

Probabilidad y Estadística
Segundo Examen Final
SOLUCIÓN

Semestre: 2008-1
Duración máxima: 2.5 h

1. En la siguiente tabla se muestra la clasificación combinada del número de millas y el volumen del motor por parte de la APA en 49 estados (todos menos California) para nueve automóviles subcompactos con transmisión estándar, de cuatro cilindros y que utilizan gasolina. El tamaño del motor se da en pulgadas cúbicas totales de cilindrada.

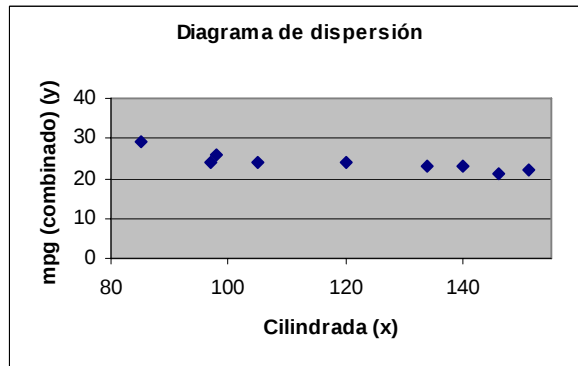
Automóvil	Cilindrada (x)	mpg (combinado) (y)
VW Rabbit	97	24
Datsun 210	85	29
Chevette	98	26
Dodge Omni	105	24
Mazda 626	120	24
Oldsmobile Starfire	151	22
Mercury Capri	140	23
Toyota Celica	134	23
Datsun 810	146	21

- Representar los datos en un diagrama de dispersión.
- Ajustar a los datos un modelo lineal de regresión, empleando el criterio de mínimos cuadrados.
- Dibujar la gráfica de la recta de los mínimos cuadrados para ver qué tan bien se ajusta a los datos.
- Utilizar la recta de los mínimos cuadrados para estimar el promedio de millas por galón para un automóvil subcompacto que tiene un volumen del motor de 125 pulgadas cúbicas.
- Calcular la covariancia, el coeficiente de correlación y de determinación. Interpretar los resultados de la relación de las variables.

20 Puntos

Resolución

- a) El diagrama de dispersión es:



- b) El ajuste de los datos a un modelo lineal de regresión por el criterio de mínimos cuadrados está dado por:

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

donde:

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

y

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

realizando los productos y las sumas, se tiene:

x	y	x ²	y ²	xy
97	24	9409	576	2328
85	29	7225	841	2465
98	26	9604	676	2548
105	24	11025	576	2520
120	24	14400	576	2880
151	22	22801	484	3322
140	23	19600	529	3220
134	23	17956	529	3082
146	21	21316	441	3066
1076	216	133336	5228	25431

de donde:

$$SS_{xx} = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} = 133336 - \frac{(1076)^2}{9} = 4694.222$$

$$SS_{xy} = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} = 25431 - \frac{(1076)(216)}{9} = -393$$

sustituyendo :

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-393}{4694.222} = -0.084$$

para encontrar los promedios, se tiene:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{9}(1076) = 119.556$$

y

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{9}(216) = 24$$

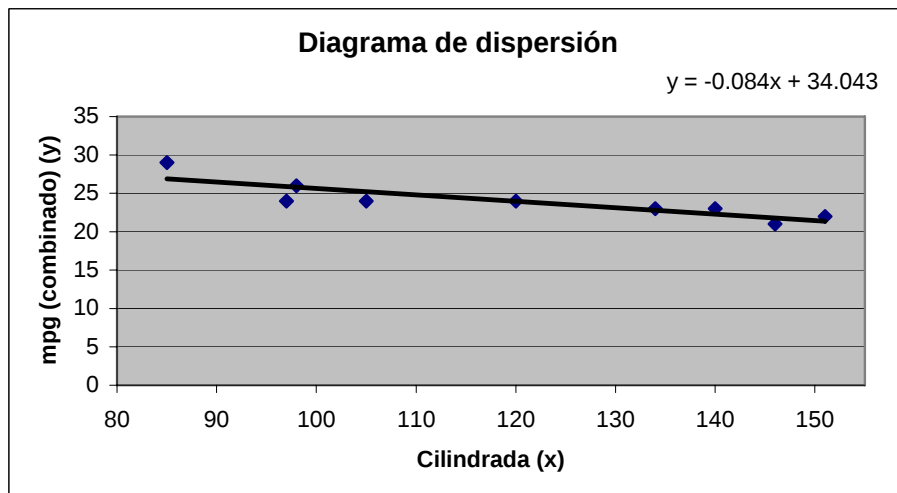
sustituyendo:

$$\hat{\beta}_0 = 24 - (-0.084)(119.556) = 34.043$$

por lo tanto el modelo lineal de regresión es:

$$\hat{y} = 34.043 - 0.084x$$

c) La gráfica es:



d) Un automóvil subcompacto que tiene un volumen del motor de 125 pulgadas cúbicas, sustituyendo en el modelo, se tiene:

$$\hat{y} = 34.043 - 0.084(125) = 23.543$$

- e) La covariancia está dada por:

$$\hat{Cov}(X, Y) = \frac{SS_{xy}}{n} = \frac{-393}{9} = -43.667$$

El coeficiente de correlación está definido por:

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

calculando:

$$SS_{yy} = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = 5228 - \frac{(216)^2}{9} = 44$$

por lo que el coeficiente de correlación es:

$$r = \frac{-393}{\sqrt{(4694.222)(44)}} = -0.865$$

El coeficiente de determinación está definido por:

$$R^2 = (r)^2 = (-0.865)^2 = 0.748$$

Se concluye que las variables tienen una buena relación lineal, puesto que: $r = -0.865$

2. Una urna contiene 40 bolas blancas y 10 bolas negras. Si dos bolas se sacan aleatoriamente, determinar la probabilidad de que las dos sean blancas si:
- la primera bola es reemplazada antes de sacar la segunda.
 - la primera bola no es reemplazada antes de sacar la segunda.

10 Puntos

Resolución

- a) Sea A el evento que representa que se selecciona aleatoriamente una bola blanca de la urna, con reemplazo, si la primera bola que se selecciona es blanca.

$$P(A|A) = \left(\frac{40}{50}\right)\left(\frac{40}{50}\right) = \left(\frac{40}{50}\right)^2 = \left(\frac{4}{5}\right)^2 = \frac{16}{25} = 0.64$$

- b) Sea B el evento que representa que se selecciona aleatoriamente una bola blanca de la urna, sin reemplazo, si la primera bola que se selecciona es blanca.

$$P(B|B) = \left(\frac{40}{50}\right)\left(\frac{39}{49}\right) = \left(\frac{4}{5}\right)\left(\frac{39}{49}\right) = \frac{156}{245} \approx 0.637$$

3. Se tiene la siguiente función de densidad:

$$f_X(x) = \begin{cases} 2e^{-2x} & ; & x > 0 \\ 0 & ; & \text{en otro caso} \end{cases}$$

- Obtener el valor esperado de la función.
- Calcular la variancia de la función.

10 Puntos

Resolución

Sea X una variable aleatoria con distribución exponencial con $\lambda = 2$

- a) El valor esperado es:

$$E(X) = \frac{1}{\lambda} = \frac{1}{2} = 0.5$$

o bien:

$$E(X) = \int_0^{\infty} 2x e^{-2x} dx = \lim_{R \rightarrow \infty} \int_0^R 2x e^{-2x} dx = \frac{1}{2} = 0.5$$

- b) La variancia es:

$$Var(X) = \frac{1}{\lambda^2} = \frac{1}{2^2} = \frac{1}{4} = 0.25$$

o bien:

$$\text{Var}(X) = \int_0^{\infty} \left(x - \frac{1}{2}\right)^2 2e^{-2x} dx = \lim_{R \rightarrow \infty} \int_0^R \left(x - \frac{1}{2}\right)^2 2e^{-2x} dx = \frac{1}{4} = 0.25$$

4. Los conductores que se fabrican para utilizarse en determinado sistema de cómputo, necesitan resistencias que varíen entre 0.12 y 0.14 ohms. Las resistencias reales medidas de los conductores que produce la compañía A tienen una distribución normal con promedio de 0.13 y desviación estándar de 0.005 ohm, respectivamente.
- ¿Cuál es la probabilidad de que un conductor seleccionado al azar de la producción de la compañía A cumpla con las especificaciones?
 - Si se usan cuatro de estos conductores en el sistema de la compañía A, ¿cuál es la probabilidad de que el cuarto conductor sea el primero que cumple con las especificaciones?
 - Si se usan cuatro de estos conductores en el sistema de la compañía A, ¿cuál es la probabilidad de que el cuarto componente sea el segundo componente que cumple con las especificaciones?

15 Puntos

Resolución

Sea Y la variable aleatoria que representa el valor de las resistencias que fabrica la compañía A.

$$Y \sim \text{Normal}(\mu_Y = 0.13, \sigma_Y = 0.005)$$

- a) Se quiere calcular:

$$P(0.12 \leq Y \leq 0.14) = P(0.12 < Y < 0.14) \approx P\left(\frac{0.12 - 0.13}{0.005} < Z < \frac{0.14 - 0.13}{0.005}\right) = P(-2 < Z < 2)$$

De la tabla de la distribución acumulativa normal estándar, se tiene:

$$F_Z(2.00) - F_Z(-2.00) = 0.9772 - 0.0228 = 0.9544$$

- b) Si se usan cuatro de estos conductores en el sistema de la compañía A, la probabilidad de que el cuarto sea el primero que cumple con las especificaciones, es:

$$Q \sim \text{Geométrica}(p = 0.9544)$$

Se quiere calcular:

$$P(Q = 4) = q^3 p = (1 - 0.9544)^3 (0.9544) = (0.0456)^3 (0.9544) = 9.05 \times 10^{-5}$$

- c) Si se usan cuatro de estos conductores en el sistema de la compañía A, la probabilidad de que el cuarto componente sea el segundo componente que cumple con las especificaciones, es:

$$R \sim \text{Pascal}(p = 0.9544, r = 2)$$

Se pide calcular:

$$P(R = 4) = \binom{3}{1} p^2 q^2 = \binom{3}{1} (0.9544)^2 (0.0456)^2 \approx 0.0057$$

5. En un conmutador telefónico de una central, se reciben en promedio dos llamadas por minuto, suponiendo que las llamadas siguen un proceso de Poisson, determinar la probabilidad de que:

- no llegue ninguna llamada en un periodo de tres minutos.
- lleguen menos de tres llamadas en un periodo de dos minutos.

15 Puntos

Resolución

Sea X la variable aleatoria que representa el número de llamadas que llegan al conmutador telefónico en un minuto.

$$X \sim \text{Poisson}(\lambda = 2)$$

- a) Se pide calcular la probabilidad de que no llegue ninguna llamada en un periodo de tres minutos, entonces:

$$X \sim \text{Poisson}(\lambda t = 2(3) = 6)$$

$$P(X = 0) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{6^0 e^{-6}}{0!} = e^{-6} \approx 0.002$$

- b) Se quiere calcular la probabilidad de que lleguen menos de tres llamadas en un periodo de dos minutos, entonces:

$$X \sim \text{Poisson}(\lambda t = 2(2) = 4)$$

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \sum_{x=0}^2 \frac{\lambda^x e^{-\lambda}}{x!}$$

$$P(X < 3) = \frac{4^0 e^{-4}}{0!} + \frac{4e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} = e^{-4} [1 + 4 + 8] \approx 0.238$$

6. Sean las distribuciones marginales de probabilidad de las variables aleatorias X e Y :

X	1	2	Y	0	1	2	3
$f_X(x)$	0.4	0.6	$f_Y(y)$	0.2	0.3	0.4	0.1

Si X y Y son estadísticamente independientes:

- obtener la distribución de probabilidad conjunta.
- demostrar que el coeficiente de correlación es cero.

15 Puntos

Resolución

Si las variables aleatorias conjuntas son independientes, entonces:

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

- a) La distribución de probabilidad conjunta es:

$f_{XY}(x, y)$		x	
		1	2
y	0	0.08	0.12
	1	0.12	0.18
	2	0.16	0.24
	3	0.04	0.06

- b) El coeficiente de correlación es cero, esto es, en particular:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

realizando los cálculos:

$$E(XY) = \sum_{\forall y} \sum_{\forall x} xy f_{XY}(x, y)$$

sustituyendo:

$$E(XY) = (1)(1)(0.12) + (1)(2)(0.16) + (1)(3)(0.04) + (2)(1)(0.18) + (2)(2)(0.24) + (2)(3)(0.06)$$

$$E(XY) = 2.24$$

Los promedios de las funciones marginales son:

$$E(X) = \sum_{\forall x} x f_X(x) = (1)(0.4) + (2)(0.6) = 1.6$$

$$E(Y) = \sum_{\forall y} y f_Y(y) = (1)(0.3) + (2)(0.4) + (3)(0.1) = 1.4$$

sustituyendo en la covariancia:

$$\text{Cov}(X, Y) = 2.24 - (1.6)(1.4) = 0$$

se verifica que son variables aleatorias conjuntas independientes.

7. Se tomó una muestra aleatoria de 16 calificaciones de una población estudiantil de 100 estudiantes, en donde se encontró una desviación estándar muestral de las calificaciones de cuatro puntos y media muestral de 20 puntos, ¿cuál es la probabilidad de seleccionar a un alumno que tenga más de 21.753 puntos como promedio?

15 Puntos

Resolución

Del enunciado se tiene:

$$n=16 \quad S_{n-1}^2=4^2 \quad \bar{x}=20$$

Se quiere determinar la probabilidad de que un alumno seleccionado al azar tenga al menos 21.753 puntos como promedio muestral, no se conoce la variancia poblacional ni la media poblacional pero se sabe que es igual a la media muestral, entonces:

$$P(\bar{X} > 21.753) = P(\bar{X} \geq 21.753) \approx P\left(\frac{\bar{X} - \mu}{\frac{S_{n-1}}{\sqrt{n}}} > \frac{21.753 - 20}{\frac{4}{\sqrt{16}}}\right) = P\left(T > \frac{(21.753 - 20)(4)}{4}\right) = P(T > 1.753) =$$

de la tabla de la distribución t de Student, con 15 grados de libertad y 1.753, se tiene:

$$P(T > 1.753) = 0.05$$

Distribución acumulativa normal estándar

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997

Distribución t de Student

	α			
v	0.001	0.002	0.003	0.004
1	318.2888	159.1444	106.0963	79.5722
2	22.3285	15.7638	12.8517	11.1130
3	10.2143	8.0524	6.9944	6.3221
4	7.1729	5.9514	5.3213	4.9076
5	5.8935	5.0303	4.5703	4.2620
6	5.2075	4.5242	4.1517	3.8982
7	4.7853	4.2071	3.8868	3.6666
8	4.5008	3.9909	3.7049	3.5067
9	4.2969	3.8345	3.5726	3.3899
10	4.1437	3.7163	3.4721	3.3010
11	4.0248	3.6238	3.3933	3.2311
12	3.9296	3.5495	3.3298	3.1747
13	3.8520	3.4887	3.2777	3.1282
14	3.7874	3.4379	3.2341	3.0893
15	3.7329	3.3948	3.1971	3.0563
16	3.6861	3.3579	3.1653	3.0279
17	3.6458	3.3259	3.1376	3.0032
18	3.6105	3.2979	3.1135	2.9815
19	3.5793	3.2732	3.0921	2.9624
20	3.5518	3.2512	3.0731	2.9453