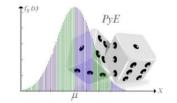


UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO FACULTAD DE INGENIERÍA DIVISIÓN DE CIENCIAS BÁSICAS COORDINACIÓN DE CIENCIAS APLICADAS DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA



SEGUNDO EXAMEN FINAL Resolución

SEMESTRE 2014 - 1 DURACIÓN MÁXIMA 2.0 HORAS TIPO 1 2 DE DICIEMBRE DE 2013

NOMBRE______Apellido paterno Apellido materno Nombre (s) Firma

Instrucciones: Leer detenidamente los siete enunciados y resolver <u>seis de los siete</u> problemas propuestos.

1. Un médico quiere hacer un estudio estadístico de una muestra semanal de mediciones de las presiones arteriales sistólicas de 90 pacientes diferentes, el médico ordenó y agrupó estos datos mediante una tabla dinámica en una hoja de cálculo electrónica y es la siguiente:

Límite superior real	Frecuencia
86.5	1
96.5	8
106.5	12
116.5	19
126.5	13
136.5	18
146.5	12
156.5	7
	86.5 96.5 106.5 116.5 126.5 136.5 146.5

Él desea calcular con datos agrupados:

- a) El coeficiente de variación e interpretar el resultado.
- b) La mediana.

Resolución

La tabla de distribución de frecuencias está dada por:

	Frontera Inferior	Frontera Superior	Frecuencia Absoluta	Marcas de Clase	Frecuencia Relativa	Frecuencia Acumulada Absoluta	Frecuencia Acumulada Relativa		
	L_{Ri}	L _{Rs}	f _i	X _i	f _i *	F _i	F _i *	$x_i f_i^*$	(x _i -media) ² f _i
	76.5	86.5	1	81.5	0.0111	1	0.0111	0.9056	1529.6790
	86.5	96.5	8	91.5	0.0889	9	0.1000	8.1333	6779.6543
	96.5	106.5	12	101.5	0.1333	21	0.2333	13.5333	4382.8148
	106.5	116.5	19	111.5	0.2111	40	0.4444	23.5389	1577.2346
Clase Mediana	116.5	126.5	13	121.5	0.1444	53	0.5889	17.5500	10.2716
	126.5	136.5	18	131.5	0.2000	71	0.7889	26.3000	2134.2222
	136.5	146.5	12	141.5	0.1333	83	0.9222	18.8667	5236.1481
	146.5	156.5	7	151.5	0.0778	90	1.0000	11.7833	6678.8642
			90					120.6111	28328.8889
									318 3021

La media= La variancia muestral= Coeficiente de Variación=

120.6111 318.3021

La muestra aleatoria no es buena

a) El coeficiente de variación se define como:

$$CV = \frac{\sqrt{S_{n-1}^2}}{\overline{X}}$$

PYE_EF2_2014-1

se requiere la media y la desviación estándar, realizando las estimaciones para los datos agrupados, se tiene:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{m} X_i f_i = \sum_{i=1}^{m} X_i f_i^*$$

$$\bar{X} = \frac{1}{90} \sum_{i=1}^{8} x_i f_i = \sum_{i=1}^{8} x_i f_i^* = 120.6111$$

la variancia mustral:

$$S_{n-1}^{2} = \frac{1}{n-1} \sum_{i=1}^{m} (X_{i} - \bar{X})^{2} f_{i}$$

$$S_{n-1}^{2} = \frac{1}{90 - 1} (28328.8889) \approx 318.3021$$

La raíz de la variancia es la desviación estándar, entonces:

$$s_{n-1} = \sqrt{s_{n-1}^2} = \sqrt{318.3021} \approx 17.8410$$

sustituyendo

$$cv = \frac{17.8410}{120.6111} \approx 0.1479$$

Este indicador para la muestra dice que hay variabilidad en los datos, según el resultado 14.79 % es buena la muestra.

b) La mediana es el valor que divide en dos partes iguales la distribución de frecuencias, con los datos agrupados, entonces se realiza una interpolación:

Fronteras de la Clase mediana	Frecuencia Acumulada Relativa			
116.5	0.4444			
Mediana	0.5			
126.5	0.5889			

$$m = 0.0144$$

$$y_2$$
- y_1 = $m(x_2$ - $x_1)$
0.5-0.4444=0.0144($x_{mediana}$ -116.5) 120.3462

2. Una revista publica tres columnas tituladas "Arte" (A), "Libros" (B) y "Cinema" (C). Los hábitos de lectura de las amas de casa son probables de acuerdo a como se muestra en la tabla siguiente:

A	В	С	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
0.5	0.3	0.5	0.15	0.25	0.1	0.05

Calcular:

- a) P(A|B)
- b) $P(A|B \cup C)$
- c) $P(A \cup B | C)$

Resolución

a) Se pide la probabilidad de que una ama de casa lea de arte dado que lee libros, entonces:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.15}{0.3} = \frac{1}{2} = 0.5$$

b) Se pide la probabilidad de que una ama de casa lea de arte dado que lee libros o cinema, entonces:

$$P(A|B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{P((AB) \cup (AC))}{P(B) + P(C) - P(BC)} = \frac{P(AB) + P(AC) - P(ABC)}{P(B) + P(C) - P(BC)}$$

$$P(A|B \cup C) = \frac{0.15 + 0.25 - 0.05}{0.3 + 0.5 - 0.1} = \frac{0.35}{0.7} = \frac{1}{2} = 0.5$$

c) Se quiere calcular la probabilidad de que una ama de casa lea de arte o libros dado que lee sobre cinema, entonces:

$$P(A \cup B|C) = \frac{P((A \cup B) \cap C)}{P(C)} = \frac{P((AC) \cup (BC))}{P(C)} = \frac{P(AC) + P(BC) - P(ABC)}{P(C)}$$
$$P(A \cup B|C) = \frac{0.25 + 0.1 - 0.05}{0.5} = \frac{0.3}{0.5} = \frac{3}{5} = 0.6$$

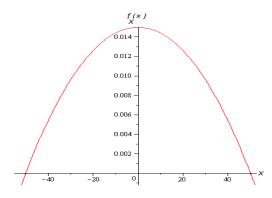
3. Se sabe que el tiempo semanal, en minutos, que un estudiante llega tarde a clase de Inferencia Estadística es una variable aleatoria X con función de densidad:

$$f_X(x) = \begin{cases} \frac{3}{500000} (2500 - x^2) & ; & -50 < x < 50 \\ 0 & ; & en \ otro \ caso \end{cases}$$

- a) Determinar el tiempo promedio semanal de retardo del estudiante.
- b) Calcular la probabilidad de que el tiempo de retardo del estudiante exceda su tiempo promedio de retardo semanal.

Resolución

La gráfica de la función de densidad es:



a) El promedio está definido por:

$$E(X) = \mu = \int_{-\infty}^{\infty} x f_X(x) dx$$
$$E(X) = \mu = \int_{-50}^{50} \frac{3}{500000} x (2500 - x^2) dx = 0$$

 La probabilidad de que el tiempo de retardo del estudiante exceda su tiempo promedio de retardo semanal.

$$P(X > \mu) = P(X > 0) = \int_0^{50} \frac{3}{500000} (2500 - x^2) dx = 0.5$$

4. Se estima que el tiempo transcurrido hasta la falla de una pantalla plana LCD, se distribuye exponencialmente con media igual a tres años. Una compañía ofrece hacer válida la garantía en la tienda por el primer año de uso, ¿qué porcentaje de pantallas hará uso efectivo de la garantía? **Resolución**

La distribución exponencial tiene media:

$$E(X) = \mu = \frac{1}{\lambda}$$

entonces:

$$\lambda = \frac{1}{E(X)} = \frac{1}{\mu}$$

$$\lambda = \frac{1}{3}$$

T es la variable aleatoria que representa el tiempo de falla de una pantalla plana LCD.

$$T \sim Exponencial \left(\lambda = \frac{1}{3} \right)$$

$$F_T(T=t) = P(T < t) = 1 - e^{-\lambda t}$$

$$F_T(T=1) = P(T<1) = 1 - e^{-\left(\frac{1}{3}\right)(1)} = 1 - e^{-\frac{1}{3}} \approx 0.2835$$

significa que el 28.35% de pantallas planas LCD presentará una falla en el primer año, los usuarios y hará válida la garantía en la tienda.

5. Un consultorio médico cuenta con dos líneas telefónicas. En un día seleccionado al azar, sea X la variable aleatoria que representa la proporción del tiempo que se utiliza la línea telefónica 1 y sea Y la variable aleatoria que representa la proporción del tiempo que se utiliza la línea telefónica 2. Si la función de densidad conjunta de estas variables aleatorias es:

$$f_{XY}(x,y) = \begin{cases} \frac{2}{3}(x+2y) & ; & 0 < x < 1 \\ 0 & ; & en \ otro \ caso \end{cases}$$

¿Cuál es la probabilidad de que la línea telefónica 2 se encuentre libre durante el 80% del día? **Resolución**

Se pide calcular $P(La\ linea\ 2\ est\'a\ desocupada\ el\ 80\%\ del\ d\'a) = P(Y<0.2)$

entonces:

$$P(Y<0.2) = \int_{0}^{1} \int_{0.2}^{0.2} \frac{2}{3} (x+2y) \, dy dx = \frac{2}{3} \int_{0}^{1} (xy+y^2) \Big|_{y=0}^{y=0.2} dx = \frac{2}{3} \int_{0}^{1} (0.2x+0.04) \, dx = \frac{2}{3} \left(\frac{0.2}{2} x^2 + 0.04x \right) \Big|_{0}^{1}$$

$$P(Y<0.2) = \frac{2}{3} \left(\frac{0.2}{2} + 0.04 \right) = \frac{7}{75} \approx 0.0933$$

- 6. Las estaturas de 1000 estudiantes están distribuidas aproximadamente en forma normal con una media de 174.5 centímetros y una desviación estándar de 6.9 centímetros. Si se extraen 200 muestras aleatorias de tamaño 25 de esta población, determinar:
 - a) El número de muestras cuyo promedio cae entre 172.5 y 175.8 centímetros, con reemplazo.

b) El número de muestras que su promedio cae abajo 172.5 centímetros, sin reemplazo.

Resolución

$$X_i \sim Normal(\mu = 174.5, \sigma = 6.9)$$
, $i = 1, 2, 3, ..., 1000$

a) Por el teorema del límite central se conoce la variancia, con reemplazo, se tiene:

$$\overline{X} \sim Normal \left(\mu_{\overline{X}} = \mu = 174.5 , \ \sigma_{\overline{X}} = \frac{\sigma_{X}}{\sqrt{n}} = \frac{6.9}{\sqrt{25}} \right)$$

$$P\left(172.5 < \overline{X} < 175.8\right) \approx P\left(\frac{172.5 - 174.5}{\frac{6.9}{\sqrt{25}}} < \frac{\overline{X} - \mu_{\overline{X}}}{\frac{\sigma_{X}}{\sqrt{n}}} < \frac{175.8 - 174.5}{\frac{6.9}{\sqrt{25}}} \right) = P\left(\frac{5\left(-2\right)}{6.9} < \frac{\overline{X} - \mu_{\overline{X}}}{\frac{\sigma_{X}}{\sqrt{n}}} < \frac{5\left(1.3\right)}{6.9} \right)$$

$$P\left(172.5 < \overline{X} < 175.8\right) = P\left(\frac{-10}{6.9} < Z < \frac{6.5}{6.9} \right) = P\left(\frac{-10}{6.9} < Z < \frac{6.5}{6.9} \right) = P\left(-1.45 < Z < 0.94\right)$$

 $P\left(172.5 < \overline{X} < 175.8\right) \approx F_{Z}\left(0.94\right) - F_{Z}\left(-1.45\right) = 0.8264 - 0.0735 \approx 0.7529$

entonces 200(0.7529)=150.58 serían 151 muestras de las doscientas.

b) Por el teorema del límite central conocida la variancia, sin reemplazo, se tiene:

$$\overline{X} \sim Normal \left(\mu_{\overline{X}} = \mu = 174.5 , \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{6.9}{\sqrt{25}} \sqrt{\frac{1000-25}{1000-1}} \right)$$

$$\overline{X} \sim Normal(\mu_{\overline{X}} = 174.5, \sigma_{\overline{X}} = 1.3633)$$

$$P\left(\overline{X} < 172.5\right) \approx P\left(\frac{\overline{X} - \mu_{\overline{X}}}{\frac{\sigma_{X}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}} < \frac{172.5 - 174.5}{\frac{6.9}{\sqrt{25}} \sqrt{\frac{1000 - 25}{1000 - 1}}}\right) = P\left(Z < -1.47\right)$$

$$P(\overline{X} < 172.5) = P(Z < -1.47) = F_Z(-1.47) \approx 0.0708$$

por lo tanto 200(0.0708)=14.16 serían 15 muestras de las doscientas.

7. Para confirmar si están relacionados el tránsito vehicular (automóviles por hora), y el contenido de ozono (porcentaje de partículas por millar) en la vegetación que crece a la orilla de las carreteras, se realizó un estudio en siete regiones de la República Mexicana, y se obtuvieron los datos siguientes:

Cantidad de automóviles	294	402	103	416	573	216	334
Cantidad de ozono	26.1	37.2	4.6	24.8	38.7	74.4	34.3

- a) Obtener la recta de regresión.
- b) Calcular el coeficiente de correlación. Proporcionar una interpretación.

Resolución

	Cantidad de autómoviles x	Cantidad de ozono y	ху	XX	уу
	294	26.1	7673.4	86436	681.21
	402	37.2	14954.4	161604	1383.84
	103	4.6	473.8	10609	21.16
	416	24.8	10316.8	173056	615.04
	573	38.7	22175.1	328329	1497.69
	216	74.4	16070.4	46656	5535.36
	334	34.3	11456.2	111556	1176.49
Sumas:	2338	240.1	83120.1	918246	10910.79

a) La recta de regresión es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

donde se sabe que los promedios están definidos por:

$$\overline{x} = \frac{1}{n} \sum_{n=1}^{7} x_i \qquad \overline{y} = \frac{1}{n} \sum_{n=1}^{7} y_i$$

sustituyendo los valores de las sumas en los promedios:

$$\overline{x} = \frac{2338}{7} = 334$$
 ,
$$\overline{y} = \frac{240.1}{7} = 34.3$$
 los estimadores se definen por:

$$\hat{\beta}_{1} = \frac{\sum_{i=1}^{7} x_{i} y_{i} - \frac{\sum_{i=1}^{7} x_{i} \sum_{i=1}^{7} y_{i}}{7}}{\sum_{i=1}^{7} x_{i}^{2} - \frac{\left(\sum_{i=1}^{7} x_{i}\right)^{2}}{7}}$$

$$\hat{\beta}_{1} = \frac{83120.1 - \frac{(2338)(240.1)}{7}}{918246 - \frac{(2338)^{2}}{7}} = \frac{2926.7}{137354} = 0.0213$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_0 = 34.3 - (0.0213)(334) = 27.1858$$

por lo tanto el modelo está dado por:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x
\hat{y} = 0.0213 x + 27.1858$$

b) El coeficiente de correlación se define como:

$$R = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

$$SS_{xx} = \sum_{i=1}^{7} x_i^2 - \frac{\left(\sum_{i=1}^{7} x_i\right)^2}{7}$$

$$SS_{xy} = 918246 - \frac{\left(2338\right)^2}{7} = 137354$$

$$SS_{yy} = \sum_{i=1}^{7} y_i^2 - \frac{\left(\sum_{i=1}^{7} y_i\right)^2}{7}$$

$$SS_{yy} = 10910.79 - \frac{\left(240.1\right)^2}{7} = 2675.36$$

$$SS_{xy} = \sum_{i=1}^{7} x_i y_i - \frac{\sum_{i=1}^{7} x_i \sum_{i=1}^{7} y_i}{7}$$

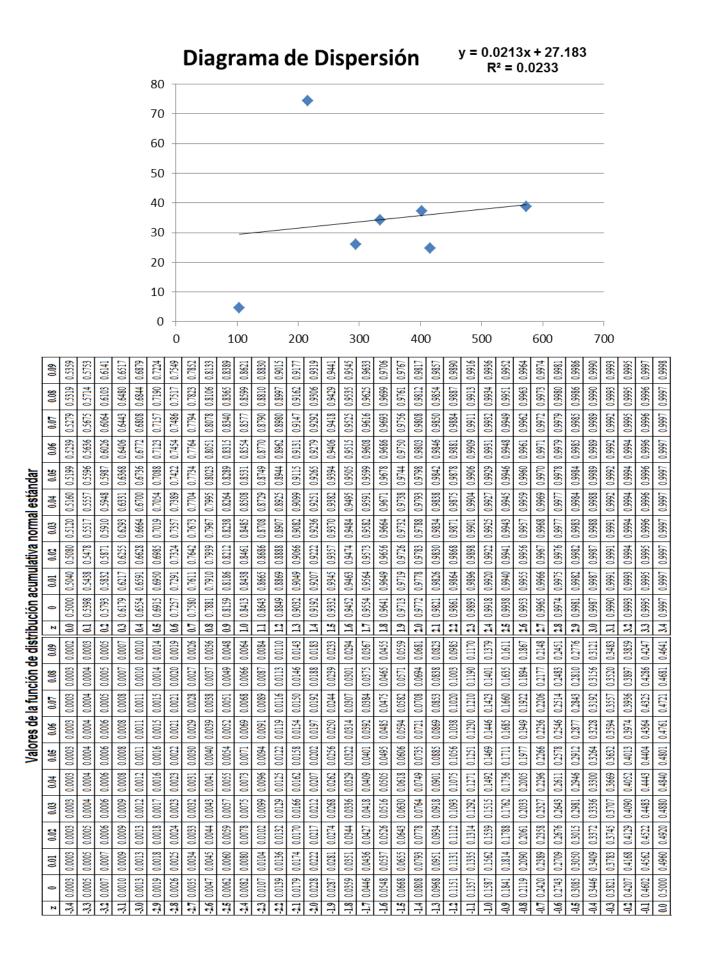
$$SS_{xy} = 83120.1 - \frac{\left(2338\right)\left(240.1\right)}{7} = 2926.7$$
sustituyendo
$$r = \frac{2926.7}{\sqrt{\left(137354\right)\left(2675.36\right)}} \approx 0.1527$$

Del resultado anterior: es muy baja la asociación lineal, prácticamente es ligera asociación lineal.

Como el coeficiente de determinación se utiliza como medida de eficacia de la regresión, éste es el cuadrado del coeficiente de correlación, se define por:

$$R^{2} = \frac{SS_{xy}^{2}}{SS_{xx}SS_{yy}}$$
$$r^{2} = \frac{(2926.7)^{2}}{(137354)(2675.36)} \approx 0.0233$$

se puede concluir que, el coeficiente de determinación 2.331 % es muy alejado al 100%, por lo que se considera que el modelo lineal no explica la cantidad de ozono emitida por los vehículos. El modelo es deficiente hay que usar otro modelo.



PYE EF2 2014-1 8