Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Segundo Semestre 2016

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

Pauta Examen

Problema 1

En regresión lineal simple, un resultado muy utilizado es el hecho que una variable aleatoria t-Student al cuadrado distribuye Fisher. Muestre este resultado indicando los parámetros de la Fisher resultante.

Solución

Sea X una variable aleatoria t-Student(ν) e $Y = g(X) = X^2$. Se pide mostrar que Y distribuye Fisher identificando sus parámetros.

Como $\Theta_X = \mathbb{R} \to \Theta_Y = \mathbb{R}_0^+$. **[1.0 Ptos.]**

Vía cambio de variable (con dos raíces $X = g^{-1}(Y) = \pm\sqrt{Y}$) **[1.0 Ptos.]** obtenemos la función de densidad de Y:

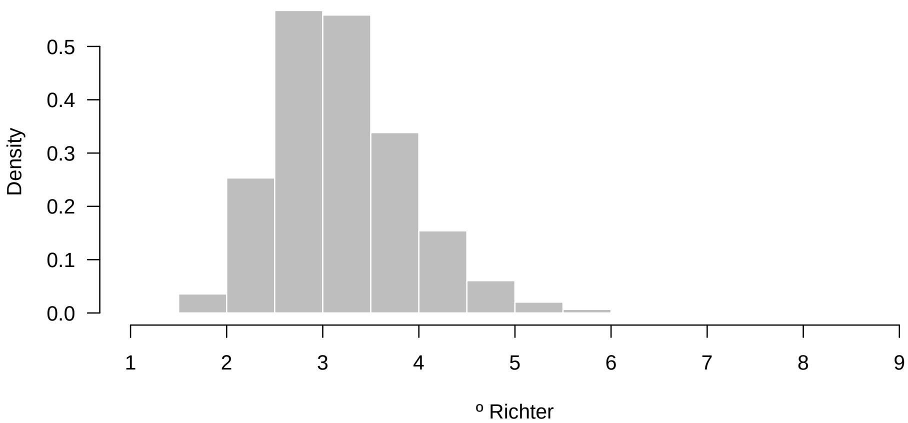
$$\begin{aligned}
f_Y(y) &= f_X(-\sqrt{y}) \left| \frac{d}{dy}\,(-\sqrt{y}) \right| + f_X(+\sqrt{y}) \left| \frac{d}{dy}\,(+\sqrt{y}) \right| && \textbf{[0.5 Ptos.]} \\
&= f_X(-\sqrt{y}) \left| -\frac{1}{2\sqrt{y}} \right| + f_X(+\sqrt{y}) \left| +\frac{1}{2\sqrt{y}} \right| && \textbf{[0.5 Ptos.]} \\
&= \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\,\nu}\,\Gamma(\nu/2)} \left(1 + \frac{y}{\nu}\right)^{-(\nu+1)/2} \frac{1}{2\sqrt{y}} + \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\,\nu}\,\Gamma(\nu/2)} \left(1 + \frac{y}{\nu}\right)^{-(\nu+1)/2} \frac{1}{2\sqrt{y}} && \textbf{[0.5 Ptos.]} \\
&= \frac{\Gamma[(1+\nu)/2]}{\sqrt{\pi\,\nu}\,\Gamma(\nu/2)} \left(1 + \frac{y}{\nu}\right)^{-(1+\nu)/2} \frac{1}{\sqrt{y}} && \textbf{[0.5 Ptos.]} \\
&= \frac{\Gamma[(1+\nu)/2]}{\Gamma(1/2)\,\Gamma(\nu/2)} \left(\frac{1}{\nu}\right)^{1/2} \left(1 + \frac{y}{\nu}\right)^{-(1+\nu)/2} y^{-1/2} && \textbf{[0.5 Ptos.]} \\
&= \frac{\Gamma[(1+\nu)/2]}{\Gamma(1/2)\,\Gamma(\nu/2)} \left(\frac{1}{\nu}\right)^{1/2} \frac{y^{1/2-1}}{\left(1 + \frac{y}{\nu}\right)^{(1+\nu)/2}}, \quad y \geq 0 && \textbf{[0.5 Ptos.]}
\end{aligned}$$

Es decir, $Y \sim \text{Fisher}(1,\, \nu)$. **[1.0 Ptos.]**

+ 1 Punto Base

Problema 2

La siguiente figura muestra las intensidades en ° Richter de los sismos ocurridos en el país en un año tomado al azar en el período 2006-2015, en total se tiene registro de dos mil diez movimientos telúricos cuyas intensidades registraron un mínimo de 1,3° y mediana de 3,2°.



Fuente: www.sismologia.cl

En base a la siguiente tabla resumen

≤ 2	$(2-3]$	$(3-4]$	> 4
140	1169	611	90

¿Podría decir que una modelo Gamma($k = 6$, $\nu = 4$) **trasladado** en γ ajusta las intensidades registradas? Use $\alpha = 5\,\%$.

Solución

Se pide contrastar las siguientes hipótesis

$$\text{H}_0 : Y \sim \text{Gamma}(k = 6\,, \nu = 4) \text{ trasladada en } \gamma \quad \text{vs} \quad \text{H}_a : Y \not\sim \text{Gamma}(k = 6\,, \nu = 4) \text{ trasladada en } \gamma$$

[0.5 Ptos.]

El parámetro γ lo estimamos con el mínimo que es igual a 1,3. **[0.5 Ptos.]**

La probabilidad acumuladas está definida como:

$$F_Y(y) = P(Y \leq y) = P(Y - \gamma \leq y - \gamma) = 1 - \sum_{x=0}^{k-1} \frac{[\nu \cdot (y-\gamma)]^x \, e^{-\nu \cdot (y-\gamma)}}{x!} \qquad \textbf{[0.5 Ptos.]}$$

y al reemplazar en los primeros tres límites se tiene que

$$F_Y(2) = 1 - \left[1 + 2{,}8 + \frac{2{,}8^2}{2} + \frac{2{,}8^3}{6} + \frac{2{,}8^4}{24} + \frac{2{,}8^5}{120}\right] \cdot e^{-2{,}8} = 0{,}065 \qquad \textbf{[0.5 Ptos.]}$$

$$F_Y(3) = 1 - \left[1 + 6{,}8 + \frac{6{,}8^2}{2} + \frac{6{,}8^3}{6} + \frac{6{,}8^4}{24} + \frac{6{,}8^5}{120}\right] \cdot e^{-6{,}8} = 0{,}673 \qquad \textbf{[0.5 Ptos.]}$$

$$F_Y(4) = 1 - \left[1 + 10{,}8 + \frac{10{,}8^2}{2} + \frac{10{,}8^3}{6} + \frac{10{,}8^4}{24} + \frac{10{,}8^5}{120}\right] \cdot e^{-10{,}8} = 0{,}958 \qquad \textbf{[0.5 Ptos.]}$$

Luego, la tabla para obtener el estadístico de prueba $X^2 \sim \chi^2(4-1-1)$ **[0.5 Ptos.]** queda como sigue:

	Observado	Probabilidad	Esperado	(O-E)^2/E
<=2	140	0.065	130.65	0.6691351
(2-3]	1169	0.608	1222.08	2.3054844
(3-4]	611	0.285	572.85	2.5406695
>4	90	0.042	84.42	0.3688273
Total	2010	1.000	2010.00	5.8841163

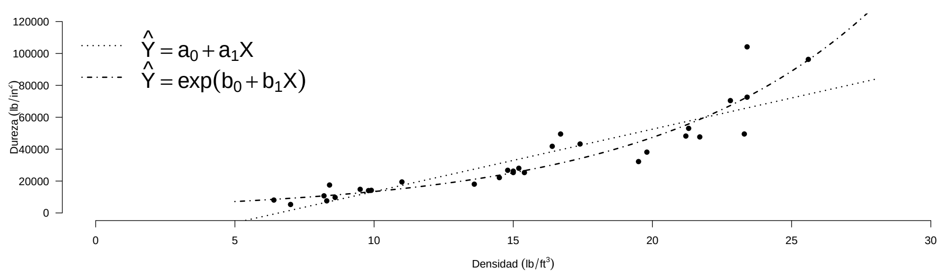
Tenemos que **[1.5 Ptos.]** $X^2 = 5{,}8841163 < 5{,}991465 = c_{0{,}95}(2)$ **[0.5 Ptos.]** o alternativamente que $5\,\% < \text{valor-p} < 10\,\%$. **[0.5 Ptos.]**

Por lo tanto, no existe suficiente evidencia (al 5 % de significancia) para rechazar la distribución Gamma trasladada, es decir, ajusta. **[0.5 Ptos.]**

+ 1 Punto Base

Problema 3

En la fabricación de productos comerciales de madera es importante estimar la relación entre la densidad de un producto de madera y su dureza. Un estudio consideró una nueva madera aglomerada que se puede formar considerablemente más fácil a la mayoría a los productos comerciales similares que en venden en el mercado y para poder competir, es necesario conocer a que nivel de densidad se comporta como la competencia. Para ellos, un investigador propone dos modelos de regresión para representar el comportamiento observado en 30 experimentos que relacionan dureza (Y) con densidad (X), los cuales se presenta en la siguiente figura:




```
Modelo 1: summary(lm(Y~X))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    X.XXX     6171.1  -4.172    X.XXX
X             3913.5      375.1   X.XXX    X.XXX

Residual standard error: 11680, Multiple R-squared:  0.7954, Adjusted R-squared:  X.XXX, F-statistic: 108.8

Modelo 2: summary(lm(log(Y)~X))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.243953   X.XXXXXX   63.99    X.XXX
X           0.126065   0.007831   X.XXX    X.XXX

Residual standard error: XXXX, Multiple R-squared:  0.9025, Adjusted R-squared:  0.899, F-statistic: XXX.X
```

Complete la información borrada de las salidas de R y comente sobre los modelos. Use $\alpha = 5\,\%$.

Solución

Del formulario tenemos que bajo $\mathrm{H}_0 : \beta_j = 0$

$$T_{b_j} = \frac{b_j}{s_{b_j}} \sim \text{t-Student}(n-1-1)$$

y por otra parte, en el caso de regresión simple el estadístico F del modelo es igual al estadístico $T_{b_1}^2$ de la pendiente de la recta.

Además, del mismo formulario se deduce que

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

con estas igualdades se puede completar lo solicitado

```
Modelo 1: summary(lm(Y~X))

Coefficients:
            Estimate Std. Error  t value Pr(>|t|)
(Intercept)-25745.83     6171.1   -4.172    <0.05 (<0.02)
X            3913.50      375.1 10.43322    <0.05 (<0.02)

Residual standard error: 11680, Multiple R-squared:  0.7954, Adjusted R-squared:  0.7880929, F-statistic: 108.8

Modelo 2: summary(lm(log(Y)~X))

Coefficients:
            Estimate Std. Error  t value Pr(>|t|)
(Intercept) 8.243953   0.128832   63.990    <0.05 (<0.02)
X           0.126065   0.007831   16.098    <0.05 (<0.02)

Residual standard error: XXXX, Multiple R-squared:  0.9025, Adjusted R-squared:  0.899, F-statistic: 259.1456
```

Asignar **[0.5 Ptos.]** por cada valor.

Aunque ambos modelos son estadísticamente significativos, el 2do explica un mayor porcentaje de la variabilidad de los datos. **[1.0 Ptos.]**

Nota: Si el alumno indica que falta información para obtener *Residual standard error*, asignar +1.0 puntos en el caso que no haya alcanzado los 6.0 puntos totales

+ 1 Punto Base

Problema 4

Un especialista afirma que la composición del grupo de estudio en el 1er semestre de la universidad es uno de los factores más importante en el buen rendimiento. Usted, con el fin de comprobar esa afirmación recolecta información de una muestra de 15 alumnos, para los cuales dispone del PPS (promedio ponderado semestral), PSU-Mat, PSU-Leng y NEM (notas enseñanza media). Además mediante una encuesta obtiene la composición del grupo de estudio del alumno seleccionado. Como usted es un estudiante avezado de probabilidad y estadística, realiza un resumen estadístico de las variable PPS y ajusta varios modelos para predecirlo:

```
    Mean   Sd  Min Median  Max
PPS 5.41 1.43 2.22   5.15 6.82

lm(formula = PPS ~ X1)
Adjusted R-squared:  0.6401

lm(formula = PPS ~ X2)
Adjusted R-squared:  0.5822

lm(formula = PPS ~ X1 + X3)
Adjusted R-squared:  0.6843

lm(formula = PPS ~ X2 + X3)
Adjusted R-squared:  0.7104

lm(formula = PPS ~ X1 + X2 + X3)
Adjusted R-squared:  0.7311
```

donde X_1: Promedio PSU, X_2: NEM, X_3: Tamaño Grupo.

(a) ¿El efecto del tamaño del grupo es significativo? Use $\alpha = 10\,\%$ y comente.

(b) ¿El efecto conjunto de NEM y tamaño del grupo es significativo? Use $\alpha = 5\,\%$.

(c) Mediante el procedimiento backward proponga un modelo de regresión para pronosticar el rendimiento. Use $\alpha = 2{,}5\,\%$.

Solución

(a) A partir de las salidas **R** podemos testear el efecto de X_3 en presencia de X_1 y en presencia de X_2.

- $PSS \sim X_1$ vs $PSS \sim X_1 + X_3$.

  Tenemos que $k = 1$ y $r = 1$.

$$\frac{SCE}{n-(k+r)-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}6843)\cdot 1{,}43^2 = 0{,}6455749 \to SCE = 7{,}746899 \qquad \textbf{[0.2 Ptos.]}$$
$$\frac{SCE_{(r)}}{n-k-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}6401)\cdot 1{,}43^2 = 0{,}7359595 \to SCE_{(r)} = 9{,}567474 \qquad \textbf{[0.2 Ptos.]}$$

  Para contrastar las hipótesis

$$\mathrm{H}_0 : \alpha_3 = 0 \quad \text{vs} \quad \mathrm{H}_a : \alpha_3 \neq 0 \qquad \textbf{[0.1 Ptos.]}$$

  bajo H_0 se tiene que el estadístico de prueba

$$F = \frac{\left(SCE_{(r)} - SCE\right)/r}{SCE/(n-(k+r)-1)} \sim F(r,\, n-(k+r)-1)$$

  Reemplazando tenemos que **[0.2 Ptos.]** $F = 2{,}820082 < 3{,}18 = F_{0{,}90}(1, 12)$ **[0.2 Ptos.]**, es decir, no existe suficiente evidencia para apoyar que el efecto tamaño de grupo es significativo en el buen desempeño en presencia del la variable X_1 (Promedio PSU). **[0.1 Ptos.]**

- $PSS \sim X_2$ vs $PSS \sim X_2 + X_3$.

  Tenemos que $k = 1$ y $r = 1$.

$$\frac{SCE}{n-(k+r)-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}7104)\cdot 1{,}43^2 = 0{,}592203 \to SCE = 7{,}106436 \qquad \textbf{[0.2 Ptos.]}$$
$$\frac{SCE_{(r)}}{n-k-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}5822)\cdot 1{,}43^2 = 0{,}8543592 \to SCE_{(r)} = 11{,}10667 \qquad \textbf{[0.2 Ptos.]}$$

  Para contrastar las hipótesis

$$\mathrm{H}_0 : \alpha_3 = 0 \quad \text{vs} \quad \mathrm{H}_a : \alpha_3 \neq 0 \qquad \textbf{[0.1 Ptos.]}$$

  bajo H_0 se tiene que el estadístico de prueba

$$F = \frac{\left(SCE_{(r)} - SCE\right)/r}{SCE/(n-(k+r)-1)} \sim F(r,\, n-(k+r)-1)$$

  Reemplazando tenemos que **[0.2 Ptos.]** $F = 6{,}754836 > 3{,}18 = F_{0{,}90}(1, 12)$ **[0.2 Ptos.]**, es decir, existe suficiente evidencia para apoyar que el efecto tamaño de grupo es significativo en el buen desempeño en presencia del la variable X_2 (NEM). **[0.1 Ptos.]**

(b) Podemos testear el efecto conjunto de X_2 y X_3 con respecto al modelo nulo y a la variable X_1.

- $PSS \sim$ NULO vs $PSS \sim X_2 + X_3$.

  Tenemos que $k = 2$.

$$\frac{SCE}{n-k-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}7104)\cdot 1{,}43^2 = 0{,}592203 \to SCE = 7{,}106436 \qquad \textbf{[0.2 Ptos.]}$$

  Para contrastar las hipótesis

$$\mathrm{H}_0 : \beta_2 = \beta_3 = 0 \quad \text{vs} \quad \mathrm{H}_a : \text{Al menos un } \beta_j \neq 0 \qquad \textbf{[0.1 Ptos.]}$$

  bajo H_0 se tiene que el estadístico de prueba

$$F = \frac{SCR/1}{SCE/(n-k-1)} = \frac{(SCT - SCE)/2}{SCE/(n-k-1)} \sim F(2,\, n-k-1)$$

  con $SCT = (n-1)\, S_Y^2$. **[0.1 Ptos.]**

  Reemplazando tenemos que **[0.2 Ptos.]** $F = 18{,}17127 > 3{,}89 = F_{0{,}95}(2, 12)$ **[0.2 Ptos.]**, es decir, existe suficiente evidencia para apoyar que el efecto conjunto del tamaño de grupo y NEM es significativo para explicar el buen desempeño. **[0.2 Ptos.]**

- $PSS \sim X_1$ vs $PSS \sim X_1 + X_2 + X_3$.

  Tenemos que $k = 1$ y $r = 2$.

$$\frac{SCE}{n-(k+r)-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}7311)\cdot 1{,}43^2 = 0{,}5498736 \to SCE = 6{,}04861 \qquad \textbf{[0.2 Ptos.]}$$
$$\frac{SCE_{(r)}}{n-k-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}6401)\cdot 1{,}43^2 = 0{,}7359595 \to SCE_{(r)} = 9{,}567474 \qquad \textbf{[0.2 Ptos.]}$$

Para contrastar las hipótesis

$$\text{H}_0 : \alpha_2 = \alpha_3 = 0 \quad \text{vs} \quad \text{H}_a : \text{Al menos un } \alpha_j \neq 0 \qquad \textbf{[0.1 Ptos.]}$$

bajo H_0 se tiene que el estadístico de prueba

$$F = \frac{\left(SCE_{(r)} - SCE\right)/r}{SCE/(n-(k+r)-1)} \sim F(r,\, n-(k+r)-1)$$

Reemplazando tenemos que **[0.2 Ptos.]** $F = 3{,}199702 < 3{,}98 = F_{0{,}95}(2, 11)$ **[0.2 Ptos.]**, es decir, no existe suficiente evidencia para apoyar que el efecto conjunto del tamaño de grupo y NEM es significativo para explicar el buen desempeño en presencia de X_1 (Promedio PSU). **[0.1 Ptos.]**

(c) De las salidas **R** se observa que eliminar X_1 del modelo completo sería la que pierde menos % de variabilidad explicada. **[0.2 Ptos.]**

Compararemos $PSS \sim X_2 + X_3$ vs $PSS \sim X_1 + X_2 + X_3$. **[0.2 Ptos.]**

Tenemos que $k = 2$ y $r = 1$.

$$\frac{SCE}{n-(k+r)-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}7311)\cdot 1{,}43^2 = 0{,}5498736 \to SCE = 6{,}04861 \qquad \textbf{[0.2 Ptos.]}$$

$$\frac{SCE_{(r)}}{n-k-1} = (1-r^2)\cdot S_Y^2 = (1-0{,}7104)\cdot 1{,}43^2 = 0{,}592203 \to SCE_{(r)} = 7{,}106436 \qquad \textbf{[0.2 Ptos.]}$$

Para contrastar las hipótesis

$$\text{H}_0 : \alpha_1 = 0 \quad \text{vs} \quad \text{H}_a : \alpha_1 \neq 0 \qquad \textbf{[0.1 Ptos.]}$$

bajo H_0 se tiene que el estadístico de prueba

$$F = \frac{\left(SCE_{(r)} - SCE\right)/r}{SCE/(n-(k+r)-1)} \sim F(r,\, n-(k+r)-1)$$

Reemplazando tenemos que **[0.2 Ptos.]** $F = 1{,}923762 < 6{,}72 = F_{0{,}975}(1, 11)$ **[0.2 Ptos.]**, es decir, no existe suficiente evidencia para apoyar que el efecto X_1 (Promedio PSU) sea significativo en presencia del tamaño de grupo y NEM para explicar el buen desempeño, así que se elimina. **[0.1 Ptos.]**

De las otras salidas, solo se puede testear si el efecto de X_3 en presencia de X_2 es significativo, lo cual ya se realizó en (a) resultando su aporte significatvo, es decir, no se elimina y paramos. **[0.2 Ptos.]**

El modelo BACKWARD sería $PSS \sim X2 + X3$. **[0.4 Ptos.]**

+ 1 Punto Base