)

Tendencia o Posición o Localización



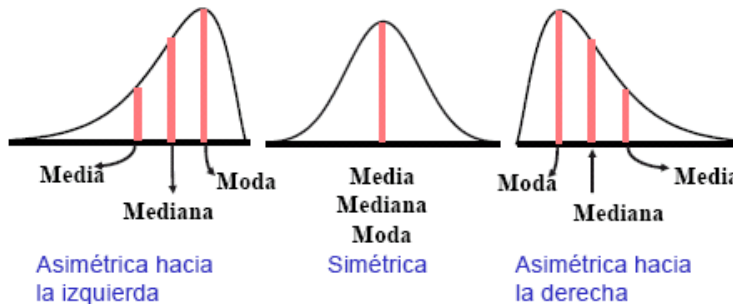

- Amplitud o Rango
- Varianza
- Desviación Típica
- Coeficiente de Variación
- Puntuaciones Z

Dispersión



- Coeficientes de Asimetría y Curtosis

Forma



Medidas descriptivas. Punto 5

5) Calcule para las variables analizadas anteriormente, las medidas descriptivas de posición, dispersión, simetría y curtosis interpretando sus resultados en términos del problema planteado. Realice los cálculos trabajando cada variable como serie simple (directamente sobre la BD utilizando las funciones estadísticas de InfoStat).

Variable número de hijos:

Medidas resumen

Resumen	Hijos
n	50.00
Media	2.90
D.E.	1.20
Var (n-1)	1.44
CV	41.36
Mín	1.00
Máx	5.00
Mediana	3.00
Q1	2.00
Q3	4.00
Asimetría	-0.10
Kurtosis	-0.91

Medidas descriptivas. Punto 5

Interpretación: □

- ▶ **Media aritmética:** el número promedio de hijos es de 2,90.
- ▶ **Mediana:** el 50% de los empleados tienen una cantidad menor o igual a 3 hijos y el otro 50% tiene una cantidad superior o igual a este valor.
- ▶ **Moda:** la cantidad de hijos más frecuente es de 3 hijos (variable con mayor frecuencia absoluta = 15).
- ▶ **Primer Cuartil:** El 25% de los empleados tienen 2 hijos o menos y el 75% restante tienen 2 hijos o más.
- ▶ **Tercer Cuartil:** El 75% de los empleados tienen 4 hijos o menos y el 25% restante tienen 4 hijos o más.
- ▶ **Asimétrica izquierda leve,** debido al coeficiente de asimetría negativo (-0,10), que puede observarse gráficamente en un gráfico de bastones o al comparar media, mediana y moda.
- ▶ **Medida de posición apropiada:** es casi simétrica por lo cual podría usarse la **media o promedio**.
- ▶ **La Desviación estándar** puede interpretarse para indicar el grado de variabilidad de los datos acompañando al promedio y emplearse en el cálculo del CV. La dispersión promedio de las cantidades de hijos con respecto al promedio de número de hijos es de 1,20. También puede interpretarse con el uso de la Regla Interpretativa que desarrollaremos en la Unidad N° 7.
- ▶ **Coeficiente de variación:** la desviación estándar representa el 41,36% de la media aritmética.
- ▶ **La diferencia, Rango o Recorrido,** entre la cantidad de hijos más baja y la más alta es de 4 hijos.
- ▶ **Curtosis o Kurtosis:** Una curtosis negativa de -0,91 indica que la distribución de las cantidades de hijos es platicúrtica es decir que existe un reducido grado de concentración alrededor de los valores centrales de la variable.

Medidas descriptivas. Punto 5

Variable salarios:

Medidas resumen

Resumen	Salarios
n	50.00
Media	137.22
D.E.	8.56
Var (n-1)	73.36
CV	6.24
Mín	117.00
Máx	153.00
Mediana	137.50
Q1	131.00
Q3	145.00
Asimetría	-0.22
Kurtosis	-0.65

Interpretación: □

- ▶ **Media:** el salario promedio de los empleados es de 137,22\$.
- ▶ **Mediana:** el 50% de los empleados tienen un salario igual o menor a 137,50 \$ y el otro 50% tiene un salario superior o igual a este valor.
- ▶ **Moda:** el salario más frecuente se encuentra entre 131\$ y 139\$ (intervalo con mayor frecuencia).
- ▶ **Primer Cuartil:** El 25% de los empleados cobran un salario menor o igual a 131\$ y el 75% restante cobran un salario igual o mayor de 131\$.
- ▶ **Tercer Cuartil:** El 75% de los empleados cobran un salario menor o igual a 145\$ y el 25% restante cobran un salario igual o mayor de 145\$.
- ▶ **Asimétrica izquierda,** debido al coeficiente de asimetría negativo (-0,22), que puede observarse gráficamente en un histograma o polígono de frecuencia o al comparar media y mediana.
- ▶ La **Desviación estándar** puede interpretarse para indicar el grado de variabilidad de los datos acompañando al promedio y emplearse en el cálculo del CV. La dispersión promedio de los salarios con respecto al salario promedio es de 8,56\$. También puede interpretarse con el uso de la Regla Interpretativa que desarrollaremos en la Unidad N° 7.
- ▶ **Coeficiente de variación:** la desviación estándar representa el 6,24% de la media aritmética.
- ▶ **Medida de posición apropiada:** por el tema de la asimetría sería la mediana pero se puede considerar el uso del promedio también ya que la asimetría es leve y el CV es bajo.
- ▶ La diferencia, **Rango o Recorrido**, entre el salario más bajo y el más alto es de aproximadamente 36\$.
- ▶ **Curtosis o Kurtosis:** Una curtosis negativa de -0,65 indica que la distribución de los salarios es platicúrtica es decir que existe un reducido grado de concentración alrededor de los valores centrales de la variable.

Medidas descriptivas. Punto 5

Variable Área, cualitativa nominal, sólo se calcula la **Moda**, que sería el área con mayor frecuencia en este caso **Administración** con una frecuencia de 21 empleados. El resto de las medidas no se calculan por no ser pertinentes.

U

Variable	Clase	Categorías	FA	FR
Área	1	Administración	21	0.42
Área	2	Comunicación	5	0.10
Área	3	Diseño	13	0.26
Área	4	Ingeniería	9	0.18
Área	5	RRHH	2	0.04

Fuente: BD- clase práctica PyE

Si el dato cualitativo fuera ordinal también se podría calcular la **Mediana**.

Propiedades de Media y Varianza. Punto 6 y 7

6) ¿Cuáles sería el valor del promedio y el de la desviación estándar correspondiente a la variable Salario, si se ha comprobado que existe un error en más (por exceso) del 3,21% en el registro de los datos?

PROPIEDADES DE MEDIA Y VARIANZA

Propiedades de Media:

$$M(c) = c$$

$$M(x + c) = M(x) + c$$

$$M(x * c) = c * M(x)$$

$$M(x + y) = M(x) + M(y)$$

Propiedades de Varianza:

$$V(c) = 0$$

$$V(x + c) = V(x)$$

$$V(x * c) = c^2 * V(x)$$

$$V(x + y) = V(x) + V(y) \text{ si } \underline{x} \text{ e } \underline{y} \text{ son independientes.}$$

Propiedades de Media y Varianza. Punto 6 y 7

Error por exceso → error en más

Tendremos en cuenta la siguiente simbología:

w = salario correcto

x = salario original con el error

Un error en más significa que está sobrando el 3,21% por lo tanto, al valor “x” le debemos restar el 3,21%.

En primer lugar debemos calcular el 3,21% de x, para ello planteamos una regla de 3 simple:

100% ____ x

3,21% ____ $(3,21\% \cdot x) / 100\% = 0,0321x$

Si al valor “x” le restamos “0,0321x” le estaremos restando el 3,21% y de esa manera obtendremos el salario correcto “w”:

$w = x - 0,0321x$

Sacamos factor común a la variable “x” obteniendo:

$w = (1 - 0,0321)x$

$w = 0,9679x$

Si a “x” lo multiplicamos por 0,9679 le estaremos restando el 3,21% que está sobrando.

Propiedades de Media y Varianza. Punto 6 y 7

Cálculo del nuevo Promedio:

Aplicamos media en ambos miembros de la igualdad

$$M(w) = M(0,9679x)$$

En el segundo miembro de la igualdad nos queda planteada la media de una constante por una variable por lo tanto corresponde aplicar la propiedad: “La media de una constante por una variable es igual a la constante por la media de la variable”: $M(cx) = cM(x)$

$$\begin{aligned} M(w) &= M(0,9679x) \\ &= 0,9679 M(x) \end{aligned}$$

Medidas resumen

Variable	n	Media	D.E.	Var(n-1)	CV	Min	Máx	Mediana	Q1	Q3	Asimetría	Kurtosis
Salarios	50	137.22	8.56	73.36	6.24	117.00	153.00	137.50	131.00	145.00	-0.22	-0.65

$M(x) = 137,22$ Media calculada con Infostat a partir de la columna Salarios en la B.D.

$$M(w) = 0,9679 * 137,22$$

$M(w) = 132,82$ Nuevo salario promedio luego de corregir el error del 3,21% en más.

Propiedades de Media y Varianza. Punto 6 y 7

Cálculo de la nueva Desviación Estándar:

Antes de calcular la nueva Desviación Estándar debemos obtener la nueva Varianza.

Aplicamos varianza en ambos miembros de la igualdad

$$V(w) = V(0,9679x)$$

En el segundo miembro de la igualdad nos queda planteada la varianza de una constante por una variable por lo tanto corresponde aplicar la propiedad: “La varianza de una constante por una variable es igual al cuadrado de la constante por la varianza de la variable”

$$\begin{aligned} V(w) &= V(0,9679x) \\ &= (0,9679)^2 V(x) \end{aligned}$$

Medidas resumen

Variable	n	Media	D.E.	Var(n-1)	CV	Mín	Máx	Mediana	Q1	Q3
Salarios	50	137.22	8.56	73.36	6.24	117.00	153.00	137.50	131.00	145.00

$V(x) = 73,36$ Esta es la varianza que debe calcularse a través del Infostat a partir de la columna Salarios en la B.D.

$$V(w) = (0,9679)^2 73,36$$

$V(w) = 68,73$ Nueva varianza luego de corregir el error del 3,21% en más.

Se pide la Desviación Estándar, debe obtenerse calculando la raíz cuadrada de la varianza:

$$DS(w) = \sqrt{68,73}$$

$DS(w) = 8,29$ Nueva Desviación Estándar luego de corregir el error del 3,21% en más.

Propiedades de Media y Varianza. Punto 6 y 7

7) ¿Cuáles sería el valor del promedio y el de la desviación estándar correspondiente a la variable cantidad de hijos si se ha comprobado que existe un error por defecto de 1 hijo en el registro de los datos?

Error por defecto = error en menos

Cabe aclarar que en este caso no corresponde el planteo de la regla de 3 porque el error no está expresado en % (es un error de 1 en menos y no de un 1% en menos) por lo tanto directamente sumamos 1:

$$w = x + 1$$

Cálculo del nuevo Promedio:

A continuación aplicamos media en ambos miembros de la igualdad:

$$M(w) = M(x + 1)$$

En el segundo miembro de la igualdad nos queda planteada la media de la suma entre una variable y una constante por lo tanto corresponde aplicar la propiedad: “La media de la suma entre una variable y una constante es igual a la media de la variable más la constante”

$$M(w) = M(x + 1)$$

$$= M(x) + 1$$

$M(x) = 2,90$ Esta es la media que debe calcularse a partir de la columna Hijos en la B.D.

Propiedades de Media y Varianza. Punto 6 y 7

Utilizando Infostat:

Medidas resumen

Variable	n	Media	D.E.	Var(n-1)	CV	Mín	Máx	Mediana	Q1	Q3
Hijos	50	2.90	1.20	1.44	41.36	1.00	5.00	3.00	2.00	4.00

$$M(w) = 2,90 + 1$$

= 3,90 Nueva cantidad de hijos promedio luego de corregir el error de 1 hijo por defecto.

Cálculo de la nueva Desviación Estándar:

Antes de calcular la nueva Desviación Estándar debemos obtener la nueva Varianza.

Aplicamos varianza en ambos miembros de la igualdad

$$V(w) = V(x + 1)$$

En el segundo miembro de la igualdad nos queda planteada la media de la diferencia entre una variable y una constante por lo tanto corresponde aplicar la propiedad: “La varianza de la diferencia entre una variable y una constante es igual a la varianza de la variable” por lo tanto, la varianza no se modifica

$$V(w) = V(x + 1)$$

= V(x) La varianza no se modifica.

Propiedades de Media y Varianza. Punto 6 y 7

Medidas resumen

Variable	n	Media	D.E.	Var(n-1)	CV	Mín	Máx	Mediana	Q1	Q3
Hijos	50	2.90	1.20	1.44	1.36	1.00	5.00	3.00	2.00	4.00

$V(x) = 1,44$ Esta es la varianza que debe calcularse a partir de la columna Hijos, utilizando el Infostat

$V(w) = 1,44$ (la varianza no se modifica si incrementamos o restamos una suma fija, solo cambia su valor cuando la modificación se expresa en porcentaje)

En consecuencia la DS tampoco se modifica:

$$DS(x) = \sqrt{1,44}$$

= 1,20 La desviación estándar no se modifica luego de corregir el error de 1 hijo por defecto.