

Pauta guía N° 1

Problema 1.

a) Datos ordenados de mayor a menor.

b) Para graficar los datos en un diagrama box-plot se necesita: Q1; Q2 = Me; Q3. Para esto, ordenamos los datos de menor a mayor y revisamos las posiciones para n=26 datos.

Para el primer cuartil, la fórmula nos pide revisar el dato en la posición 6,5, por se puede promediar el 6 y el 7:

$$0,5 \cdot (35,6 + 35,9) = 35,75 = Q_1$$

Nota: también es válido si se aproxima a la posición N°7

Para el segundo cuartil, revisamos el dato en la posición 13, Q2 = 36,9 = Me

Y para el tercer cuartil, la fórmula nos dice que revisemos la posición 20,5, por lo que promediaremos los datos 20 y 21

$$0,5 \cdot (39,3 + 39,2) = 39,15 = Q_3$$

Nota: también es válido si se aproxima a la posición N°21

Como tenemos los cuartiles, debemos determinar los bigotes. Para esto, necesitamos el rango intercuartil

$$RIQ = Q_3 - Q_1 = 39,15 - 35,75 = 3,4$$

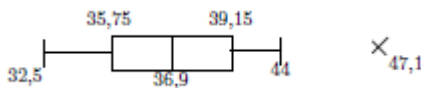
Luego, nuestros bigotes son:

$$Bi = 32,5$$

$$Bs = 44$$

Como el bigote superior es menor a nuestro dato más grande, nos damos cuenta de que este último es un dato extremo.

Por lo tanto el diagrama box-plot se ve así:



y podemos concluir que sí existen datos extremos, éste es 47,1

c) Obtendremos la media y la moda, ya que la mediana ya la obtuvimos.

$$\bar{x} = \frac{1}{26} \sum_{i=1}^{26} x_i = 37,25$$

Y por inspección podemos ver que tenemos varias modas: 32;5; 35;6; 35;9; 36;4; 37;3.

Como tenemos presencia de datos extremos, la medida de tendencia central más adecuada será la mediana.

d) Nos piden determinar el dato que es sobrepasado por el 58% del resto de los datos. Si lo vemos de otra manera, ese dato sobrepasa al 42% del resto, y eso es lo mismo que el percentil 42.

Entonces, de manera similar a los cuartiles, el percentil 42 es la posición 10,92; la que se aproximará al dato de la posición 11.

e) Sabemos que la media es:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

es decir, la suma de todos los datos entre la cantidad de estos. Luego, con un dato adicional, la cantidad de datos aumenta en 1, por lo tanto la nueva cantidad de datos es $n + 1$, entonces el nuevo promedio será:

$$\frac{1}{n+1} \sum_{i=1}^{n+1} x_i = \frac{1}{n+1} \left(\sum_{i=1}^n x_i + x_{n+1} \right) = \frac{1}{n+1} (n\bar{x} + x_{n+1})$$

f) Siguiendo el razonamiento anterior, tenemos que la muestra añadida es de tamaño i y su media es $\bar{x}'$, además nuestra cantidad total de datos es $n + i$. Entonces el nuevo promedio es:

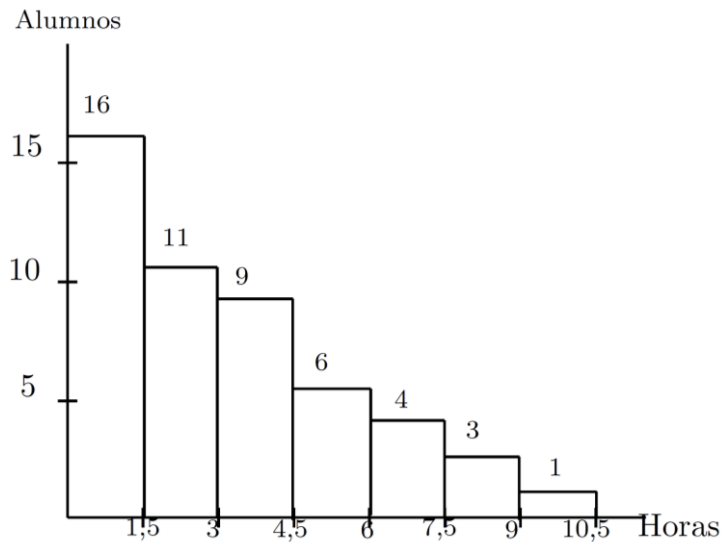
$$\frac{1}{n+i} \sum_{k=1}^{n+i} x_k = \frac{1}{n+i} \left(\sum_{k=1}^n x_k + \sum_{j=1}^i x_j \right) = \frac{1}{n+i} (n\bar{x} + i\bar{x}')$$

Comprobado esto, nuestro nuevo promedio será 37,4

$$\frac{1}{31} (26 \cdot 37,25 + 5(38,7 + 36,9 + 40,1 + 40 + 34,3)) = 37,4$$

Problema 2

a) Como vemos, la variable en estudio es cuantitativa continua, ya que toma valores reales. Para representar gráficamente los datos, como están agrupados, es fácil hacer un histograma.



Podemos ver que hay una clara tendencia a los números más bajos, siendo el intervalo modal $[0; 1,5)$

b) El promedio nos da 3,27. La mediana es 2,73. La moda es 1,14.

c) Usando la ecuación de los percentiles, tenemos la siguiente:

$$2 = 1,5 + \frac{1}{16}(i/100 \cdot 50) \cdot 1,5$$

Despejando, $i = 10,67$. En otras palabras, aprobó el 10,67% de los alumnos.

d) Como vimos en b), el tiempo más frecuente es 1,14.

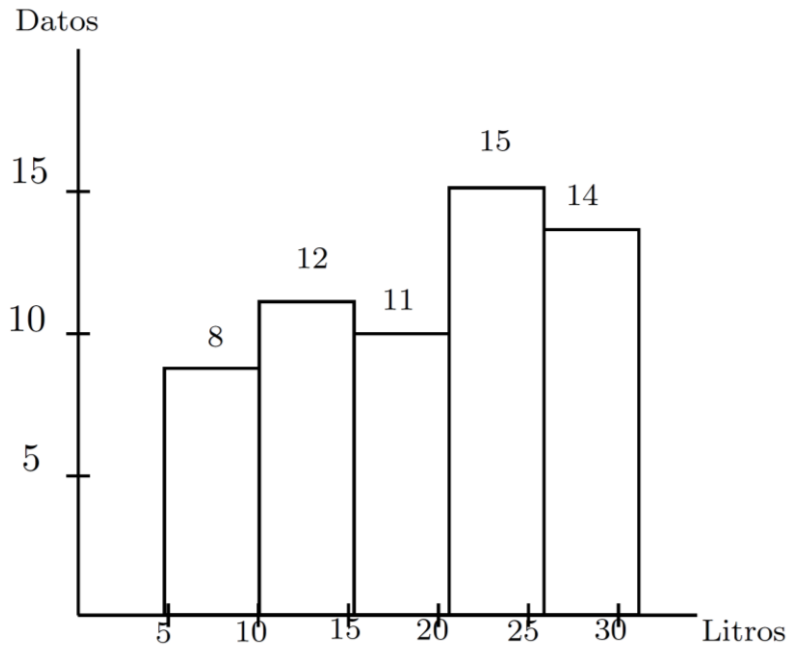
e) Usando la ecuación de c)

$$4,5 + \frac{1}{6}(84 \cdot 50/100 - 36) \cdot 1,5 = 6$$

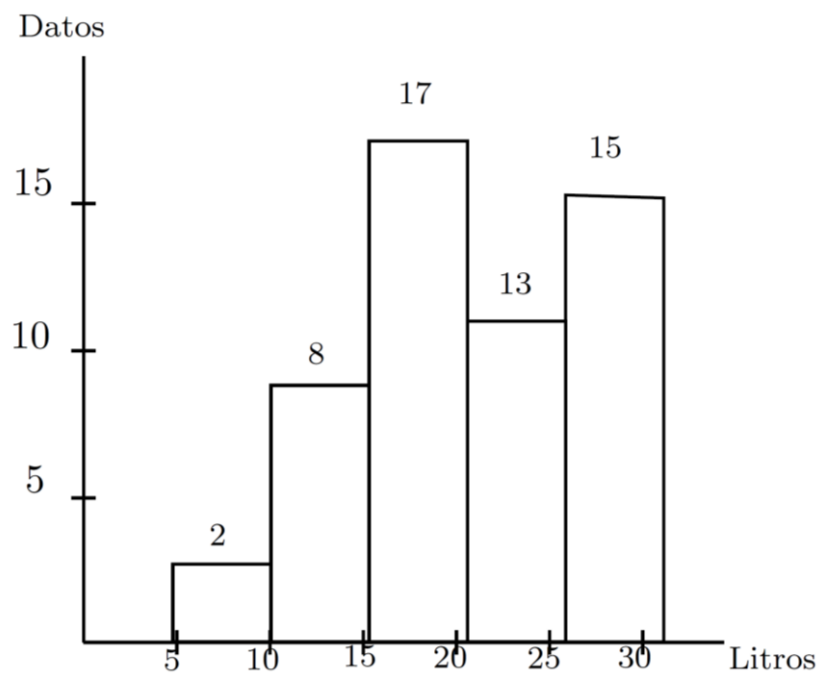
Por lo tanto el 84% de las probetas se demoraron a lo más 6 hrs.

Problema 3.

a) Modelo A:



Modelo B:



Litros	Máquinas	Litros	Máquinas
[5; 10)	8	[5; 10)	2
[10; 15)	12	[10; 15)	8
[15; 20)	11	[15; 20)	17
[20; 25)	15	[20; 25)	13
[25; 30]	14	[25; 30]	15
Modelo A		Modelo B	

b) Usando la tabla, para A se tiene una media de 18,75; para B, se tiene una media de 20,32. Por lo que, en promedio, la mejor máquina es la B.

c) Como ya tenemos la media, calcularemos la mediana. Usando la tabla, obtenemos una mediana de $Me = 19,55$, que sabemos que es el percentil 50. Después, usando la ecuación para el percentil, vemos que 18,75 es el percentil 47,08. Como necesitamos saber lo que hay entre estos ambos percentiles, los restamos y el resultado de la resta es el porcentaje de datos entre ambos. $50 - 47,08 = 2,92$ Por lo que hay un 2,92% de aplanadoras entre la media y la mediana.

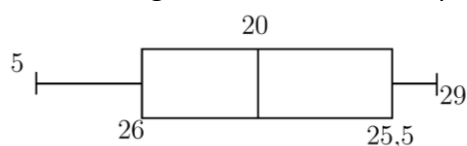
d) Con "el rendimiento frecuente" entendemos que nos piden la moda. Usando la tabla obtenemos una moda de 18,46. Por lo tanto, necesitamos saber qué porcentaje de las máquinas se encuentran entre 13,46 y 23,46. Para esto podemos ocupar la ecuación del percentil. Así nos damos cuenta de que 13,46 es el percentil 12,56 y 23,46 es el percentil 59,99. Restando igual que en el punto anterior: $59,99 - 13,46 = 47,43$, por lo que hay un 47,43% de datos entre la moda más menos 5 litros.

e) Para comparar que tan homogéneos son los modelos, calcularemos la varianza de cada uno y con esto el coeficiente de variación. Usando la tabla, la varianza de A es 46,3542, y la de B es 32,5124.

Por lo tanto los coeficientes de variación son 0,36 y 0,28 respectivamente (Nota, resultado puede variar levemente dependiendo la cantidad de decimales utilizados en pasos anteriores) . Concluimos entonces que el modelo B es el más homogéneo

Para la forma, calculamos $\gamma_3 = -0,26$ y $\gamma_4 = -0,84$. Por lo que se puede concluir que es una muestra asimétrica negativa y platocúrtica.

Por último, graficamos en un box-plot:



Problema 4.

a) Para la máquina 1:

$$\bar{x} = 1,0436 ; S^2 \approx 0,017$$

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \approx 0,0004 \implies \gamma_3 = m_3/S^3 \approx 0,185$$

$$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \approx 0,0005 \implies \gamma_4 = m_4/S^4 - 3 \approx -1,02$$

Para la máquina 2:

$$\bar{x} = 0,9726 ; S^2 \approx 0,0016$$

$$m_3 \approx 0,000006 \implies \gamma_3 \approx 0,095$$

$$m_4 \approx 0,000005 \implies \gamma_4 \approx -1,01$$

b) Como $\gamma_3 \approx 0$ para cada máquina, podemos usar la media:

$$\bar{x}_1 = 1,044$$

$$\bar{x}_2 = 0,973$$

c)

$$S_1^2 \approx 0,017 ; S_2^2 \approx 0,0016, \text{ luego } CV_1 = 0,1241 ; CV_2 = 0,041.$$

Por lo que la máquina 2 es menos dispersa que la 1.

e) La máquina más exacta es la segunda, ya que su media es más cercana a 1.
La más precisa es la 2 ya que como vimos en d), es más homogénea.