Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2025

Curso	:	**Probabilidad y Estadística**
Sigla	:	**EYP1113**
Profesores	:	**Ricardo Aravena, Maritza Márquez , Bladimir Morales , Mauricio Toro, Ricardo Olea, Jose Ordoñez, Felipe Ossa, Nixón Jerez, Daniel Saavedra, Cristian Capetillo y Hernán Robledo**

INTERROGACIÓN 2

Pregunta 1

No hay duda de que, hoy en plena campaña, el impacto de las redes sociales en la comunicación política es un factor relevante. Un indicador importante es el número de retuits (retweets) que obtienen los candidatos por parte de sus adherentes. Diversos estudios sugieren que existe una probabilidad $Y \in [0, 1]$ de que un adherente retuitee un mensaje publicado por su candidato. Esta probabilidad se considera aleatoria y se modela mediante una distribución Beta$(q = 2,\, r = 1)$, dado que las personas difieren en su propensión a compartir publicaciones.

Como parte de un experimento, se envía un único mensaje del candidato a n adherentes, y se define X como el número de adherentes que retuitean dicho mensaje, suponiendo que la decisión de retuitear es independiente entre los adherentes.

(a) **[3.0 ptos]** Obtenga la covarianza entre el número de adherentes que retuitean el mensaje y la probabilidad de que un adherente retuitee un mensaje del candidato, Cov(X, Y).

(b) **[3.0 ptos]** Si se envía un mensaje del candidato a 100 adherentes, determine la probabilidad de que exactamente 75 retuiteen dicho mensaje.

Solución

(a)

Se pide la covarianza entre X y Y, la cual, según formulario, está dada por

$$\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) \tag{1}$$

Del enunciado se tiene información respecto a la distribución marginal de Y, la cual es $Y \sim$ Beta$(q = 2, r = 1)$. Por lo tanto, del formulario se tiene que E$(Y) = 2/3$, Var$(Y) = 2/(3^2{*}4) = 1/18$, y E$(Y^2) = 1/18{+}(2/3)^2 = 1/2$.

Por otro lado, el enunciado no da información respecto a la distribución marginal de X. Sin embargo, la distribución $X|Y = y$ corresponde a una distribución Binomial de tamaño n y probabilidad y, ya que esta variable aleatoria representa el número retuits de un total de n tuits, donde se asume que todos los adherentes tienen probabilidad $Y = y$ de retuitear, y estos son independientes entre si (la decisión de retuitear un tuit es independiente entre los adherentes). Así, se conoce lo siguiente

$$\text{E}(X|Y = y) = ny \qquad \text{Var}(X|Y = y) = ny(1 - y)$$

Aplicando Teorema de Esperanza Iterada para calcular las esperanzas de (**??**), se obtienen todos los elementos para calcular Cov(X, Y):

$$\mathsf{E}(X) = \mathsf{E}(\mathsf{E}(X|Y)) = \mathsf{E}(nY) = \frac{2n}{3}$$
$$\mathsf{E}(XY) = \mathsf{E}(\mathsf{E}(XY|Y)) = \mathsf{E}(Y\,\mathsf{E}(X|Y)) = \mathsf{E}(Y\,nY) = n\mathsf{E}(Y^2) = \frac{n}{2}$$
$$\mathsf{Cov}(X,Y) = \frac{n}{2} - \frac{2n}{3}\cdot\frac{2}{3} = \frac{n}{18}$$

(b)

Para calcular la probabilidad de que el tuit sea retuiteado cierto números de veces se debe obtener la función de probabilidad de X, $p_X(x)$, para así calcular $p_X(75)$. Del enunciado se tiene que $Y \sim$ Beta$(2,1)$ y que $X|Y=y \sim$ Binomial(n,y). Por ello, la distribución marginal de X se obtiene marginalizando la conjunta $p(x,y)$ integrando respecto a y. Es decir,

$$p_X(x) = \int_Y P\left(X=x|Y=y\right) f(y)dy$$

Obtenemos $p_X(75)$ a partir de la expresión anterior.

$$\begin{aligned}
P_X(75) &= \int_0^1 \binom{100}{75} y^{75}(1-y)^{100-75}\cdot\frac{1}{B(2,1)}y^{2-1}(1-y)^{1-1}\,dy \\
&= \binom{100}{75}\cdot\frac{1}{B(2,1)}\int_0^1 y^{75+2-1}(1-y)^{25+1-1}\,dy \\
&= \binom{100}{75}\cdot\frac{1}{B(2,1)}\,B(77,26) \\
&= \binom{100}{75}\cdot\frac{B(77,26)}{B(2,1)} \\
&= \binom{100}{75}\cdot\frac{\Gamma(77)\Gamma(26)}{\Gamma(103)}\cdot\frac{\Gamma(3)}{\Gamma(2)\Gamma(1)} = \binom{100}{75}\cdot 2\cdot\frac{\Gamma(77)\Gamma(26)}{\Gamma(103)} \\
&= 2\frac{100!}{25!\,75!}\frac{76!\,25!}{102!} = \frac{2\cdot 76}{102\cdot 101} = 0.01475
\end{aligned}$$

Asignación de Puntaje:

(a)

Logro 1 [0.6 ptos]: por obtener

- [0.2 pto] E$(Y) = 2/3$
- [0.2 pto] Var$(Y) = 1/18$.
- [0.2 pto] E$(Y^2) = 1/2$.

Logro 2 [1.0 ptos]: por calcular E$(X) = 2n/3$.

- [0.5 pto] por indicar uso de Teorema de esperanza iterada para calcular E(X).
- [0.5 pto] por obtener el valor correcto E$(Y) = 2n/3$.

Logro 3 [1.0 ptos]: por calcular E$(XY) = n/2$.

- [0.5 pto] por indicar uso de Teorema de esperanza iterada para calcular E(XY).
- [0.5 pto] por obtener el valor correcto E$(XY) = n/2$.

Logro 4 [0.4 ptos]: por obtener la covarianza, Cov$(X, Y) = n/18$.

(b)

Logro 5 [1 ptos]: por definir variables aleatorias del enunciado.

- [0.5 pto] por notar que se solicita la probabilidad marginal de X, $P(X = 75)$ o $P(X = x)$.
- [0.5 pto] por reconocer que $X|Y = y \sim$ Binomial$(100, y)$

Logro 6 [1 ptos]: por construir correctamente la integral de probabilidad marginal.

- [1 pto] por plantear $\int_0^1 P(X = x \mid Y = y) f_{\mathcal{Y}}(y)\, dy$.

Logro 7 [1 ptos]: por resolver la integral y obtener el resultado final.

- [0.8 pto] por desarrollar la integral correctamente.
- [0.2 pto] por obtener que $P(X = 75) = (2 \cdot 76)/(101 \cdot 102) = 0.01475$.

Pregunta 2

De acuerdo con la *Cuenta Pública de Carabineros 2024*, la Región de Magallanes se ubica entre las zonas con menor número de delitos del país. En particular, presenta una de las tasas más bajas de hurtos, con un total de 451 casos registrados en el año, de los cuales en 97 de estos casos se logró la detención de los responsables de los delitos. Suponga que la ocurrencia de estos hurtos sigue un proceso de Poisson, considerando que el año 2024 tuvo 366 días.

(a) **[1.2 Ptos]** Si usted decide vacacionar durante tres días en dicha región, determine la probabilidad de que en ese período ocurran al menos dos hurtos.

(b) **[1.2 Ptos]** Calcule la probabilidad de que en la semana previa a su llegada se registren menos de tres días con al menos un hurto diario.

(c) **[1.2 Ptos]** Obtenga el número esperado de días hasta que se observe exactamente un día con tres hurtos.

(d) **[1.2 Ptos]** A partir de la lista de los 451 hurtos de 2024, se seleccionó una muestra aleatoria sin reemplazo de 20 casos. Determine la probabilidad de que exactamente en tres de ellos se haya logrado la detención de los responsables.

[1.2 Ptos] por planteamiento basal del ejercicio.

Solución

La tasa de ocurrencia de hurtos es $\nu = 451$ por año. Si X_t representa el número de hurtos diarios, entonces X_t sigue un proceso de Poisson con parámetro proporcional al tiempo. Es decir, $X_t \sim \text{Poisson}(\lambda \cdot t)$, donde λ es la tasa de hurtos diarios.

(a)

La probabilidad solicitada corresponde a

$$P(X_3 \geq 2) = 1 - P(X_3 = 0) - P(X_3 = 1).$$

Según el enunciado, la tasa diaria está dada por $\lambda_{\text{día}} = 451/366 = 1.232$. Así, en un período de tres días, el número de hurtos sigue

$$X_3 \sim \text{Poisson}(\lambda_{3d}), \quad \lambda_{3d} = 3 \cdot \lambda_{\text{día}} \approx 3.696.$$

Así, la probabilidad solicitada es igual a

$$P(X_{\text{3día}} \geq 2) = 1 - e^{-\lambda_{3d}}\,(1 + \lambda_{3d}) = 0.883.$$

(b)

Sea Y el número de días, en una semana de $n = 7$ días, en que se registra al menos un hurto. Dado que los días son independientes (modelo de Poisson con tasa homogénea), Y sigue una distribución Binomial$(n = 7, p)$, donde p está dada por la probabilidad de que en un día se registre al menos un hurto.

Se pide calcular la probabilidad $P(Y < 3)$, la cual corresponde a

$$\begin{aligned} P(Y < 3) &= P(Y = 0) + P(Y = 1) + P(Y = 2) = \\ &= \binom{7}{0}(1-p)^7 + \binom{7}{1}p(1-p)^6 + \binom{7}{2}p^2(1-p)^5 \\ &= (1-p)^7 + 7p(1-p)^6 + 21p^2(1-p)^5 \end{aligned}$$

Para obtener este valor se requiere calcular p. Si X_t corresponde al proceso Poisson definido en (a), entonces la probabilidad de que en un día ocurra al menos un hurto corresponde a $p = P(X_1 > 0)$. Así,

$$P(X_1 > 0) = 1 - P(X_1 = 0) = 1 - e^{-\lambda_{\text{día}}} \approx 1 - e^{-1.2322404} \approx 0.708.$$

Finalmente, sustituimos los valores obtenidos:

$$P(Y < 3) = (1-p)^7 + 7p(1-p)^6 + 21p^2(1-p)^5 \approx 0.0254610.$$

(c)

Sea W el número de días hasta observar un día con exactamente tres hurtos. Como cada día es independiente, entonces W sigue una distribución Geométrica(p), donde p corresponde a la probabilidad de tener exactamente tres hurtos en un día. Se solicita el valor esperado de W, E$(W) = 1/p$ (según formulario), por lo que basta con calcular p para calcular lo solicitado.

Si X_t corresponde al proceso Poisson definido en (a), entonces p corresponde a

$$p = P(\text{3 hurtos en un día}) = P(X_1 = 3) = \frac{e^{-\lambda_{\text{día}}} \lambda_{\text{día}}^3}{3!} = 0.091$$

Por lo tanto, el número esperado de días hasta observar exactamente tres hurtos es

$$E[W] = \frac{1}{0.091} \approx 10.996 \text{ días} \approx 11 \text{ días}.$$

(d)

Sea $N = 451$ el total de hurtos de 2024, de los cuales $K = 97$ resultaron con captura de los responsables. Si se toma una muestra sin reemplazo de $n = 20$ hurtos, y se define H como el número de casos con captura en la muestra, entonces H sigue una distribución Hipergeométrica$(N = 451, m = 97, n = 20)$.

Se solicita calcular la probabilidad de que exactamente 3 casos en la muestra tengan captura:

$$P(H = 3) = \frac{\binom{97}{3}\binom{451-97}{20-3}}{\binom{451}{20}} = \frac{\binom{97}{3}\binom{354}{17}}{\binom{451}{20}} = 0.186.$$

Es posible obtener un valor aproximado para $P(H = 3)$ si se utiliza una distribución Binomial de parámetros $n = 20$ y $p = m/N = 97/451 = 0.215$. Esto se justifica en que para poblaciones grandes realizar un muestreo sin reemplazo genera cambios menores en las probabilidades de extracción del caso favorable. Así,

$$P(H = 3) = \binom{20}{3} p^3 (1-p)^1 7 = 0.1849$$

Asignación de Puntaje:

Logro 1 [1.2 ptos]: por plantear correctamente el Proceso Poisson X_t con tasa diaria $\lambda = 1.232$.

(a)

Logro 2 [0.6 ptos]: por reconocer que se solicita calcular $P(X_3 \geq 2)$.

Logro 3 [0.6 ptos]: por obtener que $P(X_3 \geq 2) = 0.883$.

- [0.4 pto] por transformar la tasa diaria $\lambda = 0.1232$ a una tasa cada tres días, $\lambda_{3d} = 3.696$, o un procedimiento equivalente.
- [0.2 pto] Por obtener $P(X_3 \geq 2) = 0.883$.

(b)

Logro 4 [0.6 ptos]: por reconocer que se solicita calcular $P(Y < 3)$.

- [0.3 pto] por definir Y como una variable aleatoria Binomial$(7, p)$.
- [0.3 pto] por reconocer que se solicita $P(Y < 3)$.

Logro 5 [0.6 ptos]: por obtener que $P(Y < 3) = 0.0254$.

- [0.3 pto] por calcular $p = P(X_1 > 0) = 0.708$ correctamente.
- [0.2 pto] por desarrollar $P(Y < 3)$ correctamente.
- [0.1 pto] por obtener $P(Y < 3) = 0.0254$.

(c)

Logro 6 [0.6 ptos]: por reconocer que se solicita E(W) donde W es una variable aleatoria Geométrica(p).

- [0.3 pto] por definir W como una variable aleatoria Geométrica(p).
- [0.3 pto] por reconocer que se solicita E(W).

Logro 7 [0.6 ptos]: por obtener que E$(W) = 10.996$.

- [0.4 pto] por calcular $p = P(X_1 = 3) = 0.091$ correctamente.
- [0.2 pto] por obtener E$(W) = 10.996$.

(d)

Logro 8 [0.6 ptos]: por reconocer que se solicita $P(H = 3)$ donde H es una variable aleatoria Hipergeométrica$(451, 97, 20)$.

- [0.4 pto] en total por definir H como una variable aleatoria Hipergeométrica [0.1 pto] de parámetros $N = 451$ [0.1 pto], $m = 97$ [0.1 pto], y $n = 20$ [0.1 pto].
- [0.2 pto] por reconocer que se solicita $P(H = 3)$.

Caso Aproximación Binomial: se asignará el Logro 8 si se utiliza la aproximación Binomial indicando explícitamente que este es un valor aproximado por asumir un muestreo con reemplazo en vez de un muestreo sin reemplazo.

Logro 9 [0.6 ptos]: por obtener que $P(H = 3) = 0.186$ o 0.1849 si usa la aproximación binomial y se le fue asignado el Logro 8.

Pregunta 3

En Chile central, la isoterma cero (altura a la cual la temperatura es 0 ℃) juega un rol fundamental en el riesgo de crecidas de ríos: cuando la isoterma se encuentra alta, la precipitación se presenta como lluvia incluso en zonas cordilleranas, lo que puede aumentar significativamente el caudal de los ríos. Para estudiar esta situación, se modelan conjuntamente dos variables meteorológicas: la altura de la isoterma cero W (en metros) y la precipitación diaria X (en mm). Ambas variables se suponen distribuidas de manera conjunta según una normal bivariada con medias $\mu_W = 2500$ y $\mu_X = 50$, desviaciones estándar $\sigma_W = 400$ y $\sigma_X = 10$, y correlación $\rho = 0.6$.

A partir de datos históricos se ajusta un modelo lluvia–caudal que supone un enlace logarítmico entre el caudal y la precipitación:

$$Q = \exp(\alpha + \beta X), \qquad \alpha = 1.5,\ \beta = 0.03,$$

lo que asegura caudales siempre positivos y captura un efecto multiplicativo de la lluvia sobre el caudal. Con esta especificación, un aumento de 10 mm en la precipitación multiplica el caudal esperado por $e^{0.3} \approx 1.35$ (aproximadamente un 35 %).

Por otra parte, para dimensionar la eficiencia energética de los eventos de crecida, se define el índice de energía hidráulica

$$H = \frac{Q}{1+Q},$$

el cual está acotado entre 0 y 1.

A partir de esta información:

(a) **[2.0 Ptos]** Obtenga una expresión para la mejor predicción del caudal Q dado que se sabe que la altura de la isoterma cero es W.

(b) **[2.0 Ptos]** Obtenga la función de densidad de H condicional a la altura de la isoterma cero.

(c) **[2.0 Ptos]** Calcule la probabilidad de que la eficiencia sea mayor al 0.95 (95 %) si la isoterma cero ocurre a los 3000 metros.

Solución

(a) La mejor predicción de Q condicional a W está dada por $\mathsf{E}(Q|W)$. Para obtener esta esperanza se debe encontrar la distribución de Q condicional a W. Para encontrar esta, primero se requiere la distribución de $X|W$, ya que Q es una función de X.

Como (X, W) siguen una distribución normal bivariada, entonces la distribución condicional de X está dada por formulario:

$$X \mid W = w \sim \text{Normal}\left(\mu_{X|w},\, \sigma^2_{X|w}\right), \qquad \mu_{X|w} \;=\; \mu_X + \rho\frac{\sigma_X}{\sigma_W}(w - \mu_W), \quad \sigma_{X|w} \;=\; \sigma_X\,\sqrt{(1-\rho^2)}.$$

Por otra parte, $\alpha + \beta\, X$ es una transformación lineal de una distribución normal, la cual es Normal$(\mu^* = \alpha + \beta\,\mu_{X|w}, \sigma^* = \beta\,\sigma_{X|w})$, y como Q es la transformación exponencial de una distribución normal, entonces Q sigue una distribución Log-Normal de parámetros $\lambda_w = \alpha + \beta\,\mu_{X|w}$ y $\zeta_w = \beta\,\sigma_{X|w}$. Así,

$$Q = e^{\alpha + \beta\, X} \mid W = w \sim \text{Log-Normal} \sim (\lambda_w,\, \zeta_w)\,.$$

Así, del formulario se extrae la esperanza de la LogNormal para obtener lo solicitado:

$$\begin{aligned}\mathsf{E}(Q|W) &= \exp\left(\lambda_W + \zeta^2_W/2\right) = \exp\left(\alpha + \beta\,\mu_{X|w} + \beta^2\,\sigma^2_{X|w}/2\right)\\ &= \exp\left(\alpha + \beta\,\left(\mu_X + \rho\frac{\sigma_X}{\sigma_W}(W - \mu_W)\right) + \beta^2\,\sigma^2_X(1-\rho^2)/2\right)\end{aligned}$$

(b) Se pide la función de densidad de H condicional a $W = w$. Como $H = g(X) = e^{\alpha+\beta\,X}/(1+e^{\alpha+\beta\,X})$, necesitamos conocer la distribución de $X|W$ para aplicarle la transformación $g(X)$ para obtener su distribución condicional. Bajo el mismo argumento de (a), se tiene que $X \mid W = w \sim$ Normal $\left(\mu_{X|w},\, \sigma^2_{X|w}\right)$, con los parámetros indicados como en (a). Con esto, aplicamos la fórmula de transformación de variables aleatorias dada en formulario. Para esto, necesitamos encontrar $g^{-1}(H) = X$ y su derivada.

$$H = \frac{e^{\alpha+\beta\,X}}{1+e^{\alpha+\beta\,X}} \quad \Longrightarrow \quad X = g^{-1}(H) = \frac{1}{\beta}\left[\log\left(\frac{H}{1-H}\right) - \alpha\right], \quad \frac{\partial}{\partial H} g^{-1}(H) = \frac{1}{\beta H(1-H)}$$

La cantidad $\partial X/\partial H$ es positiva para $H \in (0,1)$, por lo que la función $g^{-1}(H)$ es inyectiva. Por esto, aplicamos la fórmula de transformación para $k = 1$.

$$\begin{aligned} f_H(h) &= f_X\left(g^{-1}(h)\right)) \cdot \left|\frac{\partial h}{\partial h} g^{-1}(h)\right| \\ &= \frac{1}{\sigma_{X|w}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{g^{-1}(h) - \mu_{X|w}}{\sigma_{X|w}}\right)\right] \cdot \frac{1}{\beta h(1-h)} \\ &= \frac{1}{\beta\sigma_{X|w}\sqrt{2\pi} h(1-h)} \exp\left[-\frac{1}{2}\left(\frac{\log(h/(1-h)) - (\alpha + \mu_{X|w})}{\beta\sigma_{X|w}}\right)\right] \end{aligned}$$

donde $\mu_{X|w} = \mu_X + \rho\frac{\sigma_X}{\sigma_W}(w - \mu_W)$ y $\sigma_{X|w} = \sigma_X\sqrt{(1-\rho^2)}$ dados por formulario.

Una solución alternativa y equivalente es considerar $H = g(Q) = Q/(1+Q)$ y aplicar la fórmula de transformación de variables aleatorias a la distribución Log-Normal.

(c) Se pide la probabilidad de superar 95 % de eficiencia cuando $W = 3000$ m.

$$\Pr(H > 0.95 \mid W = w) = P\left(Q > \frac{0.95}{1-0.95} \mid W = w\right) = 1 - \Phi\left(\frac{\ln(0.95/0.05) - \lambda_w}{\zeta}\right)$$

Reemplazando

$$\mu_{X|3000} = 50 + 0.6 \cdot \frac{10}{400}\,(3000 - 2500) = 57.5, \qquad \sigma_{X|w} = \sqrt{10^2(1-0.6^2)} = \sqrt{64} = 8.$$

Luego

$$\lambda_w = \alpha + \beta\,\mu_{X|w} = 1.5 + 0.03 \cdot 57.5 = 3.225, \qquad \zeta_w = \beta\,\sigma_{X|w} = 0.03 \cdot 8 = 0.24.$$

$$P(H > 0.95 \mid W = 3000) \;=\; 1 - \Phi(-1.169) \;\approx\; 0.879.$$

Asignación de Puntaje:

(a)

Logro 1 [0.6 ptos]: por reconocer que el mejor predictor es E$(Q|W)$.

Logro 2 [1.2 ptos]: por obtener que $Q|W$ sigue una distribución Log-Normal de parámetros $\lambda_w = \alpha + \beta\,\mu_{X|w}$ y $\zeta_w = \beta\,\sigma_{X|w}$.

- [0.6 pto] por reconocer que $X \mid W = w \sim$ Normal $\left(\mu_{X|w},\, \sigma^2_{X|w}\right)$. Se descuenta [0.1 pto] por cada parámetro incorrecto.
- [0.6 pto] por reconocer que $Q|W = w$ es una Log-Normal con los parámetros indicados en la solución.

Logro 3 [0.2 ptos]: por obtener la expresión correcta para E$(Q|W)$.

(b)

Logro 4 [1.0 ptos]: por reconocer que se requiere transformar la función de densidad de $X|W$ o de $Q|W$ con la fórmula de transformación de variables.

- [0.5 pto] por reconocer uso de fórmula de transformación de variables.
- [0.5 pto] por identificar que la función $g(X)$ a reconocer del teorema de transformación de cambio de variables es $g(X) = e^{\alpha+\beta\,X}/(1+e^{\alpha+\beta\,X})$.

Logro 5 [1.0 ptos]: por obtener la expresión para $f_X(h)$.

- [0.4 pto] por obtener $\partial g^{-1}(H)/\partial H$.
- [0.4 pto] por desarrollo de $f_H(h)$.
- [0.2 pto] por obtener la $f_H(h)$ correcta.

(c)

Logro 6 [0.5 ptos]: por reconocer que se solicita calcular $P(H > 0.95|W = w)$.

Logro 7 [0.6 ptos]: por reconocer que $Q|W$ sigue una distribución Log-Normal de parámetros $\lambda_w = \alpha + \beta\,\mu_{X|w}$ y $\zeta_w = \beta\,\sigma_{X|w}$.

- Asignar este puntaje sólo si estudiante reconoce que requiere la distribución de Q para hacer el cálculo.

Logro 8 [0.6 ptos]: por obtener que $P(H > 0.95|W = w) = 1 - \Phi\left((\log(0.95/0.05) - \lambda_w)/\zeta_w\right)$.

Logro 9 [0.3 ptos]: por obtener que $P(H > 0.95|W = w) = 0.879$.

Desarrollo alternativo: si estudiante desarrolla la probabilidad solicitada como $\int_{0.95}^{1} f_H(h)dh$, asignar el logro 6 y [0.2 ptos] en el logro 7.