Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Temporada Académica de Verano 2019

Curso	**:**	**Probabilidad y Estadística**
Sigla	**:**	**EYP1113**
Profesores	**:**	**Ricardo Aravena C. (Sec 1), Ricardo Olea O. (Sec 2) y Alonso Molina N. (Sec 3)**

Pauta Interrogación 4

Pregunta 1

Como han de saber, los puntajes de la PSU Mat siguen una distribución Normal(μ = 500, σ = 110). El director de un establecimiento de la Capital indica que sus alumnos son "Normales", es decir, distribuyen según la normativa, y para probar lo anterior muestra el rendimiento de la última promoción en esta prueba. Los resultados son los siguentes:

Categoría	Número de Casos
<400	12
[400-500)	18
[500-600)	22
>=600	8

¿Tiene razón la afirmación del Director? Sea explícito: plantee hipótesis, indique y obtenga el test, concluya para nivel de significancia del 5 %. En sus resultados utilice dos decimales.

Solución

$$\text{H}_0 : X \sim \text{Normal}(500,\, 110) \qquad \text{vs} \qquad \text{H}_a : X \not\sim \text{Normal}(500,\, 110)$$

[0.5 Ptos]

	Observado	Probabilidad	Esperado	(O-E)^2/E
<400	12	0.18	10.8	0.13
[400-500)	18	0.32	19.2	0.07
(500-600]	22	0.32	19.2	0.41
>=600	8	0.18	10.8	0.73
Total	60	1.00	60.0	1.34

[3.0 Ptos]

Bajo H_0 se tiene que

[0.5 Ptos] $X^2 = 1{,}34 \sim \chi^2(4{-}1{-}0) \to [70\,\%\text{valor-p} < 80\,\%]$ **[1.0 Ptos]** o $\left[X^2 = 1{,}34 < 7{,}815 = c_{0{,}95}(4-1-0)\right]$

Por lo tanto, al 5 % de significancia, no existe evidencia para rechazar la hipótesis de "Normalidad" de los alumnos del establecimiento en la PSU Mat. **[1.0 Ptos]**

+ 1 Punto Base

Pregunta 2

Se afirma que el rendimiento del 1er año en las carreras de ingeniería está fuertemente influenciado por las notas de enseñanza media (NEM) y el rendimiento en la PSU Matemática (PMAT). Usted, buscando validar la afirmación en términos estadísticos selecciona una muestra al azar de 26 alumnos de ingeniería, para los cuales dispone del PPA (Promedio Ponderado Acumulado) del 1er año. Una descripción de este PPA entrega una nota promedio de 4,86 con una desviación estándar de 1,09. Se ajustan dos modelos de regresión lineal simple, los cuales se presentan a continuación:

```
lm(formula = PPA ~ NEM)
Coefficients:
            Estimate  Std. Error     t value   Pr(>|t|)
(Intercept)  -3.1602      2.5671        XXXX    0.23024
NEM           XXXXXX      0.0036        3.13    XXXXXXX

Residual standard error: 0.9373 on XXX degrees of freedom
Multiple R-squared:  XXXXXX,    Adjusted R-squared:  0.2606
F-statistic: XXXXXX on 1 and XXX  DF,  p-value: 0.0045
```

```
lm(formula = PPA ~ PMAT)
Coefficients:
            Estimate    Std. Error    t value   Pr(>|t|)
(Intercept) -4.8173        XXXXXX     -2.017    XXXXXXX
PMAT         0.0139        0.0034     XXXXXX    0.00045

Residual standard error:  XXXXXX on XXX degrees of freedom
Multiple R-squared:  0.4074,    Adjusted R-squared:  XXXXXX
F-statistic:  16.5 on XXX   and XXX DF,  p-value:  XXXXXXX
```

Completar salida R e indique justificadamente cuál ajusta mejor.

Solución

```
lm(formula = PPA ~ NEM)
Coefficients:
            Estimate  Std. Error     t value   Pr(>|t|)
(Intercept)                            -1.23
NEM           0.0113                             0.0045

Residual standard error:        on 24  degrees of freedom
Multiple R-squared:  0.29014,    Adjusted R-squared:
F-statistic: 9.7969 on   and 24   DF,  p-value:
```

```
lm(formula = PPA ~ PMAT)
Coefficients:
            Estimate    Std. Error    t value   Pr(>|t|)
(Intercept)                2.3884                  >0.05
PMAT                                   4.088

Residual standard error:  0.8564 on 24  degrees of freedom
Multiple R-squared:       ,    Adjusted R-squared:  0.3827
F-statistic:       on 1 and 24  DF,  p-value:  0.00045
```

+[0.3 Ptos] por cada valor

En el modelo PPA $\sim$ PMAT, se observa que el t-value de la pendiente es más significativo, lo mismo ocurre con el F-statistic. Los coeficientes de determinación son mayores en este modelo. Por lo tanto, el 2do modelo de regresión lineal ajusta mejor. **[1,2 Ptos]**

+ 1 Punto Base

Pregunta 3

Se dispone de información relativa al rendimiento de 24 alumnos de cierta Universidad. Usted está interesado en determinar si las variables X_1: NEM, X_2: PSU Leng y X_3: PSU Mat influyen en el rendimiento del 1er semestre. Se ajustan diversos modelos de regresión lineal para explicar el rendimiento del 1er semestre y algunos resultados son:

```
---------------------
Modelo  Variables SCE
---------------------
     0       NULO 240
     1         X1 180
     2         X2 195
     3         X3 170
     4     X2, X3 150
     5 X1, X2, X3 135
---------------------
```

(a) **[2.0 Ptos]** Obtenga TODO lo que pueda del modelo de regresión lineal simple que utiliza PSU Mat (X_3) para predecir rendimiento en el 1er semestre.

(b) **[2.0 Ptos]** ¿Es significativo el aporte de NEM (X_1)?

(c) **[2.0 Ptos]** ¿Es significativo el aporte conjunto de NEM (X_1), PSU Leng (X_2) y PSU Mat (X_3)?

Solución

(a) Con la información disponible, para el modelo 3, es posible obtener

$$\text{SCT} = 240 \quad \textbf{[0.2 Ptos]}$$
$$\text{Multiple R-squared} = 0{,}2917 \quad \textbf{[0.2 Ptos]}$$
$$\text{Adjusted R-squared} = 0{,}2595 \quad \textbf{[0.2 Ptos]}$$
$$\textbf{[0.2 Ptos]} \quad \text{F-statistic} = 9{,}06 \sim F(1,\,22) \quad \textbf{[0.2 Ptos]} \rightarrow \text{p - value} < 0{,}05 \quad \textbf{[0.2 Ptos]}$$
$$\textbf{[0.2 Ptos]} \quad \text{t-value} = 3{,}01 \sim \text{t-Student}(22) \quad \textbf{[0.2 Ptos]} \rightarrow \text{p - value} < 0{,}05 \quad \textbf{[0.2 Ptos]}$$
$$\text{Residual Stand Error} = 2{,}78 \quad \textbf{[0.2 Ptos]}$$

(b) Aporte NEM modelo 0 vs modelo 1:

$$\text{H}_0 : \beta_1 = 0 \quad \text{vs} \quad \text{H}_a : \beta_1 \neq 0$$

$$\textbf{[0.4 Ptos]} \quad F = \frac{(SCE_{\text{mod0}} - SCE_{\text{mod1}})/1}{SCE_{\text{mod1}}/(24-(0+1)-1)} = 7{,}33 \sim F(1, 22) \rightarrow F = 7{,}33 > 4{,}30 = F_{0{,}95}(1, 22) \quad \textbf{[0.4 Ptos]}$$

Es decir, el aporte de X_1 en un modelo simple es significativo. **[0.2 Ptos]**

Aporte NEM modelo 4 vs modelo 5:

$$\text{H}_0 : \beta_1 = 0 \quad \text{vs} \quad \text{H}_a : \beta_1 \neq 0$$

$$\textbf{[0.4 Ptos]} \quad F = \frac{(SCE_{\text{mod4}} - SCE_{\text{mod5}})/1}{SCE_{\text{mod5}}/(24-(2+1)-1)} = 2{,}22 \sim F(1, 20) \rightarrow F = 2{,}22 < 4{,}35 = F_{0{,}95}(1, 20) \quad \textbf{[0.4 Ptos]}$$

Es decir, el aporte de X_1 no es significativo en presencia de X_2 y X_3. **[0.2 Ptos]**

(c) Aporte NEM (X_1), PSU Leng (X_2) y PSU Mat (X_3) modelo 0 vs modelo 5:

$$\mathrm{H}_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs} \quad \mathrm{H}_a : \text{Al menos un } \beta_j \neq 0 \quad \textbf{[0.3 Ptos]}$$

$$\textbf{[1.0 Ptos]} \quad F = \frac{(SCE_{\text{mod0}} - SCE_{\text{mod5}})/3}{SCE_{\text{mod5}}/(24-(0+3)-1)} = 5{,}18 \sim F(3,20) \to F = 5{,}18 > 3{,}10 = F_{0{,}95}(3,20) \quad \textbf{[0.5 Ptos]}$$

Es decir, el aporte conjunto de NEM (X_1), PSU Leng (X_2) y PSU Mat (X_3) es significativo, cuando no hay más variables involucradas en el modelo. **[0.2 Ptos]**

+ 1 Punto Base