Examen final

Nombres y apellidos:

Carlos Alonso Aznarán Laos

Curso:

Curso Taller Software R

Fecha:

23/09/2018

# Pregunta 1

En el desarrollo de un nuevo método para la determinación de niveles de alcohol en la sangre, se analizó una muestra de sangre, con los resultados siguientes:

* Método nuevo: 64.5, 66.0, 63.9, 65.1 y 64.0 mg/100ml.
* Método de análisis estándar (aplicado a la misma muestra): 66.2, 65.8, 66.3 y 65.6 mg/100ml.

# Solución 2

Dado que no conocemos la muestra utilizada, pero conocemos las concentraciones de sangre (en mg/100ml), podemos analizar las medias de cada uno de los métodos:

La sintaxis en R es la siguiente:

newmethod <- c(64.5, 66.0, 63.9, 65.1, 64.0); newmethod # mg/100 ml

standarmethod <- c(66.2, 65.8, 66.3, 65.6); standarmethod # mg/100 ml

# method <- data.frame(newmethod, standarmethod, stringsAsFactors = TRUE)

mean\_new <- mean(newmethod)

mean\_standar <- mean(standarmethod)

difference <- mean\_new - mean\_standar

La diferencia es de 1.275.

# Pregunta 2

Evaluar la relación entre el consumo a los 120Km/h (litros/100km) de un vehículo

todo terreno con la cilindrada (capacidad de un motor medida en cm3) que posee,

mediante el ajuste de un modelo de regresión lineal. Utilizar el archivo

*vehiculos.sav*.

1. ¿Cuál es el modelo que se utilizará en el análisis de estos datos?
2. ¿La variable implícita en el modelo es realmente relevante para explicar la variable respuesta y cuál sería el modelo ajustado?
3. ¿Existe evidencia de que este modelo sirve para explicar la variable respuesta?, ¿Qué porcentaje de la variabilidad explica?
4. ¿Podemos concluir que el ajuste del modelo es estadísticamente válido?

# Solución 2

1. El modelo que se empleará es el de *regresión lineal simple*.
2. La variable de respuesta $y\_i$ es el consumo medio (a los 120 km/h) y la variable independiente $x\_i$ la cilindrada del auto. Ahora mostraremos la nube de dispersión

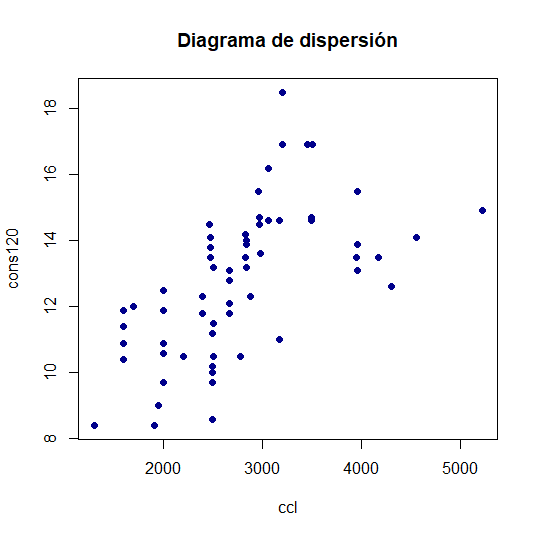
Luego, se puede observar que podría haber una regresión lineal de la forma

\begin{equation}\label{eq:1}

y\_i=\hat{\alpha} + \hat{\beta}x\_i + \epsilon

\end{equation}

Con los estimadores $\hat{\alpha}$ (intercepto o variable independiente de \eqref{eq:1}), $\hat{\beta}>0$, donde $\epsilon $ es el error aleatorio debido a las asunciones del modelo, $y\_i$ es la posición $i$-ésima cuando $i\in\{1,\ldots,110\}$ de la variable de respuesta (consumo medio a 120km/h) y $x\_i$.



La sintáxis en R es la siguiente:

setwd("D:/Clase\_de\_R\_8")

library(foreign)

vehiculos <-read.spss(file="vehiculos.sav", to.data.frame=TRUE)

fix(vehiculos)

attach(vehiculos)

# Análisis univariado

summary(vehiculos)

# Correlación, prueba de independencia y diagrama de dispersión

cor.test(cons120, ccl, alternative = "two.sided", method = "pearson")

plot(ccl, cons120, col="blue5", main="Diagrama de dispersión", pch= 19)

# Formulación modelo. Y=consumo medio, X=cilindrada

# Prueba de normalidad en la variable dependiente

shapiro.test(cons120)

Pearson's product-moment correlation

data: cons120 and ccl

t = 8.7503, df = 108, p-value = 3.179e-14

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.5194973 0.7418606

sample estimates:

cor

0.6440875

Shapiro-Wilk normality test

data: cons120

W = 0.97437, p-value = 0.03221

A

Call:

lm(formula = cons120 ~ ccl)

Residuals:

Min 1Q Median 3Q Max

-3.4656 -1.3146 0.3904 1.1812 4.9928

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.956744 0.628020 11.08 < 2e-16 \*\*\*

ccl 0.002048 0.000234 8.75 3.18e-14 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.774 on 108 degrees of freedom

Multiple R-squared: 0.4148, Adjusted R-squared: 0.4094

F-statistic: 76.57 on 1 and 108 DF, p-value: 3.179e-14

2.5 % 97.5 %

(Intercept) 5.7118981 8.201589081

ccl 0.0015838 0.002511493

Call:

aov(formula = regresion)

Terms:

ccl Residuals

Sum of Squares 240.9088 339.8061

Deg. of Freedom 1 108

Residual standard error: 1.773796

Estimated effects may be unbalanced

Estimate Std. Error t value Pr(>|t|)

ccl 0.00454406 9.168151e-05 49.56354 1.351909e-76