Modelado estadístico de datos: Práctica 1

David Alarcón Rubio

Diciembre 2019

Ejercicio 1.

1. (CALC) (2 puntos) Se ha realizado un estudio para ver si influye la metodología docente a la hora de aprobar. Para ello 50 estudiantes han recibido la metodología 1 y 50 la metodología 2. De cada estudiante se ha registrado si al final aprobaban (1) o no (2). Los datos experimentales se dan en la tabla siguiente, donde el numero de individuos con perfil aprobar = 1 y metodología = 1 es 35, con perfil aprobar = 1 y metodología = 2 es 15, con perfil aprobar = 2 y metodología = 1 es 40 y con perfil aprobar = 2 y metodología = 2 es 10: ¿Hay diferencias estadadisticamente significativas entre las dos metodología?

Calcularemos el test z de diferencia de dos proprociones que esta enmarcado en el esquema D <- D que indica que se esta intentado explicar la variable de respuesta Y (aprobados) dicotómica a través de la variable explicativa X (metodología) también dicotómica.

Comenzamos realizando la tabla resumen de Y=aprobar por M=metodología.

	M=1	M=2	r
Y=1	35	15	50
Y=2	40	10	50
n	75	25	100

Tabla de resultados

Ecuación

$$\begin{aligned} IC = (\frac{35}{75} - \frac{15}{25}) \pm 1{,}96\sqrt{\frac{35}{75} \cdot (1 - \frac{35}{75}) \cdot \frac{1}{50} + \frac{15}{25} \cdot (1 - \frac{35}{75}) \cdot \frac{1}{50}} \\ = (0{,}4666 - 0{,}6) \pm 1{,}96 \cdot 0{,}0988 \\ = (-0{,}3288, 0{,}0588) \end{aligned} \tag{1}$$

Ecuación

$$\begin{aligned} Z = \frac{\frac{35}{75} - \frac{15}{25}}{\sqrt{\frac{35+14}{75+25} \cdot (1 - \frac{35+14}{75+25}) \cdot (\frac{1}{50} + \frac{1}{50})}} \\ = \frac{-0{,}1334}{\sqrt{0{,}0099}} = \frac{-0{,}1334}{0{,}0999} = -1{,}33 \end{aligned} \tag{2}$$

Podemos comprobar que |Z| = 1.33 <1.96 luego comprobamos que no es estadísticamente significativa la diferencia de aprobados entre las dos metodologías.

Ejercicio 2.

2. (CALC) (1 punto) En el modelo de regresión lineal, se define la matriz H (matriz "hat") como aquella matriz que pone el sombrero a la y, es decir que $\hat{y}$ = Hy, entonces se verifica que H es simétrica e idempotente.

- a) Verdadero.
- b) Falso.

Si $\hat{y}$ = Hy, y dado que:

Equation

$$H = X(X^tX)^{-1}X^t \tag{3}$$

Una matriz es simétrica si es una matriz cuadrada, la cual tiene la característica de ser igual a su traspuesta.

Dado que se comprueba que H es una matriz cuadrada nXn:

Equation

$$\begin{aligned} H = X_{nm}(X_{mn}^tX_{nm})^{-1}X_{mn}^t = H_{nn} \\ Sea, W_{mm} = (X_{mn}^tX_{nm})^{-1} \\ Entonces, H = X_{nm}W_{mm}X_{mn}^t = H_{nn} \end{aligned} \tag{4}$$

Aplicado las siguientes propiedades matriciales:

- $(B \cdot C)^t = C^t \cdot B^t$
- $(A \cdot B \cdot C)^t = C^t \cdot B^t \cdot A^t$
- $(B^t)^t = B$
- $(A^{-1})^t = (A^t)^{-1}$
- $A^t \cdot A = (A^tA)^t$

Obtenemos que:

Equation

$$H^t = (X(X^tX)^{-1}X^t)^t = (X^t)^t((X^tX)^{-1})^tX^t = X((X^tX)^t)^{-1}X^t = X(X^tX)^{-1}X^t = H \quad (5)$$

Por lo que comprobamos que H es una matriz Simétrica.

Una matriz idempotente es una matriz que es igual a su cuadrado, es decir: A es idempotente si A $\times$ A $= A^2$

Aplicado la siguiente propiedad matricial:

- $A \cdot A^{-1} = I$

Obtenemos que:

Equation

$$\begin{gathered} H^2 = (X(X^tX)^{-1}X^t)^2 = (X(X^tX)^{-1}X^t)(X(X^tX)^{-1}X^t) = \\ X(X^tX)^{-1}(X^tX)(X^tX)^{-1}X^t = X(X^tX)^{-1}IX^t = X(X^tX)^{-1}X^t = H \end{gathered} \quad (6)$$

Concluimos que la matriz H es una matriz Simétrica e Idempotente, por lo que la respuesta a la pregunta es Verdadero.

Ejercicio 3.

3. (CALC) (1 punto) En el modelo de regresión lineal, se defi
ne la matriz H (matriz "hat") como aquella matriz que pone el sombrero a la y, es decir que $\hat{y} =$ Hy, entonces se veri
ca que los elementos h_{ii} de la diagonal de H vienen dados por $h_{ii} = x_i^t(X^tX)^{-1}x_i$; siendo $x_i^t = (1x_{i1}...x_{ip})$

- a) Verdadero.
- b) Falso.

Siendo: $x_i^t = (1 x_{i1} ... x_{ip})$ Entonces:

$$X_{np+1}^t = \begin{pmatrix} 1 & x_{11} & ... & x_{1p} \\ 1 & x_{21} & ... & x_{1p} \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ 1 & x_{n1} & ... & x_{np} \end{pmatrix} \tag{7}$$

Si se denota por $(q_{ij}) = (X^t X)^{-1}$, que tiene dimensión (p + 1) x (p + 1) y se realiza el producto matricial, se tiene que:

$$H = \begin{pmatrix} 1 & x_{11} & ... & x_{1p} \\ 1 & x_{21} & ... & x_{2p} \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ 1 & x_{n1} & ... & x_{np} \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} & ... & q_{1p} \\ q_{21} & q_{22} & ... & q_{2p} \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ q_{p+11} & q_{p+12} & ... & q_{p+1p+1} \end{pmatrix} \begin{pmatrix} 1 & 1 & ... & 1 \\ x_{11} & x_{21} & ... & x_{n1} \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ x_{1p} & x_{2p} & ... & x_{np} \end{pmatrix}$$
$$= \begin{pmatrix} x_1^t q_{11} & x_1^t q_{12} & ... & x_1^t q_{1p} \\ x_2^t q_{21} & x_2^t q_{22} & ... & x_2^t q_{2p} \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ x_n^t q_{p+11} & x_n^t q_{p+12} & ... & x_n^t q_{p+1p+1} \end{pmatrix} \begin{pmatrix} 1 & 1 & ... & 1 \\ x_{11} & x_{21} & ... & x_{n1} \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ x_{1p} & x_{2p} & ... & x_{np} \end{pmatrix} \tag{8}$$

Por lo que:

$$h_{ii} = \begin{pmatrix} x_i^t q_{11} & x_i^t q_{12} & ... & x_i^t q_{p+1} \end{pmatrix} x_i = x_i^t \begin{pmatrix} q_{11} & q_{12} & ... & q_{p+1} \end{pmatrix} x_i = x_i^t (X^t X)^{-1} x_i \tag{9}$$

Ejercicio 4.

4.(CALC) (2 puntos) El siguiente código en R

```
rm(list=ls())
datos=read.table('c_d_1.txt',header=T)
```

```
attach(datos)
ind1=which(exp==1)

ind2=which(exp==2)
n1=length(rta[ind1]); n1
n2=length(rta[ind2]); n2
tapply(rta,exp,mean)
tapply(rta,exp,sd)
t.test(rta[ind1],rta[ind2],var.equal=TRUE)
```

proporciona el siguiente resultado

```
> n1=length(rta[ind1]); n1
[1] 7
> n2=length(rta[ind2]); n2
[1] 10
> tapply(rta,exp,mean)
1 2
25.85714 26.20000
> tapply(rta,exp,sd)
1 2
9.856108 8.866917
Two Sample t-test
data: rta[ind1] and rta[ind2]
t = -0.075009, df = 15, p-value = 0.9412
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-10.085497 9.399783
sample estimates:
mean of x mean of y
25.85714 26.20000
```

A continuación se escribe el siguiente código:

```
exp2=1*(exp==2)
summary(lm(data = datos,formula = rta ~ exp2))
```

que proporciona el siguiente resultado.

	Estimate	Std. Error	t value	Pr(>jtj)
(Intercept)	xxx	xxx	xxx	xxx
exp2	xxx	xxx	xxx	xxx

Cuadro 1: Tabla 2: Coefi
cientes de RL con p = 1 sin informacion rellenada

Residual standard error: xxx on xxx degrees of freedom
Multiple R-squared: xxx, Adjusted R-squared: xxx
F-statistic: xxx on xxx and xxx, p-value: xxx
Se pide rellenar el mayor numero posible de valores marcados con xxx.

Con los datos aportados en el ejercicio sabemos que:

$\mathrm{n}_1 = 7; n_2 = 10$

$\overline{y}_1 = \frac{1}{n_1} \sum_{i_1=1}^{n_1} y_{i1} = 25{,}8571; \overline{y}_2 = \frac{1}{n_2} \sum_{i_2=1}^{n_2} y_{i2} = 26{,}2000$

$\mathrm{s}_1 = \sqrt{\frac{1}{n_1} - 1 \sum_{i_1}^{n_1} (y_{i1} - \bar{y})^2} = 9{,}8561$

$\mathrm{s}_2 = \sqrt{\frac{1}{n_2} - 1 \sum_{i_2}^{n_2} (y_{i2} - \bar{y})^2} = 8{,}8669$

$\mathrm{t} = \frac{\overline{y}_1 - \overline{y}_2}{\sqrt{\hat{s_c}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -0{,}0750$

Por lo que podemos deducir que:

$$\hat{s_c}^2 = \frac{(n_2 - 1)(s_2)^2 + (n_1 - 1)(s_1)^2}{(n_2 - 1) + (n_1 - 1)} = \frac{(10 - 1)(8{,}8609)^2 + (7 - 1)(9{,}8561)^2}{(10 - 1) + (9 - 1)} = 86{,}0305$$

$$gl = n_1 + n_2 - 2 = 7 + 10 - 2 = 15$$

Si

p-valor asosiado de $= 2\left(1 - pt\left(abs\left(-0{,}0750\right), n_1 \dot{+} n_2 - 2\right)\right) = 0{,}9412 > 0{,}05$

Dado el modelo de Regresión Lineal:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Siendo una estimación de este:

$$\hat{y} = \hat{\beta_0} + \hat{\beta_1} x$$

Donde $\hat{\beta_0}$ es el coeficiente de intersección - esto es, el valor esperado de Y cuando X = 0, y $\hat{\beta_1}$ el coeficiente de pendiente - el incremento medio en Y asociado a una unidad de incremento en X. Dado que la variable X fue codificada como exp2=1* (exp==2), es decir, su valor es 0 cuando los valores de exp son igual a 1 y toma un valor de 1 cuando los valores de exp son igual a 2. Podemos asumir que si valor de $\overline{y}_1 = 25{,}8571$ entonces $\hat{\beta_0} = 25{,}85714$.

Del mismo modo, el incremento medio en Y asociado a una unidad de incremento en X viene dado por la diferencia entre las medias de $\overline{y}_2 - \overline{y}_1 = 0{,}34286$, por lo que $\hat{\beta_1} = 0{,}34286$.

Sabemos que el EE viene dado por:

$$EE_{exp2} = \sqrt{\hat{s_c}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{86{,}0304\left(\frac{1}{7} + \frac{1}{10}\right)} = 4{,}5703$$

$$EE_{Intercept} = \sqrt{\hat{s_c}^2\left(\frac{1}{n_1}\right)} = \sqrt{86{,}0304\left(\frac{1}{7}\right)} = 3{,}5057$$

Por lo que podemos calcular los valores del estadístico:

$$t_{Intercept} = \frac{\overline{y}_1}{\sqrt{\hat{s_c}^2\left(\frac{1}{n_1}\right)}} = \frac{25{,}85714}{\sqrt{86{,}0304\left(\frac{1}{7}\right)}} = 7{,}3757$$

Con p-valor asociado de $= 2\left(1 - pt\left(abs\left(7{,}3757\right), \dot{n_1} - 2\right)\right) = 0{,}000 < 0{,}05$ por lo tanto significativo.

$$t_{exp2} = \frac{\hat{\beta_1} - 0}{\sqrt{\hat{s_c}^2\left(\frac{1}{n_2} + \frac{1}{n_1}\right)}} = \frac{0{,}34286}{\sqrt{86{,}0304\left(\frac{1}{10} + \frac{1}{7}\right)}} = 0{,}07501$$

	Estimate	Std. Error	t value	Pr(>jtj)
(Intercept)	25.85714	3.05057	7.3757	0.0000
exp2	0.34286	4.5703	0.07501	0.9412

Cuadro 2: Tabla 2: Coefi
cientes de RL con p = 1 sin informacion rellenada

Con p-valor asociado de $= 2\left(1 - pt\left(abs\left(-0{,}0750\right), n_1 \dot{+} n_2 - 2\right)\right) = 0{,}9412 > 0{,}05$ por lo tanto no significativo.

Dado que para el calculo del Error estándar residual (RSE) y el valor del estadístico F conocemos:

$$RSE = \sqrt{\hat{s_c}^2} = \sqrt{86{,}0304} = 9{,}275$$
$$F = t^2 = \left(-0{,}075009\right)^2 = 0{,}005626$$

Con p-valor asociado de $= pf\left(0{,}005626, df1 = 1, df2 = 15, lower.tail = F\right) = 0{,}9412006 > 0{,}05$ por lo tanto no significativo.

Finalmente, podemos calcular:

$$R^2 = \frac{t^2}{(t^2 + df)} = \frac{0{,}00562635}{(0{,}00562635 + 15)} = 0{,}0003749494$$

$$R^2_{adj} = 1 - \left[\frac{\left(1 - R^2\right)\left(\left(n - 1\right)\right)}{n - k - 1}\right] = 1 - \left[\frac{15{,}994}{17 - 1 - 1}\right] = -0{,}06626672$$

Por lo que los datos solicitados son:

Residual standard error: 9.275 on 15 degrees of freedom
Multiple R-squared: 0.0003749494, Adjusted R-squared: -0.06626672
F-statistic: 0.005626 on 1 and 15, p-value: 0.9412

Ejercicio 5.

4.(CALC) (2 puntos) El siguiente código en R

```
rm(list=ls())
datos=read.table('c_n_1.txt',header=T)
attach(datos)
ind1=which(exp==1);

ind2=which(exp==2);
ind3=which(exp==3);
n1=length(rta[ind1]); n1
n2=length(rta[ind2]); n2
n3=length(rta[ind3]); n3
tapply(rta,exp,mean); tapply(rta,exp,sd)
summary(aov(rta~factor(exp)))
```

proporciona el siguiente resultado

```
> n1=length(rta[ind1]); n1
[1] 7
> n2=length(rta[ind2]); n2
[1] 10
> n3=length(rta[ind3]); n3
[1] 5
> tapply(rta,exp,mean); tapply(rta,exp,sd)
1 2 3
25.85714 26.20000 22.60000
1 2 3
9.856108 8.866917 8.876936
Df Sum Sq Mean Sq F value Pr(>F)
factor(exp) 2 46.7 23.35 0.276 0.762
Residuals 19 1605.7 84.51
```

A continuación se escribe el siguiente código:

```
exp2=1*(exp==2)
exp3=1*(exp==3)
summary(lm(data=datos, formula=rta ~ exp2+exp3))
```

que proporciona el siguiente resultado donde se pide rellenar el mayor numero posible de valores marcados con xxx.

	Estimate	Std. Error	t value	Pr(>jtj)
(Intercept)	xxx	xxx	xxx	xxx
exp2	xxx	xxx	xxx	xxx
exp3	xxx	xxx	xxx	xxx

Cuadro 3: Tabla 2: Coefi
cientes de RL con p = 2 sin informacion rellenada

Residual standard error: xxx on xxx degrees of freedom Multiple R-squared: xxx, Adjusted R-squared: xxx F-statistic: xxx on xxx and xxx, p-value: xxx

Con los datos aportados en el ejercicio sabemos que:

$n_1 = 7$; $n_2 = 10$; $n_3 = 5$

$\overline{y}_1 = \frac{1}{n_1}\sum_{i_1=1}^{n_1} y_{i1} = 25{,}8571; \overline{y}_2 = \frac{1}{n_2}\sum_{i_2=1}^{n_2} y_{i2} = 26{,}2000; \overline{y}_3 = \frac{1}{n_3}\sum_{i_3=1}^{n_3} y_{i3} = 22{,}6000;$

$s_1 = \sqrt{\frac{1}{n_1} - 1\sum_{i_1}^{n_1} (y_{i1} - \bar{y})^2} = 9{,}856108$

$s_2 = \sqrt{\frac{1}{n_2} - 1\sum_{i_2}^{n_2} (y_{i2} - \bar{y})^2} = 8{,}866917$

$s_3 = \sqrt{\frac{1}{n_3} - 1\sum_{i_3}^{n_3} (y_{i3} - \bar{y})^2} = 8{,}876936$

Por lo que podemos deducir que:

$$\hat{s_c}^2 = \frac{(n_3 - 1)(s_3)^2 + (n_2 - 1)(s_2)^2 + (n_1 - 1)(s_1)^2}{(n_3 - 1) + (n_2 - 1) + (n_1 - 1)} =$$

$$\frac{(5-1)(8{,}876936)^2 + (10-1)(8{,}8609)^2 + (7-1)(9{,}8561)^2}{(5-1) + (10-1) + (9-1)} = 84{,}50827$$

$$gl = n_1 + n_2 + n_3 - 3 = 7 + 10 + 5 - 3 = 19$$

Dado el modelo de Regresión Lineal:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

Siendo una estimación de este:

$$\hat{y} = \hat{\beta_0} + \hat{\beta_1} x_1 + \hat{\beta_2} x_2$$

Donde $\hat{\beta_0}$ es el coeficiente de intersección - esto es, el valor esperado de Y cuando X = 0, $\hat{\beta_1}$ y $\hat{\beta_2}$ el coeficiente de la pendiente - el incremento medio en Y asociado a una unidad de incremento en X. Dado que la variable X fue codificada como exp2=1* (exp==2), es decir, su valor es 0 cuando los valores de exp son igual a 1 y toma un valor de 1 cuando los valores de exp son igual a 2; y exp3=1* (exp==3), es decir, su valor es 0 cuando los valores de exp son igual a 1 y toma un valor de 1 cuando los valores de exp son igual a 3. Podemos asumir que si valor de $\overline{y}_1 = 25{,}8571$ entonces $\hat{\beta_0} = 25{,}85714$.

Del mismo modo, para exp2 el incremento medio en Y asociado a una unidad de incremento en X viene dado por la diferencia entre las medias de $\overline{y}_2 - \overline{y}_1 = 0{,}34286$, por lo que $\hat{\beta_1} = 0{,}34286$. Y para exp3 el incremento medio en Y asociado a una unidad de incremento en X viene dado por la diferencia entre las medias de $\overline{y}_3 - \overline{y}_1 = -3{,}25714$, por lo que $\hat{\beta_2} = -3{,}25714$.

Sabemos que el EE viene dado por:

$$EE_{Intercept} = \sqrt{\hat{s_c}^2 \left(\frac{1}{n_1}\right)} = \sqrt{84{,}50827 \left(\frac{1}{7}\right)} = 3{,}474566$$

$$EE_{exp2} = \sqrt{\hat{s_c}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{84{,}50827 \left(\frac{1}{7} + \frac{1}{10}\right)} = 4{,}53028$$

$$EE_{exp3} = \sqrt{\hat{s_c}^2 \left(\frac{1}{n_1} + \frac{1}{n_3}\right)} = \sqrt{84{,}50827 \left(\frac{1}{7} + \frac{1}{5}\right)} = 5{,}382775$$

Por lo que podemos calcular los valores del estadístico:

$$t_{Intercept} = \frac{\overline{y}_1}{\sqrt{\hat{s_c}^2 \left(\frac{1}{n_1}\right)}} = \frac{25{,}85714}{\sqrt{84{,}50827 \left(\frac{1}{7}\right)}} = 7{,}441833$$

Con p-valor asociado de $= 2\left(1 - pt\left(abs\left(7{,}441833\right), n_1 - 2\right)\right) = 0{,}000 < 0{,}05$ por lo tanto significativo.

$$t_{exp2} = \frac{\hat{\beta}_1 - 0}{\sqrt{\hat{s_c}^2\left(\frac{1}{n_2} + \frac{1}{n_1}\right)}} = \frac{0{,}34286}{\sqrt{84{,}50827\left(\frac{1}{10} + \frac{1}{7}\right)}} = 0{,}07568186$$

Con p-valor asociado de $= 2\left(1 - pt\left(abs\left(0{,}07568186\right), n_1 + n_2 - 2\right)\right) = 0{,}9412 > 0{,}05$ por lo tanto no significativo.

$$t_{exp3} = \frac{\hat{\beta}_2 - 0}{\sqrt{\hat{s_c}^2\left(\frac{1}{n_2} + \frac{1}{n_1}\right)}} = \frac{-3{,}25714}{\sqrt{84{,}50827\left(\frac{1}{5} + \frac{1}{7}\right)}} = -0{,}6051043$$

Con p-valor asociado de $= 2\left(1 - pt\left(abs\left(-0{,}6051043\right), n_1 + n_2 - 2\right)\right) = 0{,}9412 > 0{,}05$ por lo tanto no significativo.

Dado que para el calculo del Error estándar residual (RSE) y el valor del estadístico F conocemos:

$$RSE = \sqrt{\hat{s_c}^2} = \sqrt{84{,}50827} = 9{,}192838$$

$$F = 0{,}2763$$

Con p-valor asociado de $= pf\left(0{,}2763, df1 = 1, df2 = 15, lower.tail = F\right) = 0{,}9412006 > 0{,}05$ por lo tanto no significativo.

Finalmente, podemos calcular:

$$R^2 = \frac{SCE}{(SCE + SCR)} = \frac{46{,}7}{(46{,}7 + 1605{,}7)} = 0{,}02826192$$

$$R^2_{adj} = 1 - \left[\frac{(1 - R^2)\left((n-1)\right)}{n - k}\right] = 1 - \left[\frac{20{,}4065}{22 - 3}\right] = -0{,}0740263$$

Por lo que los datos solicitados son:

	Estimate	Std. Error	t value	Pr(>jtj)
(Intercept)	25.8571	3.4746	7.442	0.0000
exp2	0.3429	4.5303	0.076	0.940
exp3	-3.2571	5.3828	-0.605	0.552

Cuadro 4: Tabla 2: Coefi
cientes de RL con p = 1 sin informacion rellenada

Residual standard error: 9.193 on 19 degrees of freedom
Multiple R-squared: 0.02827, Adjusted R-squared: -0.07402
F-statistic: 0.2763 on 2 and 19, p-value: 0.7615

Ejercicio 6.

6. (2 puntos) Se ha realizado un estudio para ver si el peso en kg (rta) de unos deportistas depende de su cintura en cm (exp1), del numero de km de entrenamiento (exp2) y del tipo de entrenamiento (exp3=1: Body building, exp3=2: Fitness). Han participado en el estudio 26 individuos. Los datos experimentales estan en el

chero c ccd.txt alojado en el curso virtual y se muestran en la tabla 6. Se pide: Interpretar los resultados del modelo de regresion lineal con todas las variables. Repetir el analisis quitando las variables no signi

cativas. >Que sucede? Crear una variable interaccion entre exp1 y exp3 e incorporarla al modelo anterior. >Que ocurre? Elegir de los tres modelos anteriores el mejor. >Se cumplen las condiciones de aplicabilidad de la regresion lineal? Elaborar otro enunciado para estos datos. En el documento que se entregue habra que incluir el codigo utilizado.

rta	exp1	exp2	exp3
69.3	83	8	1
69.6	84	7	1
71.5	86.5	4	1
71.5	84.5	32	1
70.6	86.4	15	1
69.2	82.5	6	1
65	82	10	2
65.4	81.8	17	2
63.7	80	6	2
69	82.5	18	1
65.8	84	0	2
68.7	87.2	3	2
64.8	84	10	2
70	86	11	1
65.9	84.2	18	2
63.9	84	4	2
62.1	79	12	2
73.1	97.2	18	2
75.4	91	0	1
72.6	89.5	9	1
69.6	89.5	11	2
72.3	87.5	7	1
67.3	87.5	15	2
68	87.5	5	2
68.1	86.5	14	2
71.3	87	9	1