Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2019

Curso	:	**Probabilidad y Estadística**
Sigla	:	**EYP1113**
Profesores	:	**Ricardo Aravena C. (Sec 01 - Sec 03) y Ricardo Olea O (Sec 02 - Sec 04).**

Pauta Examen

Problema 1

Un especialista afirma que el tiempo de estudio de un estudiante Y (en horas), previo a un examen de un curso específico, depende de su nota de presentación (X_1), distancia (en horas) del próximo examen (X_2), si rinde el curso por primera vez (X_3) y de su PPA (X_4).

Con el fin de analizar el efecto de estas variables, se procede a selecciona una muestra de 42 alumnas/os al azar y consultada respecto al tiempo dedicado al estudio para el examen y junto a las otras característica.

(a) **[2.0 Ptos]** Para el siguiente modelo de regresión lineal múltiple, complete la salida de **R**:

```
sd(Y)
9.7716

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   X.XXXX    0.37676  10.535   <2e-16
X1            2.3177    X.XXXXX   4.277  2.5e-05
X2            0.4627    0.48502   X.XXX  0.34066
X3            1.6141    0.53625   3.010  0.00284
X4            0.6152    0.48942   1.257  X.XXXXX

Residual standard error: 9.144 on XX degrees of freedom
Multiple R-squared:   X.XXXX,        Adjusted R-squared:  X.XXXX
F-statistic: XX.XX on XX and XX DF,  p-value: 4.2419e-08
```

(b) **[4.0 Ptos]** Debido al mal ajuste, un especialista indica que en lugar de usar el tiempo en horas (Y), se debe utilizar $\ln(Y)$. Se ajustan diversos modelos, los cuales arrojaron los siguientes resultados:

Mod	Variables	Adjusted R-squared
1	X1	0.34
2	X1, X3	0.45
3	X2, X3	0.40
4	X1, X2, X4	0.64
5	X1, X2, X3, X4	0.72

¿Es significativo el aporte de la variable X_3?, ¿y el aporte conjunto de X_1, X_4?. Sea explícito, indique qué modelos contrasta, hipótesis correspondiente, test y decisión final. Utilice un nivel de sifnificancia del 5 %.

Solución

(a) La salida completa queda como sigue:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.9692    0.37676  10.535  <2e-16
X1            2.3177    0.54200   4.277  2.5e-05
X2            0.4627    0.48502   0.954  0.34066
X3            1.6141    0.53625   3.010  0.00284
X4            0.6152    0.48942   1.257  0.20760

Residual standard error: 9.144 on 37 degrees of freedom
Multiple R-squared: 0.2097601, Adjusted R-squared: 0.1243288
F-statistic: 2.455307 on 4 and 37 DF,  p-value: 4.2419e-08
```

Logro 1

- Obtener al menos dos valores entre Estimate, Std. Error y t value. **[0.5 Ptos]**
- Obtener valor-p. **[0.5 Ptos]**

Logro 2

- Obtener al menos un R-squared. **[0.5 Ptos]**
- Obtener F-statistic y degrees of freedom. **[0.5 Ptos]**

(b) **Logro 3** *(Obtener a partir de r^2 los R^2)* **[1.0 Ptos]**

Mod	Variables	Adjusted R-squared	Multiple R-squared
1	X1	0.34	0.3560976
2	X1, X3	0.45	0.4768293
3	X2, X3	0.40	0.4292683
4	X1, X2, X4	0.64	0.6663415
5	X1, X2, X3, X4	0.72	0.7473171

Logro 4 *(Mod 1 vs Mod 2)*

$$F = \frac{(R^2_{\text{Mod2}} - R^2_{\text{Mod1}})/1}{(1 - R^2_{\text{Mod2}})/(n-1-1-1)} \sim F(1, n-3)$$

Como

$$F = 9{,}000045 > 4{,}09 = F_{0{,}95}(1, 39), \quad \textbf{[0.5 Ptos]}$$

el aporte de X_3 en presencia de X_1 resulta significativo. **[0.5 Ptos]**

Logro 5 *(Mod 4 vs Mod 5)*

$$F = \frac{(R^2_{\text{Mod5}} - R^2_{\text{Mod4}})/1}{(1 - R^2_{\text{Mod5}})/(n-3-1-1)} \sim F(1, n-5)$$

Como

$$F = 11{,}85714 > 4{,}11 = F_{0{,}95}(1, 37), \quad \textbf{[0.5 Ptos]}$$

el aporte de X_3 en presencia de X_1, X_2 y X_4 resulta significativo. **[0.5 Ptos]**

Logro 6 *(Mod 3 vs Mod 5)*

$$F = \frac{(R^2_{\text{Mod5}} - R^2_{\text{Mod3}})/2}{(1 - R^2_{\text{Mod5}})/(n-2-2-1)} \sim F(2, n-5)$$

Como

$$F = 23{,}28574 > 3{,}25 = F_{0{,}95}(2, 37), \quad \textbf{[0.5 Ptos]}$$

el aporte conjunto de X_1 y X_4 en presencia de X_2 y X_3 resulta significativo. **[0.5 Ptos]**

+ 1 Punto Base

Problema 2

Dado que la expectativa de salario (en UF) al egresar es un factor importante ante la opción de cursar un magister, usted decide realizar un estudio para contrastar los salarios al año de egreso para una muestra de egresados con el título profesional de ingeniero y otra muestra de egresados de algún magister. Parte de los resultados son:

	Profesional	Magister
N	21	13
mean	44	51
sd	10	17

(a) **[3.0 Ptos]** Asumiendo que el salario se comporta como una distribución Normal, ¿existe evidencia que permita afirmar que cursar un magister permite tener un salario mayor que no cursarlo? Indique explícitamente sus hipótesis, supuestos, test y decisiones para un nivel de significancia del 10 %.

(b) **[3.0 Ptos]** Un experto afirma que la conclusión anterior es incorrecta porque los salarios se comportan fuertemente asimétricos, tal como una distribución Exponencial. Lleve a cabo el test correspondiente, indique explícitamente sus hipótesis, supuestos, test y decisiones para un nivel de significancia del 5 %.

Solución

(a) **Logro 1** *Test de Varianzas Completo*

$$\text{H}_0 : \sigma_p = \sigma_m \quad \text{vs} \quad \text{H}_a : \sigma_p \neq \sigma_m]$$

Como

$$F_0 = \frac{17^2}{10^2} = 2{,}89 > F_{0{,}95}(12, 20) = 2{,}28, \quad \textbf{[0.5 Ptos]}$$

se rechaza H_0, es decir, podemos considerar varianzas son desconocidas y distintas. **[0.5 Ptos]**

Logro 2 *Test de comparación de medias + Estadístico de Prueba*

$$\text{H}_0 : \mu_p = \mu_m \quad \text{vs} \quad \text{H}_a : \mu_p < \mu_m$$

Como

$$T_0 = \frac{44 - 51}{\sqrt{\dfrac{10^2}{21} + \dfrac{17^2}{13}}} = -1,347 \sim \text{t-Student}(\nu) \quad \textbf{[0.5 Ptos]}$$

con $v = [17,2] = 17$. **[0.5 Ptos]**

Logro 3 *Valor-p (o valor crítico) + Conclusión*

$$\text{valor-p} = P(T < -1,388) \to 5\,\% < \text{valor-p} < 10\,\% \quad \textbf{[0.5 Ptos]}$$

o

$$T_0 = -1,388 < -1{,}333 = t_{0{,}10}(17) = -t_{0{,}90}(17)$$

Para un nivel de significancia del 10 %, hay evidencia para afirmar que el salario de los magister es estadísticamente superior a los profesionales. **[0.5 Ptos]**

(b) **Logro 4** *Test de medias (Exponencial): Hipótesis y Estadístico de Prueba*

$$\mathrm{H}_0 : \frac{1}{\nu_p} = \frac{1}{\nu_m} \quad \text{vs} \quad \mathrm{H}_a : \frac{1}{\nu_p} < \frac{1}{\nu_m} \quad \textbf{[0.5 Ptos]}$$

Como

$$Z_0 = \frac{44 - 51}{\hat{\nu}\sqrt{\dfrac{1}{21} + \dfrac{1}{13}}} = -0{,}36 \overset{\text{aprox}}{\sim} \text{Normal}(0,1) \quad \textbf{[0.5 Ptos]}$$

con $\hat{\nu} = \dfrac{21 + 13}{21 \cdot 41 + 13 \cdot 51} = 46{,}67647$ estimador de ν bajo H_0.

Logro 5 *Valor-p o valor crítico*

$$\text{valor-p} = \Phi(-0{,}36) = 0{,}3594 \ (35{,}94\,\%)$$

o

$$Z_0 = -0,36 > -1{,}674 = k_{0{,}05} = -k_{0{,}95}$$

[1.0 Ptos]

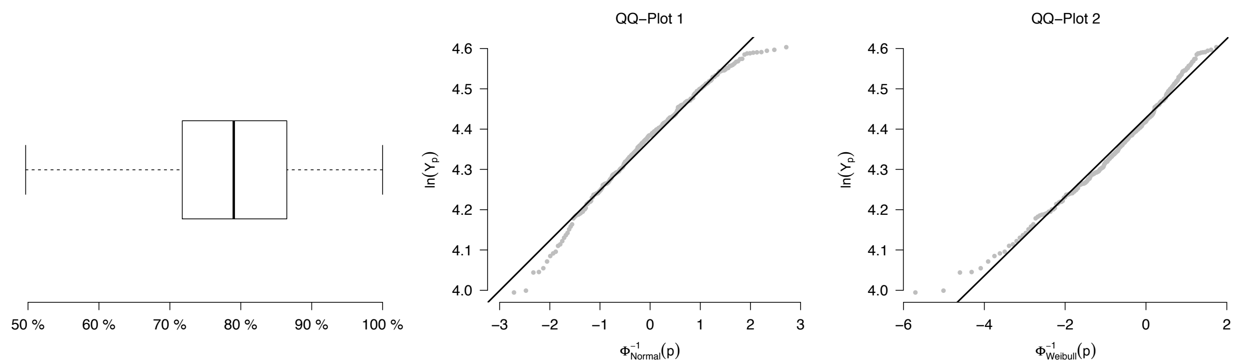
Logro 6 *Conclusión*

Para un nivel de significancia del 5 %, NO hay evidencia para afirmar que el salario de los magister es estadísticamente superior a los profesionales. **[1.0 Ptos]**

+ 1 Punto Base

Problema 3

Un análisis a los % de humedad en la Región Metropolitana durante el mes de junio, en los últimos 5 años, que se registraron en distintas estaciones de monitoreo se presenta a continuación:



La recta de los gráficos de probabilidad son:

	Intercepto	Pendiente
QQ-Plot 1:	4.372228	0.12435090
QQ-Plot 2:	4.428114	0.09805834

¿Entre los dos modelos ajustados por gráfico de probabilidad cuál ajusta mejor? Realice una prueba de bondad de ajuste χ^2 a la siguiente tabla de frecuencia:

	<60%	[60%-70%)	[70%-80%)	[80%-90%)	>=90%
Frecuencia	8	40	98	107	47

Si fuese necesario colapsar (unir) intervalos, hágalo.

Solución

Logro 1 *Hipótesis + Estimación de parámetros*

Se pide

$$\mathrm{H}_0 : X \sim \text{Log-Normal} \quad \text{vs} \quad \mathrm{H}_a : X \not\sim \text{Log-Normal} \quad \textbf{[0.5 Ptos]}$$

con $\hat{\lambda} = 4{,}372228$ y $\hat{\zeta} = 0{,}12435090$. **[0.5 Ptos]**

Logro 2 *Tabla + Estadístico de prueba*

O	p	E	(O-E)^2/E
8	0.01271958	3.815873	4.58792034
40	0.14714239	44.142716	0.38878659
98	0.37154295	111.462886	1.62609554
107	0.31614570	94.843709	1.55809410
47	0.15244939	45.734817	0.03499936
300	1.00000000	300.000000	8.19589593

[0.5 Ptos]

donde

$$X^2 = 8{,}19589593 \sim \chi^2(5 - 1 - 2) \quad \textbf{[0.5 Ptos]}$$

Logro 3 *Colapsar Tabla + Estadístico de prueba*

O	p	E	(O-E)^2/E
48	0.1471424	47.95859	3.575762e-05
98	0.3715430	111.46289	1.626096e+00
107	0.3161457	94.84371	1.558094e+00
47	0.1524494	45.73482	3.499936e-02
300	0.9872804	300.00000	3.219225e+00

[0.5 Ptos]

donde

$$X^2 = 3{,}219225 \sim \chi^2(4-1-2) \quad \textbf{[0.5 Ptos]}$$

Logro 4 *Hipótesis + Estimación de parámetros*

Se pide

$$\text{H}_0 : X \sim \text{Weibull} \quad \text{vs} \quad \text{H}_a : X \not\sim \text{Weibull} \quad \textbf{[0.5 Ptos]}$$

con $\hat{\beta} = 10{,}19801$ y $\hat{\eta} = 83{,}77327$. **[0.5 Ptos]**

Logro 5 *Tabla + Estadístico de prueba*

O	p	E	(O-E)^2/E
8	0.03270074	9.810222	0.33402947
40	0.11526891	34.580673	0.84929262
98	0.31677173	95.031518	0.09272594
107	0.41001364	123.004093	2.08229640
47	0.12524498	37.573495	2.36493810
300	1.00000000	300.000000	5.72328254

[0.5 Ptos]

donde

$$X^2 = 5{,}72328254 \sim \chi^2(5-1-2) \quad \textbf{[0.5 Ptos]}$$

Logro 6 *Conclusión*

- Test 1 Log-Normal:

$$1\,\% < \text{valor-p} < 2{,}5\,\%$$

- Test 2 Log-Normal:

$$5\,\% < \text{valor-p} < 10\,\%$$

- Test 1 Weibull:

$$5\,\% < \text{valor-p} < 10\,\%$$

Considerando 4 intervalos para el modelo Log-Normal, ambos modelos ajustan bien al 5 % de significancia. **[0.5 Ptos]**.

Considerando los 5 intervalos, el modelo Weibull obtendría un mejor ajuste. **[0.5 Ptos]**.

+ 1 Punto Base