Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2022

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

PAUTA EXAMEN

Pregunta 1

Un grupo de n electores, indecisos e independientes, conforman un panel de un programa en el cual se entrevistan a representantes de las dos opciones que tendrá el plebiscito de septiembre. Al final del programa, los electores del panel optarán por una de las dos opciones: apruebo o rechazo.

Suponga que para el representante de un sector, la probabilidad que los electores que se encuentran en el panel, decidan votar por la opción que él promueve, se comporta de manera aleatoria de acuerdo a una distribución Beta$(2, 1)$. Determine la función de probabilidad marginal del número de personas del panel que votaran por la opción que promueve dicho representante.

Solución

Definamos como Y al número de electores que votará por la opción del representante y X a la probabilidad a que los electores decidan votar por él.

[0.1 Ptos.] $Y \mid X = x \sim \text{Binomial}(n, p = x)$ y $X \sim \text{Beta}(2, 1)$ **[0.1 Ptos.]**

La marginal de Y está dada por el teorema de probabilidades totales:

$$
\begin{aligned}
p_Y(y) &= \int_{-\infty}^{+\infty} p_{Y\mid X=x}(y) \cdot f_X(x)dx \quad \textbf{[0.1 Ptos.]} \\
&= \int_0^1 \binom{n}{y} x^y(1-x)^{n-y} \cdot \frac{1}{B(2,1)} \frac{(x-0)^{2-1}(1-x)^{1-1}}{1^{2+1-1}} dx \quad \textbf{[0.1 Ptos.]} \\
&= \binom{n}{y} \cdot \frac{B(2+y, n-y+1)}{B(2,1)} \int_0^1 \frac{1}{B(2+y, n-y+1)} \frac{(x-0)^{(2+y)-1}(1-x)^{(n-y+1)-1}}{1^{(2+y)+(n-y+1)-1}} dx \quad \textbf{[0.2 Ptos.]} \\
&= \frac{n!}{y!(n-y)!} \cdot \frac{B(2+y, n-y+1)}{B(2,1)} \cdot 1, \quad \text{por área bajo la densidad Beta sobre todo su soporte} \quad \textbf{[0.1 Ptos.]} \\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \cdot \frac{\Gamma(y+2)\Gamma(n-y+1)}{\Gamma(n+3)} \cdot \frac{\Gamma(3)}{\Gamma(2)\Gamma(1)} \quad \textbf{[0.1 Ptos.]} \\
&= \frac{2(y+1)}{(n+2)(n+1)}, \quad \textbf{[0.1 Ptos.]} \quad y = 0, 1, \ldots, n \quad \textbf{[0.1 Ptos.]}
\end{aligned}
$$

Pregunta 2

Suponga que en una zona urbana, el número de fallas en el tendido eléctrico que afecta a clientes durante un día de lluvia se comporta como una variable aleatoria Poisson (λ) y que el número de clientes afectados por x fallas se distribuye Poisson (ν^x). ¿Determine el número esperado de clientes se verán afectados el próximo día que llueva?

Solución

Definamos como X al número de fallas en el tendido eléctrico que afecta a clientes durante un día de lluvia e Y al número de clientes afectados.

Del enunciado se tiene que

$$\textbf{[0.1 Ptos.]} \quad X \sim \text{Poisson}(\lambda) \qquad Y \mid X = x \sim \text{Poisson}(\nu^x) \quad \textbf{[0.1 Ptos.]}$$

Se pide $\mathsf{E}(Y)$.

Aplicando teorema de esperanzas iteradas se tiene que

$$\mathsf{E}(Y) = \mathsf{E}[\mathsf{E}(Y \mid X)] = \mathsf{E}\left[\nu^X\right] \quad \textbf{[0.3 Ptos.]}$$

Alternativa 1: Por definición

$$\begin{aligned}
\mathsf{E}\left[\nu^X\right] &= \sum_{x=0}^{\infty} \nu^x \cdot \frac{\lambda^x\, e^{-\lambda}}{x!} \quad \textbf{[0.1 Ptos.]} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\nu \cdot \lambda)^x}{x!} \quad \textbf{[0.2 Ptos.]} \\
&= e^{-\lambda} \cdot e^{\nu \cdot \lambda} \quad \textbf{[0.1 Ptos.]} \\
&= e^{\lambda\,(\nu - 1)} \quad \textbf{[0.1 Ptos.]}
\end{aligned}$$

Alternativa 2: Usando la f.g.m de una Poisson

$$\begin{aligned}
\mathsf{E}\left[\nu^X\right] &= \mathsf{E}\left[e^{X\,\ln(\nu)}\right] \quad \textbf{[0.1 Ptos.]} \\
&= M_X[\ln(\nu)] \quad \textbf{[0.2 Ptos.]} \\
&= e^{\lambda\,\left(e^{\ln(\nu)} - 1\right)} \quad \textbf{[0.1 Ptos.]} \\
&= e^{\lambda\,(\nu - 1)} \quad \textbf{[0.1 Ptos.]}
\end{aligned}$$

Pregunta 3

Para una encuesta política, se cuenta con n encuestadores para el terreno. Ellos contactarán a cien personas durante la jornada de manera independiente. Algunos de los contactados accederán a responder la encuesta y otros no. Para los que no contesten, de todas maneras el encuestador tratara en lo posible caracterizar a la persona registrando por ejemplo su sexo, rango de edad y nivel socioeconómico. Suponga que p corresponde a probabilidad que una persona no acceda a responder la encuesta, y por lo tanto el encuestador deberá, en lo posible, obtener la mayor cantidad de información para poder caracterizarlo posteriormente. Obtenga el estimador de momentos para p y determine, justificadamente, la distribución aproximada de este estimador.

Solución

Alternativa 1: Sea $X_1, \ldots, X_n$ el número de personas que no acceden a contestar la encuesta a lo encuestadores, los cuales se pueden considerar como una secuencia iid Binomial$(100,\, p)$.

El estimador de momento se obtiene igualando el primer momento teórico al empírico:

$$\textbf{[0.2 Ptos.]} \quad 100\,p = \overline{X}_n \to \hat{p} = \frac{\overline{X}_n}{100} \quad \textbf{[0.2 Ptos.]}$$

Por teorema del limite central se tiene:

$$\overline{X}_n \overset{\text{aprox}}{\sim} \text{Normal}\left(100\,p,\, \sqrt{\frac{100\,p\,(1-p)}{n}}\right) \quad \textbf{[0.3 Ptos.]}$$

y aplicando teorema de transformación:

$$\hat{p} = \frac{\overline{X}_n}{100} \overset{\text{aprox}}{\sim} \text{Normal}\left(p,\, \sqrt{\frac{p\,(1-p)}{100\,n}}\right) \quad \textbf{[0.3 Ptos.]}$$

Alternativa 2: Sea $X_1, \ldots, X_n$ el número de personas que no acceden a contestar la encuesta a lo encuestadores, los cuales se pueden considerar como una secuencia iid Hipergeométrica$(100, N,\, N \cdot p)$, con N la población de electores que según SERVEL son más de 15 millones.

El estimador de momento se obtiene igualando el primer momento teórico al empírico:

$$\textbf{[0.2 Ptos.]} \quad 100 \cdot \frac{N \cdot p}{N} = \overline{X}_n \to \hat{p} = \frac{\overline{X}_n}{100} \quad \textbf{[0.2 Ptos.]}$$

Por teorema del limite central se tiene:

$$\overline{X}_n \overset{\text{aprox}}{\sim} \text{Normal}\left(100\,p,\, \sqrt{\frac{\left(\frac{N-100}{N-1}\right) 100\,p\,(1-p)}{n}}\right) \quad \textbf{[0.3 Ptos.]}$$

y aplicando teorema de transformación:

$$\hat{p} = \frac{\overline{X}_n}{100} \overset{\text{aprox}}{\sim} \text{Normal}\left(p,\, \sqrt{\frac{\left(\frac{N-100}{N-1}\right) p\,(1-p)}{100\,n}}\right) \quad \textbf{[0.3 Ptos.]}$$

Pregunta 4

Un estudio respecto a la efectividad de la presencialidad en el aprendizaje fue llevado a cabo. Para esto, se seleccionaron dos grupos de estudiantes que realizaron un mismo curso, pero uno en formato presencial y resto en línea. La siguiente tabla resume el rendimiento final de los alumnos en cada uno de los formatos:

	Presencial	En línea
N° Alumnos	74	58
N° Aprobados	52	33

¿Existe suficiente evidencia al 10 % de significancia para afirmar que, en el formato en línea, se observa una mayor tasa de reprobación? Establezca hipótesis y responda usando valor-p.

Solución

Definamos como X al modo presencial e Y en línea.

Se pide contrastar las siguientes hipótesis

$$H_0 : p_X = p_Y \quad \text{vs} \quad H_a : p_X < p_Y, \qquad \textbf{[0.2 Ptos.]}$$

con p_X y p_Y la proporción de reprobados en ambos formatos.

Bajo H_0 se tiene que $p_X = p_Y = p$, por lo tanto

$$Z_0 = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}\,(1-\hat{p})} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \overset{\text{aprox}}{\sim} \text{Normal}(0, 1), \qquad \textbf{[0.2 Ptos.]}$$

con $\hat{p} = \dfrac{n \cdot \hat{p}_X + m \cdot \hat{p}_Y}{n+m}$. **[0.1 Ptos.]**

Del enunciado se tiene que

$$\textbf{[0.1 Ptos.]} \quad n = 74, \quad m = 58, \quad \hat{p}_X = \frac{22}{n}, \quad \hat{p}_Y = \frac{25}{m} \to \hat{p} = 0.3560606 \quad \textbf{[0.1 Ptos.]}$$

Reemplazando

$$\textbf{[0.1 Ptos.]} \quad Z_0 = -1.592614 \to \text{valor-p} \approx 1 - \Phi(1.59) = 1 - 0.9441 = 0.0559 \quad \textbf{[0.1 Ptos.]}$$

Por lo tanto, existe suficiente evidencia al 10 % de significancia para rechazar H_0 y apoyar la afirmación de que en el formato en linea la tasa de reprobación es mayor al modo presencial. **[0.1 Ptos.]**

Nota: Si el alumno define p_X y p_Y como la proporción de aprobados en ambos formatos, H_a cambia de dirección, Z_0 es igual a 1.592614, el valor p se mantiene y la conclusión sería la misma.

Pregunta 5

En poco más de dos meses, los mayores de 18 años tendrán que pronunciarse sobre si aprueban o rechazan la propuesta que la convención constitucional somete para convertirse en la nueva constitución de Chile. Una consulta a 200 personas muestra que 60 % aún no decide su voto.

(a) ¿Cuál es el error de estimación bajo criterio de varianza máxima, al 90 % de confianza, de la proporción de personas que aún no deciden su voto?

Usted ahora quiere hace un nuevo estudio con mayor precisión, tal que el error de estimación calculado en (a) disminuya en un 50 %.

(b) Considerando que el 60 % de los consultados anteriormente aún no deciden su voto y manteniendo un nivel de confianza del 90 %, ¿cuál sería el tamaño muestra propuesto?

Solución

(a) Se pide

$$e.e. = k_{1-\alpha/2} \cdot \sqrt{\frac{0.5 \cdot (1 - 0.5)}{n}} \quad \textbf{[0.2 Ptos.]}$$

Del enunciado se tiene que

$$\alpha = 0.10 \to k_{0.95} = 1.645 \quad \textbf{[0.1 Ptos.]}, \quad n = 200 \quad \textbf{[0.1 Ptos.]},$$

Reemplazando

$$e.e. = \frac{1.645}{2 \cdot \sqrt{200}} = 0.05815953 \quad \textbf{[0.1 Ptos.]}$$

(b) Se pide

$$n = \left(\frac{k_{1-\alpha/2} \cdot \sqrt{p \cdot (1 - p)}}{e.e./2} \right)^2 \quad \textbf{[0.2 Ptos.]}$$

Del enunciado se tiene que

$$\alpha = 0.10 \to k_{0.95} = 1.645 \quad \textbf{[0.1 Ptos.]}, \quad p = 0.6 \quad \textbf{[0.1 Ptos.]},$$

Reemplazando

$$n = \left(\frac{1.645 \cdot \sqrt{0.6 \cdot (1 - 0.6)}}{e.e./2} \right)^2 = 768 \quad \textbf{[0.1 Ptos.]}$$

Pregunta 6

Desde la estación de monitoreo de Parque O'higgins se descargó la temperatura registrada al mediodía durante el último mes. A partir de gráficos de probabilidad y test de bondad de ajuste χ^2, se desea evaluar el ajuste de los siguientes modelos: Normal, Log-Normal y Weibull.

Usando la siguiente información:

```
Tabla de Frecuencia:
(-Inf,10]   (10,12]   (12,14]   (14,16] (16, Inf]
        4         8         7         6         6

                          QQ-Normal QQ-LogNormal QQ-Weibull
Intercepto                13.130912    2.5433260  2.6656586
Pendiente                  3.515223    0.2791428  0.2277532
chisq.test(...)$statistic  1.407988    1.1006010  1.6009058
```

Estime los parámetros de estos modelos y determine justificadamente el mejor ajuste. Para el mejor ajuste, ¿cuál es el rango, más preciso posible, del valor-p asociado al estadístico χ^2?

Solución

Estimación QQ Normal:

$$x_p = \mu + \sigma \cdot \Phi^{-1}(p) \rightarrow \quad \textbf{[0.1 Ptos.]} \quad \hat{\mu} = 13.130912 \quad \text{y} \quad \hat{\sigma} = 3.515223 \quad \textbf{[0.1 Ptos.]}$$

Estimación QQ Log-Normal:

$$\ln(x_p) = \lambda + \zeta \cdot \Phi^{-1}(p) \rightarrow \quad \textbf{[0.1 Ptos.]} \quad \hat{\lambda} = 2.5433260 \quad \text{y} \quad \hat{\zeta} = 0.2791428 \quad \textbf{[0.1 Ptos.]}$$

Estimación QQ Weibull:

$$\ln(x_p) = \ln(\eta) + \frac{1}{\beta} \cdot \Phi^{-1}(p) \rightarrow \quad \textbf{[0.1 Ptos.]} \quad \hat{\eta} = 14.37742 \quad \text{y} \quad \hat{\beta} = 4.390718 \quad \textbf{[0.1 Ptos.]}$$

En los tres casos el estadístico de prueba distribuye $\chi^2(5 - 1 - 2)$ por lo tanto el estadístico de prueba más pequeño es el mejor ajuste, es decir, modelo Log-Normal. **[0.2 Ptos.]**

A partir de la tabla $\chi^2(2)$ se tiene que

$$\textbf{[0.1 Ptos.]} \quad c_{0.4}(2) < 1.1006010 < c_{0.6}(2) \rightarrow 40\,\% < \text{valor-p} < 60\,\% \quad \textbf{[0.1 Ptos.]}$$