

## Trabajo Práctico N°1: Estadística Descriptiva

1. En el último Censo de Población y Viviendas realizado, se obtuvo información referida a personas y viviendas; para cada caso identifique tipo de variable y escala de medición estadística.
  - a) En lo referente a personas se indagó:
    - I) Edad.
    - II) Sexo.
    - III) Estado civil.
    - IV) Si sabe leer y escribir.
    - v) Nivel de educación alcanzado.
  - b) En lo referente a viviendas se indagó sobre:
    - I) Tipo de vivienda (casa, departamento, rancho,...)
    - II) Material predominante en las paredes (ladrillo, madera, chapa,.....)
    - III) Cantidad de habitaciones.
    - IV) Combustible utilizado para cocinar.
    - v) Cantidad de personas que la habitan.
2. Los siguientes datos corresponden a la producción anual citrícola de la Provincia de Corrientes, año 2.013, expresada en miles de toneladas. Los datos fueron extraídos de la página “Corrientes exporta” del Ministerio de Producción, Trabajo y Turismo.

Naranja	142
Mandarinas	60
Limonos	32
Pomelos	8
Total Corrientes	242

- a) Confeccione la tabla de frecuencias correspondiente.
  - b) Realice un gráfico adecuado.
3. Las notas de un examen de matemáticas de 30 alumnos de una clase de 4to año de una escuela de la ciudad de Corrientes en el mes de junio de 2015 son las siguientes:

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

- a) Determine población, unidad elemental y escala de medición.
  - b) Construya la tabla de frecuencias simples y acumuladas.
  - c) Construya el gráfico de bastones.
  - d) Interprete cada una de las medidas descriptivas obtenidas utilizando el contexto del problema.

$$\bar{X} = 5,83$$

$$Mo = 8$$

$$Me = 7$$

$$Q_1 = 4$$

$$Q_2 = Me = 7$$

$$Q_3 = 8$$

$$R = 9$$

$$RI = 4$$

$$\hat{S}^2 = 8,83$$

$$\hat{S} = 2,97$$

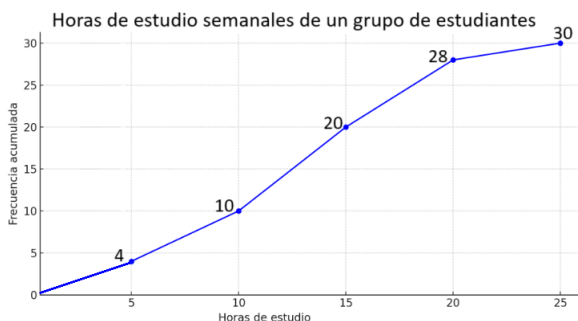
4. Una empresa de informática dedicada al análisis de virus en ordenadores, contabiliza los virus detectados con su producto en 20 ordenadores de domicilios particulares. Los resultados obtenidos son los siguientes:
  - a) Determine unidad elemental y escala de medición.
  - b) Complete la tabla con todas las frecuencias simples y acumuladas en todos los términos.

Nro. de Virus	$f_i$	$F_i$	$r_i$	$R_i$
28-36	3		0,15	
36-44		5		
44-52				0,5
52-60	7			
60-68				

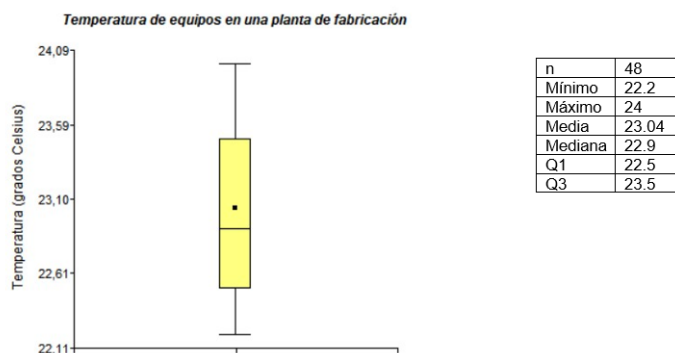
- c) Grafique el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas u ojiva.
- d) Interprete los siguientes resultados:  $Me = 52$  y  $\bar{X} = 50$ .
- e) Se calculó analíticamente la moda y se obtuvo que  $Mo = 54,67$ , interprete este resultado y compare con el valor de la moda obtenido gráficamente.
- f) Interprete las siguientes medidas descriptivas:  $\hat{S}^2 = 106,95$ ,  $\hat{S} = 10,34$  y  $Ap = -0,58$ .
5. Sea  $X$  el tiempo en minutos que debe esperar un vehículo para cruzar una intersección transitada que cuenta con semáforos. Los datos siguientes se obtienen de una muestra aleatoria de 36 vehículos:

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

- a) Agrupe los datos en seis intervalos de clase de amplitud 1 y extremos enteros.
- b) Calcule e interprete la media, la desviación típica y la moda.
6. Supongamos que se midió el tiempo de estudio semanales de un grupo de estudiantes y se obtuvieron los siguientes resultados que se presentan en el gráfico:



- a) Determine variable, tipo de variable y escala de medición.
- b) Construya la tabla de frecuencias simples y acumuladas.
- c) Calcule e interprete la moda.
- d) ¿Por debajo de qué tiempo se encuentra el 50% de los alumnos?
7. Un ingeniero trabaja en una planta de fabricación, donde se utiliza un sensor de temperatura (en grados Celsius) para monitorear las condiciones en tiempo real. Este sensor mide la temperatura cada 10 minutos durante un período de tiempo (8 horas de trabajo) para asegurarse de que los equipos se mantengan dentro de un rango de temperaturas específicas. En un determinado día de octubre de 2020 se registraron los datos y se presenta a continuación el Box-plot correspondiente:



- a) ¿Cuál es la temperatura mínima por encima de la cual se encuentra el 25 % de los datos?
- b) Determine el intervalo donde se encuentra el 50 % central de los datos.
- c) ¿La distribución es simétrica? Justifique.
8. Una empresa multinacional con sucursales en Argentina y Estados Unidos realizó un relevamiento de los salarios mensuales netos (en moneda local) de sus empleados. Los resultados fueron agrupados en las siguientes tablas de frecuencias:

Intervalo salarial (ARS)	Nro. de Empleados
400.000 – 449.999	4
450.000 – 499.999	9
500.000 – 549.999	12
550.000 – 599.999	7
600.000 – 649.999	5
650.000 – 699.999	3

Sueldos en ARS

Intervalo salarial (USD)	Nro. de Empleados
2.500 – 2.999	3
3.000 – 3.499	6
3.500 – 3.999	10
4.000 – 4.499	8
4.500 – 4.999	5
5.000 – 5.499	3

Sueldos en USD

- a) Indique en cuál de las dos sucursales se presenta mayor variabilidad. Justifique su respuesta.
- b) Si la cotización del dólar es 1 USD = 1050 ARS, ¿sería necesario convertir los salarios estadounidenses a pesos antes de comparar? Justificar su respuesta desde el punto de vista estadístico.
9. Para estudiar la relación entre el hábito de fumar y la aparición de Cardiopatía Coronaria, se observaron 80 fumadores y 70 no fumadores durante un período de tiempo, con el siguiente resultado: 35 de los primeros y 16 de los segundos presentaron dicha enfermedad. Este estudio se realizó en el Hospital Llano de la Ciudad de Corrientes en 2.013. Los datos fueron suministrados por el Ministerio de Salud de la Provincia de Corrientes.
- a) Confeccione una tabla estadística.
- b) Represente gráficamente los datos.
10. Un médico clínico, con el fin de evaluar el nivel de cefalea de sus pacientes, los encuesta en la primera consulta. Los siguientes datos corresponden a los pacientes que consultaron por primera vez en el 2007.

Tiene cefalea	Cantidad de pacientes
1: nunca	12
2: a veces	45
3: muchas veces	29
4: siempre	8

- a) Determine:
- i) Población.

II) Muestra.

III) Tipo de variable y su clasificación.

b) Presente los datos en una tabla de distribución de frecuencias.

c) Represente gráficamente los datos.

d) En base a esta tabla, se obtuvieron las siguientes medidas:

Medida	Valor
Media	2.34
Mediana	2
Moda	2
Cuartil 1 (Q1)	2
Cuartil 3 (Q3)	3
Desvío estándar	0.7925

Interprete las medidas descriptivas que correspondan a esta situación.

11. Se pretende estudiar si existe alguna relación entre el nivel de nutrición de un niño que viven en zonas peri urbanas de la ciudad de Corrientes y su rendimiento escolar. Para ello, de un total de 180 niños en edad escolar que asistieron en el año 2007 al tercer grado de cierta escuela municipal, se seleccionaron en forma aleatoria un grupo de 52 niños. Entre otras variables, se analizó la cantidad de veces que el niño no asistió a la escuela durante ese año, obteniéndose:

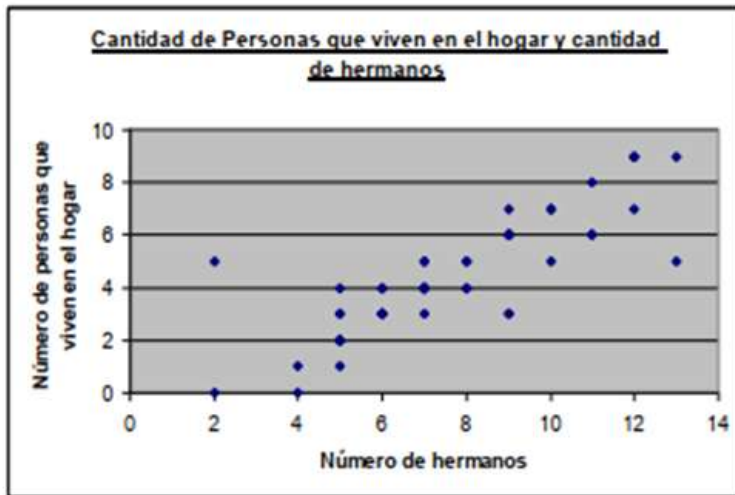
<i>n° faltas - TODOS</i>		<i>n° faltas - VARONES</i>		<i>n° faltas - MUJERES</i>	
Media	14,135	Media	13,920	Media	14,333
Error típico	1,627	Error típico	2,732	Error típico	1,902
Mediana	11	Mediana	9	Mediana	11
Moda	4	Moda	4	Moda	5
Desviación estándar	11,734	Desviación estándar	13,662	Desviación estándar	9,884
Varianza de la muestra	137,687	Varianza de la muestra	186,660	Varianza de la muestra	97,692
Curtosis	3,044	Curtosis	4,042	Curtosis	-0,452
Coefficiente de asimetría	1,563	Coefficiente de asimetría	1,916	Coefficiente de asimetría	0,745
Rango	58	Rango	58	Rango	35
Mínimo	1	Mínimo	1	Mínimo	3
Máximo	59	Máximo	59	Máximo	38
Suma	735	Suma	348	Suma	387
Cuenta	52	Cuenta	25	Cuenta	27

a) Considerando el grupo de niños varones ¿cuál es el promedio de inasistencias en un año?

b) Justifique comparando la variable entre ambos sexos, ¿Qué promedio es más representativo?

c) ¿Que pude concluir acerca de la dispersión de la variable considerando el grupo completo? ¿Y en cuanto a la simetría?

d) Otras de las variables analizadas fueron la “cantidad de personas que conviven en el hogar del niño” y la “cantidad de hermanos” menores de edad que viven con él. Teniendo en cuenta la información que se presenta a continuación, ¿Qué conclusión obtiene del análisis conjunto de estas variables?



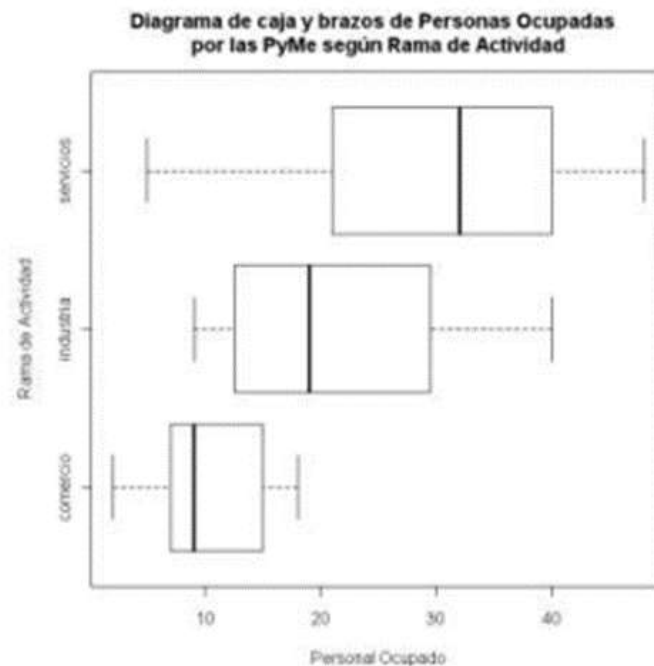
Coefficiente de correlación

	n° pers_hogar	n° hnos
n° pers_hogar	1	
n° hnos	0,836992094	1

Covarianza

	n° pers_hogar	n° hnos
n° pers_hogar	7,154955621	
n° hnos	4,957100592	4,902366864

12. En un corporativo se procesaron los datos del personal ocupado por las empresas según la rama de actividad. El presidente del mismo requiere un informe a partir de la siguiente información. Establezcan qué conclusiones creen pertinentes para poder evaluar la ocupación del personal.



<b>Resumen:</b>						<b>Desviación estándar:</b>		
> tapply(Personal,Rama,summary)						> tapply(Personal, Rama, sd)		
\$Industria						Comercio	Industria	Servicios
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	5.385	12.134	15.065
9.00	12.50	19.00	21.71	29.50	40.00			
\$Comercio						<b>Curtosis</b>		
> tapply(Personal, Rama, kurtosis)						Comercio	Industria	Servicios
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	-1.515	-1.706	-1.544
2.00	7.00	9.00	10.33	15.00	18.00			
\$Servicio						<b>Asimetría:</b>		
> tapply(Personal, Rama, skewness)						Comercio	Industria	Servicio
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	0.097	0.421	-0.318
5.00	21.00	32.00	28.78	40.00	48.00			