ESTADÍSTICA DESCRIPTIVA

Ejercicio 1.

a.

Diagrama de tallo y hojas para los pesos

5	6
6	04556
7	0019
8	049
9	9

Diagrama de tallo y hojas para las estaturas

16	1489
17	34457
18	0345
19	
20	5

b. El objetivo del ejercicio es trabajar con datos sin agrupar ya que en cada caso se tienen pocas observaciones. Recordar que para buscar la mediana y los cuartiles deben ordenarse los datos **de menor a mayor**.

Peso (kg):

Media: 73.714 kg (con Alcula)

Mediana: 70 kg (promedio entre datos séptimo y octavo)

Desvío estándar: 12.009 kg (con Alcula)

Primer cuartil: 65 kg (cuarto dato)

Tercer cuartil: 80 kg (décimo primer dato)

Estatura (cm):

Media: 176.57 cm (con Alcula)

Mediana: 174.5 (promedio entre datos séptimo y octavo)

Desvío estándar: 10.91 cm (con Alcula)

Primer cuartil: 169 cm (cuarto dato)

Tercer cuartil: 183 cm (décimo primer dato)

c. Para responder a esta pregunta no alcanza con comparar los desvíos típicos calculados para cada conjunto de datos porque no refieren a magnitudes similares e incluso presentan unidades de medidas diferentes y distintas medias. Se emplea aquí el ***coeficiente de variación***.

Peso: CV=0.1629; Estatura: CV=0.0618.

El conjunto de mediciones de pesos es el que tiene mayor variabilidad (mayor CV).

Ejercicio 2.

a) Variable: Nro. de virus detectados con el software en PC de domicilios particulares.

- Media: $\overline{x_I} = 46{,}33$
- Mediana:

Datos ordenados: 29 – 35 – 37 – 46 – 47 – 53 – 54 – 55 – 61

$n = 9 \rightarrow impar \ \rightarrow \tilde{x} = x_{\frac{2}{4}(9+1)} = x_{(5)} = 47$

- Rango: $R = x_{máx} - x_{mín} = 61 - 27 = 32$
- Desvío estándar: $s = 10{,}665$

b) Por debajo del tercer cuartil Q_3 encontramos el 75% de los datos.

Por lo tanto, calculamos la posición del mismo y luego lo determinamos en los datos ordenados.

Aquí $n = 9 \rightarrow impar \rightarrow Q_3 = x_{\frac{3}{4}(9+1)} = x_{(8)} = 55$

En 75% de los ordenadores analizados con este software se detectaron menos de 55 virus.

Ejercicio 3. Resuelto en clase.

Ejercicio 4.

a) Calificación de estudiantes en prueba inicial

- Media: $\overline{x_I} = \sum_{i=1}^{n} \frac{x_i}{n} = 2{,}7$
- Mediana:

Datos ordenados: 1 – 1 – 1 – 2 – 2 – 3 – 3 – 4 – 5 - 5

Aquí $n = 10 \rightarrow \widetilde{x_I} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{2+3}{2} = 2{,}5$

- Moda:

Nota más frecuente. $\widehat{x_I} = 1$

- Desvío estándar: $s_I = 1{,}567$
- Varianza: $s_I^2 = 1{,}567^2 = 2{,}455$

Calificación de estudiantes en prueba final

- Media: $\overline{x_F} = \sum_{i=1}^{n} \frac{x_i}{n} = 6{,}3$
- Mediana:

Datos ordenados: 3 – 4 – 5 – 6 – 6 – 6 – 7 – 8 – 9 – 9

Aquí $n = 10 \rightarrow \widetilde{x_F} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{6+6}{2} = 6$

- Moda:

Nota más frecuente. $\widehat{x_F} = 6$

- Desvío estándar: $s_F = 2{,}003$
- Varianza: $s_F^2 = 2{,}003^2 = 4{,}012$

b) Como los datos tienen diferente media, para averiguar cuál conjunto de datos es más disperso calculamos los coeficientes de variación:

Para las notas de la prueba inicial: $CV_I = \frac{s_I}{\overline{x_I}} = \frac{1{,}567}{2{,}7} = 0{,}581$

Para las notas de la prueba final: $CV_F = \frac{s_F}{\overline{x_F}} = \frac{2{,}003}{6{,}3} = 0{,}318$

Como $CV_I > CV_F$ concluimos que el conjunto de calificaciones de la prueba inicial presenta mayor dispersión.

c) Definimos: Incremento de la calificación obtenida = $x_F - x_I$

Resultando: 2 – 3 – 4 - 4 – 1 – 3 – 5 – 5 – 3 – 6

- Media: $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 3{,}6$
- Mediana:

Datos ordenados: 1 – 2 – 3 – 3 – 3 – 4 – 4 – 5 – 5 – 6

Aquí $n = 10 \rightarrow \tilde{x} = \frac{x_{(5)}+x_{(6)}}{2} = \frac{3+4}{2} = 3{,}5$

- Desvío estándar: $s = 1{,}5055$

Ejercicio 5. Resuelto en clase.

Ejercicio 6.

a. Tabla completa.

Inter. de clase	X_{PM}	fa	Fa	fr	Fr
[0, 2)	1	6	6	0.10	0.10
[2, 4)	3	9	15	0.15	0.25
[4, 6)	5	23	38	0.38	0.63
[6 ,8)	7	12	50	0.20	0.83
[8 , 10)	9	10	60	0.17	1

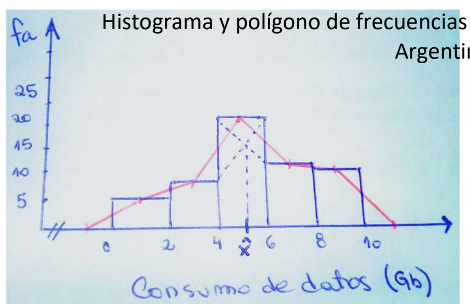
P(X<4)= 0,10+0,15 *Se suman las fr de los intervalos [0,2) y [2,4)

P(X>6)=0,20+0,17

P(2≤ X<8)=0.10+0,15+0,38+0,20

b. Gráfica. La moda aproximadamente es 5.5 Gb (determinada gráficamente).

Histograma y polígono de frecuencias absolutas para el consumo de datos en Argentina (2022)



c. El consumo promedio mensual de datos móviles es de 5.37 Gb. S=2.37 Gb.

d. El 50% de los usuarios consume menos de 5.3 Gb de datos mensualmente.

e. Como los datos de consumo en Argentina y Uruguay tienen diferentes medias, comparamos Coef. De Variación.

CV_{Arg}=0.44; CV_{Uru}=0.34. Como CV_{Arg}>CV_{Uru} → Los datos de Argentina presentan mayor variabilidad que los de Uruguay.

Ejercicio 7. Anulado porque no vimos gráficos para variables discretas.

Ejercicio 8.

a) Concentración de plomo en una misma muestra analizada por el laboratorio [g/L]

Se completó la tabla observando la salida del software. La columna IQR es el rango intercuartil (RI=Q3-Q1) y la columna CV es el coeficiente de variación (CV=S/$\bar{X}$).

Lab.	Media ± DS	Mínimo	Cuartil 1	Mediana	Cuartil 3	Máximo	IQR	CV
A	3,433 ±0,993	2,300	2,450	3,400	4,300	4,900	1,850	0,289
B	5,080 ±1,215	4,000	4,100	4,400	6,400	6,500	2,300	0,239
C	2,833 ±1,098	1,600	1,675	2,750	4,100	4,100	2,425	0,388
D	3,743 ±1,003	2,100	2,800	3,800	4,800	4,800	2,000	0,268
E	7,075 ±1,497	5,500	5,650	7,150	8,425	8,500	2,775	0,212

b) Mirando la columna "N" de la salida del software se observa que se realizaron 6 análisis en laboratorio A, 5 en B, 6 en C, 7 en D y 4 en E. Por lo tanto, se realizó más cantidad de análisis en laboratorio D.

c) Para comparar dispersión se utiliza COEFICIENTE DE VARIACIÓN, debido a que los conjuntos de mediciones tienen diferentes medias.

Comparando los coeficientes de variación de los laboratorios (columna "CV") se observa que los análisis realizados en el laboratorio C presentan mayor dispersión (mayor CV: CV_{labC}=0,388).

Ejercicio 9. Resuelto en clase.

Ejercicio 10. Anulado porque no vimos gráficos para variables discretas.