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

RESOLUCIÓN

SEMESTRE 2013 - 1 — **TIPO 1**
DURACIÓN MÁXIMA 2.0 HORAS — **29 DE NOVIEMBRE DE 2012**

NOMBRE____
Apellido paterno **Apellido materno** **Nombre (s)** **Firma**

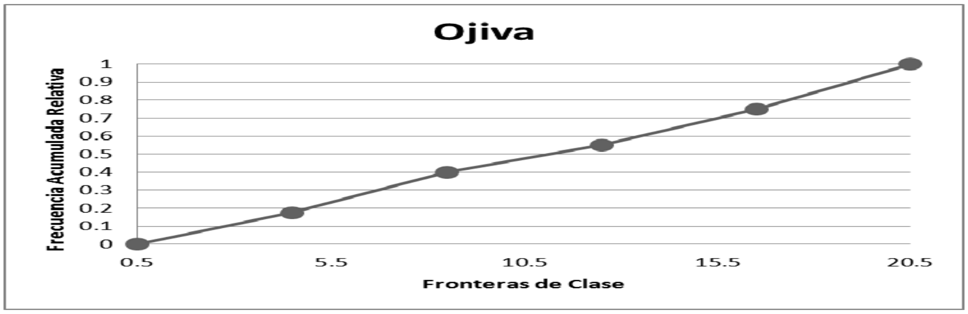
Problema 1
En la tabla siguiente se presentan los ingresos familiares en miles de pesos, en las áreas metropolitanas más grandes del país.

Intervalo de clase	Frecuencia
1-4	14
5-8	18
9-12	12
13-16	16
17-20	20

a) Bosquejar la ojiva.
b) Obtener la media, la mediana y la moda.
c) A partir del resultado obtenido en el inciso b), determinar el tipo de sesgo de la distribución de los datos. Justificar su respuesta.
d) Calcular el coeficiente de variación.
15 Puntos

Resolución
a) La tabla de distribución de frecuencias es

Clase	Límites	Fronteras	Marcas de Clase (x_i)	Frecuencia Absoluta (f_i)	Frecuencia Relativa (f^*_i)	Frecuencia Acumulada (F_i)	Frecuencia Acumulada Relativa (F^*_i)
1	1 - 4	0.5 - 4.5	2.5	14	0.175	14	0.175
2	5 - 8	4.5 – 8.5	6.5	18	0.225	32	0.400
3	9 - 12	8.5 – 12.5	10.5	12	0.150	44	0.550
4	13 - 16	12.5 – 16.5	14.5	16	0.200	60	0.750
5	17 - 20	16.5 – 20.5	18.5	20	0.250	80	1



b) La media se define por

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{m} x_i f_i = \sum_{i=1}^{m} x_i f_i^*$$

al sustituir, se tiene

$$\overline{x} = \frac{1}{80}\left[(2.5)(14) + \ldots + (10.5)(12)\right] = \frac{880}{80} = 11$$

La moda se define como la marca de clase con mayor frecuencia, entonces $x_{mo} = 18.5$

La moda también puede aproximarse utilizando la marca de clase de intervalo con frecuencia máxima

$$x_{mo} = Fr_{\inf erior clase \bmod al} + \frac{\Delta_1}{\Delta_1 + \Delta_2} c$$

con los valores respectivos $x_{mo} = 16.5 + \frac{4}{4+20}(4) = 17.167$

La mediana es el valor que divide en dos partes iguales a la muestra, entonces se realiza una interpolación lineal como

Fronteras	Frecuencia Acumulada Relativa (F^*_i)
8.5	32/80
$\tilde{x}$	40/80
12.5	44/80

La pendiente dados dos puntos es $m = \frac{y_1 - y_0}{x_1 - x_0} = \frac{\frac{44}{80} - \frac{32}{80}}{12.5 - 8.5} = \frac{3}{80} = 0.038$

la ecuación de la recta es

$$y - y_0 = m(x - x_0)$$

$$y - \frac{44}{80} = \frac{3}{80}(x - 12.5)$$

entonces $\frac{\mathbf{40}}{\mathbf{80}} - \frac{44}{80} = \frac{3}{80}(\tilde{\boldsymbol{x}} - 12.5)$

se tiene la mediana $\tilde{x} = 12.5 - \frac{4}{3} = 11.167$

c) El sesgo se puede determinar en este caso de forma empírica, ya que,

$\bar{x} = 11 \qquad \tilde{x} = 11.167 \qquad x_{mo} = 17.167$

$\bar{x} = 11 < \tilde{x} = 11.167 < x_{mo} = 17.167$ Por lo tanto el sesgo es negativo.

d) El coeficiente de variación está definido por $CV = \dfrac{S_{n-1}}{\bar{X}}$

para lo cual se requiere de la variancia de la muestra, entonces de la definición para datos agrupados, se tiene

$$S_{n-1}^2 = \frac{1}{n-1}\sum_{m=1}^{5}\left(X_i - \bar{X}\right)^2 f_i$$

al sustituir en la variancia muestral, se tiene

$$s_{n-1}^2 = \frac{1}{79}\left[\left(2.5-11\right)^2\left(14\right) + \ldots + \left(18.5-11\right)^2\left(20\right)\right] = \frac{2700}{79} = 34.177$$

por lo que el coeficiente de variación es $cv = \dfrac{\sqrt{34.177}}{11} = 0.531$

Problema 2
Un alumno contesta una pregunta que ofrece cuatro soluciones posibles en un examen de opción múltiple. Supóngase que la probabilidad de que el alumno conozca la respuesta correcta es 0.8 y la probabilidad de que tenga que contestar al azar es de 0.2. Si el alumno contesta correctamente la pregunta, ¿cuál es la probabilidad de que efectivamente conozca la respuesta correcta?
15 Puntos
Resolución
Sean los eventos:
A el alumno contesta correctamente la pregunta.
B el alumno conoce la respuesta correcta.
Del enunciado
$P\left(B\right) = 0.8$
$P\left(\bar{B}\right) = 0.2$
$P\left(A \middle| B\right) = 1$
$P\left(A \middle| \bar{B}\right) = 0.25$
de acuerdo con los eventos, se pide calcular $P\left(B \middle| A\right)$, del Teorema de Bayes, se tiene

$$P\left(B \middle| A\right) = \frac{P\left(A \cap B\right)}{P\left(A\right)} = \frac{P\left(B\right)P\left(A \middle| B\right)}{P\left(A\right)} = \frac{P\left(B\right)P\left(A \middle| B\right)}{P\left(B\right)P\left(A \middle| B\right) + P\left(\bar{B}\right)P\left(A \middle| \bar{B}\right)}$$

sustituyendo

$$P\left(B \middle| A\right) = \frac{\left(0.8\right)\left(1\right)}{\left(0.8\right)\left(1\right) + \left(0.2\right)\left(0.25\right)} = \frac{0.8}{0.85} \approx 0.941$$

Problema 3
Si X denota el tiempo de digestión (medido en horas) de una comida. Entonces X es una variable aleatoria. Supóngase que su función de densidad es

$$f_X(x) = \begin{cases} 9x\, e^{-3x} & ; \quad x \geq 0 \\ 0 & ; \quad en\ otro\ caso \end{cases}$$

a) Determinar la función distribución acumulativa y usarla para determinar la probabilidad de que la comida sea completamente digerida cuando mucho en dos horas.

b) ¿Cuál es la probabilidad de que el tiempo para digerir toda la comida tome más de tres horas?

c) ¿Cuál es el tiempo promedio de digestión de una comida?

15 Puntos

Resolución

a) La función de distribución acumulativa se define por

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

al sustituir en la función de densidad e integrando por partes, se tiene

$$F_X(x) = \int_0^x 9t\, e^{-3t}dt = 9\int_0^x t\, e^{-3t}dt = 9\left[\frac{-t}{3}e^{-3t} - \frac{1}{9}e^{-3t}\right]\Bigg|_0^x =$$

$$F_X(x) = -3e^{-3t}\left[t + \frac{1}{3}\right]\Bigg|_0^x = 1 - 3xe^{-3x} - e^{-3x}$$

por lo tanto la función está dada por

$$F_X(x) = \begin{cases} 0 & ; \quad x < 0 \\ 1 - 3xe^{-3x} - e^{-3x} & ; \quad x \geq 0 \end{cases}$$

La probabilidad de que la comida sea digerida en máximo dos horas, es

$$P(X < 2) = F_X(2) = 1 - 3(2)e^{-3(2)} - e^{-3(2)} = 1 - 6e^{-6} - e^{-6} = 1 - 7e^{-6} \approx 0.983$$

b) Se pide calcular la probabilidad de, usando la función acumulativa y sus propiedades

$$P(X > 3) = 1 - P(X \leq 3) = 1 - F_X(3) = 1 - \left[1 - 3(3)e^{-3(3)} - e^{-3(3)}\right] =$$

$$= 1 - 1 + 9e^{-9} + e^{-9} = 10e^{-9} \approx 0.0012$$

c) El tiempo promedio de digestión está dado por el valor esperado, que se define por

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\,dx \qquad \text{si existe}$$

al sustituir

$$E(X) = \int_0^{\infty} x\left(9xe^{-3x}\right)dx = 9\int_0^{\infty} x^2 e^{-3x}dx$$

se tiene una integral impropia y se resuelve por partes, entonces

$$E(X) = 9 \lim_{R\to\infty} \int_0^R x^2 e^{-3x}dx = 9 \lim_{R\to\infty}\left[-\frac{x^2}{3}e^{-3x} - \frac{2x}{9}e^{-3x} - \frac{2}{27}e^{-3x}\right]\Bigg|_0^R = -3\lim_{R\to\infty}\left[e^{-3x}\left(x^2 + \frac{2x}{3} + \frac{2}{9}\right)\right]\Bigg|_0^R$$

$$\mathrm{E}(X)=-3\,\underset{R\to\infty}{lím}\left[\frac{1}{9}e^{-3x}\left(9x^2+6x+2\right)\right]\Bigg|_0^R=-\frac{1}{3}\underset{R\to\infty}{lím}\left[e^{-3R}\left(9R^2+6R+2\right)-2\right]=-\frac{1}{3}(0-2)=\frac{2}{3}$$

Problema 4

Se sabe que el 7% de los paquetes que se envían de una ciudad a un algún destinatario se pierden y nadie se hace responsable. Una persona tiene dos libros que desea enviar cuyo costo es de $20.00 por cada uno. El costo del envío por un paquete con los dos libros es de $5.20 pero si los envía por separado cada paquete tiene un costo de $3.30. El señor desea minimizar el valor esperado de su desembolso que es el gasto del correo más la posible pérdida. ¿Qué es mejor, enviar los libros en un solo paquete o separados?

15 Puntos

Resolución

Sea X la variable aleatoria el gasto de envío más la posible pérdida. La función de distribución de probabilidad para enviar un solo paquete

	X	$f_X(x)$
No se pierde	5.2	0.93
Se pierde	45.20	0.07

Para este caso el valor esperado es

$$E(X)=\sum_{\forall\, x} x\, f_X(x)=(5.20)(0.93)+(45.20)(0.07)=8$$

La función de distribución de probabilidad para enviar dos paquetes es

Paquetes perdidos	X	$f_X(x)$
0	6.6	0.8649
1	26.6	0.1302
2	46.6	0.0049

Para este caso el valor esperado es

$$E(X)=\sum_{\forall\, x} x\, f_X(x)=(6.60)(0.8649)+(26.6)(0.1302)+(46.60)(0.0049)=9.4$$

Las probabilidades para este caso se obtuvieron aplicando una función de distribución binomial porque es independiente, entonces

$$f(x)=\binom{n}{x}p^x q^{n-x}$$

$$f(x)=\binom{2}{0}(0.07)^0(0.93)^{2-0}=(1)(0.07)^0(0.93)^2=0.8649$$

$$f(x)=\binom{2}{1}(0.07)^1(0.93)^{2-1}=(2)(0.07)(0.93)=0.1302$$

$$f(x)=\binom{2}{2}(0.07)^2(0.93)^0=(1)(0.07)^2(0.93)^0=0.0049$$

Problema 5

Sea X la temperatura ambiente (en °C) y Y el tiempo (en minutos), requeridos para que el motor diesel de una camioneta esté listo para ponerla en movimiento. Supóngase que la densidad conjunta de X y Y está dada por

$$f_{XY}(x,y)=\begin{cases}\dfrac{1}{6640}(4x+2y+1) & ; \quad 0\le x\le 40 \quad , \quad 0\le y\le 2\\ 0 & ; \quad \textit{en otro caso}\end{cases}$$

a) Determinar la probabilidad de que en un día seleccionado aleatoriamente la temperatura ambiente sea mayor a 20°C y se requiera por lo menos un minuto para poner en movimiento la camioneta.

b) ¿Son independientes X y Y? Justifique la respuesta sobre bases matemáticas.

15 Puntos

Resolución

a) Del enunciado se pide calcular

$$P(X>20\ ,\ 1\le Y\le 2)=\int_1^2\int_{20}^{40}\frac{1}{6640}(4x+2y+1)\,dxdy=\frac{1}{6640}\int_1^2\left(2x^2+2xy+x\right)\Big|_{20}^{40}dy=$$

$$=\frac{1}{6640}\int_1^2\left(2(40)^2+2(40)y+40-2(20)^2-2(20)y-20\right)dy=\frac{1}{6640}\int_1^2(2420+40y)\,dy=$$

$$=\frac{1}{6640}\left(2420y+20y^2\right)\Big|_1^2=\frac{1}{6640}\left((2420)(2)+(20)(4)-2420-20\right)=\frac{2480}{6640}\approx 0.3735$$

b) Las variables aleatorias conjuntas son independientes, si cumple con

$$f_{XY}(x,y)=f_X(x)\,f_Y(y)$$

o bien,

$$f_{X|Y}(X|Y)=f_X(x)$$

o bien

$$f_{Y|X}(Y|X)=f_Y(y)$$

Entonces al calcular las funciones marginales, de la definición

$$f_X(x)=\int_{-\infty}^{\infty}f_{XY}(x,y)\,dy$$

$$f_Y(y)=\int_{-\infty}^{\infty}f_{XY}(x,y)\,dx$$

sustituyendo

$$f_X(x)=\int_0^2\frac{1}{6640}(4x+2y+1)\,dy=\frac{1}{6640}\left[4xy+y^2+y\right]\Big|_0^2=\frac{1}{6640}[8x+6]$$

$$f_X(x)=\begin{cases}\dfrac{1}{6640}[8x+6] & ; \quad 0\le x\le 40\\ 0 & ; \quad \textit{en otro caso}\end{cases}$$

$$f_Y(y)=\int_0^{40}\frac{1}{6640}(4x+2y+1)\,dx=\frac{1}{6640}\left[2x^2+2xy+x\right]\Big|_0^{40}=\frac{1}{6640}[80y+3240]$$

$$f_Y(y)=\begin{cases}\dfrac{1}{6640}[80y+3240] & ; \quad 0\le x\le 2\\ 0 & ; \quad \textit{en otro caso}\end{cases}$$

sustituyendo en la condición

$$\frac{1}{6640}(4x+2y+1) \neq \left(\frac{1}{6640}[8x+6]\right)\left(\frac{1}{6640}[80y+3240]\right)$$

Por lo tanto, las variables aleatorias conjuntas no son independientes.

Problema 6

Se debe colocar un cordón de soldadura a 40 piezas. Basado en su experiencia, un soldador sabe que el tiempo promedio requerido para colocar el cordón de soldadura en una de las piezas es de 5 minutos y su desviación estándar es de 3 minutos. El soldador comienza a aplicar los cordones a las 6:00 p.m. y sabe que el valor medio de aplicación de los 40 cordones debe ser de máximo 4.5 minutos, si desea colocar todos los cordones de soldadura antes de las 9:00 p.m. (hora de término de su turno de trabajo). ¿Cuál es la probabilidad de que el soldador termine de colocar los 40 cordones antes de que termine su turno de trabajo?

15 Puntos

Resolución

Sea X la variable aleatoria que representa el tiempo para colocar el cordón de soldadura a una pieza, entonces X tiene un promedio de 5 minutos y desviación estándar de 3 minutos, no se especifica qué distribución tiene, pero es una muestra grande.

Se pide calcular para que en promedio las 40 piezas, cada una tarde a lo más 4.5 minutos, con $X_1, X_2, X_3, \ldots, X_{40}$ por lo que, se pide calcular $P\left(\bar{X} < 4.5\right)$

Por el Teorema del Límite Central

$$\bar{X} \sim Normal\left(\mu_{\bar{X}} = \mu_X = 5\min \ , \ \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} = \frac{3^2}{40}\right)$$

al realizar cálculos y haciendo la estandarización y al usar la tabla de la distribución acumulativa normal estándar, se tiene

$$P\left(\bar{X} < 4.5\right) \approx P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\frac{\sigma_{\bar{X}}}{\sqrt{n}}} < \frac{4.5-5}{\frac{3}{\sqrt{40}}}\right) = P\left(Z < \sqrt{40}\,\frac{(-0.5)}{3}\right) = P(Z < -1.05) = F_Z(-1.05) = 0.1469$$

Problema 7

El departamento de publicidad de una compañía que fabrica detergentes para trastos desea saber qué tan fuerte es la relación y el porcentaje de explicación entre las ventas y el número de comerciales televisivos transmitidos por día para una muestra aleatoria de siete ciudades. Ventas (en cientos de unidades por mes, y); Comerciales (número transmitido por día x). Se sabe que $r^2 = 0.736$, interpretar el resultado.

10 Puntos

Resolución

La relación lineal es relativamente fuerte, según el decisor, $r = \sqrt{0.736} \approx 0.858$ ya que no es tan cercana a uno.

La relación es muy ligera ya que por cada comercial transmitido, y (las ventas) se incrementan ligeramente, además, el modelo explica el 73.6% de la variabilidad de las ventas, y.