Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

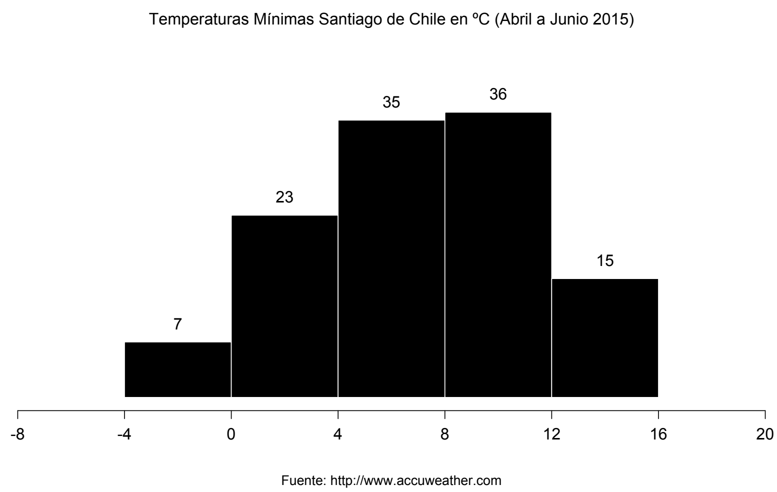
Primer Semestre 2015

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

Pauta Examen

Problema 1

Un modelo muy usado para ajustar valores extremos, como son el mínimo o máximo, es la distribución Sev(μ, σ). A Usted se le pide realizar una prueba de hipótesis para ver si las temperaturas mínimas registradas entre abril y lo que va de junio de este año ajustan o no dicha distribución. Estos datos registraron una media igual a 7.4 grados con un c.o.v igual a 61.82 %.



Considere un nivel de significancia del 5 %.

Nota: Si $Y \sim$ Sev(μ, σ) con $\Theta_Y = \mathbb{R}$, entonces

$$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right), \qquad f_Y(y) = \frac{1}{\sigma}\,\phi\left(\frac{y-\mu}{\sigma}\right), \qquad \mathrm{E}(Y) = \mu - \sigma\,\gamma, \qquad \mathrm{Var}(Y) = \frac{\sigma^2\,\pi^2}{6}$$

donde $\Phi(z) = 1 - \exp\left[-\exp(z)\right]$, $\phi(z) = \exp\left[z - \exp(z)\right]$, $\gamma \approx 0.5772$, $\pi \approx 3.1416$, $\mu \in \mathbb{R}$ y $\sigma > 0$.

Solución

Tenemos que

$$\textbf{[0.5 Ptos.]} \qquad \mu - \sigma\,\gamma = 7.4 \qquad \text{y} \qquad \frac{\sigma\,\pi}{\sqrt{6}\,(\mu - \sigma\,\gamma)} = 0.6182 \qquad \textbf{[0.5 Ptos.]}$$

resolviendo el sistema

$$\textbf{[0.5 Ptos.]} \qquad \mu = 9.458794 \qquad \text{y} \qquad \sigma = 3.566863 \qquad \textbf{[0.5 Ptos.]}$$

Alternativa 1: (Cinco intervalos)

La probabilidad estimada para los cinco intervalos se calcula como sigue

$$p_1 = F_Y(0) = 1 - \exp\left[-\exp\left(\frac{0-\mu}{\sigma}\right)\right] = 0.06809136 \qquad \textbf{[0.5 Ptos.]}$$

$$p_2 = F_Y(4) - F_Y(0) = \exp\left[-\exp\left(\frac{0-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{4-\mu}{\sigma}\right)\right] = 0.12653187 \qquad \textbf{[0.5 Ptos.]}$$

$$p_3 = F_Y(8) - F_Y(4) = \exp\left[-\exp\left(\frac{4-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{8-\mu}{\sigma}\right)\right] = 0.29075558 \qquad \textbf{[0.5 Ptos.]}$$

$$p_4 = F_Y(12) - F_Y(8) = \exp\left[-\exp\left(\frac{8-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{12-\mu}{\sigma}\right)\right] = 0.38445946 \qquad \textbf{[0.5 Ptos.]}$$

$$p_5 = 1 - p_1 - p_2 - p_3 - p_4 = 0.13016172 \qquad \textbf{[0.5 Ptos.]}$$

los valores esperado para cada uno de ellos son

```
Intervalo  Observado Probabilidad   Esperado    (O-E)^2/E
---------------------------------------------------------
<0                 7   0.06809136   7.898597 0.1022304818
[0-4)             23   0.12653187  14.677697 4.7187730318
[4-8)             35   0.29075558  33.727648 0.0479986111
[8-12)            36   0.38445946  44.597298 1.6573544647
>= 12             15   0.13016172  15.098760 0.0006459774
---------------------------------------------------------
Total            116   1.00000000 116.000000 6.5270025669
```

siendo el estadístico de pruebas X^2 igual a 6.527 **[0.5 Ptos.]**. Al compara con una $\chi^2(5-1-2)$ se tiene que

$$2.5\,\% < \text{valor-p} < 5\,\% \qquad \textbf{[0.5 Ptos.]}$$

Por lo tanto, existe suficiente evidencia para rechazar que las temperaturas mínimas en los últimos 116 días se ajustan a un modelo Sev. **[0.5 Ptos.]**

Alternativa 2: (Siete intervalos)

La probabilidad estimada para los siete intervalos se calcula como sigue

$$p_1 = F_Y(-4) = 1 - \exp\left[-\exp\left(\frac{-4-\mu}{\sigma}\right)\right] = 0.06809136 \qquad \textbf{[0.5 Ptos.]}$$

$$p_2 = F_Y(0) - F_Y(-4) = \exp\left[-\exp\left(\frac{-4-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{0-\mu}{\sigma}\right)\right] = 0.045376854 \qquad \textbf{[0.3 Ptos.]}$$

$$p_3 = F_Y(4) - F_Y(0) = \exp\left[-\exp\left(\frac{0-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{4-\mu}{\sigma}\right)\right] = 0.126531873 \qquad \textbf{[0.3 Ptos.]}$$

$$p_4 = F_Y(8) - F_Y(4) = \exp\left[-\exp\left(\frac{4-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{8-\mu}{\sigma}\right)\right] = 0.290755584 \qquad \textbf{[0.3 Ptos.]}$$

$$p_5 = F_Y(12) - F_Y(8) = \exp\left[-\exp\left(\frac{8-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{12-\mu}{\sigma}\right)\right] = 0.38445946 \qquad \textbf{[0.3 Ptos.]}$$

$$p_6 = F_Y(16) - F_Y(12) = \exp\left[-\exp\left(\frac{12-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(\frac{16-\mu}{\sigma}\right)\right] = 0.128246904 \qquad \textbf{[0.3 Ptos.]}$$

$$p_7 = 1 - p_1 - p_2 - p_3 - p_4 - p_5 - p_6 = 0.001914816 \qquad \textbf{[0.5 Ptos.]}$$

los valores esperado para cada uno de ellos son

Intervalo	Observado	Probabilidad	Esperado	(O-E)^2/E
< -4	0	0.022714503	2.6348824	2.63488237
[-4 - 0)	7	0.045376854	5.2637151	0.57272956
[0 - 4)	23	0.126531873	14.6776973	4.71877303
[4 - 8)	35	0.290755584	33.7276478	0.04799861
[8 - 12)	36	0.384459464	44.5972979	1.65735446
[12 - 16)	15	0.128246904	14.8766409	0.00102291
>= 16	0	0.001914816	0.2221187	0.22211870
Total	116	1.000000000	116.0000000	9.85487964

siendo el estadístico de pruebas X^2 igual a 9.855 **[0.5 Ptos.]**. Al compara con una $\chi^2(7-1-2)$ se tiene que

$$2.5\,\% < \text{valor-p} < 5\,\% \qquad \textbf{[0.5 Ptos.]}$$

Por lo tanto, existe suficiente evidencia para rechazar que las temperaturas mínimas en los últimos 116 días se ajustan a un modelo Sev. **[0.5 Ptos.]**

+ 1 Punto Base

Problema 2

Se ha llevado a cabo un análisis de regresión múltiple que busca explicar el rendimiento de los alumnos/as durante el primer semestre (Y) en función de NEM (X_1), PSU Lenguaje (X_2), PSU Matemáticas (X_3) y Test de habilidades sociales (X_4) que corresponde a un puntaje en escala 0 (nada sociable) a 100 (muy sociable). Para una muestra de 34 alumnos/as, se han ajustado varios modelos

Modelo	Variables Incluidas	SCE
0	Constante	245
1	X_1	185
2	X_3	160
3	X_1 X_3	138
4	X_1 X_2	168
5	X_1 X_4	170
6	X_1 X_2 X_4	162
7	X_1 X_2 X_3 X_4	126

(a) **[3.0 Ptos.]** ¿Es significativo el aporte conjunto de X_2 y X_4? Puede compararse con más de un modelo. Use $\alpha = 5\,\%$

(b) **[3.0 Ptos.]** Para el Modelo 3, obtenga el estadístico F y coeficiente de determinación R^2. Interprete. Use $\alpha = 10\,\%$

Solución

(a) Comparación de modelos que incluyen a X_2 y X_4 vs. los que no lo incluyen:

1. $\mathrm{H}_0 : \mathrm{Mod}_1$ vs $\mathrm{H}_a : \mathrm{Mod}_6$ **[0.5 Ptos.]**

   Test F:
$$F = \frac{(185 - 162)/2}{162/30} = 2.1296 \qquad \textbf{[0.5 Ptos.]}$$
   Comparando con $F_{0.95}(2,\, 30) = 3.316$, concluimos que no hay evidencia contra H_0, es decir, X_2 y X_4 no aportan en presencia de X_1. **[0.5 Ptos.]**

2. $\mathrm{H}_0 : \mathrm{Mod}_3$ vs $\mathrm{H}_a : \mathrm{Mod}_7$ **[0.5 Ptos.]**

   Test F:
$$F = \frac{(138 - 126)/2}{126/29} = 1.3809 \qquad \textbf{[0.5 Ptos.]}$$
   Comparando con $F_{0.95}(2,\, 29) = 3.328$, concluimos que no hay evidencia contra H_0, es decir, X_2 y X_4 no aportan en presencia de X_1 y X_3. **[0.5 Ptos.]**

(b) Para el modelo 3, tenemos que
$$F = \frac{(245 - 138)/2}{138/31} = 12.01812 \qquad \textbf{[0.5 Ptos.]}$$
Comparando con $F_{0.90}(2,\, 31) = 2.482$, concluimos que existe suficiente evidencia contra H_0, es decir, X_1 y X_3 explican de manera significativa el comportamiento de Y. **[1.0 Ptos.]**

Por otra parte
$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{138}{245} = 0.4367347 \qquad \textbf{[0.5 Ptos.]}$$
Es decir, el 43,7 % de la variabilidad del rendimiento (o de Y) puede ser explicado por el modelo (o equivalentemente, por X_1-nem y X_3-PsuMat) **[1.0 Ptos.]**

+ 1 Punto Base

Problema 3

Un estudio busca conocer la opinión y valoración que tienen los alumnos respecto al cambio del sistema de toma de ramos (Pucmático vs Banner). Parte de la encuesta utilizada es la que sigue:

- P_1: ¿Cuál es el grupo de prioridad académica? : ____
- P_2: ¿Está de acuerdo con el cambio propuesto?: Si ____ No ____

Si en P_2 su respuesta es Si responda:

- P_3: ¿Cuál es tu valoración del sistema anterior (Pucmático)?: ____ (nota entre 1 y 7)

Un breve resumen estadístico se presenta a continuación: Categoría

P_1: Grupo	Número de Encuestados	P_2: ¿Acuerdo?		P_3: Valoración	
		Si	No	Promedio	Desviación Estándar
1 a 5	58	22	36	4,3	1,6
6 a 10	82	21	61	4,9	1,0
Total	140	43	97	4,6	1,1

(a) **[2.0 Ptos.]** ¿Existe evidencia - para un nivel de significancia del 10 % - que permita afirmar que las proporciones de "acuerdo" con el cambio a Banner difieren según grupo? (Sea explícito, hipótesis, test y valor-p)

(b) **[1.0 Ptos.]** Se puede afirmar que las varianzas de las valoraciones son iguales? Use $\alpha = 10\,\%$.

(c) **[3.0 Ptos.]** ¿Existe evidencia que permita afirmar que los de menor prioridad académica (grupo 6 a 10) tenían una mejor valoración del sistema anterior en relación a los alumnos de mejor prioridad académica? (Sea explícito, supuestos, hipótesis, test y valor-p). Use $\alpha = 5\,\%$

Solución

(a) Hipótesis:

$$\mathrm{H}_0 : p_A = p_B \quad \text{vs} \quad \mathrm{H}_a : p_A \neq p_B \qquad \textbf{[0.5 Ptos.]}$$

Los estimadores son

$$\hat{p}_A = \frac{22}{58} = 0.3793103 \quad \text{y} \quad \hat{p}_B = \frac{21}{82} = 0.2560976$$

Bajo H_0 el estadístico de prueba sería

$$Z_0 = \frac{\hat{p}_A - \hat{p}_B}{\hat{p}\,(1-\hat{p})\sqrt{\frac{1}{58}+\frac{1}{82}}} \sim \text{Normal}(0,1)$$

donde

$$\hat{p} = \frac{22+21}{58+82} = 0.3071429 \qquad \textbf{[0.3 Ptos.]}$$

reemplazando $Z_0 = 1.554$ **[0.5 Ptos.]**, lo que implica un valor-p igual a

$$\text{valor-p} = 2 \cdot [1 - \Phi(|1.554|)] = 2 \cdot (1 - 0.94) = 0.12 > 0.10 = \alpha \qquad \textbf{[0.5 Ptos.]}$$

por tanto, NO se rechaza H_0, es decir, no se puede afirmar que difieran. **[0.2 Ptos.]**

(b) Hipótesis, por mayor prioridad:

$$\mathrm{H}_0 : \sigma_A^2 = \sigma_B^2 \quad \text{vs} \quad \mathrm{H}_a : \sigma_A^2 \neq \sigma_B^2 \qquad \textbf{[0.2 Ptos.]}$$

Bajo H_0 se tiene que

$$F = \frac{s_A^2}{s_B^2} = \frac{1.6^2}{1.0^2} = 2.56 \sim F(22-1,\, 21-1) \qquad \textbf{[0.4 Ptos.]}$$

Al comparar con $F_{1-\alpha/2}(21,\, 20) = 2.12$ **[0.2 Ptos.]**, se tiene que $F > F_{1-\alpha/2}$ por lo cual se rechaza H_0, es decir, las varianzas difieren. **[0.2 Ptos.]**

(c) Por (b) se realizará un test de media con con varianzas desconocidas y iguales **[0.5 Ptos.]**

Hipótesis:

$$\mathrm{H}_0 : \mu_A = \mu_B \quad \text{vs} \quad \mathrm{H}_a : \mu_A < \mu_B \qquad \textbf{[0.5 Ptos.]}$$

Bajo H_0 se tiene que

$$T = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{s_A^2}{n} + \frac{s_B^2}{m}}} \sim t - \text{Student}(\nu) \qquad \textbf{[0.3 Ptos.]}$$

con

$$\nu = \left[\frac{\left(S_X^2/n + S_Y^2/m\right)^2}{\frac{\left(S_X^2/n\right)^2}{n-1} + \frac{\left(S_Y^2/m\right)^2}{m-1}} \right] = 35.46767 \qquad \textbf{[0.2 Ptos.]}$$

reemplazando $T_0 = -1.482$ **[0.5 Ptos.]**, lo que implica un valor-p (aproximando por una Normal, ya que $\nu > 30$) igual a

$$\text{valor-p} \approx \Phi(-1.482) = 1 - \Phi(1.482) = 1 - 0.9306 = 0.0694 \qquad \textbf{[0.5 Ptos.]}$$

Como valor-p no es menor que el nivel de significancia $\alpha = 5\,\%$, no se rechaza H_0, es decir, no existe evidencia que permite afirmar que los de menor prioridad tenían una mejor valoración. **[0.5 Ptos.]**

+ 1 Punto Base

Problema 4

Un grupo de investigadores tiene pensado realizar una encuesta de salud, pero tienen dudas si el tamaño muestral lo calculan a partir de un intervalo de confianza o una prueba de hipótesis. Para ambos casos, los investigadores tomaran como referencia lo reportado en la Encuesta Nacional de Salud 2009 sobre nivel de colesterol total de los chilenos mayores a 15 años, es decir un promedio de 189 mg/dL con un c.o.v igual a 22 %, además del supuesto de Normalidad.

(a) **[3.0 Ptos.]** Con un nivel de confianza del 95 % y un error de estimación igual a 5 mg/dL, ¿cuál debe ser el tamaño de la muestra?

(b) **[3.0 Ptos.]** Para contrastar las siguientes hipótesis

$$\mathrm{H}_0 : \mu = 189 \quad \text{vs} \quad \mathrm{H}_a : \mu < 189$$

¿Cuál debería ser el tamaño de muestra, si se queré cometer una probabilidad de error tipo I igual a 10 % (α) con una potencia del 80 % ($1 - \beta$), para detectar una diferencia de un 5 % con respecto a la media de referencia (Δ)?

Solución

(a) Tenemos que $\sigma = 189 \cdot .22 = 41.58$ **[0.5 Ptos.]**, mientras que el error de estimación está dado por:

$$EE = k_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad \textbf{[1.0 Ptos.]}$$

Para un nivel de confianza del 95 % y un $EE = 5$, se tiene que

$$\textbf{[1.0 Ptos.]} \qquad n = \left(\frac{41.48 \times 1.96}{5} \right)^2 = 265.6691 \approx 266 \qquad \textbf{[0.5 Ptos.]}$$

(b) Tenemos que $\alpha = 10\,\%$, $\beta = 20\,\%$, $\sigma = 41.58$ y $\Delta = 189 \cdot 0.05 = 9.45$ **[0.5 Ptos.]**

La potencia para las hipótesis propuesta esta dada por

$$1 - \beta = \Phi\left(k_\alpha - \frac{\Delta\sqrt{n}}{\sigma} \right) \qquad \textbf{[1.0 Ptos.]}$$

despejando y reemplazado

$$\textbf{[1.0 Ptos.]} \qquad n = \left[\frac{(k_\beta + k_\alpha) \cdot \sigma}{\Delta} \right]^2 = \left[\frac{-(0.84 + 1.28) \cdot 189 \cdot 0.22}{189 \cdot 0.05} \right]^2 = 87.01158 \approx 87 \qquad \textbf{[0.5 Ptos.]}$$

+ 1 Punto Base