

MÉTODOS ESTADÍSTICOS

Nombre(s): Equipo 4:

- Diana Zepeda Martínez
- José Juan García Romero

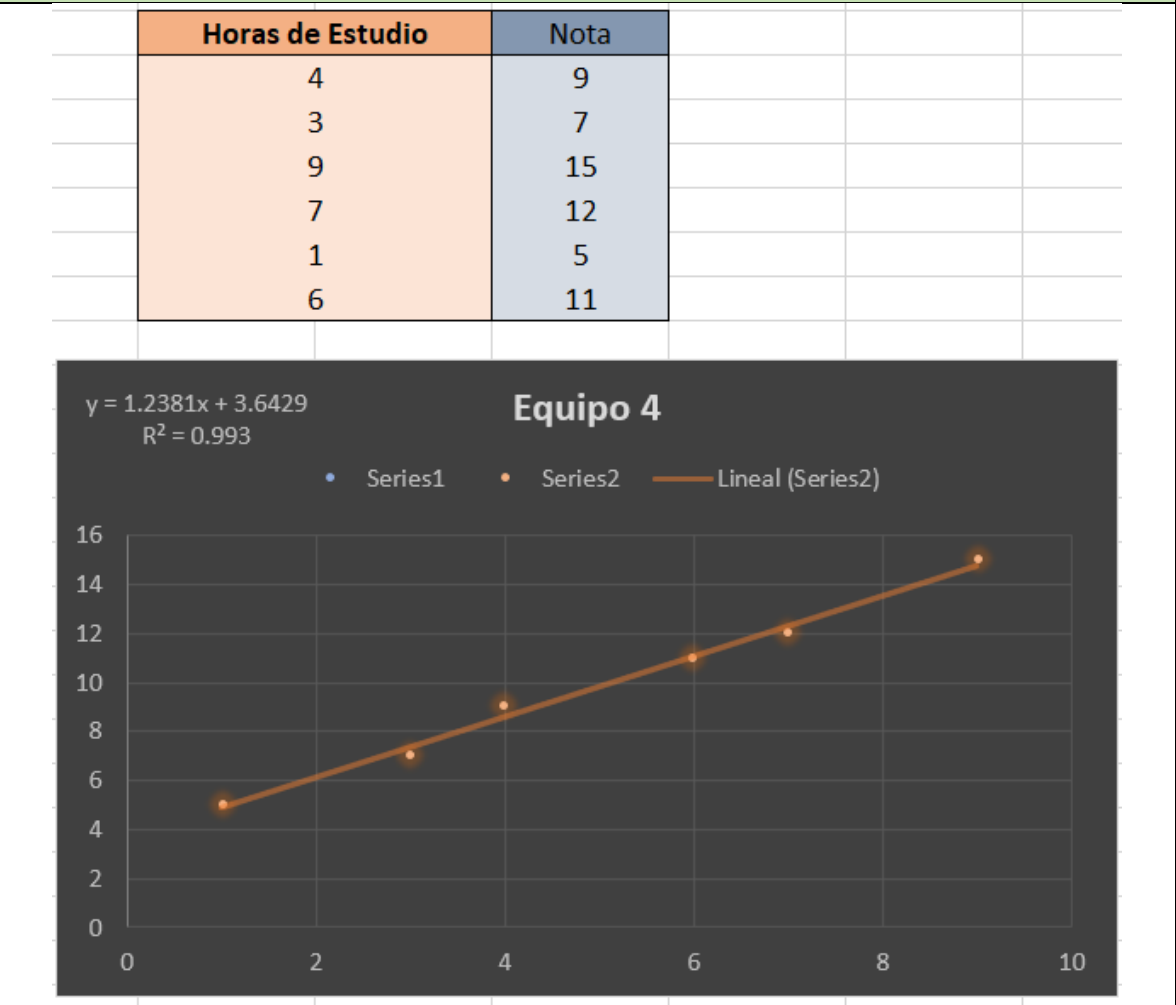
Nº
19-1

Descripción:

- Se desea predecir la **nota de un alumno** en función a la **cantidad de horas a la semana** que estudió para un examen.
- Para ello se tomó una muestra de 6 alumnos, cuyos resultados se muestran en el siguiente cuadro:

Horas de estudio	Nota
4	9
3	7
9	15
7	12
1	5
6	11

A) En Excel obtener:
1. La gráfica y el modelo de regresión lineal del problema

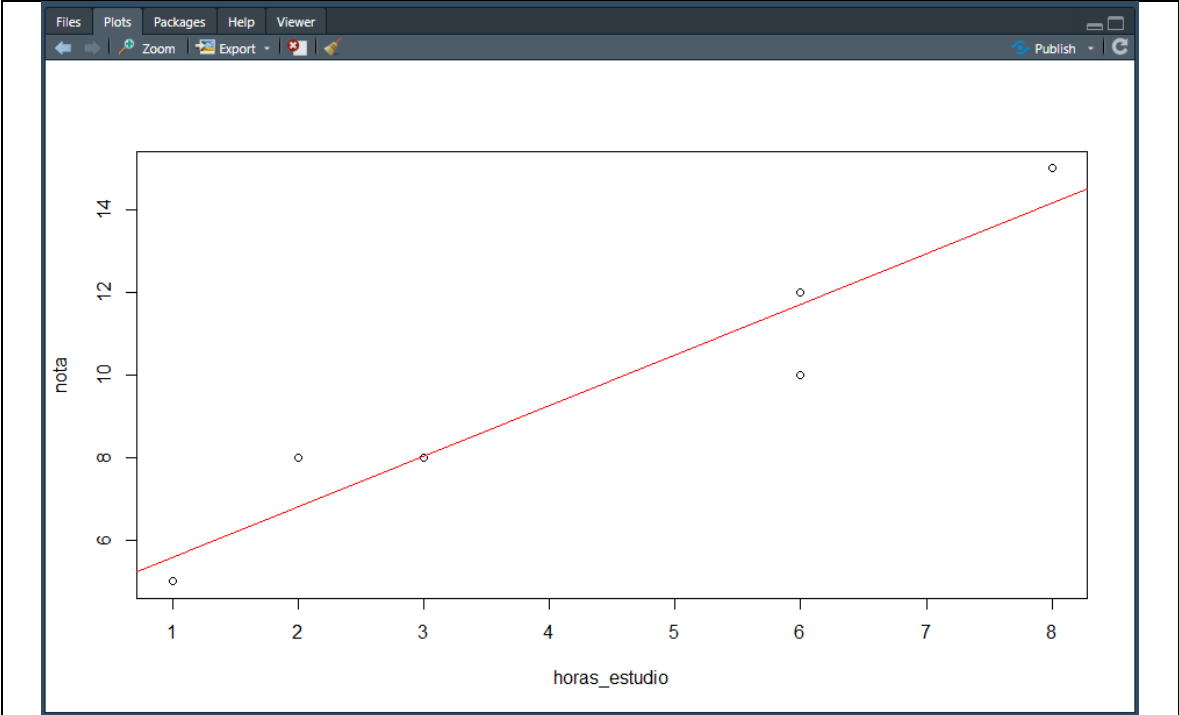


Nombre(s): Equipo 4: <ul style="list-style-type: none"> Diana Zepeda Martínez José Juan García Romero 	
Nº 19-2	Descripción: B) Aplicando la secuencia de pasos, <div> <div>RESUMEN:</div> <div> <div>Regresión Lineal Simple</div> <div>Diagrama de dispersión y correlación bivariada</div> <div>TABLA ANOVA</div> <div>COEFICIENTE DE DETERMINACIÓN – R²</div> <div>TABLA DE COEFICIENTES</div> <div>$y = a + bx$</div> </div> </div> <p>en RStudio obtener:</p>

1-5. El modelo de regresión lineal del problema con sus respectivas interpretaciones	
SOLUCIÓN EN R	INTERPRETACIÓN
<div> <div> <div>Actividad 19.R</div> <div> <div>Source on Save</div> <div> <div> <pre> 1 # Diana Zepeda Martínez 2 # José Juan García Romero 3 4 #Regresión Lineal 5 library(graphics) 6 library(psych) 7 8 horas_estudio = c(3,6,8,2,1,6) #variable independientes 9 nota = c(8,10,15,8,5,12) #variable dependiente 10 11 ##### 12 # A) En Excel obtener: 13 # 1. La gráfica y el modelo de regresión lineal del pro 14 15 ##### 16 # B) En RStudio obtener: 17 # 1. DIAGRAMA DE DISPERSIÓN Y CORRELACIÓN BIVARIADA 18 cor.test(x=horas_estudio, y=nota, method = "pearson") 19 estudio_notas = cbind.data.frame(horas_estudio, nota) 20 pairs.panels(estudio_notas, 21 method = "pearson", 22 density = F, 23 ellipses = F, 24 smooth = F) </pre> </div> <div> <pre> # Diana Zepeda Martínez # José Juan García Romero # Regresión Lineal library(graphics) library(psych) horas_estudio = c(3,6,8,2,1,6) #variable independientes nota = c(8,10,15,8,5,12) #variable dependiente # En estudio obtener: # 1. DIAGRAMA DE DISPERSIÓN Y CORRELACIÓN BIVARIADA cor.test(x=horas_estudio, y=nota, method = "pearson") </pre> </div> <div> <pre> data: horas_estudio and nota t = 6.3894, df = 4, p-value = 0.00308 alternative hypothesis: true correlation is not equal to 0 95 percent confidence interval: 0.6331731 0.9931511 sample estimates: cor 0.9543389 </pre> </div> <div> <pre> > estudio_notas = cbind.data.frame(horas_estudio, nota) > pairs.panels(estudio_notas, + method = "pearson", + density = F, + ellipses = F, + smooth = F) </pre> </div> </div> </div> <div> <div> <div>horas_estudio</div> <div> <div>nota</div> <div>0.95</div> </div> </div> </div> </div> </div>	<p>Se utilizaron los datos proporcionados de las horas de estudio y las notas para obtener el coeficiente de correlación con el método de Pearson = 0.95</p>
SOLUCIÓN EN R	INTERPRETACIÓN
<div> <div> <div>Actividad 19.R</div> <div> <div>modelo1</div> <div>Show Attributes</div> <div> <div> <div>Name</div> <div>Type</div> <div>Value</div> </div> <div> <div>modelo1</div> <div>list [12] (S3: lm)</div> <div>List of length 12</div> </div> <div> <div>coefficients</div> <div>double [2]</div> <div>4.37 1.22</div> </div> <div> <div>residuals</div> <div>double [6]</div> <div>-0.0397 -1.7054 0.8482 1.1875 -0.5893 0.2946</div> </div> <div> <div>effects</div> <div>double [6]</div> <div>-23.678 -7.474 1.700 0.794 -1.191 0.731</div> </div> <div> <div>rank</div> <div>integer [1]</div> <div>2</div> </div> <div> <div>fitted.values</div> <div>double [6]</div> <div>8.04 11.71 14.15 6.81 5.59 11.71</div> </div> <div> <div>assign</div> <div>integer [2]</div> <div>0 1</div> </div> <div> <div>cj</div> <div>list [5] (S3: cj)</div> <div>List of length 5</div> </div> <div> <div>df.residual</div> <div>integer [1]</div> <div>4</div> </div> <div> <div>xlevels</div> <div>list [0]</div> <div>List of length 0</div> </div> <div> <div>call</div> <div>language</div> <div>lm(formula = nota ~ horas_estudio)</div> </div> <div> <div>terms</div> <div>formula</div> <div>nota ~ horas_estudio</div> </div> <div> <div>model</div> <div>list [6 x 2] (S3: data.frame)</div> <div>A data.frame with 6 rows and 2 columns</div> </div> </div> </div> </div> </div>	<p>Se determinó si existen diferencias significativas entre las medias, esto con ayuda del análisis de varianza.</p>
SOLUCIÓN EN R	INTERPRETACIÓN
<div> <div> <div>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</div> <div>Residual standard error: 1.17 on 4 degrees of freedom</div> <div>Multiple R-squared: 0.9108, Adjusted R-squared: 0.8885</div> <div>F-statistic: 40.82 on 1 and 4 DF, p-value: 0.00308</div> </div> </div>	<p>Con ayuda del p-valor, rechazamos H_0 y aceptamos H_1</p>

SOLUCIÓN EN R	INTERPRETACIÓN
<pre>> summary(modelo1) Call: lm(formula = nota ~ horas_estudio) Residuals: 1 2 3 4 5 6 -0.03571 -1.70536 0.84821 1.18750 -0.58929 0.29464 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 4.3661 0.9572 4.561 0.01033 * horas_estudio 1.2232 0.1914 6.389 0.00308 ** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 1.17 on 4 degrees of freedom Multiple R-squared: 0.9108, Adjusted R-squared: 0.8885 F-statistic: 40.82 on 1 and 4 DF, p-value: 0.00308</pre>	Se utiliza la función summary para obtener la tabla de coeficientes. De esta forma tenemos la significancia con un intercepto de 4.3661.
SOLUCIÓN EN R	INTERPRETACIÓN
<pre>> # 5. CONSTRUIR LA ECUACIÓN > modelo1\$coefficients (Intercept) horas_estudio 4.366071 1.223214</pre>	Con esta función se obtienen los coeficientes del modelo 1

6. La gráfica del modelo obtenido



7. Calcular las notas si las horas de estudio fueran 5, 10 y 12

```
41 # 7. Calcular las notas si las horas de estudio fueran (5, 10 y 12)
42 # para ejemplificar se hará con otras horas 15
43 nota_5 = 4.366071 + 1.223214*(5)
44 nota_5 = round(nota_5)
45 nota_5
46
47 nota_10 = 4.366071 + 1.223214*(10)
48 nota_10 = round(nota_10)
49 nota_10
50
51 nota_12 = 4.366071 + 1.223214*(12)
52 nota_12 = round(nota_12)
53 nota_12
54
```

54:1 (Top Level) ▾

Console Terminal Jobs

R 4.1.3 · ~/

```
> # 7. Calcular las notas si las horas de estudio fueran (5, 10 y 12)
> # para ejemplificar se hará con otras horas 15
> nota_5 = 4.366071 + 1.223214*(5)
> nota_5 = round(nota_5)
> nota_5
[1] 10
> nota_10 = 4.366071 + 1.223214*(10)
> nota_10 = round(nota_10)
> nota_10
[1] 17
> nota_12 = 4.366071 + 1.223214*(12)
> nota_12 = round(nota_12)
> nota_12
[1] 19
>
```

DESCRIPCIÓN DE FUNCIONES

color.test: Sirve para hacer la prueba de correlación.
pairs.panels: Crea líneas de regresión, histogramas, intervalos de confianza, etc.
lm: Crea los modelos lineales.
summary: Crea un resumen general sobre las variables, en este caso es sobre un modelo lineal. Ya que esta se puede aplicar de diferentes formas.
plot: Crea gráficos pasando por dos arreglos.
abline: Dibuja una línea en la gráfica.