

Clase_4_actualizada.R

Usuario

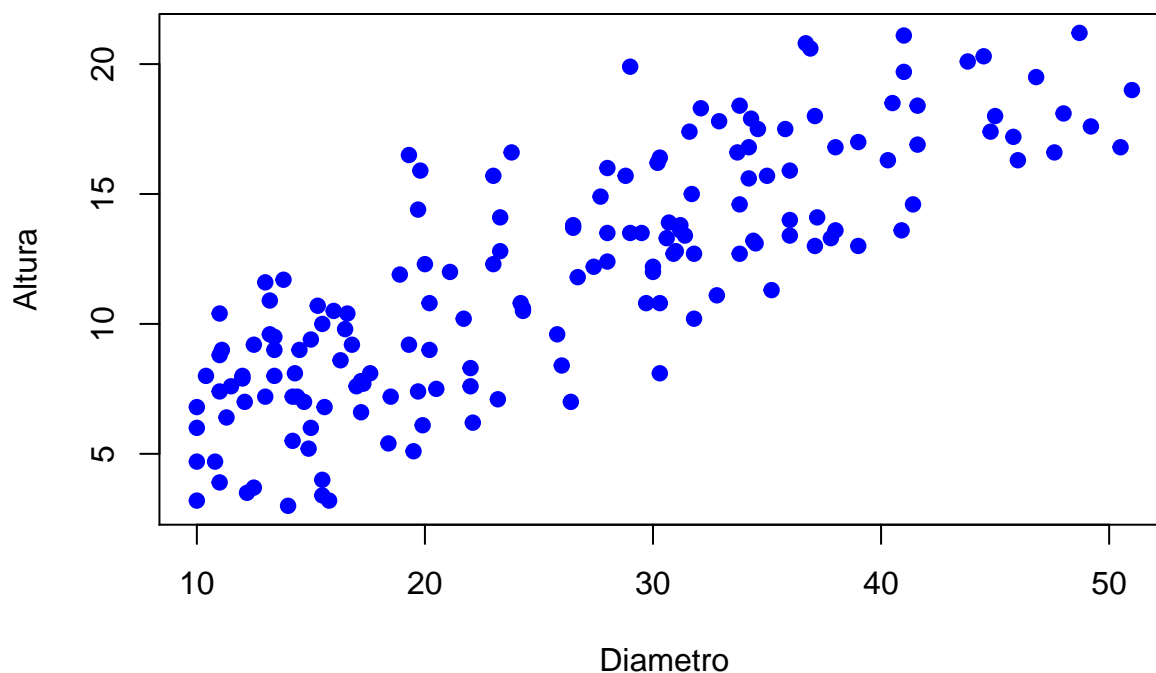
2019-08-10

```
#César A. Martínez Gauna
#09/08/2019
#Clase_4_correlación_regresión

# Ejercicio 1 -----

# Establecimiento de hipótesis -----
#H0.- No existen diferencias estadísticamente significativas en las variables diametro y altura.
#H1.- Sí existen diferencias estadísticamente significativas en las variables diametro y altura.

#Importa datos de Ebanos
ebanos <- read.csv("C:/MCF2019/MCF2/MCF202/ebanos.csv", header= T)
plot(ebanos$diametro, ebanos$altura, pch=19, col= "blue",
     xlab = "Diametro",
     ylab = "Altura")
```



```
library(pastecs)
stat.desc(ebanos$altura, basic = FALSE, norm= TRUE)
```

##	median	mean	SE.mean	CI.mean.0.95	var
----	--------	------	---------	--------------	-----

```
## 12.000000000 11.885365854 0.357428221 0.705786566 20.951809068
##      std.dev      coef.var      skewness      skew.2SE      kurtosis
## 4.577314613 0.385121894 0.053516314 0.141163547 -0.932366816
##      kurt.2SE      normtest.W      normtest.p
## -1.236840496 0.977187792 0.008242431
```

#Con el comando de stat.desc se aplica a la variable altura debido a que es la variable dependiente.

```
# Prueba Shapiro-Wilk para normalidad de datos -----
shapiro.test(log(ebanos$altura))
```

```
##
## Shapiro-Wilk normality test
##
## data:  log(ebanos$altura)
## W = 0.94218, p-value = 3.15e-06
```

```
shapiro.test(ebanos$altura)
```

```
##
## Shapiro-Wilk normality test
##
## data:  ebanos$altura
## W = 0.97719, p-value = 0.008242
```

De acuerdo con el resultado obtenido con la prueba de shapiro, se observa el valor obtenido de p-value de (0.008242), al compararlo con el valor estándar de alfa (0.05) es menor así que aceptamos H1 obteniendo diferencias significativas.

```
shapiro.test(ebanos$diametro)
```

```
##
## Shapiro-Wilk normality test
##
## data:  ebanos$diametro
## W = 0.94921, p-value = 1.215e-05
```

#Si existen diferencias significativas según la prueba de shapiro que nos da un valor de p-value de 1.215e-05.

```
cor.test(ebanos$diametro, ebanos$altura)
```

```
##
## Pearson's product-moment correlation
##
## data:  ebanos$diametro and ebanos$altura
## t = 18.354, df = 162, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.7648115 0.8659458
## sample estimates:
##      cor
## 0.8217467
```

#Al realizar la correlacion aceptamos nos damos cuenta que p-value (2.2e-16) es menor que (0.05) indicando que la hipotesis alternativa H1 es aceptada, es decir, si hay diferencias significativas entre las variables diámetro y altura.

```
# EJERCICIO 2 Correlación -----

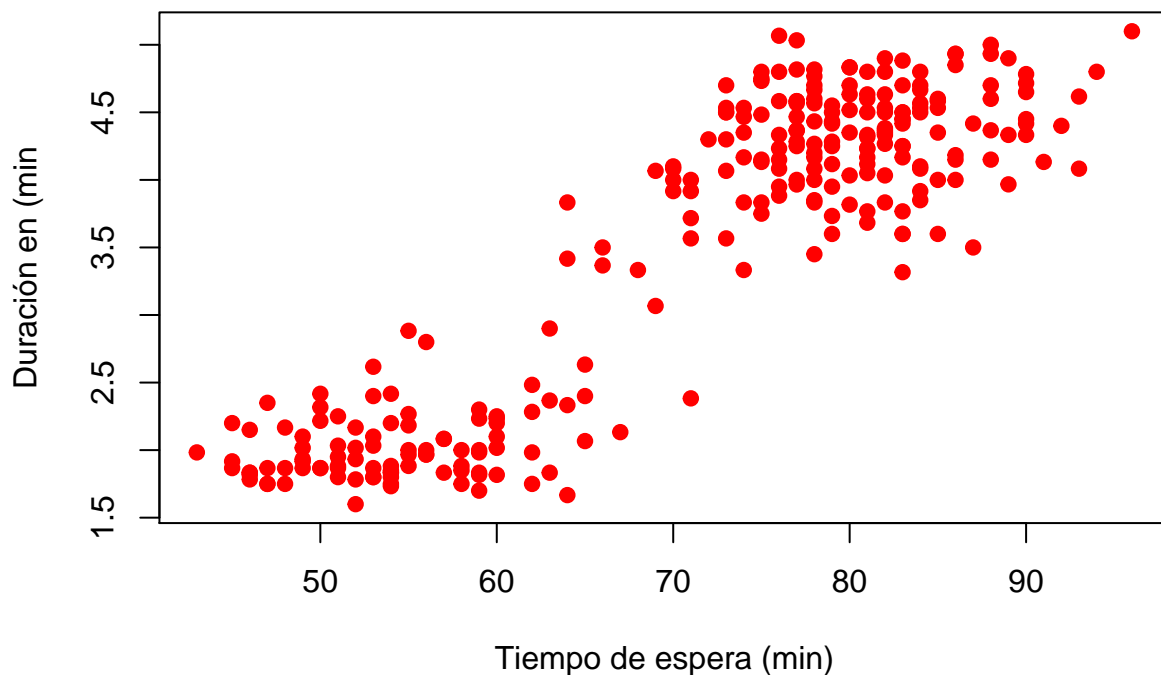
library(repmis)

## Registered S3 method overwritten by 'R.oo':
##   method      from
##   throw.default R.methodsS3

erupciones <- source_data("https://dl.dropboxusercontent.com/s/liir6sil7hkqlxs/erupciones.csv=1")

## Downloading data from: https://dl.dropboxusercontent.com/s/liir6sil7hkqlxs/erupciones.csv=1
## SHA-1 hash of the downloaded data file is:
## b07708389ddf62ee20d19c759c88d7dc2d0da3ac

plot(erupciones$waiting, erupciones$eruptions, pch= 19, col= "red",
     xlab = "Tiempo de espera (min)",
     ylab = "Duración en (min)")
```



```
library(pastecs)
stat.desc(erupciones$eruptions, basic = FALSE, norm = TRUE)
```

	median	mean	SE.mean	CI.mean.0.95	var
##	4.000000e+00	3.487783e+00	6.920580e-02	1.362494e-01	1.302728e+00
##	std.dev	coef.var	skewness	skew.2SE	kurtosis
##	1.141371e+00	3.272483e-01	-4.135498e-01	-1.399854e+00	-1.511605e+00
##	kurt.2SE	normtest.W	normtest.p		

```
## -2.567516e+00 8.459156e-01 9.036119e-16
shapiro.test(erupciones$eruptions)

##
## Shapiro-Wilk normality test
##
## data: erupciones$eruptions
## W = 0.84592, p-value = 9.036e-16
#Segun la prueba de shapiro los datos no son de distribucion normal ya que se encuentran
#por debajo de el alfa establecido 0.05.

shapiro.test(log(erupciones$eruptions))

##
## Shapiro-Wilk normality test
##
## data: log(erupciones$eruptions)
## W = 0.81727, p-value < 2.2e-16

shapiro.test(erupciones$waiting)

##
## Shapiro-Wilk normality test
##
## data: erupciones$waiting
## W = 0.92215, p-value = 1.015e-10

cor.test(erupciones$eruptions, erupciones$waiting)

##
## Pearson's product-moment correlation
##
## data: erupciones$eruptions and erupciones$waiting
## t = 34.089, df = 270, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.8756964 0.9210652
## sample estimates:
## cor
## 0.9008112

#Si hay una correlacion es significativa porque esta por debajo de 0.05 por lo cual se acepta H1

# Regresión Lineal -----

#Hipotesis general: Que el tiempo de espera nos ayudara a predecir la duracion de la
#proxima erupción

#H0= No es significativa para predecir la próxima erupción.
#H1= sí es significativa para predecir la próxima erupción.

#Comando "lm" para realizar la regresión lineal.
lm.erup <- lm(erupciones$eruptions ~ erupciones$waiting)
```

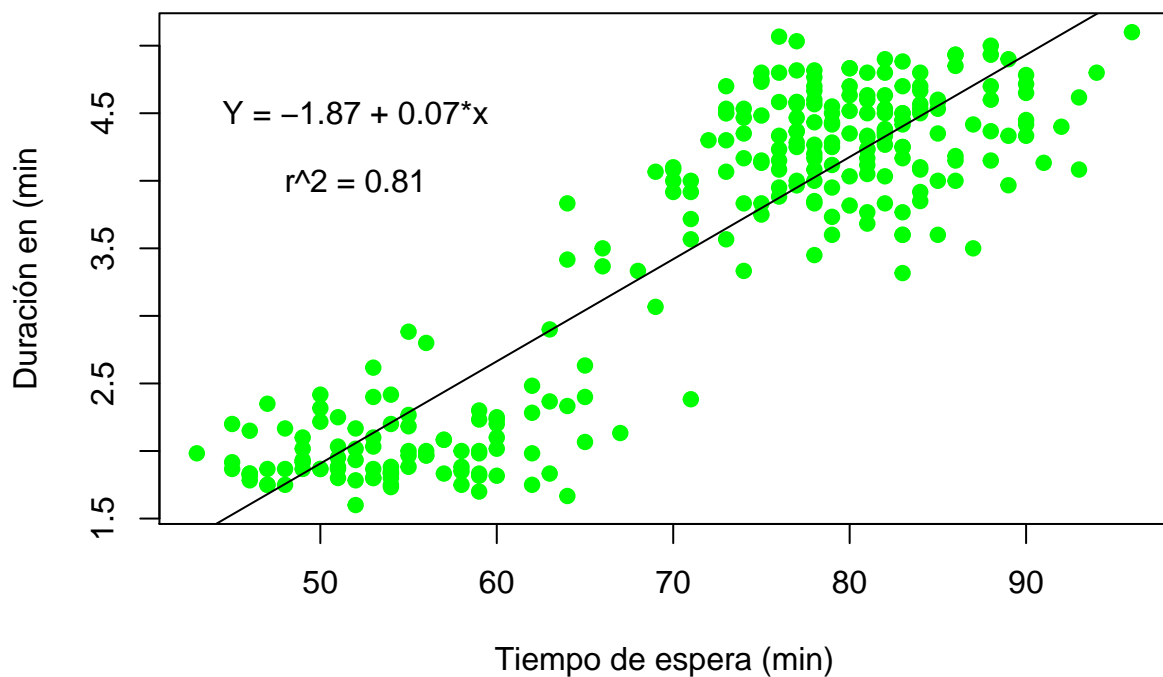
```

#Grafica
plot(erupciones$waiting, erupciones$eruptions, pch= 19, col= "green",
      xlab = "Tiempo de espera (min)",
      ylab = "Duración en (min)")

abline(lm.erup, col= "black")

text(52, 4.5, "Y = -1.87 + 0.07*x")
text(52, 4, "r^2 = 0.81")

```



```
lm.erup
```

```

##
## Call:
## lm(formula = erupciones$eruptions ~ erupciones$waiting)
##
## Coefficients:
##      (Intercept)  erupciones$waiting
##          -1.87402           0.07563

```

```
summary(lm.erup)
```

```

##
## Call:
## lm(formula = erupciones$eruptions ~ erupciones$waiting)
##
## Residuals:

```

```
##      Min      1Q   Median      3Q      Max
## -1.29917 -0.37689  0.03508  0.34909  1.19329
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.874016   0.160143  -11.70  <2e-16 ***
## erupciones$waiting  0.075628   0.002219   34.09  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4965 on 270 degrees of freedom
## Multiple R-squared:  0.8115, Adjusted R-squared:  0.8108
## F-statistic: 1162 on 1 and 270 DF, p-value: < 2.2e-16
```

```
length(erupciones$eruptions)
```

```
## [1] 272
```

```
sqrt(0.90)
```

```
## [1] 0.9486833
```

```
(0.90)^2
```

```
## [1] 0.81
```

```
#Para saber la duración en tiempo de espera del 60 min.
```

```
y.60 <- -1.87 + 0.07*60
```

```
y.60
```

```
## [1] 2.33
```

```
# Datos de regresión -----
```

```
#Comando para asignar la fórmula a un nuevo directorio para no escribir toda la fórmula.
```

```
espera <-erupciones$waiting
```

```
duracion <- erupciones$eruptions
```

```
res <- resid(lm.erup)
```

```
res
```

```
##      1      2      3      4      5
## -0.500591902 -0.409893203 -0.389452162 -0.531916787 -0.021359589
##      6      7      8      9     10
##  0.597478849 -0.081243433 -0.954359589 -0.033009359 -0.204359589
##     11     12     13     14     15
## -0.376893203 -0.561731642  0.175036046  0.069502433  0.296896306
##     16     17     18     19     20
##  0.108362693 -1.064916787  0.321268358 -0.458637307  0.149408098
##     21     22     23     24     25
## -0.183009359  0.069502433 -0.574963954 -0.277312422  0.810547838
##     26     27     28     29     30
## -0.803103694 -0.318521151  0.209291942 -0.174963954  0.332408098
##     31     32     33     34     35
##  0.653175786  0.517663994  0.249571422 -0.143219850  0.110547838
##     36     37     38     39     40
## -0.041637307  0.110874485  0.656780150 -0.755032943 -0.149499329
##     41     42     43     44     45
```

##	0.173780150	-0.629404995	0.088268358	-0.762404995	0.886175786
##	46	47	48	49	50
##	-1.086103694	0.866827317	-0.034265255	0.305524254	-0.588032943
##	51	52	53	54	55
##	1.001919890	-0.216499329	-0.376893203	0.656780150	-0.476893203
##	56	57	58	59	60
##	0.479896306	0.221431682	-1.299172683	0.617663994	0.065152202
##	61	62	63	64	65
##	-0.355032943	0.021268358	-0.006125515	0.472524254	-0.846660891
##	66	67	68	69	70
##	-0.683755225	0.142036046	0.675036046	-0.974800630	1.053175786
##	71	72	73	74	75
##	-0.294475746	-0.394149099	0.399408098	0.504431682	-0.831916787
##	76	77	78	79	80
##	1.193291942	-0.646660891	0.542036046	0.009291942	-0.803103694
##	81	82	83	84	85
##	0.334919890	0.005524254	0.680059630	-0.408800630	0.420175786
##	86	87	88	89	90
##	0.151756567	0.076291942	0.340780150	0.410874485	-0.629987537
##	91	92	93	94	95
##	-0.463660891	-0.599499329	-0.040381411	0.792036046	-1.057544735
##	96	97	98	99	100
##	0.728803734	0.188268358	-0.048080110	-0.116009359	0.572524254
##	101	102	103	104	105
##	-0.331916787	-0.414243433	0.268246537	0.096896306	-0.201847798
##	106	107	108	109	110
##	0.186502433	0.221268358	-0.275637307	0.220012463	-0.568847798
##	111	112	113	114	115
##	0.934919890	-0.288032943	0.043128619	0.316408098	-0.888032943
##	116	117	118	119	120
##	0.381152202	0.409618589	0.045640411	-0.771032943	-0.288615485
##	121	122	123	124	125
##	0.482734745	0.722687578	0.300663994	-0.394149099	-0.181243433
##	126	127	128	129	130
##	-0.484847798	0.387758329	0.172524254	-0.018521151	-0.282499329
##	131	132	133	134	135
##	0.337758329	-0.236103694	0.438850901	-0.523871381	0.228130381
##	136	137	138	139	140
##	0.055524254	-0.100009359	0.303012463	-0.101265255	-0.367591902
##	141	142	143	144	145
##	-0.018847798	-0.430660891	0.205524254	0.867663994	0.459291942
##	146	147	148	149	150
##	-0.605032943	0.456780150	0.185246537	-0.286267017	-0.334265255
##	151	152	153	154	155
##	1.083663994	0.050663994	-0.641800630	0.348152202	0.071431682
##	156	157	158	159	160
##	0.580059630	0.248152202	-1.076383173	-0.334265255	-0.889871381
##	161	162	163	164	165
##	0.670758329	-0.479987537	-0.512404995	-0.191963954	0.382571422
##	166	167	168	169	170
##	0.709291942	-0.523544735	0.218756567	-0.125637307	-0.542383173
##	171	172	173	174	175
##	0.085246537	-0.353777047	0.633663994	0.064315526	-0.084847798
##	176	177	178	179	180

```
## 0.081152202 0.853175786 0.509618589 -0.554359589 0.444547838
##      181      182      183      184      185
## -0.402521151 0.633663994 -0.153103694 -0.636103694 0.049990641
##      186      187      188      189      190
## 0.408036046 -0.395731642 0.228130381 0.013896306 -0.102521151
##      191      192      193      194      195
## 0.548152202 -0.603777047 0.926291942 -0.378731642 0.016663994
##      196      197      198      199      200
## -0.018847798 -1.205615485 0.416663994 0.266990641 0.642036046
##      201      202      203      204      205
## -0.563660891 0.022524254 -0.875127277 -0.267265255 0.575036046
##      206      207      208      209      210
## 0.178130381 0.417663994 -0.628731642 0.101246537 0.096896306
##      211      212      213      214      215
## -1.112568318 0.523780150 0.035246537 0.034919890 0.450827317
##      216      217      218      219      220
## 0.359291942 0.265734745 -0.435011121 -0.285521151 0.276291942
##      221      222      223      224      225
## -0.040381411 -0.060475746 -0.459893203 0.684919890 -0.024963954
##      226      227      228      229      230
## 0.016408098 0.058036046 0.242036046 0.497059630 0.449408098
##      231      232      233      234      235
## 0.663059630 0.207106797 -0.446987537 0.309618589 -0.482499329
##      236      237      238      239      240
## -0.326893203 -0.359893203 0.333663994 -0.150591902 -0.633172683
##      241      242      243      244      245
## 0.351919890 0.669502433 0.303012463 0.009455265 0.028640411
##      246      247      248      249      250
## -0.494475746 -0.353777047 0.039524254 -1.060056526 0.627547838
##      251      252      253      254      255
## -0.009893203 0.046896306 -0.079824214 0.853175786 -0.631243433
##      256      257      258      259      260
## -0.359219850 0.421431682 0.046896306 -0.361149099 0.182408098
##      261      262      263      264      265
## 0.742036046 0.054268358 -0.662404995 -0.153103694 0.605014224
##      266      267      268      269      270
## -0.413660891 0.951919890 -0.134847798 0.545130381 -0.515499329
##      271      272
## 0.212130381 0.744547838
```

```
sum(res)
```

```
## [1] 6.973588e-16
```

```
pre <- fitted(lm.erup)
```

```
res.2 <- res^2
```

```
cuadro <- round(data.frame(espera, duracion, pre, res,
                           res.2),4)
```

```
SSE <- sum(cuadro$res.2)
```

```
SSE
```

```
## [1] 66.5612
```

```
SSE <- sum((duracion - pre)^2)
```

```
SSE
```



```
## [1] 66.56178
vari <- SSE/(length(erupciones$waiting)-2)
vari

## [1] 0.2465251
# Prueba de hipotesis de la regresión -----

an.erup <- anova(lm.erup)
an.erup

## Analysis of Variance Table
##
## Response: erupciones$eruptions
##           Df Sum Sq Mean Sq F value    Pr(>F)
## erupciones$waiting  1 286.478  286.478  1162.1 < 2.2e-16 ***
## Residuals          270  66.562    0.247
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#aceptamos la hipótesis alternativa que el modelo de regresion aplicado son
#significativos, entonces podemos decir que la regresión se puede aplicar.
```