Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2014

Curso	**:**	**Probabilidad y Estadística**
Sigla	**:**	**EYP1113**
Profesores	**:**	**Ricardo Aravena C. y Ricardo Olea O.**

Pauta Examen

Problema 1 (Teorema de Probabilidades Totales)

Como han de saber, hoy en día los cajeros automáticos se encuentran en verdaderos "tándemes", conjunto de cajeros en unidades comerciales, tales como stripcenter, mall, paseos, etc. Suponga que el número de cajeros en las unidades comerciales se puede considerar como una variable aleatoria Poisson cuyo valor esperado es ν. A su vez, un cajero, independiente de los otros, puede estar fuera de servicio con probabilidad p. Suponiendo que el número de cajeros en las unidades comerciales son independientes entre si, ¿cuantas unidades comerciales espera recorrer hasta encontrar al menos un cajero en servicio?

Solución

Definamos Y_i al número de cajeros y X_i al número de cajeros operativos en la en la i-ésima unidad comercial. **[0.5 Ptos.]**

Del enunciado se tiene que

$$Y_i \sim \text{Poisson}(\nu) \quad \text{y} \quad X_i \,|\, Y_i = y \sim \text{Binomial}(y,\, 1-p) \qquad \textbf{[0.5 Ptos.]}$$

Por teorema de probabilidades totales se tiene que (resultado visto en clase y evaluado en la I_2):

$$X_i \sim \text{Poisson}(\nu\,(1-p)) \qquad \textbf{[2.0 Ptos.]}$$

Nota: Basta con que el alumno ponga el resultado para asignar puntaje.

Definamos como

$$Z_i = \begin{cases} 1, & \text{si } X_i \geq 1 \\ 0, & \text{si } X_i = 0 \end{cases} \qquad \textbf{[0.5 Ptos.]}$$

Es decir,

$$Z_i \sim \text{Bernoulli}\left(1 - e^{-\nu\,(1-p)}\right) \qquad \textbf{[0.5 Ptos.]}$$

Por la independencia entre las unidades comerciales, tenemos que el número esperado unidades comerciales a recorrer hasta encontrar al menos un cajero operativo, corresponde al valor esperado de un modelo Geométrico **[1.0 Ptos.]**, es decir, se espera recorrer $\left(1 - e^{\nu\,(1-p)}\right)^{-1}$ unidades comerciales. **[1.0 Ptos.]**

+ 1 Punto Base

Problema 2 (Esperanza y Varianza Condicional)

Si a un cajero se abastece con un total de MM$3 (3 millones de pesos), determine la probabilidad que éste se quede sin dinero durante un día, en el cual concurre N personas de acuerdo a un proceso Poisson con media 30. Suponga que los giros de dinero se comportan como variables aleatorias independientes con distribución Normal de media M$120 y desviación estándar de M$40.

Solución

Definamos como X_i los giros de dinero, los cuales se comportan como variables aleatorias independientes con distribución Normal($\mu = 120$, $\sigma = 40$). **[0.3 Ptos.]**

Además tenemos que $N \sim$ Poisson(30). **[0.3 Ptos.]**

Se pide $P(T > 3000)$, donde

$$T = \sum_{i=1}^{N} X_i \sim \text{Normal}(\mu_T,\, \sigma_T) \qquad \textbf{[1.0 Ptos.]}$$

por ser una suma de Normales.

Por otra parte

$$\begin{aligned}
\mu_T = \text{E}(T) = \text{E}\left[\text{E}\left(T \mid N\right)\right] &= \text{E}\left[\text{E}\left(\sum_{i=1}^{N} X_i \mid N\right)\right] = \text{E}\left[\sum_{i=1}^{N} \text{E}\left(X_i \mid N\right)\right], \quad \text{por linealidad del operador esperanza} \qquad \textbf{[0.3 Ptos.]}\\
&= \text{E}\left[\sum_{i=1}^{N} \text{E}\left(X_i\right)\right], \quad \text{por la independencia de } X_i \text{ con respecto a } N \qquad \textbf{[0.3 Ptos.]}\\
&= \text{E}\left(\sum_{i=1}^{N} \mu\right) = \text{E}\left(N \cdot \mu\right) = \mu \cdot \text{E}\left(N\right) = \mu \cdot \mu_N = 120 \cdot 30 = 3600 \qquad \textbf{[0.3 Ptos.]}
\end{aligned}$$

y

$$\begin{aligned}
\sigma_T^2 &= \text{Var}(T) = \text{Var}\left[\text{E}\left(T \mid N\right)\right] + \text{E}\left[\text{Var}\left(T \mid N\right)\right] \qquad \textbf{[0.3 Ptos.]}\\
&= \text{Var}\left[\text{E}\left(\sum_{i=1}^{N} X_i \mid N\right)\right] + \text{E}\left[\text{Var}\left(\sum_{i=1}^{N} X_i \mid N\right)\right] \qquad \textbf{[0.3 Ptos.]}\\
&= \text{Var}\left[\sum_{i=1}^{N} \text{E}\left(X_i \mid N\right)\right] + \text{E}\left[\text{Var}\left(\sum_{i=1}^{N} X_i \mid N\right)\right], \quad \text{por linealidad el operdor esperanza} \qquad \textbf{[0.3 Ptos.]}\\
&= \text{Var}\left[\sum_{i=1}^{N} \text{E}\left(X_i \mid N\right)\right] + \text{E}\left[\sum_{i=1}^{N} \text{Var}\left(X_i \mid N\right)\right], \quad \text{por independencia entre las variables } X\text{'s} \qquad \textbf{[0.3 Ptos.]}\\
&= \text{Var}\left[\sum_{i=1}^{N} \text{E}\left(X_i\right)\right] + \text{E}\left[\sum_{i=1}^{N} \text{Var}\left(X_i\right)\right], \quad \text{por la independencia de } X_i \text{ con respecto a } N \qquad \textbf{[0.3 Ptos.]}\\
&= \text{Var}\left(N \cdot \mu\right) + \text{E}\left(N \cdot \sigma^2\right) \qquad \textbf{[0.3 Ptos.]}\\
&= \mu^2 \cdot \text{Var}\left(N\right) + \sigma^2 \cdot \text{E}\left(N\right) \qquad \textbf{[0.3 Ptos.]}\\
&= \mu^2 \cdot \sigma_N^2 + \sigma^2 \cdot \mu_N \qquad \textbf{[0.3 Ptos.]}\\
&= 120^2 \cdot 30 + 40^2 \cdot 30 \qquad \textbf{[0.3 Ptos.]}\\
&= 433600 \qquad \textbf{[0.3 Ptos.]}
\end{aligned}$$

Luego

$$P(T > 3000) = 1 - \Phi\left(\frac{3000 - 3600}{\sqrt{433600}}\right) = 1 - \Phi(-0{,}9111851) = \Phi(0{,}9111851) \approx 0{,}8186 \qquad \textbf{[0.5 Ptos.]}$$

+ 1 Punto Base

Problema 3 (Esperanza, Varianza y Covarianza)

Suponga que usted, en este momento, ha decidido reunirse con un conocido en un cervecería del barrio Bellavista, acordando que el que llegue tarde paga!. Usted sabe que el concurrirá en vehículo y de acuerdo a su distancia, congestión, disponibilidad de estacionamiento el tiempo de llegada se comporta como una distribución normal de media 55 min con una desv.estándar de 10 min. En cambio usted, como una forma de reducir y asegurar su desplazamiento ha decidido ir en metro, por tanto su viaje se divide en tres tramos: Colectivo - Metro Estación 1 - Metro Estación 2 (Baquedano) y finalmente caminata desde la estación Baquedano a la cervecería. Los tiempos de cada una de las etapas se pueden asumir normales con medias 15, 30 y 10 min, y desviación estándar 5, 2 y 3 min, respectivamente. Además, existe correlación entre el tiempo de viaje en colectivo y la caminata final y es de $r = -0{,}3$. ¿Cuál es la probabilidad que usted arribe primero?

Solución

Definamos como T_1 el tiempo de llegada del conocido e T_2 su tiempo de llegada, el cual se descompone en la suma de los tiempos de traslado de los tres tramos: X_1, X_2 y X_3.

Del enunciado se tiene que

$T_1 \sim \text{Normal}(55,\, 10); \quad X_1 \sim \text{Normal}(15,\, 5); \quad X_2 \sim \text{Normal}(30,\, 2); \quad X_3 \sim \text{Normal}(10,\, 3)$ **[0.5 Ptos.]**

$$\text{Corr}(X_1,\, X_3) = \frac{\text{Cov}(X_1,\, X_3)}{\sigma_{X_1} \cdot \sigma_{X_3}} = -0{,}3$$

$$\text{Cov}(X_1,\, X_3) = -4{,}5; \quad \text{Cov}(X_1,\, X_2) = 0; \qquad \text{Cov}(X_2,\, X_3) = 0 \qquad \textbf{[0.5 Ptos.]}$$

y

$$T_2 \sim \text{Normal}(\mu_{T_2},\, \sigma_{T_2}) \qquad \textbf{[0.5 Ptos.]}$$

donde

$$\mu_{T_2} = \text{E}(X_1 + X_2 + X_3) = \text{E}(X_1) + \text{E}(X_2) + \text{E}(X_3) = 15 + 30 + 10 = 55 \qquad \textbf{[1.0 Ptos.]}$$

$$\sigma^2_{T_2} = \text{Var}(X_1 + X_2 + X_3) = \sum_{i=1}^{3}\sum_{j=1}^{3} \text{Cov}(X_i,\, X_j) = 5^2 + 0 - 4{,}5 + 0 + 2^2 + 0 - 4{,}5 + 0 + 3^2 = 29 \qquad \textbf{[1.5 Ptos.]}$$

Se pide

$$P(T_2 < T_1) = P(T_2 - T_1 < 0) = P(Z < 0) \qquad \textbf{[0.5 Ptos.]}$$

con

$$Z \sim \text{Normal}(\mu_Z = 55 - 55,\, \sigma_Z = \sqrt{29 + 10^2}) \qquad \textbf{[0.5 Ptos.]}$$

Por lo tanto

$$P(Z < 0) = \Phi\left(\frac{0 - 0}{\sqrt{129}}\right) = \Phi(0) = \frac{1}{2} \qquad \textbf{[1.0 Ptos.]}$$

+ 1 Punto Base

Problema 4 (Bondad de Ajuste)

Un test para detectar contaminación en la leche, consiste en colocar 0.01 ml sobre una lamina de 1 cm^2. Un estudio consideró 400 muestras, cuyos resultados fueron los siguientes:

Nº de Bacterias por cm^2	0	1	2	3	4	5	6	7	8	9	10	19
Frecuencia	56	104	80	62	42	27	9	9	5	3	2	1

¿El modelo Poisson es razonable para ajustar el número de bacterias por cm^2? Por simplicidad considere al número de bacterias por cm^2 mayor o igual a 7 como un solo grupo.

Solución

Definamos como $X_1, \ldots, X_n$ al número de bacterias por cm^2 de una muestra de tamaño n.

Del enunciado se propone que $X_i \sim \text{Poisson}(\nu)$.

Recordemos que el estimador máximo verosímil del parámetro ν es $\hat{\nu} = \overline{X}_n = 2{,}44$. **[1.0 Ptos.]**

Luego las probabilidad estimadas para los 8 grupos las obtenemos a partir de

$$\hat{p}_x = \frac{\hat{\nu}^x \, e^{-\hat{\nu}}}{x!}, \quad x = 0, 1, 2, \ldots$$

mientras que los valores esperados son iguales a $n \cdot \hat{p}_x$.

Se pide contrastar las siguientes hipótesis:

$$\text{H}_0 : X \sim \text{Poisson} \quad \text{vs} \quad \text{H}_1 : X \nsim \text{Poisson}$$ **[0.5 Ptos.]**

Bajo el supuesto que H$_0$ es correcta, se tiene que

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(k - 1 - 1)$$ **[0.5 Ptos.]**

Luego, se rechaza el ajuste Poisson si $X^2 > c_{1-\alpha}(k - 2)$.

Reemplazando, tenemos que

```
x                  0       1       2      3      4      5      6     +7   suma
Observado     56.000 104.000  80.000 62.000 42.000 27.000  9.000 20.000 400.00
Probabilidad   0.087   0.213   0.259  0.211  0.129  0.063  0.026  0.012   1.00
Esperado      34.800  85.200 103.600 84.400 51.600 25.200 10.400  4.800 400.00
(O-E)^2/E     12.915   4.148   5.376  5.945  1.786  0.129  0.188 48.133  78.62
```

[3.0 Ptos.]

Para $k = 8$, se tiene que $X^2 = 78{,}62 > 18{,}55 = c_{0{,}995}(6)$ (valor-p $< 0{,}005$). **[0.5 Ptos.]**

Por lo tanto, existe suficiente evidencia para rechazar la hipótesis que los datos distribuyen Poisson. **[0.5 Ptos.]**

Nota: Si el alumno concluye con un nivel de significancia mayor a 0.5 % no descontar puntaje.

+ 1 Punto Base

Problema 5 (Regresión Simple)

El control de la contaminación del agua es muy relevante para la industria de la curtiembre, proceso que convierte las pieles de los animales en cuero, ya que los desechos del proceso son químicamente complejos, los cuales se caracterizan por los altos valores de demanda bioquímica de oxígeno, sólidos volátiles y otras medidas de contaminación. Un estudio consideró 33 muestras de desechos químicamente tratados, registrando lecturas de reducción porcentual X de sólidos totales y la reducción porcentual Y de la demanda química de oxígeno. Un modelo de regresión lineal arrojó un coeficiente de determinación R^2 igual a 12,5796 %. ¿Es este modelo significativo? Utilice un nivel α de significancia del 5 %.

Solución

Tenemos que

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT} \qquad \textbf{[0.5 Ptos.]}$$

y necesitamos obtener

$$\begin{aligned}
F &= \frac{SCR/k}{SCE/(n-k-1)} \qquad \textbf{[1.0 Ptos.]}\\
&= \frac{(n-k-1)}{k} \cdot \frac{SCR}{SCE} \qquad \textbf{[0.5 Ptos.]}\\
&= \frac{(n-k-1)}{k} \cdot \frac{SCR/SCT}{SCE/SCT} \qquad \textbf{[1.0 Ptos.]}\\
&= \frac{(n-k-1)}{k} \cdot \frac{R^2}{(1-R^2)} \qquad \textbf{[1.0 Ptos.]}\\
&= \frac{(33-1-1)}{1} \cdot \frac{0{,}125796}{(1-0{,}125796)} \qquad \textbf{[0.5 Ptos.]}\\
&= 4{,}460831 \qquad \textbf{[0.5 Ptos.]}
\end{aligned}$$

Como $F = 4{,}460831 > 4{,}160 = F_{0{,}95}(1,\, 31)$, **[0.5 Ptos.]** existe suficiente evidencia para rechazar la hipótesis que la regresión es no significativa, es decir, se apoya la significancia del modelo. **[0.5 Ptos.]**

+ 1 Punto Base

Problema 6 (Regresión Multiple)

Interesa modelar la permeabilidad de un material (Y) em función de la densidad (X_1), grosor (X_2) y la cantidad de aglomerante (X_3). A partir de una muestra de 24 ensayos, se ajustaron los siguientes modelos:

Modelo	SCE	SCT
Nulo	220	220
X_1	180	220
X_2	190	220
X_1, X_2, X_3	120	220

(a) Es significativo el aporte de las variables en términos individuales.

(b) Construya una clase jerárquica y mediante el método de selección Backward (hacía atrás) encuentre el mejor modelo.

Use un nivel de significancia del 5 %.

Solución

(a) Solo es posible cuantificar la significancia del aporte individual de X_1 y X_2.

$$\mathrm{H}_0 : \beta_1 = 0 \quad \text{vs} \quad \mathrm{H}_1 : \beta_1 \neq 0$$

$$F = \frac{SCR/1}{SCE/(n-1-1)} = (n-2) \cdot \frac{(SCT - SCE)}{SCE} = 22 \cdot \frac{(220-180)}{180} = 4{,}888889 \qquad \textbf{[0.5 Ptos.]}$$

Como $F = 4{,}888889 > 4{,}301 = F_{0{,}95}(1,\, 22)$ **[0.5 Ptos.]**, existe suficiente evidencia para rechazar la hipótesis que la regresión es no significativa, es decir, se apoya que el aporte de X_1 es significativo. **[0.5 Ptos.]**

$$\mathrm{H}_0 : \beta_2 = 0 \quad \text{vs} \quad \mathrm{H}_1 : \beta_2 \neq 0$$

$$F = \frac{SCR/1}{SCE/(n-1-1)} = (n-2) \cdot \frac{(SCT - SCE)}{SCE} = 22 \cdot \frac{(220-190)}{190} = 3{,}473684 \qquad \textbf{[0.5 Ptos.]}$$

Como $F = 3{,}473684 < 4{,}301 = F_{0{,}95}(1,\, 22)$ **[0.5 Ptos.]**, no existe suficiente evidencia para rechazar la hipótesis que la regresión es no significativa, es decir, el aporte de X_2 es no significativo. **[0.5 Ptos.]**

(b) **Alternativa 1**: Para la clase jerárquica (X_1, X_2, X_3), (X_2) y (NULO).

Probaremos primero si el aporte de las variables (X_1, X_3) en presencia de X_2 es significativo, y luego decidir si salen o no del modelo completo.

$$\mathrm{H}_0 : \beta_1 = \beta_3 = 0 \quad \text{vs} \quad \mathrm{H}_1 : \text{Al menos uno es} \neq 0 \qquad \textbf{[0.5 Ptos.]}$$

$$F = \frac{\left(SCE_{(m)} - SCE\right)/m}{SCE/(n-k-1)} = F = \frac{(190-120)\,/2}{120/(24-3-1)} = 5{,}833333 \qquad \textbf{[1.5 Ptos.]}$$

Como $F = 5{,}833333 > 3{,}493 = F_{0{,}95}(2,\, 20)$ **[0.5 Ptos.]**, existe suficiente evidencia para rechazar la hipótesis que el aporte conjunto de (X_1, X_3) es no significativa, es decir, se apoya el modelo (X_1, X_2, X_3). **[0.5 Ptos.]**

Alternativa 2: Para la clase jerárquica (X_1, X_2, X_3), (X_1) y (NULO)

Probaremos primero si el aporte de las variables (X_2, X_3) en presencia de X_1 es significativo, y luego decidir si salen o no del modelo completo.

$$\mathrm{H}_0 : \beta_2 = \beta_3 = 0 \quad \text{vs} \quad \mathrm{H}_1 : \text{Al menos uno es} \neq 0 \qquad \textbf{[0.5 Ptos.]}$$

$$F = \frac{\left(SCE_{(m)} - SCE\right)/m}{SCE/(n-k-1)} = F = \frac{(180-120)/2}{120/(24-3-1)} = 5 \qquad \textbf{[1.5 Ptos.]}$$

Como $F = 5 > 3{,}493 = F_{0{,}95}(2,\, 20)$ **[0.5 Ptos.]**, existe suficiente evidencia para rechazar la hipótesis que el aporte conjunto de (X_2, X_3) es no significativa, es decir, se apoya el modelo $(X_1,\, X_2,\, X_3)$. **[0.5 Ptos.]**

+ 1 Punto Base

Formulario

Propiedades función $\Gamma(\cdot)$

$$(1)\quad \Gamma(k)=\int_0^\infty u^{k-1}\,e^{-u}\,du;\quad (2)\quad \Gamma(a+1)=a\,\Gamma(a);$$

$$(3)\quad \Gamma(n+1)=n!,\quad \text{si } n\in\mathbb{N}_0;\quad (4)\quad \Gamma(1/2)=\sqrt{\pi}$$

Propiedades función $B(\cdot,\,\cdot)$

$$(1)\quad B(q,\,r)=\int_0^1 x^{q-1}\,(1-x)^{r-1}\,dx;\quad (2)\quad B(q,\,r)=\frac{\Gamma(q)\,\Gamma(r)}{\Gamma(q+r)}$$

Igualdades

$$(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^k\,b^{n-k},\qquad \sum_{k=x}^{\infty}\phi^k=\frac{\phi^x}{1-\phi}\quad \text{si } |\phi|<1,\qquad \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\exp(\lambda)$$

Propiedad distribución Gamma

$$\text{Si } T\sim\text{Gamma}(k,\,\nu),\text{ con } k\in\mathbb{N}\longrightarrow F_T(t)=1-\sum_{x=0}^{k-1}\frac{(\nu\,t)^x\,e^{-\nu\,t}}{x!}$$

Transformación

Sea $Y=g(X)$ una función cualquiera, con k raíces:

$$f_Y(y)=\sum_{i=1}^{k}f_X\left(g_i^{-1}(y)\right)\cdot\left|\frac{d}{dy}\,g_i^{-1}(y)\right|$$

$$p_Y(y)=\sum_{i=1}^{k}p_X\left(g_i^{-1}(y)\right)$$

Sea $Z=g(X,Y)$ una función cualquiera:

$$p_Z(z)=\sum_{g(x,y)=z}p_{X,Y}(x,y)$$

Sea $Z=g(X,Y)$ una función invertible para X o Y fijo:

$$\begin{aligned} f_Z(z)&=\int_{-\infty}^{\infty}f_{X,Y}(g^{-1},y)\left|\frac{\partial}{\partial z}g^{-1}\right|dy\\ &=\int_{-\infty}^{\infty}f_{X,Y}(x,g^{-1})\left|\frac{\partial}{\partial z}g^{-1}\right|dx\end{aligned}$$

Esperanza y Varianza Condicional

$$\mathrm{E}(Y)=\mathrm{E}[\mathrm{E}(Y\,|\,X)]\quad \text{y}\quad \mathrm{Var}(Y)=\mathrm{Var}[\mathrm{E}(Y\,|\,X)]+\mathrm{E}[\mathrm{Var}(Y\,|\,X)]$$

Teorema del Límite Central

Sean $X_1,\,\ldots,\,X_n$ variables aleatorias independientes e idénticamente distribuidas, entonces

$$Z_n=\frac{\displaystyle\sum_{i=1}^{n}X_i-n\cdot\mu}{\sqrt{n}\,\sigma}=\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\longrightarrow Z\sim\text{Normal}(0,1),$$

cuando $n\to\infty$, $\mathrm{E}(X_i)=\mu$ y $\mathrm{Var}(X_i)=\sigma^2$.

Propiedades Esperanza, Varianza y Covarianza

Sean X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_m variables aleatorias y a_0, a_1, ..., a_n, b_0, b_1, ..., b_m constantes conocidas.

- $\mathrm{E}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i\right) = a_0 + \sum_{i=1}^{n} a_i \cdot \mathrm{E}(X_i)$.
- $\mathrm{Cov}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i,\, b_0 + \sum_{j=1}^{m} b_j \cdot Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i \cdot b_j \cdot \mathrm{Cov}\,(X_i, Y_j)$.
- $\mathrm{Var}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \cdot a_j \cdot \mathrm{Cov}\,(X_i, X_j)$.
- Si X_1, ..., X_n son variables aleatorias independientes, entonces

$$\mathrm{Var}\left(a_0 + \sum_{i=1}^{n} a_i \cdot X_i\right) = \sum_{i=1}^{n} a_i^2 \cdot \mathrm{Var}\,(X_i)$$

Mínimo y Máximo

Sean X_1, ..., X_n variables aleatorias continuas independientes con idéntica distribución (f_X y F_X), entonces para:

$$Y_1 = \text{mín}\{X_1, \ldots, X_n\} \longrightarrow f_{Y_1} = n\,[1 - F_X(y)]^{n-1}\, f_X(y) Y_n \quad = \text{máx}\{X_1, \ldots, X_n\} \longrightarrow f_{Y_n} = n\,[F_X(y)]^{n-1}\, f_X(y)$$

Mientras que la distribución conjunta entre Y_1 e Y_n está dada por:

$$f_{Y_1,Y_n}(u, v) = n\,(n-1)\,[F_X(v) - F_X(u)]^{n-2}\, f_X(v)\, f_X(u), \qquad u \leq v$$

Estimador Máximo Verosímil

Sea X_1,...,X_n una muestra aleatoria independiente e idénticamente distribuida con función de probabilidad p_X o de densidad f_X, determinada por un parámetro θ. Si $\hat{\theta}$ es el estimador máximo verosímil del parámetro θ, entonces:

- $\mathrm{E}(\hat{\theta}) \to \theta$, cuando $n \to \infty$.
- $\mathrm{Var}(\hat{\theta}) = \dfrac{1}{I_n(\theta)}$, con $I_n(\theta) = -\mathrm{E}\left[\dfrac{\partial^2}{\partial\,\theta^2}\,\ln L(\theta)\right]$.
- $\hat{\theta} \,\dot{\sim}\, \text{Normal}\left(\theta,\, \sqrt{\dfrac{1}{I_n(\theta)}}\right)$, cuando $n \to \infty$.
- El estimador máximo verosímil de $g(\theta)$ es $g(\hat{\theta})$, cuya varianza está dada por: $\mathrm{Var}[g(\hat{\theta})] = \dfrac{[g'(\theta)]^2}{I_n(\theta)}$.

Error Cuadrático Medio

El error cuadrático medio de un estimador $\hat{\theta}$ de θ se define como:

$$\mathrm{ECM}(\hat{\theta}) = \mathrm{E}\left\{\left[\hat{\theta} - \mathrm{E}(\hat{\theta})\right]^2\right\} = \mathrm{Var}(\hat{\theta}) + \text{Sesgo}^2$$

Distribuciones Muestrales

Sean X_1,...,X_n variables aleatorias independientes e idénticamente distribuidas Normal(μ, σ), entonces

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1), \qquad \frac{\overline{X}_n - \mu}{s/\sqrt{n}} \sim \text{t-student}(n-1), \qquad \frac{s^2\,(n-1)}{\sigma^2} \sim \chi^2_{n-1}$$

con $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

Regresión Lineal Simple

Para el modelo de regresión lineal simple $Y = \alpha + \beta\, x$, se tiene que

$$\hat{\alpha} = \overline{y} - \hat{\beta}\,\overline{x}, \quad \hat{\beta} = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2}, \quad r^2 = 1 - \frac{s^2_{Y|x}}{s^2_Y}, \qquad s^2_{Y|x} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - y_i')^2$$

$$\langle \mu_{Y\,|x_i} \rangle_{1-\alpha} = \overline{y}_i \pm t_{(1-\alpha/2),\,n-2} \cdot s_{Y|x}\sqrt{\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\displaystyle\sum_{j=1}^{n}(x_j - \overline{x})^2}}$$

Regresión Lineal Múltiple

Para el siguiente modelo de regresión

$$y_i = \beta_0 + \beta_1 \cdot x_{i\,1} + \cdots \beta_k \cdot x_{i\,k} + \varepsilon_i$$

con $\mathrm{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$ y $\mathrm{E}(\varepsilon_i \cdot \varepsilon_j) = 0$ para $i \neq j$, se define el modelo de regresión multiple estimado como

$$\hat{y}_i = b_0 + b_1 \cdot x_{i\,1} + \cdots b_k \cdot x_{i\,k}$$

Descomposición de cuadrados

$$SCT = SCR + SCE$$

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Coeficiente de determinación

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT} = 1 - \frac{(n-k-1)}{(n-1)}\,\frac{s^2_{Y\,|\,x}}{s^2_Y}$$

Coeficiente de determinación ajustado

$$r^2 = 1 - \frac{(n-1)}{(n-k-1)}\,\frac{SCE}{SCT} = 1 - \frac{s^2_{Y\,|\,x}}{s^2_Y}$$

Distribuciones

$$T_{b_j} = \frac{b_j - \beta_j}{s_{b_j}} \sim \text{t-Student}(n-k-1) \qquad F = \frac{SCR/k}{SCE/(n-k-1)} \sim F(k,\, n-k-1)$$

$$F = \frac{\left(SCE_{(m)} - SCE\right)/m}{SCE/(n-k-1)} \sim F(m,\, n-k-1)$$

donde $SCE_{(m)}$ y SCE son suma de errores al cuadrado de dos modelos anidados, con $0 < m \leq k$.

Aproximación de Momentos

Sean $X_1, \ldots, X_n$ variables aleatorias con valores esperados $\mu_{X_1}, \ldots, \mu_{X_n}$ y varianzas $\sigma^2_{X_1}, \ldots, \sigma^2_{X_n}$ e Y una función de ellas.

Aproximación de primer orden:

$$Y \approx g(\mu_{X_1}, \ldots, \mu_{X_n}) + \sum_{i=1}^{n}(X_i - \mu_{X_i})\,\frac{\partial}{\partial X_i} g(\mu_{X_1}, \ldots, \mu_{X_n})$$

$$\mathrm{E}(Y) \approx g(\mu_{X_1}, \ldots, \mu_{X_n})$$

$$\mathrm{Var}(Y) \approx \sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{ij}\,\sigma_{X_i}\,\sigma_{X_j}\,\left[\frac{\partial}{\partial X_i}g(\mu_{X_1}, \ldots, \mu_{X_n}) \cdot \frac{\partial}{\partial X_j}g(\mu_{X_1}, \ldots, \mu_{X_n})\right]$$

Distribución	Densidad de Probabilidad	Esperanza y Varianza
Binomial	$\binom{n}{x}\, p^x\, (1-p)^{n-x}$	$\mu_X = n\,p$ $\sigma_X^2 = n\,p\,(1-p)$ $M(t) = [p\,e^t + (1-p)]^n, \quad t \in \mathbb{R}$
Geométrica	$p\,(1-p)^{x-1}$	$\mu_X = 1/p$ $\sigma_X^2 = (1-p)/p^2$ $M(t) = p\,e^t/[1-(1-p)\,e^t], \quad t < -\ln(1-p)$
Binomial-Negativa	$\binom{x-1}{r-1}\, p^r\, (1-p)^{x-r}$	$\mu_X = r/p$ $\sigma_X^2 = r\,(1-p)/p^2$ $M(t) = \left\{p\,e^t/[1-(1-p)\,e^t]\right\}^r, \quad t < -\ln(1-p)$
Poisson	$\dfrac{(\nu\,t)^x\, e^{-\nu\,t}}{x!}$	$\mu_X = \nu\,t$ $\sigma_X^2 = \nu\,t$ $M(t) = \exp\left[\lambda\left(e^t - 1\right)\right], \quad t \in \mathbb{R}$
Exponencial	$\nu\, e^{-\nu\,x}$	$\mu_X = 1/\nu$ $\sigma_X^2 = 1/\nu^2$ $M(t) = \nu/(\nu - t), \quad t < \nu$
Gamma	$\dfrac{\nu^k}{\Gamma(k)}\, x^{k-1}\, e^{-\nu\,x}$	$\mu_X = k/\nu$ $\sigma_X^2 = k/\nu^2$ $M(t) = [\nu/(\nu - t)]^k, \quad t < \nu$
Normal	$\dfrac{1}{\sqrt{2\,\pi}\,\sigma}\, \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$	$\mu_X = \mu$ $\sigma_X^2 = \sigma^2$ $M(t) = \exp(\mu\,t + \sigma^2\,t^2/2), \quad t \in \mathbb{R}$
Log-Normal	$\dfrac{1}{\sqrt{2\,\pi}\,(\zeta\,x)}\, \exp\left[-\dfrac{1}{2}\left(\dfrac{\ln x - \lambda}{\zeta}\right)^2\right]$	$\mu_X = \exp\left(\lambda + \dfrac{1}{2}\,\zeta^2\right)$ $\sigma_X^2 = \mu_X^2\left(e^{\zeta^2} - 1\right)$ $E(X^r) = e^{r\,\lambda}\, M_Z(r\,\zeta)$, con $Z \sim$Normal(0,1)
Uniforme	$\dfrac{1}{(b-a)}$	$\mu_X = (a+b)/2$ $\sigma_X^2 = (b-a)^2/12$ $M(t) = [e^{t\,b} - e^{t\,a}]/[t\,(b-a)], \quad t \in \mathbb{R}$
Beta	$\dfrac{1}{B(q,\, r)}\, \dfrac{(x-a)^{q-1}\,(b-x)^{r-1}}{(b-a)^{q+r-1}}$	$\mu_X = a + \frac{q}{q+r}\,(b-a)$ $\sigma_X^2 = \frac{q\,r\,(b-a)^2}{(q+r)^2\,(q+r+1)}$
Hipergeométrica	$\dfrac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$\mu_X = n\,\frac{m}{N}$ $\sigma_X^2 = \left(\frac{N-n}{N-1}\right)\, n\,\frac{m}{N}\,\left(1 - \frac{m}{N}\right)$

Tablas de Percentiles p

Distribución Normal Estándar k_p

k_p	0,00	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

Distribución t-student $t_p(\nu)$

ν	$t_{0,90}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$
1	3,078	6,314	12,706	31,821
2	1,886	2,920	4,303	6,965
3	1,638	2,353	3,182	4,541
4	1,533	2,132	2,776	3,747
5	1,476	2,015	2,571	3,365
6	1,440	1,943	2,447	3,143
7	1,415	1,895	2,365	2,998
8	1,397	1,860	2,306	2,896
9	1,383	1,833	2,262	2,821
10	1,372	1,812	2,228	2,764
11	1,363	1,796	2,201	2,718
12	1,356	1,782	2,179	2,681
13	1,350	1,771	2,160	2,650
14	1,345	1,761	2,145	2,624
15	1,341	1,753	2,131	2,602
16	1,337	1,746	2,120	2,583
17	1,333	1,740	2,110	2,567
18	1,330	1,734	2,101	2,552
19	1,328	1,729	2,093	2,539
20	1,325	1,725	2,086	2,528
21	1,323	1,721	2,080	2,518
22	1,321	1,717	2,074	2,508
23	1,319	1,714	2,069	2,500
24	1,318	1,711	2,064	2,492
25	1,316	1,708	2,060	2,485
26	1,315	1,706	2,056	2,479
27	1,314	1,703	2,052	2,473
28	1,313	1,701	2,048	2,467
29	1,311	1,699	2,045	2,462
30	1,310	1,697	2,042	2,457
∞	1,282	1,645	1,960	2,326

Distribución Chi-Cuadrado $c_p(\nu)$

ν	$c_{0,025}$	$c_{0,05}$	$c_{0,10}$	$c_{0,90}$	$c_{0,95}$	$c_{0,975}$	$c_{0,99}$	$c_{0,995}$
1	0,00	0,00	0,02	2,71	3,84	5,02	6,63	7,88
2	0,05	0,10	0,21	4,61	5,99	7,38	9,21	10,60
3	0,22	0,35	0,58	6,25	7,81	9,35	11,34	12,84
4	0,48	0,71	1,06	7,78	9,49	11,14	13,28	14,86
5	0,83	1,15	1,61	9,24	11,07	12,83	15,09	16,75
6	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	1,69	2,17	2,83	12,02	14,07	16,01	18,48	20,28
8	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,95
9	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
10	3,25	3,94	4,87	15,99	18,31	20,48	23,21	25,19
11	3,82	4,57	5,58	17,28	19,68	21,92	24,72	26,76
12	4,40	5,23	6,30	18,55	21,03	23,34	26,22	28,30
13	5,01	5,89	7,04	19,81	22,36	24,74	27,69	29,82
14	5,63	6,57	7,79	21,06	23,68	26,12	29,14	31,32
15	6,26	7,26	8,55	22,31	25,00	27,49	30,58	32,80
16	6,91	7,96	9,31	23,54	26,30	28,85	32,00	34,27
17	7,56	8,67	10,09	24,77	27,59	30,19	33,41	35,72
18	8,23	9,39	10,86	25,99	28,87	31,53	34,81	37,16
19	8,91	10,12	11,65	27,20	30,14	32,85	36,19	38,58
20	9,59	10,85	12,44	28,41	31,41	34,17	37,57	40,00
21	10,28	11,59	13,24	29,62	32,67	35,48	38,93	41,40
22	10,98	12,34	14,04	30,81	33,92	36,78	40,29	42,80
23	11,69	13,09	14,85	32,01	35,17	38,08	41,64	44,18
24	12,40	13,85	15,66	33,20	36,42	39,36	42,98	45,56
25	13,12	14,61	16,47	34,38	37,65	40,65	44,31	46,93

Tabla de Percentil 95 % $F_{\mathtt{df1,df2}}$

	df2 = 20	df2 = 21	df2 = 22	df2 = 23	df2 = 24	df2 = 25	df2 = 26	df2 = 27	df2 = 28	df2 = 29	df2 = 30	df2 = 31	df2 = 32	df2 = 33
df1 = 1	4,351	4,325	4,301	4,279	4,260	4,242	4,225	4,210	4,196	4,183	4,171	4,160	4,149	4,139
df1 = 2	3,493	3,467	3,443	3,422	3,403	3,385	3,369	3,354	3,340	3,328	3,316	3,305	3,295	3,285
df1 = 3	3,098	3,072	3,049	3,028	3,009	2,991	2,975	2,960	2,947	2,934	2,922	2,911	2,901	2,892
df1 = 4	2,866	2,840	2,817	2,796	2,776	2,759	2,743	2,728	2,714	2,701	2,690	2,679	2,668	2,659
df1 = 5	2,711	2,685	2,661	2,640	2,621	2,603	2,587	2,572	2,558	2,545	2,534	2,523	2,512	2,503
df1 = 6	2,599	2,573	2,549	2,528	2,508	2,490	2,474	2,459	2,445	2,432	2,421	2,409	2,399	2,389
df1 = 7	2,514	2,488	2,464	2,442	2,423	2,405	2,388	2,373	2,359	2,346	2,334	2,323	2,313	2,303
df1 = 8	2,447	2,420	2,397	2,375	2,355	2,337	2,321	2,305	2,291	2,278	2,266	2,255	2,244	2,235
df1 = 9	2,393	2,366	2,342	2,320	2,300	2,282	2,265	2,250	2,236	2,223	2,211	2,199	2,189	2,179
df1 = 10	2,348	2,321	2,297	2,275	2,255	2,236	2,220	2,204	2,190	2,177	2,165	2,153	2,142	2,133