Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2017

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

EXAMEN

Problema 1

Las encuestas el domingo pasado se vieron fuertemente cuestionadas. Usted, preparándose para el domingo 17 de diciembre lleva a cabo ciertas simulaciones que le permitirían determinar el potencial ganador de la elección. Para ello, selecciona 34 comunas "muy representativas" y recoge información de las elecciones recientes y anteriores. Con esta información construye varios modelos de regresión que permitan explicar el caudal de votos (Y) de su "candidato" en una 2da vuelta.

Los predictores que dispone son:

- X_1: Tamaño de la comuna.
- X_2: Número de votos obtenidos en la elección de alcaldes por el sector de su candidato.
- X_3: % de abstención en la elección reciente.
- X_4: Si el alcalde actual es del su sector versus (1) vs. Si el alcalde no es de su sector (0).

Un resumen de los modelos ajustado es el siguiente:

```
-----------------------
Modelo   Variables   R2
-----------------------
0                  0.00
1               X1 0.21
2            X1+X2 0.38
3            X1+X3 0.24
4         X1+X2+X3 0.42
5      X1+X2+X3+X4 0.44
-----------------------
```

Por último, los residuos del modelo 5 tienen una varianza igual a 124.

Usando un nivel de significancia del 5 %:

(a) **[3.0 Ptos.]** Para el modelo 1, entregue los coeficientes de determinación, los estadísticos t y F del modelo junto a sus hipótesis. Calcule valores-p y comente.

(b) **[2.0 Ptos.]** ¿En qué modelos el aporte de X_3 es significativo?

(c) **[1.0 Ptos.]** ¿El aporte conjunto de X_3 y X_4 es significativo?

Solución

(a) Del enunciado tenemos que $n = 34$, mientras que la SCT la podemos obtener de la varianza de los residuos y del R^2 del modelo 5 ($k = 4$):

$$\textbf{[0.3 Ptos.]} \quad SCE_{\text{mod5}} = 124\times(n-k-1) = 124\times 29 = 3596 \to SCT = \frac{SCE_{\text{mod5}}}{(1-R^2)} = \frac{3596}{1-0.44} = 6421.429 \quad \textbf{[0.3 Ptos.]}$$

y la SCE del modelo 1 la obtenemos a partir de su R^2:

[0.3 Ptos.] $SCE_{\text{mod1}} = (1 - R^2) \times SCT = (1 - 0.21) \times 6421.429 = 5072.929 \to SCR_{\text{mod1}} = SCT - SCE_{\text{mod1}}$

$$= 6421.429 - 5072.929$$

$$= 1348.5$$ **[0.3 Ptos.]**

Se pide

[0.3 Ptos.] $R^2 = 0.21, \qquad r^2 = 1 - \dfrac{SCE_{\text{mod1}}/(n-1-1)}{SCT/(n-1)} = 0.1853125$ **[0.3 Ptos.]**

[0.3 Ptos.] $F = \dfrac{SCR_{\text{mod1}}/(1)}{SCE_{\text{mod1}}/(n-1-1)} = 8.506328 \to T_{X_1} = \pm 2.916561 \sim \text{t-Student}(32)$ **[0.3 Ptos.]**

$$\overset{\text{aprox}}{\sim} \text{Normal}(0,1)$$

$\to$ valor-p $\approx 2 \times [1 - \Phi(|2.92|)] = 0.0036$ **[0.3 Ptos.]**

El valor-p es el mismo para ambos estadísticos.

Por lo tanto, el tamaño de la comuna es una variable significativa para explicar el caudal de votos.

[0.3 Ptos.]

(b) El aporte X_3 lo podemos testear comparando:

- Mod1 vs Mod3

  $SCE_{\text{mod3}} = (1 - R^2) \times SCT = (1 - 0.24) \times 6421.429 = 4880.286$ **[0.2 Ptos.]**

  [0.2 Ptos.] $F = \dfrac{(SCE_{\text{mod1}} - SCE_{\text{mod3}})/1}{SCE_{\text{mod3}}/(n-1-1-1)} = 1.223685 \sim F(1,31)$ **[0.2 Ptos.]**

  Como $F = 1.223685 < 4.16 = F_{0.95}(1,31)$, entonces no existe suficiente evidencia para rechazar que el aporte de X_3 en presencia de X_1 resulte NO significativo.

  [0.3 Ptos.]

- Mod2 vs Mod4

  $SCE_{\text{mod2}} = (1 - R^2) \times SCT = (1 - 0.38) \times 6421.429 = 3981.286$ **[0.2 Ptos.]**

  $SCE_{\text{mod4}} = (1 - R^2) \times SCT = (1 - 0.42) \times 6421.429 = 3724.429$ **[0.2 Ptos.]**

  [0.2 Ptos.] $F = \dfrac{(SCE_{\text{mod2}} - SCE_{\text{mod4}})/1}{SCE_{\text{mod4}}/(n-2-1-1)} = 2.068964 \sim F(1,30)$ **[0.2 Ptos.]**

  Como $F = 2.068964 < 4.17 = F_{0.95}(1,30)$, entonces no existe suficiente evidencia para rechazar que el aporte de X_3 en presencia de X_1 y X_2 resulte NO significativo.

  [0.3 Ptos.]

(c) El aporte conjunto de X_3 y X_4 lo podemos testear comparando mod2 vs mod5

[0.4 Ptos.] $F = \dfrac{(SCE_{\text{mod2}} - SCE_{\text{mod5}})/2}{SCE_{\text{mod5}}/(n-2-2-1)} = 1.553573 \sim F(2,29)$ **[0.2 Ptos.]**

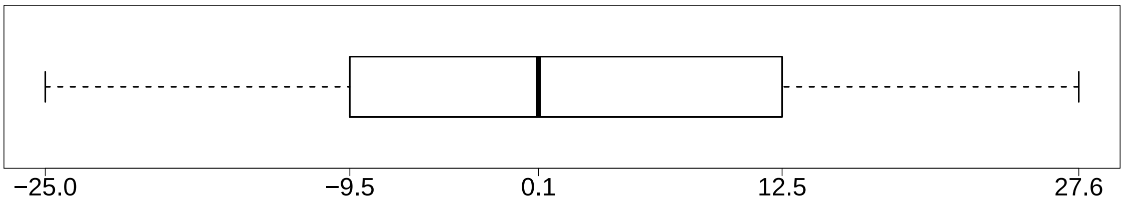
Como $F = 1.553573 < 3.33 = F_{0.95}(2,29)$, entonces no existe suficiente evidencia para rechazar que el aporte conjunto de X_3 y X_4 en presencia de X_1 y X_2 resulte NO significativo.

[0.4 Ptos.]

+ 1 Punto Base

Problema 2

La siguiente figura muestra el diagrama de caja, en base al **summary** que entrega **R**, para los residuos del modelo 5 visto en el Problema 1.



Recuerde que el R^2 y varianza de los residuos del modelo 5 fueron 44 % y 124 respectivamente.

¿Entre un modelo Normal(μ, σ) y una Gamma$(k = 5, \nu)$ cuál ajusta mejor?

Considere la siguiente tabla de frecuencia:

Intervalo	≤ -10	$(-10; 0]$	$(0; +10]$	$> +10$
Frecuencia	6	13	10	5

Por simplicidad, no colapse intervalos si el valor esperado es inferior a 5.

Solución

Para el caso Normal, tenemos que

$$\textbf{[0.5 Ptos.]} \quad \mu = 0.1 \quad \text{y} \quad \sigma = \sqrt{124} = 11.13553 \quad \textbf{[0.5 Ptos.]}$$

mientras que para el caso Gamma, para poder ajustarla se debe trasladar en α, reemplazando

$$\textbf{[0.5 Ptos.]} \quad \alpha = -25 \quad \text{y} \quad \nu = \sqrt{\frac{5}{124}} = 0.2008048 \quad \textbf{[0.5 Ptos.]}$$

Observación: Los parámetros μ, σ y ν, podrían también obtenerse igualando los percentiles 25 %, 50 % y 75 % teóricos a los empíricos.

Test 1:

$$\text{H}_0 : X \sim \text{Normal} \quad \text{vs} \quad \text{H}_a : X \not\sim \text{Normal}$$

Considerando $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ se construye la siguiente tabla:

	Observado	Probabilidad	Esperado	(O-E)^2/E
<= -10	6	0.1822	6.1948	0.006125628
(-10, 0]	13	0.3142	10.6828	0.502622518
(0, +10]	10	0.3166	10.7644	0.054281461
>10	5	0.1870	6.3580	0.290054105
Total	34	1.0000	34.0000	0.853083713

Tenemos que $X^2 = 0.8531 \sim \chi^2(4 - 2 - 1)$ **[0.5 Ptos.]** $\to 30\,\% < \text{valor-p} < 40\,\%$. **[0.5 Ptos.]**

Test 2:

$$\text{H}_0 : X \sim \text{Gamma}(5, \nu) \text{ trasladad en } \alpha \quad \text{vs} \quad \text{H}_a : X \not\sim \text{Gamma}(5, \nu) \text{ trasladad en } \alpha$$

Como $k = 5$, entonces $F_X(x) = P(X \leq x) = P(X - \alpha \leq x - \alpha) = 1 - \displaystyle\sum_{y=0}^{5-1} \frac{[\nu\,(x-\alpha)]^y\, e^{-\nu\,(x-\alpha)}}{y!}$ **[0.5 Ptos.]**, luego la tabla queda como sigue:

	Observado	Probabilidad	Esperado	(O-E)^2/E
<= -10	6	0.1868	6.3512	0.019420179
(-10, 0]	13	0.3762	12.7908	0.003421572
(0, +10]	10	0.2666	9.0644	0.096569807
>10	5	0.1704	5.7936	0.108706324
Total	34	1.0000	34.0000	0.228117882

En este caso, tenemos que $X^2 = 0.2281 \sim \chi^2(4-2-1)$ **[1.5 Ptos.]** $\to 60\,\% <$ valor-p $< 70\,\%$. **[0.5 Ptos.]**

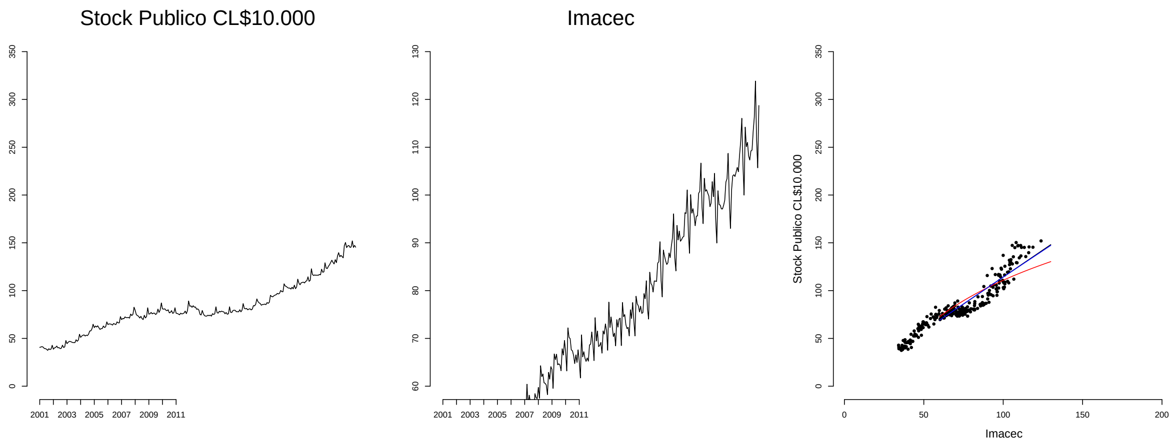
Por lo tanto, el mejor ajuste se logra con la Gamma(5, ν) trasladad en α. **[0.5 Ptos.]**

Observación: Si no calcula valores-p, pero argumenta que no es necesario ya que ambas estadísticos distribuyen igual, siendo el mejor ajuste el estadístcio más cercano a cero asignar **[1.0 Ptos.]** en reemplazo.

+ 1 Punto Base

Problema 3

Para el Banco Central de Chile es de suma importancia poder predecir el Stock Público de billetes para poder planificar con tiempo la elaboración de nuevas piezas, cuyo tiempo total muchas veces supera los 9 meses. Para esto, usted ajusta un modelo lineal simple y otro no lineal (cuadrático) considerando como regresor el Imacec para explicar el Stock Público (en MM piezas) de billetes de diez mil pesos.




```
Stock ~ Imacec:
            Estimate Std. Error t value
(Intercept) -209.5387    14.1387  -14.82
Imacec         X.XXXX     0.1585   XX.XX

Residual standard error: 20.58 on 118 degrees of freedom
Multiple R-squared:  X.XXXX,Adjusted R-squared:  X.XXXX
F-statistic: 641.6 on 1 and 118 DF,  p-value: < 2.2e-16
```

```
Stock ~ Imacec + Imacec^2:
             Estimate Std. Error t value
(Intercept) 289.65768  104.08443   2.783
Imacec       -7.47337    2.38102  -3.139
Imacec2       0.06490    X.XXXXX   X.XXX

Residual standard error: 18.87 on 117 degrees of freedom
Multiple R-squared:  0.8705,Adjusted R-squared:  0.8683
F-statistic: 393.3 on 2 and 117 DF,  p-value: < 2.2e-16
```

```
mean(Stock) mean(Stock^2)   n
   145.4267      23829.45 120
```

Complete la información faltante e indique si el efecto cuadrático del Imacec es significativo para explicar el Stock Público. Justifique su respuesta.

Respuesta

Tenemos que

$$SCT = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\,(\overline{Y})^2 = 120 \cdot 23829.45 - 120 \cdot (145.4267)^2 = 321663 \quad \textbf{[0.6 Ptos.]}$$

con $n = 120$.

Recordemos que

$$SCT = SCR + SCE \qquad R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT} \qquad r^2 = 1 - \frac{(n-1)}{(n-2)} \cdot \frac{SCE}{SCT}$$

y que en regresión simple bajo $\text{H}_0 : \beta_1 = 0$ se tiene que $T_{b_1} \sim$ t-Student$(n-2)$ y que el estadístico

$$F = \frac{SCR/1}{SCE/(n-2)} = T_{b_1}^2 \sim F(1,\, n-2).$$

Además $\sqrt{SCE/(n-2)}$ = Residual standard error.

Por último, en el caso del modelo múltiple el estadístico

$$T_{b_2}^2 = F = \frac{(SCE_{(1)} - SCE)/1}{SCE/(n-1-1-1)}$$

Reemplazando

```
Stock ~ Imacec:
             Estimate Std. Error t value
(Intercept) -209.5387    14.1387  -14.82
Imacec         4.0148     0.1585   25.33

Residual standard error: 20.58 on 118 degrees of freedom
Multiple R-squared:  0.8446,Adjusted R-squared:  0.8433
F-statistic: 641.6 on 1 and 118 DF,  p-value: < 2.2e-16
```

[2.4 Ptos.]

```
Stock ~ Imacec + Imacec^2:
             Estimate Std. Error t value
(Intercept) 289.65768  104.08443   2.783
Imacec       -7.47337    2.38102  -3.139
Imacec2       0.06490    0.01343   4.834

Residual standard error: 18.87 on 117 degrees of freedom
Multiple R-squared:  0.8705,Adjusted R-squared:  0.8683
F-statistic: 393.3 on 2 and 117 DF,  p-value: < 2.2e-16
```

[2.0 Ptos.]

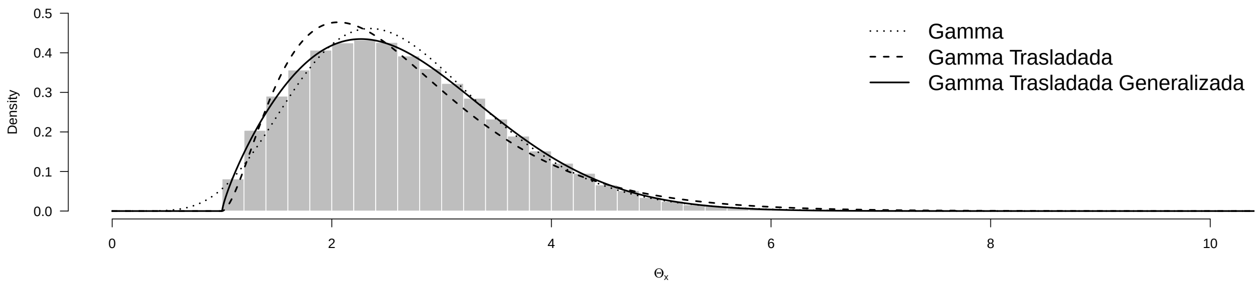
Por lo tanto, podemos afirmar que el efecto cuadrático resulta ser significativo, ya que valor-p ≈ 0.

[1.0 Ptos.]

+ 1 Punto Base

Problema 4

Una distribución de probabilidad que permite mayor grado de libertad en su forma para modelar datos con soporte distinto a $\mathbb{R}_0^+$ como lo hace clásica Gamma, es la Gamma-Generalizada(k, ν, β) trasladada en α. La figura muestra el ajuste de extensiones de una Gamma a un conjunto de mediciones, donde la Gamma-Generalizada trasladada es la que logra un mejor resultado.



Esta este modelo tiene una función de densidad dada por:

$$f(x) = \frac{\beta\,\nu^{\beta\,k}}{\Gamma(k)}\,(x-\alpha)^{\beta\,k-1}\,\exp\left\{-[\nu\,(x-\alpha)]^{\beta}\right\}, \quad x \geq \alpha$$

con $k > 0$, $\nu > 0$, $\beta > 0$, $\alpha \in \mathbb{R}$ y $\mathrm{E}[(X-\alpha)^r] = \dfrac{\Gamma\left(k + \frac{r}{\beta}\right)}{\nu^r\,\Gamma(k)}$.

Para una muestra aleatoria X_1, X_2,..., X_n y asumiendo que los parámetros k, α y β son conocidos:

(a) **[2.0 Puntos]** Obtenga el estimador de momento $\hat{\nu}$ de ν.

(b) **[2.0 Puntos]** Obtenga el estimador máximo verosímil $\widetilde{\nu}$ de ν.

(c) **[2.0 Puntos]** Calcule el sesgo aproximado de 1er orden que tienen los estimadores obtenidos en (a) y (b). Comente.

Solución

(a) Consideremos una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una población Gamma-Generalizada(k, ν, β) trasladada en α, con k, ν y β parámetros conocidos. Se pide estimar α por método de momentos.

Tenemos que el 1er momento de una Gamma-Generalizada trasladada está dado por

$$\mathrm{E}(X) = \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{\nu\,\Gamma(k)} + \alpha \quad \textbf{[1.0 Ptos.]}$$

igualando al 1er momento empírico $\overline{X}_n$, se tiene que el estimador de momento $\hat{\nu}$ de ν es

$$\hat{\nu} = \frac{\Gamma\left(k + \frac{1}{\beta}\right)}{(\overline{X}_n - \alpha)\cdot\Gamma(k)} \quad \textbf{[1.0 Ptos.]}$$

(b) La verosimilitud, su logaritmo natural y la primera derivada con respecto a ν son:

$$L(\nu) = \frac{\beta^n\,\nu^{n\,\beta\,k}}{[\Gamma(k)]^n} \times \left[\prod_{i=1}^n (X_i - \alpha)\right]^{\beta\,k-1} \times \exp\left[-\nu^{\beta}\sum_{i=1}^n (X_i-\alpha)^{\beta}\right] \quad \textbf{[0.5 Ptos.]}$$

$$\ln\,L(\nu) = n\,\ln(\beta) + n\,\beta\,k\,\ln(\nu) + (\beta\,k-1)\sum_{i=1}^n \ln(X_i-\alpha) - \nu^{\beta}\sum_{i=1}^n (X_i-\alpha)^{\beta} - n\,\ln[\Gamma(k)] \quad \textbf{[0.5 Ptos.]}$$

$$\textbf{[0.5 Ptos.]}\ \frac{\partial}{\partial\nu}\ln L(\nu)=\frac{n\,\beta\,k}{\nu}-\beta\,\nu^{\beta-1}\sum_{i=1}^{n}(X_i-\alpha)^\beta=0\to\widetilde{\nu}=\left[\frac{n\,k}{\displaystyle\sum_{i=1}^{n}(X_i-\alpha)^\beta}\right]^{1/\beta}\quad\textbf{[0.5 Ptos.]}$$

(c) Las esperanzas aproximadas de 1er orden son:

$$\mathrm{E}(\hat{\nu})\approx\frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\left[\frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\nu\,\Gamma(k)}+\alpha-\alpha\right]\cdot\Gamma(k)}=\nu\quad\textbf{[0.5 Ptos.]}$$

y

$$\mathrm{E}(\widetilde{\nu})\approx\left\{\frac{n\,k}{\displaystyle\sum_{i=1}^{n}\left[\frac{\Gamma\left(k+\frac{1}{\beta}\right)}{\nu\,\Gamma(k)}+\alpha-\alpha\right]^\beta}\right\}^{1/\beta}=\nu\cdot\frac{k^{1/\beta}\,\Gamma(k)}{\Gamma\left(k+\frac{1}{\beta}\right)}\quad\textbf{[0.5 Ptos.]}$$

Por lo tanto el sesgo en (a) es cero **[0.5 Ptos.]** y en (b) es $\nu\cdot\left[1-k^{1/\beta}\,\dfrac{\Gamma(k)}{\Gamma\left(k+\frac{1}{\beta}\right)}\right]$.

[0.5 Ptos.]

+ 1 Punto Base

Problema 5

Un grupo de n electores independientes (público) conforma a un panel de un programa de entrevistas a los dos candidatos presidenciales que quedan para la segunda vuelta. Al final de la entrevista, los electores indican si votarían por el candidato invitado, es decir, los convenció.

Para su candidato, que es invitado al programa, usted asume que la probabilidad que los electores decidan votar por él (digamos, atracción) se comporta de acuerdo a una distribución Beta(2, 1). Si el panel está conformado por 10 personas y 8 han decidido votar por su candidato, determine la probabilidad que la atracción sea mayor a 0.7.

Solución

Definamos como Y al número de electores que votarían por su candidato y X a la probabilidad a que los electores decidan votar por él.

$$\textbf{[0.5 Ptos.]} \quad Y \,|\, X = x \sim \text{Binomial}(n,\, p = x) \qquad \text{y} \qquad X \sim \text{Beta}(2,\, 1) \quad \textbf{[0.5 Ptos.]}$$

Se pide $P(X > 0.7 \,|\, Y = 8)$. **[0.2 Ptos.]**

Obtengamos $f_{X\,|\,Y=y}(x)$.

Primero calculemos la marginal de Y:

$$\begin{aligned}
p_Y(y) &= \int_{-\infty}^{+\infty} p_{Y\,|\,X=x}(y) \cdot f_X(x)\, dx \quad \textbf{[0.3 Ptos.]}\\
&= \int_0^1 \binom{n}{y} x^y\,(1-x)^{n-y} \cdot \frac{1}{B(2,1)}\,\frac{(x-0)^{2-1}\,(1-x)^{1-1}}{1^{2+1-1}}\,dx \quad \textbf{[0.3 Ptos.]}\\
&= \binom{n}{y} \cdot \frac{B(2+y,\, n-y+1)}{B(2,\,1)} \int_0^1 \frac{1}{B(2+y,\, n-y+1)}\,\frac{(x-0)^{(2+y)-1}\,(1-x)^{(n-y+1)-1}}{1^{(2+y)+(n-y+1)-1}}\,dx \quad \textbf{[0.4 Ptos.]}\\
&= \frac{n!}{y!\,(n-y)!} \cdot \frac{B(2+y,\, n-y+1)}{B(2,\,1)} \cdot 1, \quad \text{por área bajo la densidad Beta sobre todo su soporte} \quad \textbf{[0.4 Ptos.]}\\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\,\Gamma(n-y+1)} \cdot \frac{\Gamma(y+2)\,\Gamma(n-y+1)}{\Gamma(n+3)} \cdot \frac{\Gamma(3)}{\Gamma(2)\,\Gamma(1)} \quad \textbf{[0.3 Ptos.]}\\
&= \frac{2\,(y+1)}{(n+2)\,(n+1)}, \quad y = 0, 1, \ldots, n \quad \textbf{[0.3 Ptos.]}
\end{aligned}$$

Luego, la condicional de X dado $Y = y$ queda como:

$$\begin{aligned}
f_{X\,|\,Y=y}(y) &= \frac{p_{Y\,|\,X=x}(y) \cdot f_X(x)}{p_Y(y)} \quad \textbf{[0.3 Ptos.]}\\
&= \frac{(n+2)\,(n+1)}{2\,(y+1)} \cdot \frac{\binom{n}{y}}{B(2,1)}\, x^{(y+2)-1}\,(1-x)^{(n-y+1)-1} \quad \textbf{[0.3 Ptos.]}\\
&= \frac{(n+2)\,(n+1)}{2\,(y+1)} \cdot \frac{\Gamma(n+1)\,\Gamma(3)}{\Gamma(y+1)\,\Gamma(n-y+1)\,\Gamma(2)\,\Gamma(1)}\, x^{(y+2)-1}\,(1-x)^{(n-y+1)-1} \quad \textbf{[0.4 Ptos.]}\\
&= \frac{\Gamma(n+3)\,\Gamma(3)}{\Gamma(y+2)\,\Gamma(n-y+1)\,\Gamma(3)\,\Gamma(1)}\, x^{(y+2)-1}\,(1-x)^{(n-y+1)-1} \quad \textbf{[0.4 Ptos.]}\\
&= \frac{1}{B(y+2,\, n+3)}\, x^{(y+2)-1}\,(1-x)^{(n-y+1)-1} \quad \textbf{[0.4 Ptos.]}
\end{aligned}$$

Es decir,

$$X \,|\, Y = y \sim \text{Beta}(10,\, 3)$$

con $0 < x < 1$.

Por lo tanto

$$\begin{aligned}
\textbf{[0.5 Ptos.]} \quad P(X > 0.7 \,|\, Y = 8) &= 1 - F_Z(0.7), \quad \text{con } Z \sim \text{Beta}(10,\, 3) \text{ y } \Theta_Z = [0,\, 1] \quad \textbf{[0.2 Ptos.]} \\
&= 1 - 0.2528153 \quad \textbf{[0.2 Ptos.]} \\
&= 0.7471847 \quad \textbf{[0.1 Ptos.]}
\end{aligned}$$

+ 1 Punto Base