Regresion lineal 2

Contents


```
## Warning: package 'TH.data' was built under R version 4.1.3

## Loading required package: survival

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

## Warning: package 'car' was built under R version 4.1.3

## Loading required package: carData

## Warning: package 'lmtest' was built under R version 4.1.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.1.3

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Warning: package 'faraway' was built under R version 4.1.3
```

```
##
## Attaching package: 'faraway'

## The following objects are masked from 'package:car':
##
##     logit, vif

## The following objects are masked from 'package:survival':
##
##     rats, solder

## Warning: package 'GGally' was built under R version 4.1.3

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.1.3

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

##
## Attaching package: 'GGally'

## The following object is masked from 'package:faraway':
##
##     happy
```

Introducción

Representación de los datos

En primer lugar vamos a realizar la regresión lineal con los datos del grupo males 1. Donde la variable dependiente es el número de huevos y la variable independiente es el peso.

A continuación podemos ver un valor de p del 0.00014. Lo que nos da a entender que el número de huevos depende del tamaño de la hembra.

```
##
## Call:
## lm(formula = eggs ~ weight, data = photinus, subset = males ==
##     1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.763  -5.627   1.164   4.263  18.355
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -12.2657     6.2231  -1.971  0.06430 .
```

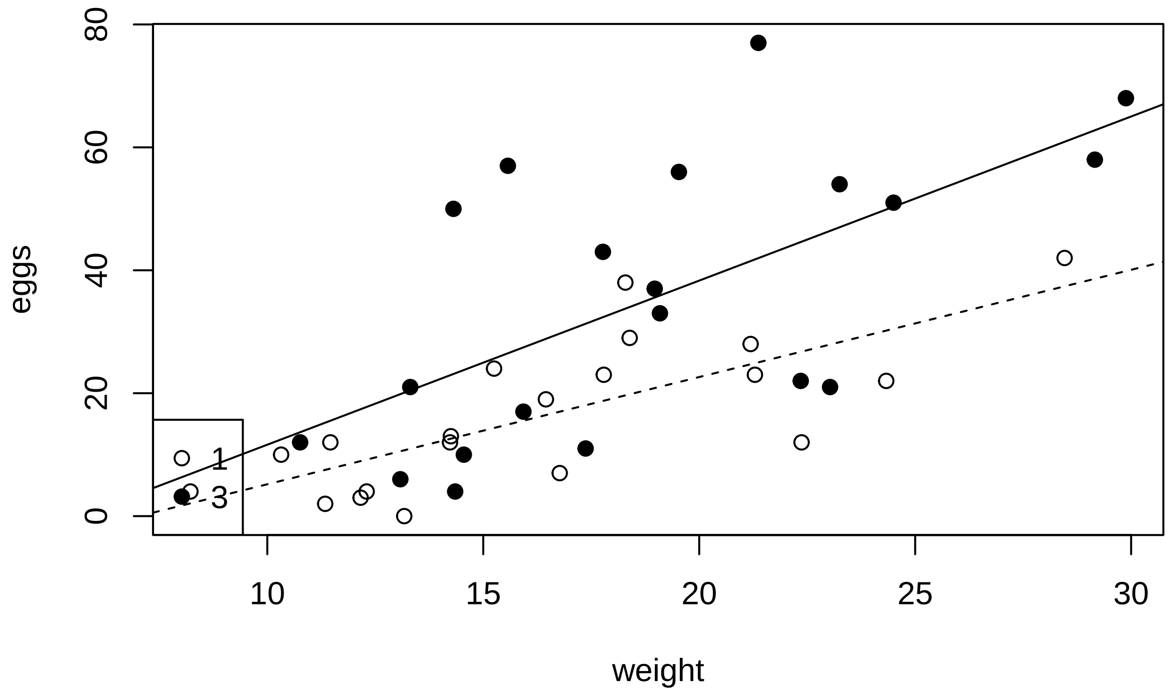
```
## weight        1.7447     0.3626   4.811  0.00014 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.191 on 18 degrees of freedom
## Multiple R-squared:  0.5626, Adjusted R-squared:  0.5383
## F-statistic: 23.15 on 1 and 18 DF,  p-value: 0.0001399
```

A continuación haremos la regresión para el segundo grupo males 3. En este caso sucede lo mismo, el p valor es inferior a 0.05 (0.00324).

```
##
## Call:
## lm(formula = eggs ~ weight, data = photinus, subset = males ==
##     3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.410 -13.787  -0.553   7.911  35.022
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.0811    15.4165  -0.978  0.34092
## weight        2.6700     0.7869   3.393  0.00324 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.07 on 18 degrees of freedom
## Multiple R-squared:  0.3901, Adjusted R-squared:  0.3562
## F-statistic: 11.51 on 1 and 18 DF,  p-value: 0.003241
```

En el siguiente gráfico podemos ver ambas regresiones representadas.

```
## The following object is masked from package:faraway:
##
##     eggs
```



Estimar y validar el modelo ANCOVA

El modelo ANCOVA o análisis de la covarianza es una fusión entre el modelo ANOVA y de la regresión lineal múltiple.

Vemos en la gráfica del apartado anterior, que ambas rectas estan casi paralelas la una de la otra. Vamos a contrastar si hay diferencias significativa entre ellas. Es decir que si en el grupo de males 3 hay más huevos, independientemente del peso. Para ellos vamos a creas un modelo en el que haya interacción entre las variables weight y males. Estudiamos el contraste.

```
## Analysis of Variance Table
##
## Response: eggs
##              Df Sum Sq Mean Sq F value    Pr(>F)
## weight        1 6999.6  6999.6 35.5722 7.758e-07 ***
## males         1 1718.9  1718.9  8.7357  0.005476 **
## weight:males  1  222.0   222.0  1.1282  0.295224
## Residuals    36 7083.8   196.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En primer lugar observamos que podemos prescindir de la interacción ya que no es significativa. Así pues, las rectas de regresión son paralelas.

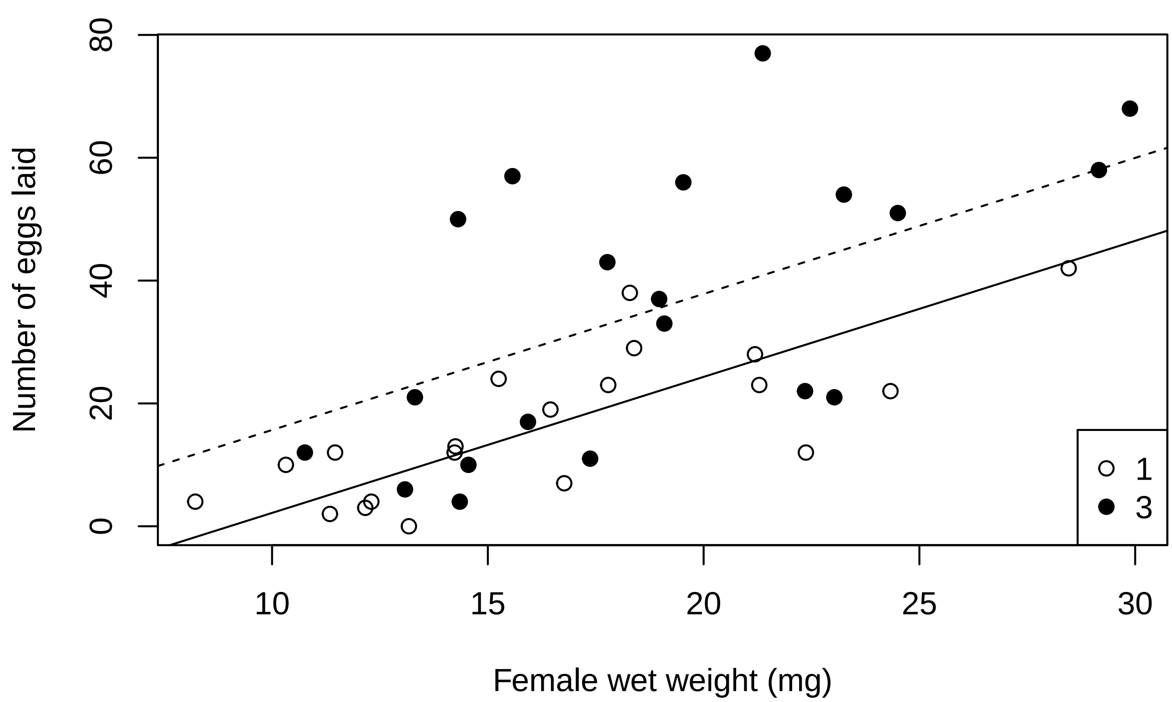
Vamos a estudiar eL modelo en su conjunto. Podemos observar que todos los valores p son significativos, lo que nos indica que hay diferencias entre los grupos maleS1 y males3 (0.00548).

```
##
## Call:
## lm(formula = eggs ~ weight + males, data = photinus)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.533  -9.436   0.174   7.402  36.143
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -19.9787     7.8150  -2.556  0.01481 *
## weight        2.2150     0.4363   5.077 1.11e-05 ***
## males3       13.5015     4.5760   2.951  0.00548 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.05 on 37 degrees of freedom
## Multiple R-squared:  0.5441, Adjusted R-squared:  0.5194
## F-statistic: 22.08 on 2 and 37 DF,  p-value: 4.891e-07
```

La diferencia es:

```
## [1] 13.50152
```

Podemos ver las paralelas que forman ambas rectas:



Intervalo de confianza al 90%

Vemos que la pendiente es la misma en las dos rectas. Por lo tanto el intervalo de confianza para el peso es el siguiente.

```
##       5 %     95 %
## 1.478924 2.950999
```