Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano 2013

Curso : **Probabilidad y Estadística**
Sigla : **EYP1113**
Profesores : **R. Aravena (Sec1), C. Wehrhahn (Sec2) y A. Iturriaga (Sec3)**

PAUTA Examen

Problema 1

Sean X e Y los tiempos de ida y vuelta (en minutos) respectivamente, de un viaje de Santiago a algún lugar de la costa. Suponga que el vector (X, Y) tiene distribución normal bivariada, esto es,

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\},$$

con $\mu_X = 70$, $\mu_Y = 45$, $\sigma_X = 10$, $\sigma_Y = 12$ y $\rho = 0{,}4$.

a) **[2.0 puntos]** Determine la probabilidad que el viaje de ida sea más corto que el viaje de regreso a Santiago. *(Indicación: Use que, en este caso, la combinación lineal de variables aleatorias normales distribuye normal).*

b) **[2.0 puntos]** Se cree que por trabajos de construcción en las vías de acceso, el tiempo de ida se ha incrementado. Para validar lo anterior se toma una muestra de 30 vehículos a los cuales se les pidió que midieran el tiempo de viaje. Los resultados se muestran a continuación:

98,5	77,4	81,6	69,3	84,8	107,9	86,7	82,2	84,6	84,1
79,2	84,8	87,0	75,3	69,0	92,2	85,4	82,1	82,3	96,4
82,9	86,1	80,1	78,3	82,3	69,3	89,7	61,5	75,1	78,2

Usando que $\bar{x} = \frac{1}{30}\sum_{i=1}^{30} x_i \approx 82{,}5$ y $s_x^2 = \frac{1}{29}\sum_{i=1}^{30}(x_i - \bar{x})^2 \approx 9{,}2$, valide la creencia anterior con un nivel de significación del 5 %.

c) **[2.0 puntos]** Dado que las variables aleatorias normales tienen soporte en todo $\mathbb{R}$, puede ser cuestionable usar este tipo de distribuciones para modelar tiempos de viaje. Use el test chi-cuadrado para validar la normalidad de X para un nivel de significación del 5 %. Suponga que la siguiente partición de $\mathbb{R}$: $I_1 = (-\infty;\ 79{,}6]$, $I_2 = (79{,}6;\ 81{,}2]$, $I_3 = (81{,}2;\ 82{,}5]$, $I_4 = (82{,}5;\ 83{,}8]$, $I_5 = (83{,}8;\ 85{,}4]$ e $I_6 = (85{,}4;\ \infty)$.

Solución Preg.1

a) Notemos que la probabilidad pedida es $P(X < Y) = P(W < 0)$ con $W = X - Y$. Usando la indicación, se tiene que $W \sim \mathcal{N}(\mu_W, \sigma_W)$, donde μ_W y σ_W son parámetros por determinar. En efecto,

$$\begin{aligned}
\mu_W &= E(W) = E(X-Y) = E(X) - E(Y) = \mu_X - \mu_Y = 70 - 45 = 25, \\
\sigma_W^2 &= Var(W) = Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y) \\
&= \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho = 10^2 + 12^2 - 2\cdot 10\cdot 12\cdot 0{,}4 = 148, \\
\sigma_W &\approx 12{,}2.
\end{aligned}$$

Entonces, $W \sim \mathcal{N}(25;\ 12{,}2)$ y $P(W < 0) = \Phi\left(\frac{-25}{12{,}2}\right) = 1 - \Phi\left(\frac{25}{12{,}2}\right) \approx 1 - 0{,}9798 = 0{,}0202.$

b) Se quieren constrastar las hipótesis

$$H_0:\ \mu_X = 70 \quad \text{vs} \quad H_1:\ \mu > 70.$$

Test: $t = \frac{\sqrt{n}(\bar{x}-70)}{s_x} \approx 22{,}6$. Además, $t_{1-\alpha,n-1} \approx 1{,}699$. Dado que $t > t_{1-\alpha,n-1}$ se rechaza la hipótesis nula, esto es, parece razonable la creencia que el tiempo de viaje ha aumentado.

c) H_0 : Datos $\sim N(\mu,\sigma)$ $\quad H_1$: Datos $\nsim N(m,s)$ con μ y σ desconocidos. **[0,5]**

I_i	n_i	e_i	$n_i - e_i$	$(n_i - e_i)^2$	$\frac{(n_i-e_i)^2}{e_i}$
I_1	10	5	5	25	5
I_2	1	5	-4	16	3,2
I_3	5	5	0	0	0
I_4	1	5	-4	16	3,2
I_5	5	5	0	0	0
I_6	8	5	3	9	1,8
				$\mathcal{X}^2$	13,2

Dado que $\mathcal{X}^2 = 13{,}2 > \chi^2_{1-\alpha,k-1-2} = \chi^2_{0{,}95;3} = 7{,}81$ (se estimaron 2 parámetros), entonces se rechaza la hipótesis nula, esto es, se refuta la normalidad.

Problema 2

Brownlee (1960) propusó que la distancia requerida por un automóvil para detenerse (Y) puede ser modelada por un modelo lineal, en términos de la velocidad (X). Para ello, mediante un experimento, recoge datos, los cuales son:

X: Velocidad (km/h)	Y: Distancia (mts)
30	6
30	5
45	10
65	18
75	35
85	60

Con:

$$\sum x_i = 330; \quad \sum y_i = 134; \quad \sum x_i^2 = 20900; \quad \sum y_i^2 = 5310; \quad \sum x_i y_i = 9675. \quad \sum (y_i - y_i^{'})^2 = 385$$

a) **[2.5 puntos]** Ajuste el modelo de regresión lineal simple. Es decir, obtenga la estimación de los parámetros y de las estadísticas asociadas.

b) **[2.5 puntos]** Asumiendo normalidad, ¿Cuál es la probabilidad que la distancia requerida sea superior a 30 mts si la velocidad es igual a 70 km/h?

c) **[1.0 puntos]** Grafique los puntos: ¿Qué opina del modelo propuesto? (¿qué hubiese propuesto usted?).

Nota: Trabaje todo con tres decimales.

Solución Preg.2

a) reemplazando los términos en las respectivas fórmulas, se tiene:

$$\hat{\alpha} = -23,77 \quad \hat{\beta} = 0,838$$

$$s_Y^2 = 463,47 \quad s_{Y/X}^2 = 96,331 \quad \rightarrow r^2 = 0,792$$

b) Sea $W = \widehat{Y/X = 70}$, entonces $W \sim N(34,91; 9,815)$ por tanto,

$$P(W > 30) = 1 - \Phi(\frac{30 - 34,91}{9,815}) = 1 - 0,3085 = 0,6914$$

c) El gráfico muestra curva, por tanto el modelo lineal no es el más adecuado (subestima los extremos), el modelo más apropiado es uno cuadrático.

Problema 3

Un edificio tiene dos ascensores, uno más rápido que el otro. El tiempo medio de espera para el ascensor lento es 3 minutos y el tiempo medio de espera para el ascensor rápido es 1 minuto. Un pasajero escoge el ascensor rápido con probabilidad 2/3 y el ascensor lento con probabilidad 1/3

a) **[2.0 puntos]** ¿Cuánto es el tiempo medio de espera para subirse a un ascensor?
Recuerde que $E(g(y)) = \sum_{y\in\Theta_Y} g(y)P(Y=y)$ *y* $E(Z) = E(E(Z|W))$

b) **[2.0 puntos]** Sea N una variable aleatoria con media μ_N que representa el número de personas que esperan un ascensor. Sean $X_i, i = 1, 2, \ldots, N$ variables aleatorias independientes que registran el tiempo de espera de cada persona. Suponiendo que N es independiente de $X_i, i = 1, 2, \ldots, N$ y que cuando llega el ascensor todas las personas se suben a el. Determinar el valor esperado del tiempo total de espera (de todo el grupo).

c) **[2.0 puntos]** (aplicacion TCL) Supongamos ahora que hay 100 personas esperando el ascensor y que el tiempo de espera de cada persona sigue una distribución exponencial con media 2 y son independientes entre si, ¿Cuál es la probabilidad aproximada de que el tiempo medio de espera del grupo sea más de 2 minutos y medio para subir al ascensor? (El ascensor es muy grande).

Solución Preg.3

a) Sea X el tiempo de espera de una persona y sea Y la variable aleatoria que indica si usamos el asensor lento o rápido, es decir

$$Y = \begin{cases} 1, & \text{usamos ascesor rápido;} \\ 0 & \text{si usamos el ascensor lento} \end{cases}$$

y nos piden $E(X)$. Entonces

$$E(X) = E(E(X|Y)) = E(X|Y=1)P(Y=1) + E(X|Y=0)P(Y=0) = 1\cdot\frac{2}{3} + 3\cdot\frac{1}{3} = \frac{5}{3}$$

b) El tiempo de espera de todo el grupo es $\sum_{i=1}^N X_i$ y nos piden encontrar $E(\sum_{i=1}^{N} X_i)$. Entonces

$$E(\sum_{i=1}^{N} X_i) = E(E(\sum_{i=1}^{N} X_i|N=n)) = E(E(\sum_{i=1}^{n} X_i)) = E(\frac{5}{3}N) = \frac{5}{3}E(N) = \frac{5}{3}\mu_N$$

c) Suponemos ahora que $X_i \sim \exp(1/2)$ independientes, entonces $E(X_i) = 2$ y $Var(X_i) = 4$. Encontramos primero la distribución aproximada de $T = \frac{1}{100}\sum_{i=1}^{100} X_i$ usando TCL.

$$\begin{aligned} E(T) &= \frac{1}{100}\sum_{i=1}^{100} E(X_i) = \frac{1}{100}\sum_{i=1}^{100} 2 = 2 \\ Var(T) &= \frac{1}{100}\sum_{i=1}^{100} Var(X_i) = \frac{1}{100}\sum_{i=1}^{100} 4 = \frac{4}{100} = 0.04 \\ \Rightarrow T &\dot{\sim} N(2, \sqrt{0.04}) \end{aligned}$$

La probabilidad pedida entonces es

$$P(T > 2.5) = 1 - P(T \leq 2.5) = 1 - \Phi(\frac{2.5-2}{\sqrt{0.04}}) = 1 - \Phi(2.5) = 0.0062$$