

Laboratorio 2 - Inferencia Estadística

Estadística Descriptiva

Laboratorista: Héctor Lira Talancón

Ago-Dic 2017

1. La siguiente información corresponde a una muestra de 13 autos usados. Se registró el valor de las variables marca, años de uso y precio (x10,000) del auto.

Marca	B	C	C	B	A	B	A	D	B	A	A	C	C
Años uso (X)	11	12	8	7	12	8	11	10	8	9	6	7	6
Precio (Y)	3.6	4.0	5.0	6.0	6.0	7.6	8.0	8.0	8.0	8.1	8.6	10.0	12.0

$$\text{Datos: } \sum_{i=1}^{13} x_i y_i = 800.9, \sum_{i=1}^{13} x_i = 115, \sum_{i=1}^{13} y_i = 94.9, \sum_{i=1}^{13} x_i^2 = 1073, \sum_{i=1}^{13} y_i^2 = 759.29$$

a) Con los trece datos de Años de uso, llene la siguiente tabla de frecuencias:

Años de uso	f	f%	F	F%
[5, 7)				
[7, 9)				
[9, 11)				
[11, 13)				

b) Usando exclusivamente la tabla anterior calcule la moda, mediana y media de los años de uso.

c) Considere las variables marca, años de uso y precio. Para cada variable diga cuál es su escala de medición y si se trata de una variable cualitativa o cuantitativa.

d) Construya un diagrama de caja y brazos con los datos de la variable precio e interprete el diagrama.

e) Calcule el grado de asociación lineal entre las variables años de uso (X) y precio (Y), interprete el resultado.

2. Suponga que x_1, \dots, x_n son datos de una muestra de tamaño n y que y_1, \dots, y_n y z_1, \dots, z_n son transformaciones definidas por $y_i = \alpha x_i$ y $z_i = \beta x_i, i = 1, \dots, n$ con $\alpha, \beta \in \mathbb{R}$. Conteste las siguientes preguntas:

a) Calcule las medias de los datos transformados \bar{y} y \bar{z} .

b) Calcule las varianzas de los datos transformados s_Y^2 y s_Z^2 .

c) Calcule la covarianza de los datos transformados s_{YZ}^2 .

d) Calcule el coeficiente de correlación muestral r_{YZ} .

3. Marque con una 'x' las afirmaciones que sean falsas.

☐ Si la desviación media con respecto a la mediana es cero, todos los datos tienen el mismo valor.

☐ En un diagrama de caja y brazos se dispone de un criterio para saber la dispersión de los datos.

☐ La distribución de frecuencias es una tabla que organiza los datos en clases. Estas clases pueden o no, ser exhaustivas y mutuamente excluyentes.

☐ La mediana de un conjunto de datos no necesariamente es igual a uno de los valores observados en el conjunto.

4. Considere la muestra compuesta por 20 mediciones de tiempos de entrega en un restaurante de comida rápida medidos en minutos:

4.14	4.20	4.35	4.58	5.26	5.29	5.41	5.55	5.75	6.19
6.37	6.56	6.71	6.81	9.32	9.32	12.58	13.61	15.27	22.64

Datos: $\sum_{i=1}^{20} x_i = 159.91$, $\sum_{i=1}^{20} x_i^2 = 1,699.746$, $m_3 = 184.3635$, $m_4 = 2695.727$

a) Construya un histograma con las clases $[4, 5]$, $(5, 6]$, $(6, 12]$, $(12, 20]$, $(20, 25]$. ¿Cómo se calcula el porcentaje de observaciones que están por arriba de 20 a partir de este histograma?

b) [Tendencia central] Calcule la media, mediana, moda, media podada y la media de Windsord. (La media podada, \bar{x}_α , se calcula eliminando el $100\alpha\%$ superior e inferior, mientras que la media de Windsord, \bar{x}_{W_α} , se calcula sustituyendo las observaciones que se eliminan en la media podada por los valores menor y mayor de las observaciones consideradas). ¿Cómo se comparan entre ellas? ¿Qué se puede concluir acerca de la distribución?

c) [Localización] Calcule el resumen de los 5 números: $(x_{min}, q_1, q_2, q_3, x_{max})$

d) [Dispersión] Calcule la amplitud, amplitud intercuartílica, varianza muestral, desviación estándar muestral y el coeficiente de variación muestral.

e) [Forma] Calcule el coeficiente de asimetría muestral y el coeficiente de curtosis muestral. Mencione cómo es la simetría de la distribución, lo apuntada de la distribución y cómo son sus colas. ¿Es consistente el resultado de la simetría con el obtenido en el inciso b)?

5. Ahora considere los datos agrupados obtenidos en la pregunta 4. inciso a). Calcule media muestral, mediana, desviación media respecto a la mediana, varianza muestral y moda muestral. ¿La moda de los datos individuales coincide con la clase modal?

6. Ahora imagine que se le proporciona más información acerca de las entregas del restaurante de comida rápida. Información adicional: número de personas realizando el pedido, tipo de promoción que pidieron las personas y propina (como %).

tiempo (Y)	4.14	4.20	4.35	4.58	5.26	5.29	5.41	5.55	5.75	6.19
personas (X)	1	1	1	1	1	1	2	2	2	2
promoción (Z)	A	A	A	B	C	A	A	B	B	B
propina (W)	10.0	12.5	11.0	12.0	15	15	14.5	13.2	12.0	13.5

tiempo (Y)	6.37	6.56	6.71	6.81	9.32	9.32	12.58	13.61	15.27	22.64
personas (X)	2	2	2	2	3	3	3	3	3	3
promoción (Z)	B	B	B	B	C	C	B	B	C	C
propina (W)	13.0	12.5	11.5	11.0	10.5	11.0	10.5	9.2	9.5	8.5

Datos: $\sum_{i=1}^{20} w_i = 235.9$, $\sum_{i=1}^{20} w_i^2 = 2,849.13$, $\sum_{i=1}^{20} x_i = 40$, $\sum_{i=1}^{20} x_i^2 = 92$, $\sum_{i=1}^{20} x_i w_i = 455.5$.

a) Construye una tabla de contingencia considerando las variables número de personas y número de alimentos en el pedido.

b) Construya la tabla de frecuencias relativas conjuntas, incluya las frecuencias marginales.

c) Constuya un diagrama esquemático por tipo de promoción considerando la variable número de personas.

d) Calcule la covarianza y el coeficiente de correlación muestral para las variables tiempo de espera y propina si se sabe que $\sum_{i=1}^{20} y_i w_i = 1,771.302$. Interprete brevemente.

e) ¿Cómo es la matriz de correlación considerando las variables Y , X y W ?

f) Defina $Q = \alpha Y + \beta W + \epsilon$, con $\alpha, \beta, \epsilon \in \mathbb{R}$. Calcule \bar{Q} y s_Q^2 .