

## Estadística (Q) - Clase 18 (10/06/2014)

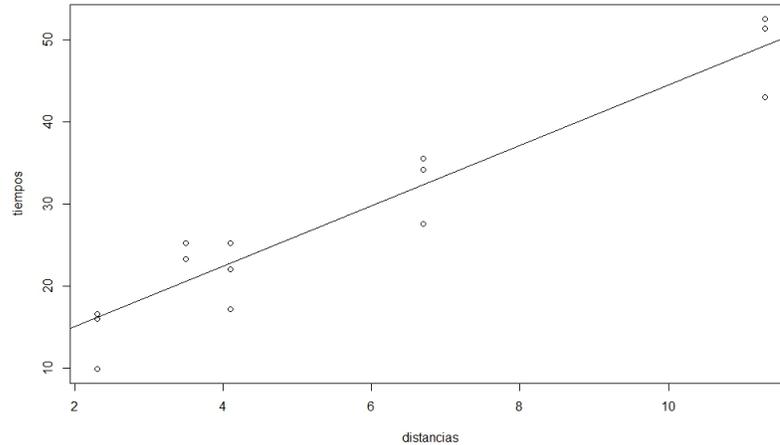
### Regresión simple

---

1. 5 curiosos estudiantes de Estadística Q desean poner en práctica sus conocimientos. Tienen en común, entre otras cosas, que todos llegan a la facultad en el colectivo 42 pero de distintas distancias. Se ponen de acuerdo en que cada uno tomará tres veces el tiempo que demora desde que llega a la parada de colectivo cercana a su casa hasta que baja en el pabellón 2. Se ponen de acuerdo que día tomará el tiempo cada uno para que haya independencia entre las observaciones. A continuación la tabla con las distancias medidas en Km. y los tiempos tomados en minutos:

	distancias	tiempos
[1,]	2.3	9.925551
[2,]	2.3	16.041055
[3,]	2.3	16.576962
[4,]	3.5	25.275232
[5,]	3.5	23.300148
[6,]	3.5	23.306799
[7,]	4.1	22.102919
[8,]	4.1	25.241391
[9,]	4.1	17.195020
[10,]	6.7	34.237753
[11,]	6.7	27.606280
[12,]	6.7	35.586040
[13,]	11.3	51.345192
[14,]	11.3	52.557606
[15,]	11.3	43.072103

- Viendo los datos en el diagrama de dispersión y la recta ajustada por la regresión lineal, decidir si es razonable suponer que existe una relación lineal entre tiempos y distancias.
- ¿Cuál es la variable dependiente y cuál la independiente en este problema? ¿Cuál de las variables es aleatoria y cual determinística?
- Escribir el modelo de regresión lineal para este problema, definiendo los parámetros involucrados junto con los supuestos del modelo. ¿Es cierto que dicho modelo asume que las variables dependientes son i.i.d.? Determinar la pendiente y la ordenada al origen de la recta de mínimos cuadrados.
- ¿Cuál es el coeficiente de determinación  $r^2$ ? Interprete el significado de esta medida en este ejemplo.
- Escriba los supuestos necesarios para que sea válida la estimación mediante mínimos cuadrados y escriba las hipótesis necesarias para que sean válidas las conclusiones respecto de los test e intervalos de confianza.
- Realizar intervalos de confianza para la ordenada al origen y la pendiente de la recta del modelo.
- Es muy razonable suponer que la cantidad de tiempo de viaje y la distancia del viaje están asociadas. Asumiendo que se verifica el modelo lineal planteado en el item c) testear esta afirmación al 5%.



```
> salida<-lm(tiempos~distancias)
> summary(salida)
```

```
Call:
lm(formula = tiempos ~ distancias)
```

```
Residuals:
    Min     1Q   Median     3Q     Max
-6.240 -2.708  1.895  2.726  4.698
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.7097     2.0545   3.753  0.00242 **
distancias    3.6765     0.3194  11.512 3.43e-08 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.96 on 13 degrees of freedom
Multiple R-squared:  0.9107,    Adjusted R-squared:  0.9038
F-statistic: 132.5 on 1 and 13 DF,  p-value: 3.431e-08
```