

Biología Examen Enero 2017

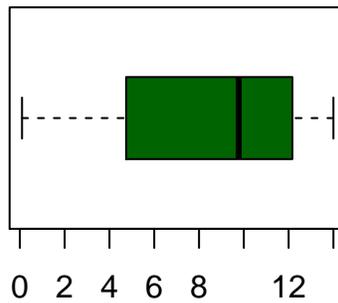
Equipo docente

12 de enero de 2017

Cuestión voluntaria 1

Supón que dispones de una muestra de una variable cuantitativa en la que la mediana es mucho mayor que la media. Dibuja, de forma aproximada, un diagrama de caja y bigotes que corresponda a esos datos, y explica porqué tiene esa forma.

La media, que es una medida numérica, es sensible a la presencia de unos pocos valores anormalmente (comparados con el resto) grandes o pequeños



Por ejemplo, para el diagrama anterior, la media vale 8.1498084 y la mediana 9.7693095

Cuestión voluntaria 2

Dispones de dos muestras pareadas de una variable cuantitativa. Se quiero probar que la media en la muestra 2 es mayor que en la muestra 1. Escribe las hipótesis nula y alternativa que definen el contraste. Como las muestras están pareadas, hay que crear una nueva variable que

$$D = X_1 - X_2$$

y la media de la segunda muestra es mayor que la de la primera muestra si $D < 0$. Entonces, fijamos

$$H_0 : D \geq 0, \quad H_1 : D < 0$$

Así, sólo supondremos que $D > 0$ cuando haya una evidencia muestral muy fuerte a su favor. En caso de haber elegido fijas las hipótesis nula y alternativa al revés, nuestra comprobación se limitaría a “a la vista de nuestra muestra, no hay evidencia de que $D < 0$ ”.

Pregunta 1: regresión

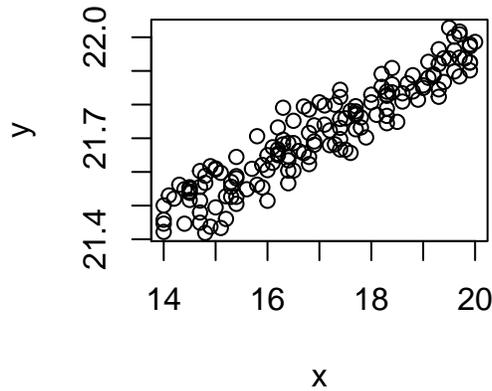
<http://epidemiologiamolecular.com/correlacion-regresion-lineal-simple/>

Los datos de la tabla enseñan la relación entre los chirridos por segundo que realiza un grillo y la temperatura del suelo.

```
##  
## Call:  
## lm(formula = y ~ x)  
##
```

```
## Coefficients:
## (Intercept)          x
## 20.30988          0.08252
```

Aunque no se pide explícitamente, **siempre** hay que representar los datos, para cercionarse de que tiene sentido aproximarlos por una recta. Suponemos que la variable explicativa es el número de chirridos por minuto y la respuesta la temperatura

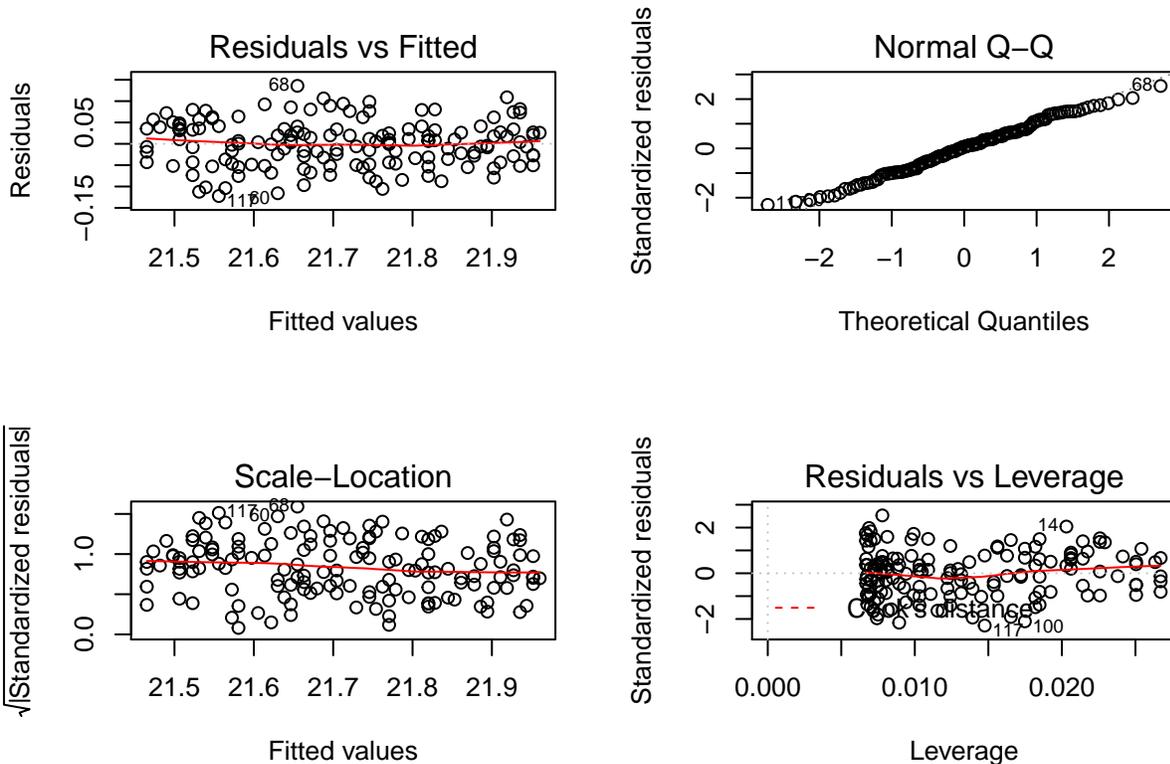


Calcular los coeficientes de la recta de regresión

```
## (Intercept)          x
## 20.30988491 0.08251706
```

¿Se cumplen las condiciones para aplicar la regresión lineal?

Puedes usar tanto criterios gráficos



Otra alternativa pasa por usar la librería `gvlma` para obtener el p-valor correspondiente a los contrastes sobre el sesgo (skewness), la kurtosis y la homocedasticidad

```
## [1] 0.9023
## [1] 0.2215
## [1] 0.1345
```

Estima, a partir de la recta calculada, la temperatura esperada para 14.6 y 30 chirridos por segundo

Los valores se obtienen al sustituir en la expresión de la recta de regresión. Hay que hacer notar que la temperatura obtenida para el valor 30 no tiene valor predictivo, dado que 30 está fuera del intervalo de observados. Las cantidades pedidas son

```
b0+b1*14.6
```

```
## [1] 21.51463
```

```
b0+b1*30
```

```
## [1] 22.7854
```

Calcula e interpreta el intervalo de confianza para la pendiente de la recta

Como no se indica, fijamos el nivel de confianza a 0.95, y obtenemos

```
##      2.5 %      97.5 %
## 0.07751947 0.08751466
```

es decir, si la variable explicativa varía una unidad, el IC nos indica, con una probabilidad de 0.95, cuánto variará la variable respuesta.

Pregunta Anova

http://www.hrc.es/bioest/Anova_4.html

Queremos comprobar si hay diferencia entre la longitud del ala de cuatro especies de águila que pueden ser vistas en Extremadura: Águila-azor Perdicera, Águila Real, Águila Pescadora, Águila Imperial ibérica. Para ello, medimos la longitud del ala de 40 especímenes de cada una de esas cuatro especies.

Las hipótesis nula y alternativa son

$H_0 : \mu_1 = \dots = \mu_4$ frente a $H_1 : \text{alguna de las medias es diferente a las demás}$

Un error muy repetido es considerar como hipótesis alternativa “todas las medias son distintas”, que no es lo contrario de “ H_0 : todas las medias son iguales”.

La tabla del contraste ANOVA es

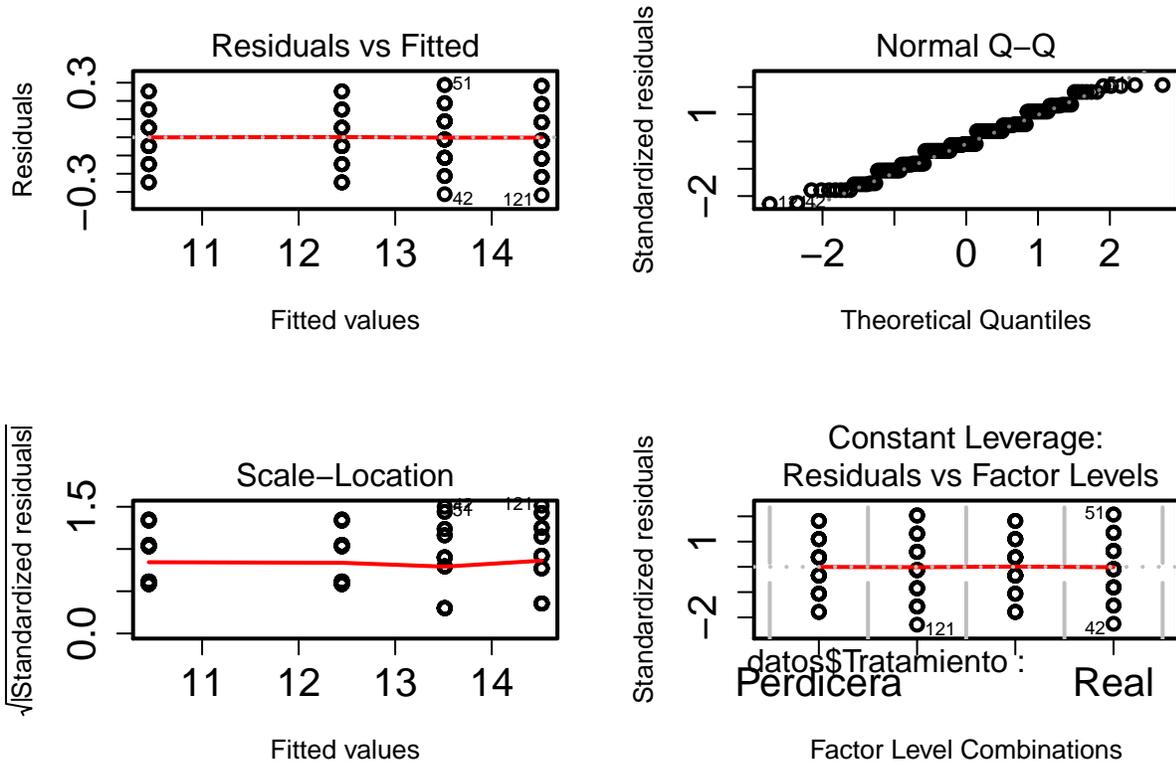
```
datos.lm = lm(datos$Respuesta ~ datos$Tratamiento)
anova(datos.lm)
```

```
## Analysis of Variance Table
##
## Response: datos$Respuesta
##              Df Sum Sq Mean Sq F value    Pr(>F)
## datos$Tratamiento  3 363.88  121.29  6141.5 < 2.2e-16 ***
## Residuals        156   3.08    0.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El p-valor es muy pequeño, por lo que el contraste es significativo: la longitud de las alas de alguna de las especies es, en promedio, diferente de las demás.

¿Se cumplen las condiciones para realizar un contraste ANOVA?

De nuevo podemos usar métodos gráficos



O, de nuevo, la función `gvlma`

```
## [1] 0.8954
## [1] 0.1781
## [1] 0.721
```

¿Hay diferencia significativa entre el tamaño de ala de estas especies? En caso afirmativo, ordena las especies de menor longitud de ala a mayor longitud. Justifica tus respuestas.

Puedes usar el ajuste de bonferroni Bonferroni

```
##
## Pairwise comparisons using t tests with non-pooled SD
##
## data: datos$Respuesta and datos$Tratamiento
##
##      Perdicera Imperial Pescadora
## Imperial <2e-16 - -
## Pescadora <2e-16 <2e-16 -
## Real      <2e-16 <2e-16 <2e-16
##
## P value adjustment method: bonferroni
```

o métodos gráficos

