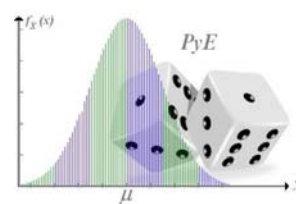




UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE INGENIERÍA
DIVISIÓN DE CIENCIAS BÁSICAS
COORDINACIÓN DE CIENCIAS APLICADAS
DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA
PRIMER EXAMEN FINAL
RESOLUCIONES



SEMESTRE 2012 - 2
DURACIÓN MÁXIMA 2.5 HORAS

TIPO 1
30 DE MAYO DE 2012

NOMBRE _____

Apellido paterno

Apellido materno

Nombre (s)

Firma

Problema 1

La siguiente tabla muestra la estimación de hundimientos de suelo en centímetros, de diez diferentes sitios del Distrito Federal durante el último año.

Hundimiento (cm)	[0,10)	[10,20)	[20,30)	[30,40]
Frecuencia	3		2	1
Frecuencia Acumulada	3		9	
Frecuencia Relativa	0.30		0.20	
Frecuencia Relativa Acumulada	0.30		0.90	

- Completar los valores faltantes en la tabla.
- Determinar el valor medio de los hundimientos registrados.
- Calcular el valor del coeficiente de variación.

Puntos 15

Resolución

- Como el total de observaciones es de 10, el dato faltante del renglón de la frecuencia debe ser 4 para tener el total de 10. Luego se completa la tabla de frecuencias.

Hundimiento (cm)	[0,10)	[10,20)	[20,30)	[30,40]
Frecuencia	3	4	2	1
Frecuencia Acumulada	3	7	9	10
Frecuencia Relativa	0.30	0.40	0.20	0.10
Frecuencia Relativa Acumulada	0.30	0.70	0.90	1.0

- Para obtener el valor medio de los hundimientos, se debe calcular la media, utilizando la fórmula correspondiente a datos agrupados. Tomando las marcas de clase de 5, 15, 25 y 35, con sus frecuencias 3, 4, 2 y 1 respectivamente. Entonces el valor medio del hundimiento será:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^k f_k m_k = 16 \text{ cm}$$

- Se tiene que el coeficiente de variación está dado por: $C.V. = \frac{s}{\bar{x}}(100)$, por lo que falta calcular S, utilizando la fórmula correspondiente a datos agrupados.

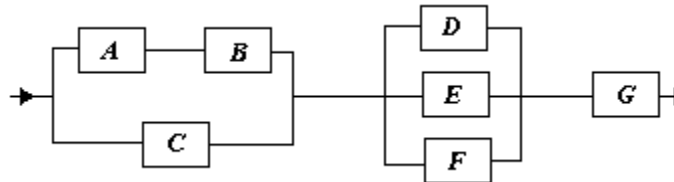
$$s^2 = \frac{\sum_{j=1}^k f_j m_j^2 - \frac{1}{n} (\sum_{j=1}^k f_j m_j)^2}{n - 1} = 98.8888$$

entonces

$S=9.9442$ y $C.V.=62.15\%$

Problema 2

El sistema contiene siete componentes. Estos se conectan como se muestra en el diagrama. Si los siete componentes operan de manera independiente y si la probabilidad de que cada uno de ellos A, B, C, D, E, F, o G, esté funcionando, es de 0.95, determinar la probabilidad de que el sistema completo funcione.



Puntos 15

Resolución

La probabilidad de que el sistema completo funcione, equivale a todas las componentes funcionen, entonces

$$P(\text{Funciona completamente}) = (0.95)^7$$

$$P(\text{Funciona completamente}) = 0.6983$$

Problema 3

Supóngase que el error en la lectura del consumo de energía eléctrica en Kw/h, es una variable aleatoria X que tiene por función densidad

$$f_X(x) = \begin{cases} \frac{x^2}{3} & ; \quad -1 \leq x \leq 2 \\ 0 & ; \quad \text{en otro caso} \end{cases}$$

a) Calcular $P(0 \leq X \leq 1)$

b) Determinar de manera analítica y gráfica la función de distribución acumulada y utilizarla para verificar el resultado del inciso a).

c) Obtener la variancia de $h(X) = 4X + 3$

15 Puntos

Resolución

a) La probabilidad de que el error en la lectura de consumo sea entre cero y uno, es por propiedades se sabe

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

entonces

$$P(0 \leq X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \left. \frac{1}{3} \frac{x^3}{3} \right|_0^1 = \frac{1}{9} [1^3 - 0^3] = \frac{1}{9}$$

b) La función de distribución se define por la integral con límite superior variable, entonces

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

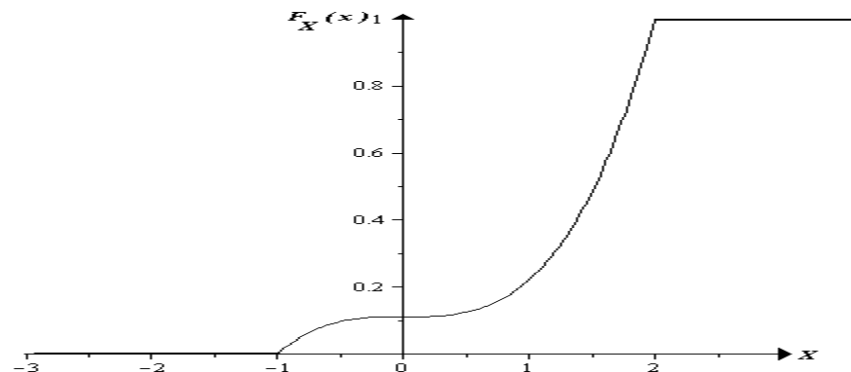
sustituyendo

$$F_X(x) = \int_{-1}^x \frac{t^2}{3} dt = \left. \frac{1}{3} \frac{t^3}{3} \right|_{-1}^x = \frac{1}{9} [x^3 - (-1)^3] = \frac{1}{9} (x^3 + 1)$$

por lo tanto

$$F_X(x) = \begin{cases} 0 & ; \quad x < -1 \\ \frac{1}{9}(x^3 + 1) & ; \quad -1 < x < 2 \\ 1 & ; \quad x > 2 \end{cases}$$

La gráfica de la función acumulativa es



Para calcular la probabilidad, por propiedades de la función de distribución acumulativa

$$P(a \leq X \leq b) = F_X(b) - F_X(a)$$

$$P(0 \leq X \leq 1) = F_X(1) - F_X(0) = \frac{1}{9}(1^3 + 1) - \frac{1}{9}(0^3 + 1) = \frac{1}{9}$$

- c) Se pide calcular la variancia, segundo momento con respecto a la media, o bien, en términos de momentos con respecto al origen, se sabe que es el segundo momento con respecto a origen menos el primer momento con respecto al origen al cuadrado, es decir

$$Var(h(X)) = Var(4X + 3) = 16 Var(X) = 16 [E(X^2) - [E(X)]^2]$$

Integrando para calcular los momentos con respecto al origen, de la definición de momentos con respecto al origen, se sabe

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

sustituyendo

$$E(X) = \int_{-1}^2 x \frac{x^2}{9} dx = \frac{1}{9} \int_{-1}^2 x^3 dx = \frac{1}{12} [x^4]_{-1}^2 = \frac{1}{12} [16 - 1] = \frac{5}{4}$$

El segundo momento con respecto al origen, se tiene

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

sustituyendo

$$E(X^2) = \int_{-1}^2 x^2 \frac{x^2}{9} dx = \frac{1}{9} \int_{-1}^2 x^4 dx = \frac{1}{15} [x^5]_{-1}^2 = \frac{1}{15} [32 + 1] = \frac{11}{5}$$

al sustituir en la variancia de $h(X)$ se tiene y se sabe que la variancia de una constante es igual a cero

$$Var(h(X)) = Var(4X + 3) = 16 Var(X) + Var(4) = 16 \left[\frac{11}{5} - \left(\frac{5}{4} \right)^2 \right] = \frac{51}{5} = 10.2$$

Otra forma de calcular la variancia de $h(X)$, al aplicar variancia en la función de la variable aleatoria, se tiene que obtener la media

$$\mu_{h(X)} = E[h(X)] = E[4X + 3] = 4E(X) + 3 = 4\left(\frac{8}{4}\right) + 3 = 8$$

entonces el segundo momento con respecto de su media de la función dada, es

$$Var(h(X)) = Var([h(X) - \mu_{h(X)}]^2) = E([4X + 3 - 8]^2) = E([4X - 5]^2)$$

por integral de acuerdo con la definición de valor esperado, se tiene

$$E([4X - 5]^2) = \int_{-1}^2 (4x - 5)^2 \left(\frac{x^2}{3}\right) dx = \frac{1}{3} \int_{-1}^2 (4x^2 - 5x)^2 dx = \frac{51}{5} = 10.2$$

Problema 4

Un examen contiene 15 preguntas tipo falso-verdadero. El examen se aprueba si por lo menos se contestan correctamente 13 preguntas. Si un alumno contesta al azar el examen, ¿cuál es la probabilidad de que lo apruebe?

10 Puntos

Resolución

Sea X la variable aleatoria que representa el número de respuestas correctas.

$X \sim \text{Binomial}(n = 15, p = 0.5)$

$P(\text{Aprobar el examen}) = P(X \geq 13)$

entonces

$P(X \geq 13) = P(X = 13) + P(X = 14) + P(X = 15)$

sustituyendo

$$P(X \geq 13) = \binom{15}{13} (0.5)^{13} (0.5)^2 + \binom{15}{14} (0.5)^{14} (0.5)^1 + \binom{15}{15} (0.5)^{15} (0.5)^0 = 3.69 \times 10^{-3}$$

Por lo tanto, la probabilidad de que el alumno apruebe el examen es de 0.00369

Problema 5

Se hacen dos pruebas diferentes para un artículo deportivo. Las siguientes distribuciones marginales corresponden a cada una de las pruebas. En donde X e Y son variables aleatorias que representan a cada una de las pruebas aplicadas a los artículos deportivos.

x	0	1	2
$f_X(x)$	0.2	0.3	0.5

y	0	1	2	3
$f_Y(y)$	0.3	0.2	0.1	0.4

Si se sabe que X y Y son variables aleatorias conjuntas estadísticamente independientes.

a) Obtener la distribución de probabilidad conjunta.

b) Si se sabe que la prueba Y es exactamente dos, ¿cuál es la probabilidad de que la prueba X sea menor a dos?

15 Puntos

Resolución

a) Si se sabe X y Y son estadísticamente independientes entonces

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

La función de probabilidad conjunta es

$f_{XY}(x, y)$		x		
		0	1	2
y	0	0.06	0.09	0.15
	1	0.04	0.06	0.1
	2	0.02	0.03	0.05
	3	0.08	0.12	0.2

b) Se pide obtener $P(X < 2 | Y=2)$

$$P(X = 0 | Y = 2) = \frac{f_{XY}(X = 0, Y = 2)}{f_Y(2)} = \frac{0.02}{0.1} = 0.2$$

$$P(X = 1 | Y = 2) = \frac{f_{XY}(X = 1, Y = 2)}{f_Y(2)} = \frac{0.03}{0.1} = 0.3$$

$$P(X < 2 | Y = 2) = 0.2 + 0.3 = 0.5$$

Otra forma, como las variables aleatorias son estadísticamente independientes, entonces se sabe que

$$f_{X|Y=y}(X | Y=2) = f_X(x)$$

x	0	1	2
$f_X(x)$	0.2	0.3	0.5

entonces

$$P(X < 2 | Y=2) = f_X(0) + f_X(1) = 0.2 + 0.3 = 0.5$$

Problema 6

En una tienda venden tres marcas diferentes de yogurt en envases de 150 g. De todos los clientes que compran un sólo envase, 50% compra el que contiene 160 calorías, 30% compra el de 200 calorías y el otro 20% compra el de 250 calorías. Sean X y Y el número de calorías de los envases comprados por dos clientes seleccionados independientemente.

a) Calcular la media μ y la variancia σ^2 de la población.

b) Determinar la distribución muestral de $\bar{X}$, calcular $E(\bar{X})$ y comparar con μ .

c) Obtener la distribución muestral de S_{n-1}^2 , calcular $E(S_{n-1}^2)$ y comparar con σ^2 .

15 Puntos

Resolución

a) La función de probabilidad para un cliente que llega a la tienda a comprar yogurt, es

x	160	200	250
$f_X(x)$	0.5	0.3	0.2

La media está definida por

$$\mu = E(X) = \sum x f_X(x)$$

$$\mu = 0.5(160) + 0.3(200) + 0.2(250) = 190$$

La variancia está definida por

$$\sigma_X^2 = Var(X) = E([X - \mu]^2) = \sum [X - \mu]^2 f_X(x)$$

$$\sigma_X^2 = 0.5(160 - 190)^2 + 0.3(200 - 190)^2 + 0.2(250 - 190)^2 = 1200$$

Los cuales son los parámetros de la población.

b) Para la distribución muestral de la media y variancia, si dos clientes entran de forma independiente a comprar a la tienda un yogurt, entonces una forma para calcular es

X	Y	Probabilidad	$\bar{X}$	S_{n-1}^2
160	160	$(0.5)(0.5)=0.25$	160	$\frac{1}{2-1} [(160-160)^2 + (160-160)^2] = 0$
160	200	$(0.5)(0.3)=0.15$	180	$\frac{1}{2-1} [(160-180)^2 + (200-180)^2] = 800$
160	250	$(0.5)(0.2)=0.1$	205	$\frac{1}{2-1} [(160-205)^2 + (250-205)^2] = 4050$
200	160	$(0.3)(0.5)=0.15$	180	$\frac{1}{2-1} [(200-180)^2 + (160-180)^2] = 800$
200	200	$(0.3)(0.3)=0.09$	200	$\frac{1}{2-1} [(200-200)^2 + (200-200)^2] = 0$
200	250	$(0.3)(0.2)=0.06$	225	$\frac{1}{2-1} [(200-225)^2 + (250-225)^2] = 1250$
250	160	$(0.2)(0.5)=0.1$	205	$\frac{1}{2-1} [(250-205)^2 + (160-205)^2] = 4050$
250	200	$(0.2)(0.3)=0.06$	225	$\frac{1}{2-1} [(250-225)^2 + (200-225)^2] = 1250$
250	250	$(0.2)(0.2)=0.04$	250	$\frac{1}{2-1} [(250-250)^2 + (250-250)^2] = 0$

la función de probabilidad para la media muestral es

$\bar{X}$	160	180	200	205	225	250
$f_{\bar{X}}(\bar{X})$	0.25	0.3	0.09	0.2	0.12	0.04

$$E(\bar{X}) = \sum_{\bar{X}} \bar{X} f_{\bar{X}}(\bar{X})$$

$$E(\bar{X}) = 0.25(160) + 0.3(180) + 0.09(200) + 0.2(205) + 0.12(225) + 0.04(250) = 190$$

Por lo tanto

$$\mu = E(\bar{X}) = 190$$

c) Para la distribución de la variancia muestral, se tiene

s_{n-1}^2	0	800	1250	4050
$f_{s_{n-1}^2}(s_{n-1}^2)$	0.38	0.3	0.12	0.2

$$E(s_{n-1}^2) = \sum_{s_{n-1}^2} s_{n-1}^2 f_{s_{n-1}^2}(s_{n-1}^2)$$

$$E(s_{n-1}^2) = 0.3(800) + 0.12(1250) + 0.2(4050) = 1200$$

Por lo tanto

$$E(s_{n-1}^2) = \sigma^2 = 1200$$

Problema 7

En la siguiente tabla se registraron los datos que se refieren al crecimiento de células de piel humana en un medio de cultivo controlado, en una investigación dirigida a la restauración de piel en víctimas de quemaduras.

Días de cultivo	3	6	9	12	15	18
Número de células	115000	147000	189000	230600	257900	286400

- a) Obtener el coeficiente de determinación para un modelo lineal, y a partir de este responder: ¿Es adecuado el modelo usado? Argumentar su respuesta. Conclusión:

_____.

- b) Con la ecuación de regresión lineal, estimar el número de células después de 17 días: _____.

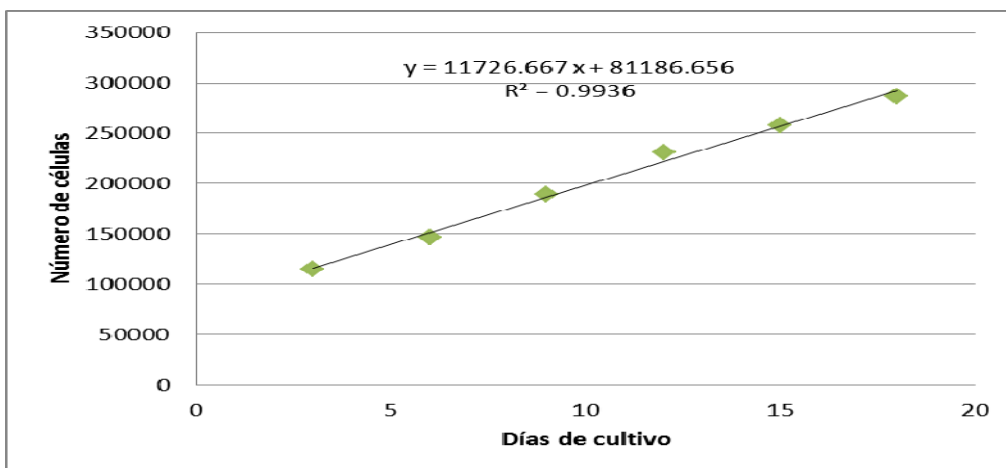
Se sabe que las sumas son:

	Días de cultivo x	Número de células y	x^2	y^2	xy
Sumas	63	1225900	819	2.72269E+11	14718900

15 Puntos

Resolución

El diagrama de dispersión está dado por la gráfica siguiente (opcional)



Días de cultivo x	Número de células y	x^2	y^2	xy
3	115000	9	13225000000	345000
6	147000	36	21609000000	882000
9	189000	81	35721000000	1701000
12	230600	144	53176360000	2767200
15	257900	225	66512410000	3868500
18	286400	324	82024960000	5155200
63	1225900	819	2.72269E+11	14718900

- a) Como el coeficiente de determinación se utiliza como medida de eficacia de la regresión, éste se calculará a partir del cuadrado del coeficiente de correlación:

Las medias son

$$\bar{x} = \frac{1}{n} \sum_{i=1}^6 x_i = 10.5$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^6 y_i = 204316.66$$

Los parámetros y el modelo, son:

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^6 x_i y_i - \frac{\sum_{i=1}^6 x_i \sum_{i=1}^6 y_i}{6}}{\sum_{i=1}^6 x_i^2 - \frac{\left(\sum_{i=1}^6 x_i\right)^2}{6}} = \frac{14718900 - \frac{(63)(1225900)}{6}}{819 - \frac{(63)^2}{6}} = \frac{1846950}{157.5} \approx 11726.667$$

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 204316.66 - (11726.667)(10.5) \approx 81186.656$$

por lo tanto el modelo está dado por

$$\hat{y} = 11726.667x + 81186.656$$

Para determinar si el modelo es válido debe obtenerse el coeficiente de determinación.

El coeficiente de correlación, está definido por

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

$$SS_{xx} = \sum_{i=1}^6 x_i^2 - \frac{\left(\sum_{i=1}^6 x_i\right)^2}{6} = 819 - \frac{(63)^2}{6} = 157.5$$

$$SS_{yy} = \sum_{i=1}^6 y_i^2 - \frac{\left(\sum_{i=1}^6 y_i\right)^2}{6} = 2.72269E+11 - \frac{(1225900)^2}{6} = 21796928333$$

$$SS_{xy} = \sum_{i=1}^6 x_i y_i - \frac{\sum_{i=1}^6 x_i \sum_{i=1}^6 y_i}{6} = 14718900 - \frac{(63)(1225900)}{6} = 1846950$$

sustituyendo

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} = \frac{1846950}{\sqrt{(157.5)(21796928333)}} \approx 0.9968$$

entonces el coeficiente de determinación es

$$r^2 = R^2 = \frac{SS_{xy}^2}{SS_{xx} SS_{yy}} = \frac{(1846950)^2}{(157.5)(21796928333)} \approx 0.9936$$

Del resultado anterior, se puede observar y concluir, que el coeficiente de determinación es $r^2 \approx 0.9936$, esto es, 99.36% y está cercano al 100%, por lo que se considera que el modelo lineal es adecuado para estos datos.

- b)** Se debe considerar los días de cultivo como variable x, y en número de células como y, entonces la ecuación de regresión que se obtiene es

$$\hat{y} = 11726.667x + 81186.656$$

Si $x=17$, entonces sustituyendo en la estimación de la recta de regresión

$$\hat{y} = 11726.667x + 81186.656$$

$$\hat{y} = 11726.667(17) + 81186.656 = 280539.995$$

células.