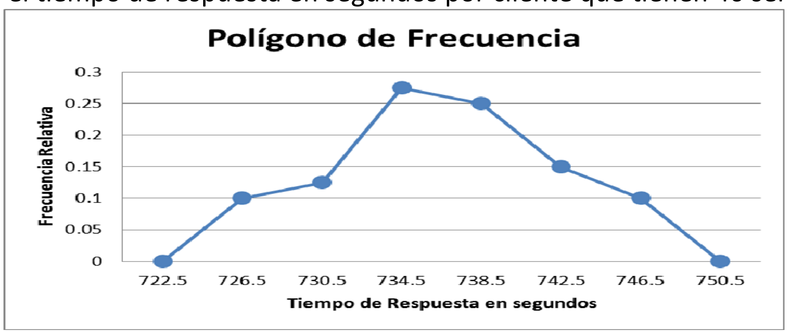
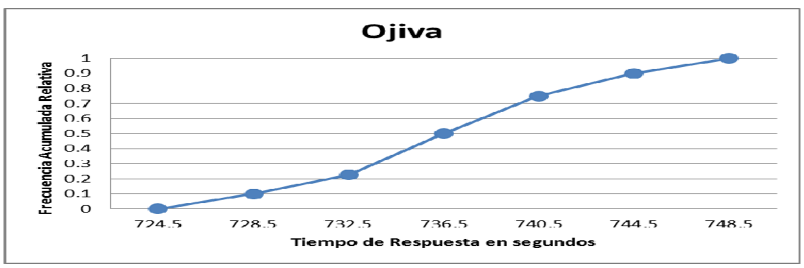
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE INGENIERÍA
DIVISIÓN DE CIENCIAS BÁSICAS
COORDINACIÓN DE CIENCIAS APLICADAS
DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA
SEGUNDO EXAMEN FINAL
RESOLUCIONES

SEMESTRE 2013 - 2 | **TIPO 1**
DURACIÓN MÁXIMA 2.0 HORAS | **7 DE JUNIO DE 2013**

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

1. Las gráficas muestran el tiempo de respuesta en segundos por cliente que tienen 40 servidores de la marca IBT.





a) Obtener media y variancia de la muestra.
b) Calcular de manera aproximada: la moda, la mediana y el sesgo.
c) Si el tiempo promedio de respuesta para 40 servidores de la marca DELLA fue de 403 y con una desviación estándar de 119 segundos, ¿cuál de las dos marcas de servidores presenta mayor variabilidad de los datos? Justificar su respuesta.

20 Puntos

Resolución

a) Para la media, de la gráfica se observa que

La media se define por $\bar{X} = \frac{1}{n}\sum_{i=1}^{m} X_i f_i = \sum_{i=1}^{m} X_i f_i^*$

sustituyendo

$$\bar{x} = \left[726.5(0.1) + \ldots + 746.5(0.1)\right] = 736.6$$

La variancia muestral, es el segundo momento con respecto de su media

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

sustituyendo

$$s_{n-1}^2 = \frac{1}{39}\left[(726.5-736.6)^2(4)+\ldots+(746.5-736.6)^2(4)\right] = \frac{1}{39}(1279.6) \approx 32.81$$

b) La moda es la marca de clase con mayor frecuencia relativa

$$x_{mo} = 734.5$$

otra forma de calcular la moda es

$$x_{mo} = L_{x_{mo\_\inf}} + \left[\frac{a}{a+b}\right] c_{x_{mo}}$$

$$a = f_{x_{mo}} - f_{x_{mo-1}}$$

$$b = f_{x_{mo-\inf}}$$

$f_{x_{mo}}$: *frecuencia absoluta de la clase que contiene a la moda.*

c_{mo} : *Longitud de la clase que contiene a la moda.*

$L_{Mo\_\inf}$: *Límite inferior de la clase que contiene a la moda.*

sustituyendo

$$x_{mo} = 732.5 + \left[\frac{11-5}{(11-5)+(11-10)}\right](4) = 735.929$$

La mediana es la medida de tendencia central que divide en dos partes iguales a la muestra, entonces

Con $y = 0.5$, $\tilde{x} = mediana$

$$\tilde{x} = 736.5$$

Se observa de la gráfica dada, que hay ligero sesgo negativo de la muestra, no es una muestra insesgada.

$$a_3 = \frac{\frac{1}{n}\sum_{i=1}^{m}\left(Xi-\bar{X}\right)^3 f_i}{\left(S_{n-1}\right)^3} = \frac{\sum_{i=1}^{m}\left(Xi-\bar{X}\right)^3 f_i^*}{\left(S_{n-1}\right)^3}$$

sustituyendo

$$a_3 = \frac{\frac{1}{40}\left[(726.5-736.6)^3(4)+\ldots+(746.5-736.6)^3(4)\right]}{\left(\sqrt{32.81}\right)^3} = \frac{-4.398}{(5.728)^3} \approx -0.023 < 0$$

c) Para hacer la comparación y tomar una decisión, al utilizar el coeficiente de variación

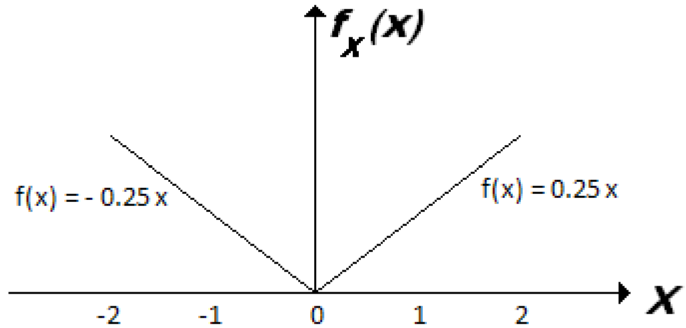
$$CV = \frac{S_{n-1}}{\bar{X}}$$

para DELLA el coeficiente de variación es $cv_{DELLA} = \frac{119}{403} \approx 0.295$

para IBT el coeficiente de variación es $cv_{IBT} = \frac{5.728}{736.6} \approx 0.008$

La marca DELLA tiene menor promedio y mayor variabilidad. La marca IBT mayor promedio y menor variabilidad, es mejor la marca IBT, tiene menor variabilidad.

2. En la siguiente figura se presenta la función de densidad de la variable aleatoria continua. Calcular la media y la variancia.



15 Puntos

Resolución

La media es igual a cero la abscisa del centroide, matemáticamente se tiene

$$E(X)=\int_{-\infty}^{\infty} x f_X(x)\ dx$$

$$E(X)=\int_{-2}^{0} x(-0.25x)\ dx+\int_{0}^{2} x(0.25x)\ dx=\frac{-0.25}{3}\left[x^3\right]\Big|_{-2}^{0}+\frac{0.25}{3}\left[x^3\right]\Big|_{0}^{2}=0$$

La variancia es el segundo momento con respecto de la media, entonces

$$Var(X)=E\left[(X-\mu)^2\right]=\int_{-\infty}^{\infty}(X-\mu)^2 f_X(x)\ dx$$

$$Var(X)=E\left[(X-0)^2\right]=\int_{-\infty}^{\infty}(x-0)^2 f_X(x)\ dx=\int_{-2}^{0}(x)^2(-0.25x)\ dx+\int_{0}^{2}(x)^2(0.25x)\ dx$$

$$Var(X)=-0.25\int_{-2}^{0} x^3\ dx+0.25\int_{0}^{2} x^3\ dx=\frac{-0.25}{4}\left[x^4\right]\Big|_{-2}^{0}+\frac{0.25}{4}\left[x^4\right]\Big|_{0}^{2}$$

$$Var(X)=-\frac{1}{16}\left[-(-2)^4\right]+\frac{1}{16}\left[(2)^4\right]=\frac{16}{16}+\frac{16}{16}=2$$

3. Una moneda corriente se lanza 14 veces. Determinar la probabilidad de que:

a) el número de caras que salga esté entre cinco y ocho, inclusive;

b) la primera cruz aparezca en el quinto lanzamiento; y

c) la tercera cara aparezca en el octavo lanzamiento.

15 Puntos

Resolución

a) Sea X la variable aleatoria que representa el número de caras que se pueden tener en el lanzamiento de la moneda legal.

$$X \sim Binomial\left(n=14\ ,\ p=\frac{1}{2}\right)$$

$$f_X(x)=\begin{cases}\binom{14}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{14-x} & ;\quad x=0,1,2,\ldots,14\\ 0 & ;\quad en\ otro\ caso\end{cases}$$

$$P(5\le X\le 8)=P(X=5)+P(X=6)+P(X=7)+P(X=8)=\sum_{x=5}^{8}\binom{14}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{14-x}\approx 0.6982$$

También se puede resolver aproximando de normal a binomial, entonces

$$X \sim Binomial\left(n = 14\ ,\ p = \frac{1}{2} \right)$$

como $np = 7 > 5$ es aceptable la aproximación, entonces

$$Y \sim Normal\left(\mu_Y = np = 7\ ,\ \sigma_Y^2 = \text{npq} = \frac{14}{4} = 3.5 \right)$$

$$P(5 \le Y \le 8) \approx P\left(\frac{5 - np - 0.5}{\sqrt{npq}} \le \frac{Y - \mu_Y}{\sigma_Y} \le \frac{8 - np + 0.5}{\sqrt{npq}} \right) = P\left(\frac{5 - 7 - 0.5}{\sqrt{3.5}} \le \frac{Y - \mu_Y}{\sigma_Y} \le \frac{8 - 7 + 0.5}{\sqrt{3.5}} \right)$$

$$P(5 \le X \le 8) \approx P(-1.34 \le Z \le 0.80) = F_Z(0.80) - F_Z(-1.34) = 0.7881 - 0.0901 = 0.6980$$

que es muy buena aproximación.

b) Sea U la variable aleatoria que representa la primera águila se pueda obtener en el quinto lanzamiento.

$$U \sim Geométrica\left(p = \frac{1}{2} \right)$$

$$f_U(u) = \begin{cases} \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{u-1} & ;\quad u = 1, 2, \ldots, 14 \\ 0 & ;\quad en\ otro\ caso \end{cases}$$

$$P(U = 5) = \frac{1}{2}\left(\frac{1}{2}\right)^4 = \frac{1}{32} \approx 0.0312$$

c) Sea T la variable aleatoria que representa la tercera cara pueda aparecer en el octavo lanzamiento.

$$T \sim Pascal\left(\text{r=3}\ ,\ p = \frac{1}{2} \right)$$

$$f_T(t) = \begin{cases} \binom{7}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{t-3} & ;\quad t = 3, 4, \ldots, 8 \\ 0 & ;\quad en\ otro\ caso \end{cases}$$

$$P(T = 8) = \binom{7}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5 = \binom{7}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5 = \frac{21}{256} \approx 0.0820$$

4. Supóngase que X y Y tienen la distribución conjunta dada por

$f_{XY}(x,y)$		y = -2	y = 1	y = 3
x	2	0.15	0.14	0.20
	4	0.25	0.15	0.11

Calcular la probabilidad de que:

a) A lo sumo X sea dos y Y cuando mucho sea tres.

b) Al menos Y sea uno y X cuatro.

c) El total de la suma sea mínimo tres.

15 Puntos

Resolución

a) Se pide determinar

$$P(X \le 2, Y \le 3) = f_{XY}(2,-2) + f_{XY}(2,1) + f_{XY}(2,3) = 0.15 + 0.14 + 0.20 = 0.49$$

b) Se pide determinar

$$P(X=4, Y\geq 1) = f_{XY}(4,1) + f_{XY}(4,3) = 0.15 + 0.11 = 0.26$$

c) Se pide determinar

$$P(X+Y\geq 3) = f_{XY}(2,1) + f_{XY}(2,3) + f_{XY}(4,1) + f_{XY}(4,3) = 0.14 + 0.20 + 0.15 + 0.11 = 0.60$$

5. En la UNAM, la media de edad de los estudiantes es de 22.3 años y la desviación estándar de cuatro años, se toma una muestra aleatoria de 64 estudiantes, ¿cuál es la probabilidad que la edad promedio de estos estudiantes sea a lo más de 22 años? ¿cuál es la probabilidad de que la edad promedio de estos estudiantes sea mayor de 23 años?

15 Puntos

Resolución

Sea X la variable aleatoria que representa la edad de los estudiantes de la UNAM.

$$X_i \sim Normal\left(\mu_{X_i} = 22.3\ ,\ \sigma^2_{X_i} = 16\right) \quad ; \quad i = 1, 2, 3, \ldots, 64$$

por el teorema del límite central, ya que, es una muestra aleatoria grande 64 estudiantes, cada uno idénticamente distribuido, se tiene

$$\bar{X} \sim Normal\left(\mu_{\bar{X}} = \mu_{X_i} = 22.3\ ,\ \sigma^2_{\bar{X}} = \frac{\sigma^2_{X_i}}{n} = \frac{16}{64}\right)$$

se pide calcular y al normalizar, se tiene

$$P\left(\bar{X} < 22\right) \approx P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\frac{\sigma_{X_i}}{\sqrt{n}}} < \frac{22 - 22.3}{\frac{4}{\sqrt{64}}}\right) = P\left(Z < \frac{8(-0.3)}{4}\right) = P\left(Z < \frac{-2.40}{4}\right) = P(Z < -0.60)$$

al utilizar tablas de la distribución acumulativa, se tiene

$$P(Z < -0.60) = F_Z(-0.60) = 0.2743$$

También se pide calcular

$$P\left(\bar{X} > 23\right) \approx P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\frac{\sigma_{X_i}}{\sqrt{n}}} > \frac{23 - 22.3}{\frac{4}{\sqrt{64}}}\right) = P\left(Z > \frac{8(0.7)}{4}\right) = P\left(Z > \frac{5.60}{4}\right) = P(Z > 1.40)$$

al utilizar tablas de la distribución acumulativa, se tiene

$$P(Z > 1.40) = P(Z < -1.40) = 0.0808$$

Por la simetría de la distribución normal estándar con respecto de su media que es cero.

6. En un centro de cómputo se ha observado que el crecimiento de la información almacenada en los últimos seis meses ha ido aumentando su tamaño de acuerdo a la siguiente secuencia por mes: 50GB, 100GB, 110GB, 125GB, 170GB y 180GB.

a) Hacer la gráfica de dispersión.

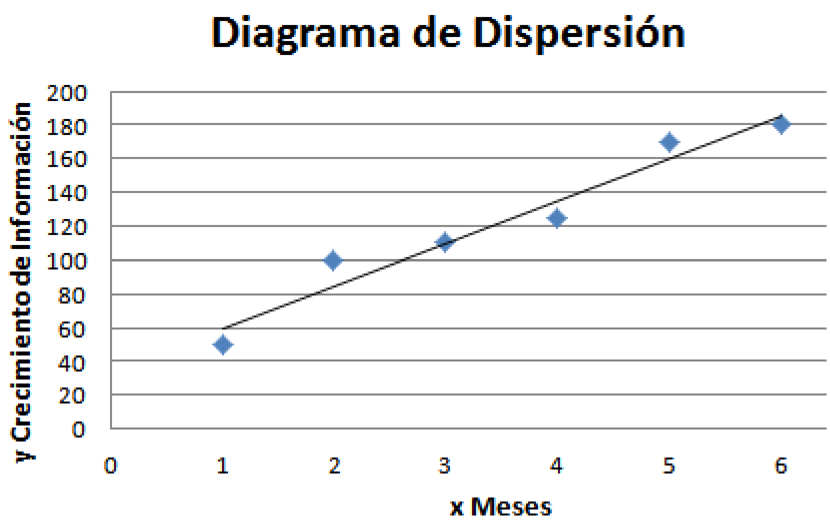
b) Obtener la recta de regresión por el método de mínimos cuadrados.

c) ¿Se podría afirmar que el ritmo de crecimiento de la información lleva una tendencia lineal? Justificar su respuesta.

20 Puntos

Resolución

a)



$\hat{y} = 25x + 35$

$\hat{R}^2 = 0.952$

$\hat{R} = 0.976$

b)

Meses x	Crecimiento de la Información y	xy	x^2	y^2
1	50	50	1	2500
2	100	200	4	10000
3	110	330	9	12100
4	125	500	16	15625
5	170	850	25	28900
6	180	1080	36	32400
21	**735**	**3010**	**91**	**101525**

Las medias son

$\bar{X} = \frac{1}{n}\sum_{i=1}^{6} x_i$ sustituyendo $\bar{x} = \frac{1}{6}\sum_{i=1}^{6} 21 = 3.5$

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{6} y_i$ sustituyendo $\bar{y} = \frac{1}{6}\sum_{i=1}^{6} 735 = 122.5$

Los parámetros y el modelo, son

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{6} x_i y_i - \frac{\sum_{i=1}^{6} x_i \sum_{i=1}^{6} y_i}{6}}{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{6}} = \frac{3010 - \frac{(21)(735)}{6}}{91 - \frac{(21)^2}{6}} = \frac{437.5}{17.5} \approx 25$$

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

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 122.5 - 25(3.5) = 35$$

por lo tanto el modelo está dado por

$$\hat{y} = 35 + 25x$$

c) Para determinar si el modelo es válido debe obtenerse el coeficiente de determinación.
El coeficiente de correlación, está definido por

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

$$SS_{xx} = \sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{6} = 91 - \frac{(21)^2}{6} = 17.5$$

$$SS_{yy} = \sum_{i=1}^{6} y_i^2 - \frac{\left(\sum_{i=1}^{6} y_i\right)^2}{6} = 101525 - \frac{(735)^2}{6} = 11487.5$$

$$SS_{xy} = \sum_{i=1}^{6} x_i y_i - \frac{\sum_{i=1}^{6} x_i \sum_{i=1}^{6} y_i}{6} = 3010 - \frac{(21)(735)}{6} = 437.5$$

sustituyendo

$$r = \frac{437.5}{\sqrt{(17.5)(11487.5)}} \approx 0.9758$$

que es el coeficiente de correlación, es fuerte la asociación lineal de las variables. Hay una tendencia lineal fuerte.

Entonces el coeficiente de determinación es, elevando al cuadrado el coeficiente de correlación

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

sustituyendo

$$r^2 = \frac{(437.5)^2}{(17.5)(11487.5)} \approx 0.9522$$

El modelo explica bien a la variable dependiente que es 95.22% se puede considerar válido el modelo.
Del coeficiente de determinación 0.9522 se puede decir que hay una buena explicación del crecimiento de la información conforme pasan los meses. Es buena la eficiencia del modelo lineal.