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.**

INTERROGACIÓN 3

Cierre Refinería Ventanas

Hoy en día se discute sobre el cierre de la refinería Ventanas de Codelco (comuna de Puchuncaví). Usted interesado en saber un poco más con el fin de formarse una opinión profesional, revisa información ambiental de los tres primeros meses del año 2019 y 2022, específicamente sobre los niveles de dióxido de azufre (SO_2).

Parte de los resultados se presenta a continuación:

```
                                2019 2022
Número de días medidos (*)        84   65
Numero de días sobre norma (**)   26   28

(*) : Hay días sin mediciones
(**): Medición sobre 150 µg/m3

Resumen para SO2 en días sobre norma:

     2019 2022
mean  165  174
sd     24   33
```

Pregunta 1

¿Es posible afirmar que durante el primer trimestre del presente año, el nivel medio de SO_2, en los días que ha sobrepasado la norma, ha sido inferior a 185 µg/m3? Considere un nivel de significancia del 2.5 % y que los niveles medios de SO_2 distribuyen Normal.

Solución

Definamos como X a la información del 1er trimestre del presente año.

Para α igual a 2.5 %, se pide contrastar las siguientes hipótesis

$$\text{H}_0 : \mu_X = \mu_0 \quad \text{vs} \quad \text{H}_a : \mu_X < \mu_0, \quad \textbf{[0.2 Ptos.]}$$

con $\mu_0 = 185$. **[0.1 Ptos.]**

Bajo H_0 se tiene que

$$T_0 = \frac{\overline{X}_n - \mu_0}{S_X/\sqrt{n}} \sim \text{t-Student}(n-1). \quad \textbf{[0.1 Ptos.]}$$

Del enunciado tenemos que

$$n = 28, \quad \overline{X}_n = 174, \quad S_X = 33. \quad \textbf{[0.1 Ptos.]}$$

Reemplazando, $T_0 = -1.763834$. **[0.2 Ptos.]**

Alternativa 2: Como $T_0 = -1.763834 > -2,052 = t_{2.5\,\%}(27)$, no existe suficiente evidencia, basado en la información del 1er trimestre de este año, para rechazar H_0 y apoyar la afirmación que durante el primer trimestre del presente año, el nivel medio de SO_2, en los días que ha sobrepasado la norma, ha sido inferior a 185 $\mathtt{µg/m3}$. **[0.3 Ptos.]**

Alternativa 1: Como $t_{2.5\,\%}(27) = -2,052 < -1.763834 < -1.703 = t_{5\,\%}(27)$, entonces $2.5\,\% <$ valor-p $< 5\,\%$, por lo tanto no existe suficiente evidencia, basado en la información del 1er trimestre de este año, para rechazar H_0 y apoyar la afirmación que durante el primer trimestre del presente año, el nivel medio de SO_2, en los días que ha sobrepasado la norma, ha sido inferior a 185 $\mathtt{µg/m3}$. **[0.3 Ptos.]**

Pregunta 2

¿Es posible afirmar que la proporción de días sobre la norma durante el presente año supera significativamente a la proporción de días sobre la norma del año 2019? Considere un nivel de significancia del 5 %.

Solución

Definamos como X al año 2019 e Y al año 2022.

Para α igual a 5 %, se pide contrastar las siguientes hipótesis

$$\text{H}_0 : p_X = p_Y \quad \text{vs} \quad \text{H}_a : p_X < p_Y. \quad \textbf{[0.1 Ptos.]}$$

Bajo H_0 se tiene que

$$Z_0 = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\hat{p}\,(1-\hat{p})} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \overset{\text{aprox}}{\sim} \text{Normal}(0, 1), \quad \textbf{[0.2 Ptos.]}$$

con $\hat{p} = \dfrac{n \cdot \overline{X}_n + m \cdot \overline{Y}_m}{n + m}$. **[0.1 Ptos.]**

Del enunciado tenemos que

$$n = 84, \quad m = 65, \quad \overline{X}_n = \frac{26}{84}, \quad \overline{Y}_m = \frac{28}{65}. \quad \textbf{[0.1 Ptos.]}$$

Reemplazando

$$\textbf{[0.1 Ptos.]} \quad \hat{p} = \frac{26 + 28}{84 + 65} \to Z_0 = -1.513754 \quad \textbf{[0.1 Ptos.]}$$

Alternativa 1: Como valor-p $\approx \Phi(-1.51) = 1 - \Phi(1.51) = 1 - 0.9345 = 0.065 > 0.05 = \alpha$, no existe suficiente evidencia, basado en la información de los primeros meses de cada año, para rechazar H_0 y apoyar la afirmación que la proporción de días sobre la norma durante el presente año supera significativamente a la proporción de días sobre la norma del año 2019. **[0.3 Ptos.]**

Alternativa 2: Como $Z_0 = -1.513754 > -1.645 = k_{5\,\%}$, no existe suficiente evidencia, basado en la información de los primeros meses de cada año para rechazar H_0 y apoyar la afirmación que la proporción de días sobre la norma durante el presente año supera significativamente a la proporción de días sobre la norma del año 2019. **[0.3 Ptos.]**

Observación: Si el alumno define X e Y de manera inversa, el estadístico de prueba cambia de signo, la dirección de Ha se invierte. El valor p será el mismo, el valor critico para comparar Z_0 tendrá signo cambiado, pero la conclusión no cambiara.

Pregunta 3

¿Es posible afirmar que el nivel medio de SO_2 durante el 2022 supera significativamente al nivel medio de SO_2 del 2019? Considere un nivel de significancia del 10 % y que los niveles medios de SO_2 distribuyen Normal.

Solución

Definamos como X a la muestra 2019 e Y a la muestra 2022.

Se pide realizar un test de comparación de medias, con varianzas desconcidas, para α igual a 10 %.

Primero realizaremos un test para ver si existe evidencia que nos permita asumir varianzas distintas:

$$\text{H}_0 : \sigma_Y = \sigma_X \quad \text{vs} \quad \text{H}_a : \sigma_Y \neq \sigma_X \quad \textbf{[0.1 Ptos.]}$$

Bajo H_0 se tiene que

$$F_0 = \frac{S_X^2}{S_Y^2} = \frac{24^2}{33^2} = 0.5289256 \sim F(26-1, 28-1) \quad \textbf{[0.2 Ptos.]}$$

Como

$$0.5289256 = F_0 > F_{0.10/2}(26-1, 28-1) = \frac{1}{F_{1-0.10/2}(28-1, 26-1)} = \frac{1}{1.94} = 0.5154639,$$

no es posible afirmar que las varianzas difieren. **[0.1 Ptos.]**

Nota: Si utiliza $F_0 = S_Y^2/S_X^2 = 1.890625 < 1.94 = F_{1-0.10/2}(28-1, 26-1)$ asignar todo el puntaje si concluye que no es posible afirmar que las varianzas difieren.

$$\text{H}_0 : \mu_X = \mu_Y \quad \text{vs} \quad \text{H}_a : \mu_X < \mu_Y \quad \textbf{[0.1 Ptos.]}$$

Bajo H_0, con varianzas desconocidas e iguales, se tiene que

$$T_0 = \frac{165 - 174}{Sp\sqrt{\frac{1}{26} + \frac{1}{28}}} = -1.138573 \sim \text{t-Student}(26 + 28 - 2) \approx \text{Normal}(0, 1), \quad \textbf{[0.2 Ptos.]}$$

donde $S_p = 29.02353$. **[0.1 Ptos.]**

Como valor-p $\approx \Phi(-1.14) = 1 - \Phi(1.14) = 1 - 0.8729 = 0.1271 > 0.10 = \alpha$, entonces NO es posible rechazar H_0 y apoyar la afirmación que los niveles medios de SO_2 en 2022 superan a lo observado en 2019. **[0.2 Ptos.]**

Observación: Si el alumno define X e Y de manera inversa, el estadístico de prueba cambia de signo, la dirección de Ha se invierte. El valor p será el mismo, el valor critico para comparar Z_0 tendrá signo cambiado, pero la conclusión no cambiara.

Pregunta 4

Dada la relación del dióxido de azufre (SO_2), en ppb, y la temperatura, en °C, que se observa en el gráfico de dispersión (Figura 1), usted ajusta un modelo de regresión lineal simple. Este modelo busca determinar el efecto de la temperatura sobre los niveles de SO_2 a partir de promedios diarios, los cuales fueron medidos en la estación de monitoreo de Puchuncaví durante enero de 2022 (31 días).

Fuente: https://sinca.mma.gob.cl/

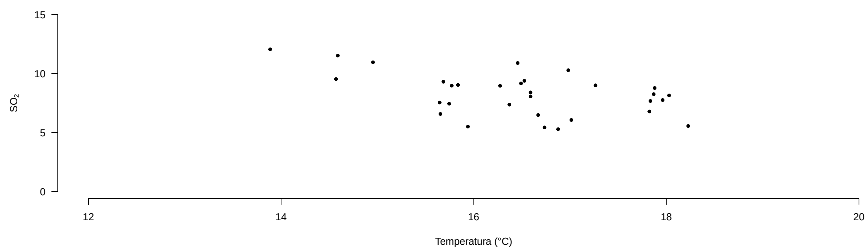


Figura 1: Relación niveles medios diarios SO_2 vs Temperatura promedio diaria

```
lm(formula = SO2 ~ TEM, data = dplyr::filter(Data, year == 2022, month == 1))

Coefficients:
            Estimate Std. Error
(Intercept)  21.6842     4.3423
TEM          -0.8148     0.2630

Residual standard error: 1.593 (Sy|x)

var(dplyr::filter(Data, year == 2022, month == 1)$SO2)
[1] 3.26559
```

(a) Considerando un nivel de significancia del 1 %, ¿es posible afirmar que la temperatura promedio diaria explica los niveles medios diarios de SO_2 en el sector de Puchuncaví?

(b) ¿Qué porcentaje de la variabilidad observada para SO_2 es explicado por este modelo? ¿Cuál sería la variación en los niveles medios esperados de SO_2 por cada dos grados de aumento en la temperatura media?

Solución

(a) Se pide verificar si existe evidencia para afirmar que $\beta_{\texttt{TEM}} \neq 0$.

Tenemos que

$$\textbf{[0.1 Ptos.]} \quad \texttt{t-value} = -\frac{0.8148}{0.2630} = -3.098099 \sim \text{t-Studente}(31-2) \quad \textbf{[0.1 Ptos.]}$$

Alternativa 1: De la tabla de percentiles t-Student se tiene que:

$$\textbf{[0.1 Ptos.]} \quad |\texttt{t-value}| = 3.098099 > t_{0.99}(29) \to \frac{\texttt{p-value}}{2} < 1\,\% \to \texttt{p-value} < 2\,\% \quad \textbf{[0.1 Ptos.]}$$

Por lo tanto con la información disponible desde la tabla, no es posible confirmar al 1 % de significancia que la temperatura promedio diaria explica lo niveles de SO_2, ya que solo tenemos certeza que el valor-p es inferior a 2 %, pero no al 1 %. **[0.1 Ptos.]**

Alternativa 2: Dado el signo negativo de la pendiente se podría buscar evidencia para afirmar que $\beta_{\texttt{TEM}} < 0$.

En este caso

$$\textbf{[0.1 Ptos.]} \quad \texttt{t-value} = -3.098099 < -t_{0.99}(29) \to \texttt{p-value} < 1\,\% \quad \textbf{[0.1 Ptos.]}$$

y si se podría afirmar al 1 % de significancia que que la temperatura promedio diaria explica inversamente los niveles de SO_2. **[0.1 Ptos.]**

Alternativa 3: Aproximar t-Student(29) por Normal(0,1).

Como

$$\texttt{p-value} \approx 2 \cdot [1 - \Phi(|-3.098|)] = 2 \cdot [1 - 0.9990] = 0.002 < 0.01 = \alpha \quad \textbf{[0.2 Ptos.]}$$

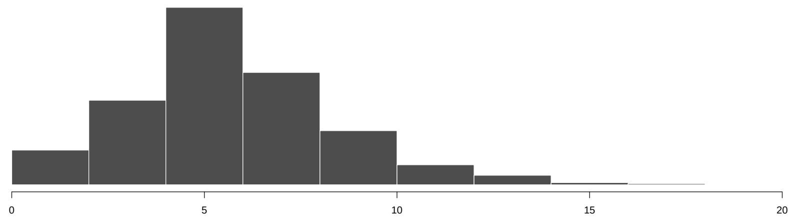
si se podría afirmar al 1 % de significancia que que la temperatura promedio diaria explica los niveles de SO_2. **[0.1 Ptos.]**

(b) El coeficiente de determinación (ajustado) es igual a $1 - \dfrac{1.593^2}{3.26559} = 0.2229126$, por lo tanto este modelo explica el 22.3 % de la variabilidad observada para SO_2. **[0.3 Ptos.]**

Notar que por cada dos grados de temperatura los niveles de SO_2 bajan en -1.6296 ppb. **[0.2 Ptos.]**

Pregunta 5

Los niveles medios diarios de SO_2 registrados entre 2018 y lo que ha transcurrido de este año presentan un comportamiento asimétrico como muestra el histograma.



Si se ajusta un modelo Weibull(η, β) con β conocido e igual a 2, ¿entre qué valores, aproximadamente, se encuentra el parámetro η con un 95 % de confianza? Utilice el estimador máximo verosímil para responder.

Puede ser se utilidad la siguiente información empírica:

```
  n mean(X) mean(X^2) mean(X^3) mean(X^4)
945    5.77    40.282   329.595  3113.951
```

Solución

Un intervalo al 95 % de confianza para η basado en el estimador máximo verosímil está dado por

$$< \eta >_{0.95} \in \ \hat{\eta} \pm 1.96 \cdot \sqrt{\text{CCR}(\hat{\eta})} \quad \textbf{[0.1 Ptos.]}$$

El EMV de η lo obtenemos de la siguiente manera:

$$L(\eta) = \left(\frac{\beta}{\eta}\right)^n \cdot \left(\eta^{\beta-1}\right)^{-n} \cdot \left(\prod_{i=1}^n X_i\right)^{\beta-1} \cdot \exp\left\{-\frac{1}{\eta^\beta}\sum_{i=1}^n X_i^\beta\right\} \quad \textbf{[0.1 Ptos.]}$$

$$\ln\, L(\eta) = n \cdot \ln(\beta) - n \cdot \ln(\eta) - (\beta - 1) \cdot n \cdot \ln(\eta) + (\beta - 1)\sum_{i=1}^n \ln(X_i) - \frac{1}{\eta^\beta}\sum_{i=1}^n X_i^\beta \quad \textbf{[0.1 Ptos.]}$$

$$\textbf{[0.1 Ptos.]} \quad \frac{\partial}{\partial\,\eta}\ln\, L(\eta) = -\frac{n\,\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}}\sum_{i=1}^n X_i^\beta = 0 \to \hat{\eta} = (\overline{X^\beta})^{1/\beta} = \sqrt{\overline{X^2}} = \sqrt{40.282} = 6.34681 \quad \textbf{[0.1 Ptos.]}$$

y la CCR como sigue:

$$\frac{\partial^2}{\partial\,\eta^2}\ln\, L(\eta) = \frac{n\,\beta}{\eta^2} - \frac{\beta\,(\beta+1)}{\eta^{\beta+2}}\sum_{i=1}^n X_i^\beta \quad \textbf{[0.1 Ptos.]}$$

$$\textbf{[0.1 Ptos.]} \quad \text{I}(\eta) = -\frac{n\,\beta}{\eta^2} + \frac{\beta\,(\beta+1)}{\eta^{\beta+2}}\sum_{i=1}^n \text{E}\left(X_i^\beta\right) = -\frac{n\,\beta}{\eta^2} + \frac{\beta\,(\beta+1)}{\eta^{\beta+2}} \cdot n \cdot \eta^\beta\,\Gamma\,(1 + \beta/\beta) = \frac{\beta^2\,n}{\eta^2} \quad \textbf{[0.1 Ptos.]}$$

$$\text{CCR}(\eta) = \frac{\eta^2}{\beta^2\,n} = \frac{\eta^2}{4\,n} \to \sqrt{\text{CCR}(\hat{\eta})} = \sqrt{\frac{6.34681^2}{4 \cdot 945}} = 0.1032309 \quad \textbf{[0.1 Ptos.]}$$

Por lo tanto

$$< \eta >_{0.95} \in [6.14451 \ - \ 6.549177] \quad \textbf{[0.1 Ptos.]}$$

Pregunta 6

Para los datos del ejercicio anterior se propone ajustar un modelo Gamma$(k,\, \nu)$, con k conocido.

(a) ¿El método de momentos entrega un estimador consistente para ν? Si lo desea puede responder en base a momentos exactos o aproximados de 1er orden por método delta.

(b) Si k es igual a 5, ¿qué valor usted estima para ν?

Solución

(a) Igualando el primer momento teórico de una Gamma$(k,\, \nu)$ a $\overline{X}_n$ se tiene que el estimador de momento de ν es:

$$\hat{\nu} = \frac{k}{\overline{X}_n} \quad \textbf{[0.1 Ptos.]}$$

Se pide probar si es consistente, es decir, chequear si su ECM tiende a cero cuando $n \to \infty$.

ECM exacto: Como $\overline{X}_n \sim$ Gamma$(n\,k,\, n\,\nu)$, **[0.1 Ptos.]** entonces

$$\text{E}(\hat{\nu}) = k \int_0^\infty \frac{1}{y} \cdot \frac{(n\,\nu)^{n\,k}}{\Gamma(n\,k)} \cdot y^{n\,k-1} \cdot e^{-n\,\nu\,y}\,dy = \frac{k\,n\,\nu}{n\,k-1} = \left(\frac{\nu}{1-\frac{1}{n\,k}}\right) \quad \textbf{[0.1 Ptos.]}$$

$$\text{E}(\hat{\nu}^2) = k^2 \int_0^\infty \frac{1}{y^2} \cdot \frac{(n\,\nu)^{n\,k}}{\Gamma(n\,k)} \cdot y^{n\,k-1} \cdot e^{-n\,\nu\,y}\,dy = \frac{k^2\,n^2\,\nu^2}{(n\,k-1)\,(n\,k-2)} = \left(\frac{\nu}{1-\frac{1}{n\,k}}\right) \cdot \left(\frac{\nu}{1-\frac{2}{n\,k}}\right) \quad \textbf{[0.1 Ptos.]}$$

$$\text{Var}(\hat{\nu}) = \left(\frac{\nu}{1-\frac{1}{n\,k}}\right) \cdot \left(\frac{\nu}{1-\frac{2}{n\,k}}\right) - \left(\frac{\nu}{1-\frac{1}{n\,k}}\right)^2 \quad \textbf{[0.1 Ptos.]}$$

Luego

$$\text{ECM}(\hat{\nu}) = \text{Var}(\hat{\nu}) + [\text{E}(\hat{\nu}) - \nu]^2 \to (\nu^2 - \nu^2) + [\nu - \nu]^2 = 0, \quad \text{cuando } n \to \infty. \quad \textbf{[0.1 Ptos.]}$$

Por lo tanto, $\hat{\nu}$ es un estimador consistente para ν. **[0.1 Ptos.]**

ECM aproximado: Tenemos que la aproximación de 1er orden para $\hat{\nu}$ está dada por:

$$\hat{\nu} \approx \nu + \sum_{i=1}^n \left(X_i - \frac{k}{\nu}\right) \cdot \left(-\frac{\nu^2}{n\,k}\right) \quad \textbf{[0.2 Ptos.]}$$

Por lo tanto

$$\textbf{[0.1 Ptos.]} \quad \text{E}(\hat{\nu}) \approx \nu \quad \text{y} \quad \text{Var}(\hat{\nu}) \approx \frac{\nu^2}{n\,k} \quad \textbf{[0.1 Ptos.]}$$

Luego

$$\text{ECM}(\hat{\nu}) = \text{Var}(\hat{\nu}) + [\text{E}(\hat{\nu}) - \nu]^2 \approx \frac{\nu^2}{n\,k} + (\nu - \nu)^2 \to 0, \quad \text{cuando } n \to \infty. \quad \textbf{[0.1 Ptos.]}$$

Por lo tanto, $\hat{\nu}$ es un estimador consistente para ν. **[0.1 Ptos.]**

(b) Si k es igual a 5, entonces el estimador de ν sería $\dfrac{5}{5.77} = 0.8665511.$ **[0.3 Ptos.]**