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 2012 - 1 — **TIPO 1**
DURACIÓN MÁXIMA 2.5 HORAS — **2 DE DICIEMBRE DE 2011**

NOMBRE____
Apellido Paterno Apellido Materno Nombre (s) Firma

Problema 1
Los siguientes datos representan el total de grasas, en porciento, en las hamburguesas y productos de pollo en una muestra aleatoria, tomada de la cadena de tiendas de comida rápida.

Hamburguesas							
19	31	34	35	39	39	38	43

Productos de pollo										
39	9	15	18	16	16	22	25	27	33	7

Para las hamburguesas y los productos de pollo, realizar por separado, el cálculo de:

a) media, mediana y moda.
b) variancia, desviación estándar y coeficiente de variación.
c) Con base en los resultados del inciso a), ¿qué conclusiones se obtienen en relación con las diferencias en la grasa total de las hamburguesas y los productos de pollo?

15 Puntos

Resolución
a) La media de la muestra para las hamburguesas

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{8} x_i$$

desarrollando

$$\bar{x} = \frac{1}{8}[19 + 31 + 34 + 35 + 39 + 39 + 38 + 43] = 34.75$$

La media de la muestra para los productos de pollo

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{11} x_i$$

desarrollando

$$\bar{x} = \frac{1}{11}[7 + 9 + 15 + 16 + 16 + 18 + 22 + 25 + 27 + 33 + 39] = 20.64$$

La mediana de la muestra de hamburguesas, es el promedio de los dos valores que están al centro de los datos ordenados en forma ascendente

19	31	34	35	38	39	39	43

entonces

$$\tilde{x} = \frac{1}{2}[35 + 38] = 36.50$$

La mediana de la muestra para los productos de pollo, es el valor central de los valores ordenados en forma ascendente

7	9	15	16	16	18	22	25	27	33	39

entonces
$\tilde{x} = 18$

La moda de la muestra de hamburguesas es el valor que más se repite, entonces el valor que define a la moda es
$x_{mo} = 39$

entonces para la muestra productos de pollo, el valor que define a la moda es
$x_{mo} = 16$
en ambos casos es unimodal.

b) La variancia de la muestra aletoria se define como

$$S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

sustituyendo los valores de las hamburguesas

$$s_{n-1}^2 = \frac{1}{8-1}[(19-34.75)^2 + (31-34.75)^2 + (34-34.75)^2 + \cdots + (43-34.75)^2] \approx 53.92$$

sustituyendo valores de la muestra de productos de pollo

$$s_{n-1}^2 = \frac{1}{11-1}[(7-20.64)^2 + (9-20.64)^2 + (15-20.64)^2 + \cdots + (39-20.64)^2] \approx 95.46$$

La desviación estándar muestral, es la raíz de la variancia muestral, por lo que

$$S_{n-1} = \sqrt{S_{n-1}^2}$$

sustituyendo valores de la muestra de las hamburguesas, se tiene

$$s_{n-1} = \sqrt{53.93} \approx 7.34$$

para la muestra de productos de pollo

$$s_{n-1} = \sqrt{95.45} \approx 9.77$$

El coeficiente de variación para las muestras aleatorias, es

$$CV = \frac{S_{n-1}}{\bar{X}}$$

sustituyendo los cálculos obtenidos, se tiene para las hamburguesas

$$CV = \frac{7.34}{34.75} \approx 0.21$$

y para los productos de pollo, se tiene

$$CV = \frac{9.77}{20.64} \approx 0.47$$

c) El promedio de contenido de grasa es menor el de la muestra dos, es preferible comer alimentos de pollo, ya que, contienen en promedio menos grasa, todas las medidas de tendencia central son menores las de pollo.
Las dos muestras de comida rápida tienen sesgo negativo.

Problema 2
Supóngase que en un centro médico, de todos los fumadores de quienes se sospecha tenían cáncer pulmonar, el 90% lo tenía mientras que únicamente el 5% de los no fumadores lo padecía. Si la proporción de fumadores es de 0.45, ¿cuál es la probabilidad de que un paciente con cáncer pulmonar, seleccionado al azar, sea fumador?
10 Puntos

Resolución
Sean los eventos:
F representa a un paciente fumador.
$\bar{F}$ representa a un paciente no fumador.
C representa a un paciente con cáncer.
$\bar{C}$ representa a un paciente que no tiene cáncer.
del enunciado, se tiene
$P(F) = 0.45$
$P(\bar{F}) = 0.55$
$P(C|F) = 0.9$
$P(C|\bar{F}) = 0.05$
se pide calcular la probabilidad de que un paciente que se sabe tiene cáncer pulmonar, sea fumador, entonces

$$P(F|C) = \frac{P(C \cap F)}{P(C)} = \frac{P(F)P(C|F)}{P(C)} = \frac{P(F)P(C|F)}{P(F)P(C|F) + P(\bar{F})P(C|\bar{F})}$$

sustituyendo los datos

$$P(F|C) = \frac{(0.45)(0.9)}{(0.45)(0.9) + (0.55)(0.05)} = \frac{0.405}{0.405 + 0.0275} = 0.9364$$

Problema 3
En una urna se encuentran cinco pelotas numeradas del uno al cinco. Se toman al azar dos pelotas, sin reemplazo, y se anotan sus números. Determinar la distribución de probabilidad para

a) el mayor de los números seleccionados.
b) la suma de los dos números seleccionados.

20 Puntos
Resolución

a) Se pide obtener la función de probabilidad para el mayor de los números seleccionados, al tomar dos pelotas sin reemplazo, las cuales están numeradas del uno al cinco.

El espacio muestra del experimento aleatorio está dado por
$S = \{12, 13, 14, 15, 21, 23, 24, 25, 31, 32, 34, 35, 41, 42, 43, 45, 51, 52, 53, 54\}$

$n(S) = 20$
La función de probabilidad se define por
$f_X(x) = P(X = x)$

Entonces, sea X la variable aleatoria que representa el mayor de los número seleccionados.
$R_X = \{2, 3, 4, 5\}$
La función de probabilidad está dada por

x	2	3	4	5
$f_X(x)$	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{6}{20}$	$\frac{8}{20}$

que es igual a

x	2	3	4	5
$f_X(x)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$

b) Sea Y la variable aleatoria que representa la suma de los números seleccionados.

El espacio muestra del experimento aleatorio es
$S = \{12, 13, 14, 15, 21, 23, 24, 25, 31, 32, 34, 35, 41, 42, 43, 45, 51, 52, 53, 54\}$

$n(S) = 20$
De la definición de distribución de probabilidad

$f_Y(y) = P(Y = y)$

$R_Y = \{3,4,5,6,7,8,9\}$

La función de probabilidad está dada por

y	3	4	5	6	7	8	9
$f_Y(y)$	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{4}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	$\frac{2}{20}$

o bien

y	3	4	5	6	7	8	9
$f_Y(y)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

Problema 4

De la experiencia en semestres pasados se sabe que en una prueba de conocimientos generales aplicada al azar acreditan el 60% de los sustentantes, ¿cuál es la probabilidad de que en un grupo de 39 estudiantes,

a) acrediten 26 alumnos?

b) reprueben 15 alumnos?

15 Puntos

Resolución

Sea X la variable aleatoria que representa el número de alumnos que acreditan de 39, si la probabilidad de acreditar es 0.6

$X \sim Binomial(n = 39, p = 0.6)$

El modelo binomial está dado por

$$f_X(x) = \begin{cases} \binom{39}{x}(0.6)^x(0.4)^{39-x} & ; \quad x = 0, 1, 2, \ldots, 39 \\ 0 & ; \quad en\ otro\ caso \end{cases}$$

a) la probabilidad de que acrediten 26 alumnos, es:

$$P(X = 26) = f_X(X = 26) = \binom{39}{26}(0.6)^{26}(0.4)^{39-26} \approx 0.093$$

b) la probabilidad de que no aprueben 15 sustentantes, está dada por:

$$P(X = 24) = f_X(X = 24) = \binom{39}{24}(0.6)^{24}(0.4)^{39-24} \approx 0.128$$

Problema 5

Una instalación de servicio telefónico opera con dos líneas de servicio. En un día seleccionado aleatoriamente, sea X la proporción de tiempo que se utiliza la primera línea y Y la proporción de tiempo que se utiliza la segunda. Supóngase que la función de densidad conjunta para las variables aleatorias es

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{2}(x^2 + y^2) & ; \quad 0 \le x \le 1 \ , \quad 0 \le y \le 1 \\ 0 & ; \quad en\ otro\ caso \end{cases}$$

a) Calcular la probabilidad de que ninguna línea se utilice más de la mitad del tiempo.

b) Obtener la probabilidad de que la primera línea esté ocupada más del 75% del tiempo.

15 Puntos

Resolución

Sea X la variable aleatoria que representa la proporción del tiempo que se utiliza la primera línea.

Sea Y la variable aleatoria que representa la proporción del tiempo que se utiliza la segunda línea.

La función de densidad está definida por

$$f_{X,Y}(x,y)=\begin{cases}\frac{3}{2}(x^2+y^2) & ; \quad 0\le x\le 1 \quad , \qquad 0\le y\le 1\\ 0 & ; \qquad\qquad en\ otro\ caso\end{cases}$$

a) La probabilidad de que ninguna línea se utilice más de la mitad del tiempo, es

$$P\left(X\le\frac{1}{2},Y\le\frac{1}{2}\right)=P\left(X<\frac{1}{2},Y<\frac{1}{2}\right)$$

como la región de definición es cuadrada, entonces la doble integral para la porción de volumen es:

$$P\left(X<\frac{1}{2},Y<\frac{1}{2}\right)=\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}}\frac{3}{2}(x^2+y^2)\,dxdy=\frac{3}{2}\int_0^{\frac{1}{2}}\left(\frac{1}{3}x^3+xy^2\right)\Big|_0^{\frac{1}{2}}dy=\frac{3}{2}\int_0^{\frac{1}{2}}\left(\frac{1}{3}\left(\frac{1}{2}\right)^3+\frac{1}{2}y^2\right)dy$$

$$=\frac{3}{2}\int_0^{\frac{1}{2}}\left(\frac{1}{24}+\frac{1}{2}y^2\right)dy=\frac{3}{2}\left(\frac{1}{24}y+\frac{1}{6}y^3\right)\Big|_0^{\frac{1}{2}}=\frac{3}{2}\left(\frac{1}{24}\left(\frac{1}{2}\right)+\frac{1}{6}\left(\frac{1}{8}\right)\right)=\frac{3}{2}\left(\frac{1}{48}+\frac{1}{48}\right)$$

$$=\frac{3}{2}\left(\frac{2}{48}\right)=\frac{3}{48}\approx 0.0625$$

b) La probabilidad que la primera línea esté ocupada más del 75% del tiempo, es

$$P\left(X\ge\frac{3}{4}\right)=P\left(X>\frac{3}{4}\right)$$

es posible obtener esta probabilidad con la función de densidad conjunta o marginal, entonces

$$P\left(X>\frac{3}{4}\right)=\int_0^1\int_{\frac{3}{4}}^1\frac{3}{2}(x^2+y^2)\,dxdy=\frac{3}{2}\int_0^1\left(\frac{1}{3}x^3+xy^2\right)\Big|_{\frac{3}{4}}^1dy=\frac{3}{2}\int_0^1\left(\frac{1}{3}+y^2-\frac{1}{3}\left(\frac{3}{4}\right)^3-\frac{3}{4}y^2\right)dy$$

$$=\frac{3}{2}\int_0^1\left(\frac{1}{4}y^2+\frac{37}{192}\right)dy=\frac{3}{2}\left(\frac{1}{12}y^3+\frac{37}{192}y\right)\Big|_0^1=\frac{3}{2}\left(\frac{1}{12}+\frac{37}{192}\right)=\frac{53}{128}\approx 0.4141$$

Problema 6

La distribución de ingresos mensuales de 6400 trabajadores tiene una media de \$950 y una desviación estándar de \$185 si se extraen muestras de 36 elementos, de esa población sin reemplazo.

a) ¿Cuál es la probabilidad que el promedio de la muestra esté entre \$975 y \$1025?

b) ¿Cuál debe ser el ingreso promedio de los trabajadores, si con base en la muestra de 36 elementos, la probabilidad de que la media muestral tenga un valor mayor de \$1050 es de 0.95? Considérese la misma desviación estándar.

Nota: Emplear el factor de corrección para la varianza $\sigma_{\bar{X}}^2$ cuando se tiene población finita.

15 Puntos

Resolución

Sea X la variable aleatoria que representa los ingresos mensuales de los trabajadores.

Por el Teorema del límite central siendo una muestra grande

$$\bar{X}\sim Normal\left(\mu_{\bar{X}}=\mu_{X_i}\ ,\sigma_{\bar{X}}^2=\frac{\sigma_{X_i}^2}{n}\left(\frac{N-n}{N-1}\right)\right)$$

$$\bar{X}\sim Normal\left(\mu_{\bar{X}}=950\ ,\sigma_{\bar{X}}^2=\frac{(185)^2}{36}\left(\frac{6400-36}{6400-1}\right)\right)$$

$$\bar{X}\sim Normal(\mu_{\bar{X}}=950\ ,\sigma_{\bar{X}}^2=945.495)$$

a) La probabilidad de que el promedio de los ingresos de la muestra esté entre \$975 y \$1025, es

$$P(975 \leq \bar{X} \leq 1025) \approx P\left(\frac{975-950}{\sqrt{\frac{(185)^2}{36}\left(\frac{6400-36}{6400-1}\right)}} \leq \frac{\bar{X}-\mu_{\bar{X}}}{\sqrt{\frac{\sigma_{X_i}^2}{n}\left(\frac{N-n}{N-1}\right)}} \leq \frac{1025-950}{\sqrt{\frac{(185)^2}{36}\left(\frac{6400-36}{6400-1}\right)}}\right) =$$

$$= P\left(\frac{25}{\sqrt{945.495}} \leq Z \leq \frac{75}{\sqrt{945.495}}\right)$$

al utilizar la tabla de valores de la función de distribución acumulativa normal estándar

$P(0.813 \leq Z \leq 2.439) = F_Z(2.439) - F_Z(0.813) = 0.9927 - 0.7910 \approx 0.2017$

b) Se pide determinar el valor promedio de ingresos de cada trabajador, esto es $\mu_X = \mu_{\bar{X}}$, entonces

$$P(\bar{X} \geq \bar{x}) = P(\bar{X} \geq 1050) = 0.95$$

$$0.95 \approx P\left(\frac{\bar{X}-\mu_{\bar{X}}}{\sqrt{\frac{\sigma_{X_i}^2}{n}\left(\frac{N-n}{N-1}\right)}} \geq \frac{1050-\mu}{\sqrt{\frac{(185)^2}{36}\left(\frac{6400-36}{6400-1}\right)}}\right) = P\left(Z \geq \frac{1050-\mu}{30.748}\right)$$

se sabe que, al utilizar la tabla de valores de la función de distribución acumulativa normal estándar

$$\frac{1050-\mu}{\sqrt{\frac{(185)^2}{36}\left(\frac{6400-36}{6400-1}\right)}} = z_0 = -1.645$$

despejando

$1050 - \mu = -1.645(30.748)$

$1050 + 1.645(30.748) = 1100.580 = \mu$

Problema 7

Una compañía produce partes para camión. El gerente de contabilidad desea desarrollar un modelo de regresión que pueda utilizarse para predecir costos. Él selecciona ocho unidades de producción fabricadas como una variable de predicción y recolecta los datos que se muestran a continuación. Los costos están en miles de dólares y las unidades en cientos.

	Unidades x	Costos y	xy	x^2	y^2
Sumas	82.2	40.1	427.46	967.2	205.23

a) Calcular e interpretar el modelo de regresión lineal por el método de mínimos cuadrados.
b) Según el modelo, ¿cuánto costaría producir 750 unidades?
c) ¿Qué tan bueno resulta el modelo de regresión lineal?

10 Puntos

Resolución

a) Para obtener la recta de regresión, con las sumas, se tiene

$$SS_{xx} = \sum_{i=1}^{8} x_i^2 - \frac{\left(\sum_{i=1}^{8} x_i\right)^2}{n}$$

sustituyendo

$$ss_{xx} = 967.2 - \frac{(82.2)^2}{8} = 122.595$$

$$SS_{xy} = \sum_{i=1}^{8} xy - \frac{(\sum_{i=1}^{8} x_i)(\sum_{i=1}^{8} y_i)}{n}$$

sustituyendo

$$ss_{xy} = 427.46 - \frac{(82.2)(40.1)}{8} = 15.433$$

$$SS_{yy} = \sum_{i=1}^{8} y_i^2 - \frac{(\sum_{i=1}^{8} y_i)^2}{n}$$

sustituyendo

$$ss_{yy} = 205.23 - \frac{(40.1)^2}{8} = 4.229$$

$$\beta_1 = \frac{SS_{xy}}{SS_{xx}}$$

sustituyendo

$$\beta_1 = \frac{15.433}{122.595} \approx 0.126$$

$$\bar{X} = \frac{\sum_{i=1}^{8} x_i}{n}$$

$$\bar{x} = \frac{82.2}{8} = 10.275$$

$$\bar{Y} = \frac{\sum_{i=1}^{8} y_i}{n}$$

$$\bar{y} = \frac{40.1}{8} = 5.013$$

El estimador

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_0 = 5.013 - 0.126(10.275) = 3.718$$

La ecuación de la recta está dada por

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

con los valores obtenidos es

$\hat{y} = 3.718 + 0.126x$

Representa el modelo de regresión lineal de los costos a partir de las unidades de producción estudiadas.

b) si $x = 7.5$ porque está en cientos, sustituyendo en la ecuación de la recta estimada

$\hat{y} = 3.719 + 0.126(7.5) = 4.664$

c) Se sabe que el coeficiente de correlación está dado por el estimador

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

sustituyendo los valores obtenidos

r= $\frac{15.433}{\sqrt{(122.595)(4.229)}} \approx 0.678$
el ajuste a una recta es de 0.678, que se puede decir que es bueno.
Elevando al cuadrado el coeficiente de correlación, se tiene

$$R^2 = \frac{(SS_{xy})^2}{SS_{xx}SS_{yy}}$$

sustituyendo

$$r^2 = \frac{(15.433)^2}{(122.595)(4.229)} \approx (0.678)^2 = 0.459$$

Indica que las unidades producidas explican sólo el 45.9% de los costos, por tanto, el modelo no es bueno pues el grado de explicación es muy bajo.