

Parte III Guías Prácticas de R

Lilian Martínez

November 26, 2015

```
#UNIDAD 5: Práctica 21 - Prueba de hipótesis estadísticas y prueba de normalidad.
```

```
# PRUEBAS DE NORMALIDAD DE UNA MUESTRA
```

```
#A continuación procedemos a contrastar normalidad para los datos del IMC en  
#los grupos de Control y de Pacientes
```

```
# se digitan los datos del grupo de control
```

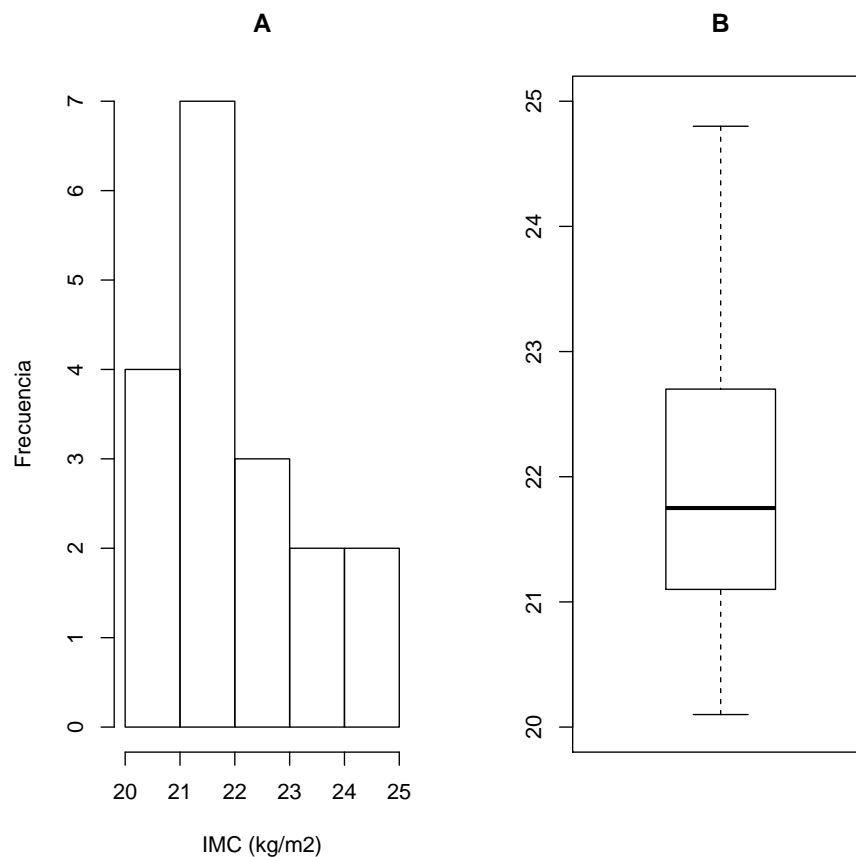
```
IMC_Control <- c(23.6, 22.7, 21.2, 21.7, 20.7, 22.0, 21.8, 24.2, 20.1, 21.3, 20.5, 21.1, 2  
par(mfrow=c(1,2))
```

```
# se genera el histograma de la variables de interés
```

```
hist(IMC_Control,main="A",xlab="IMC (kg/m2)",ylab="Frecuencia")
```

```
# se genera el diagrama de caja de la variable de interés y se muestra en la misma ventana
```

```
boxplot(IMC_Control,main="B", lab="IMC (kg/m2)",ylim=c(20,25))
```



```
# los commands para contrastar normalidad son los siguientes
```

```
sw <- shapiro.test(IMC_Control)
```

```
sw
```

```
##
##  Shapiro-Wilk normality test
##
## data:  IMC_Control
## W = 0.95321, p-value = 0.4776

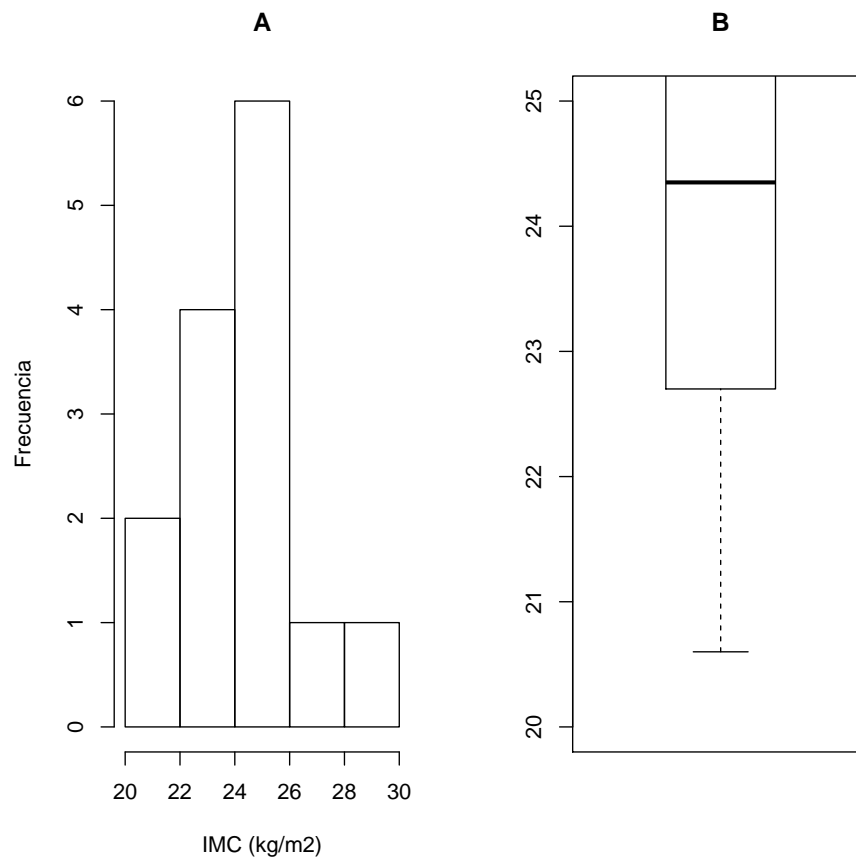
#La prueba de Kolmogorov
ks <- ks.test(IMC_Control,"pnorm",mean=mean(IMC_Control),sd=sd(IMC_Control))
ks

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  IMC_Control
## D = 0.11172, p-value = 0.9595
## alternative hypothesis: two-sided

# luego se digitan los datos para pacientes y se ejecutan las mismas instrucciones
IMC_Pacientes <- c(25.6, 22.7, 25.9, 24.3, 25.2, 29.6, 21.3, 25.5, 27.4, 22.3, 24.4, 23.7,

# se genera el histograma de la variables de interés
hist(IMC_Pacientes,main="A",xlab="IMC (kg/m2)",ylab="Frecuencia")

# se genera el diagrama de caja de la variable de interés y se muestra en la misma ventana
boxplot(IMC_Pacientes,main="B", lab="IMC (kg/m2)",ylim=c(20,25))
```



```
# los comandos para contrastar normalidad son los siguientes
sw <- shapiro.test(IMC_Pacientes)
sw

##
##  Shapiro-Wilk normality test
##
## data:  IMC_Pacientes
## W = 0.97437, p-value = 0.929

#La prueba de Kolmogorov
ks <- ks.test(IMC_Pacientes,"pnorm",mean=mean(IMC_Pacientes),sd=sd(IMC_Pacientes))
ks

##
##  One-sample Kolmogorov-Smirnov test
##
## data:  IMC_Pacientes
## D = 0.12172, p-value = 0.9695
## alternative hypothesis: two-sided
```

#UNIDAD 5: Práctica 22 - Prueba de hipótesis estadísticas. Una población

#1.PRUEBA DE HIPÓTESIS ACERCA DEL VALOR DE UNA PROPORCIÓN

*#Una muestra de 100 empleados que habían estado en contacto con sangre o derivados de ésta,
#examinada por presentar evidencia serológica de hepatitis B. Se encontró que 23 de ellos
#reacción positiva. ¿Puede concluirse a partir de estos datos que la proporción de los pos
#mayor que 0.15? Tome un nivel de significancia del 5%.*

Construyendo una función en R para realizar la prueba de hipótesis.

```
Prueba.prop <- function(x, n, po, H1="Distinto", alfa=0.05)
{
  op <- options();
  options(digits=2)
  pe=x/n #calcula la proporción muestral
  SE <- sqrt((po * (1-po))/n) # calcula la varianza de la proporción muestral
  Zo <- (pe-po)/SE #calcula el estadístico de prueba
  # Si lower.tail = TRUE (por defecto), P[X <= x], en otro caso P[X > x]
  if (H1 == "Menor" || H1 == "Mayor")
  {
    Z <- qnorm(alfa, mean=0, sd=1, lower.tail = FALSE, log.p = FALSE)
    #calcula los valores críticos de la distribución N(0;1) en el caso de una prueba unila
    valores <- rbind(Prop_Estimada=pe, Prop_Hipotetica=po, Z_critico=Z, Estadistico= Zo)
  }
  else
  {
    Z <- qnorm(alfa/2, mean=0, sd=1, lower.tail = FALSE, log.p = FALSE)
    #calcula los valores críticos de la distribución N(0;1) en el caso de una prueba bilate
    valores <- rbind(Prop_Estimada=pe, Prop_Hipotetica =po, Z_critico_menor=-Z,
                    Z_critico_mayor =Z, Zo)
  } # esto es para encontrar los valores críticos
  if (H1 == "Menor")
  {
    if (Zo < -Z) decision <- paste("Como Estadistico <", round(-Z,3), ", entonces rechazamos
    else decision <- paste("Como Estadistico>=", round(-Z,3), ", entonces aceptamos Ho")
  }
  if (H1 == "Mayor")
  {
    if (Zo > Z) decision <- paste("Como Estadistico >", round(Z,3), ", entonces rechazamos
    else decision <- paste("Como Estadistico <=", round(Z,3), ", entonces aceptamos Ho")
  }
  if (H1 == "Distinto")
  {
    if (Zo < -Z) decision <- paste("Como Estadistico <", round(-Z,3), ", entonces rechazamos
    if (Zo > Z) decision <- paste("Como Estadistico >", round(Z,3), ", entonces rechazamos
    else decision <- paste("Como Estadistico pertenece a [", round(-Z,3), ",", round(Z,3),
                          entonces aceptamos Ho")
  } # esto para llevar a cabo los contraste de hipótesis
  print(valores)
```

```

    print(decision)
    options(op) # restablece todas las opciones iniciales
}

Prueba.prop(23, 100, 0.15, H1="Menor", alfa=0.05)

##           [,1]
## Prop_Estimada 0.23
## Prop_Hipotetica 0.15
## Z_critico      1.64
## Estadistico    2.24
## [1] "Como Estadistico >= -1.645 , entonces aceptamos Ho"

Prueba.prop(23, 100, 0.15, H1="Mayor", alfa=0.05)

##           [,1]
## Prop_Estimada 0.23
## Prop_Hipotetica 0.15
## Z_critico      1.64
## Estadistico    2.24
## [1] "Como Estadistico > 1.645 , entonces rechazamos Ho"

Prueba.prop(23, 100, 0.15, H1="Distinto", alfa=0.05)

##           [,1]
## Prop_Estimada 0.23
## Prop_Hipotetica 0.15
## Z_critico_menor -1.96
## Z_critico_mayor 1.96
## Zo              2.24
## [1] "Como Estadistico > 1.96 , entonces rechazamos Ho"

prop.test(x=23, n=100, p=0.15, alternative="less", conf.level=0.95)

##
## 1-sample proportions test with continuity correction
##
## data: 23 out of 100, null probability 0.15
## X-squared = 4.4118, df = 1, p-value = 0.9822
## alternative hypothesis: true p is less than 0.15
## 95 percent confidence interval:
##  0.0000000 0.3111509
## sample estimates:
##      p
## 0.23

prop.test(x=23, n=100, p=0.15, alternative="greater", conf.level=0.95)

##
## 1-sample proportions test with continuity correction
##

```

```

## data: 23 out of 100, null probability 0.15
## X-squared = 4.4118, df = 1, p-value = 0.01785
## alternative hypothesis: true p is greater than 0.15
## 95 percent confidence interval:
## 0.1640827 1.0000000
## sample estimates:
## p
## 0.23

prop.test(x=23, n=100, p=0.15, alternative="two.sided", conf.level=0.95)

##
## 1-sample proportions test with continuity correction
##
## data: 23 out of 100, null probability 0.15
## X-squared = 4.4118, df = 1, p-value = 0.03569
## alternative hypothesis: true p is not equal to 0.15
## 95 percent confidence interval:
## 0.154215 0.326941
## sample estimates:
## p
## 0.23

#PRUEBA DE HIPÓTESIS SOBRE UNA MEDIA, VARIANZA DESCONOCIDA.
#Los siguientes datos corresponden a la longitud medida en centímetros de 18 pedazos de ca
#sobrantes en cada rollo utilizado: 9.0 3.41 6.13 1.99 6.92 3.12 7.86 2.01
#5.98 4.15 6.87 1.97 4.01 3.56 8.04 3.24 5.05 7.37

#Una solución podría ser la siguiente:
Datos = c(9.0,3.41,6.13,1.99,6.92,3.12,7.86,2.01,5.98,4.15,6.87,1.97,4.01,3.56,8.04,3.24,5.05,7.37)

# digitamos las observaciones
t.test(Datos,mu=4,alternative="greater")

##
## One Sample t-test
##
## data: Datos
## t = 1.9291, df = 17, p-value = 0.03529
## alternative hypothesis: true mean is greater than 4
## 95 percent confidence interval:
## 4.101959 Inf
## sample estimates:
## mean of x
## 5.037778

#UNIDAD 5: Práctica 23 - Prueba de hipótesis estadísticas. Dos poblaciones.

#PRUEBAS SOBRE DOS MUESTRAS INDEPENDIENTES
#Volviendo al problema de la importancia del estadonutricional (introducido en la practica

```

#pacientes diabéticos (pacientes) y saludables (grupo control) con complicaciones. Los datos muestran en los siguientes cuadros.

#Las hipótesis a contrastar son:

#H0: $\mu_1 = \mu_2$

#H1: $\mu_1 \neq \mu_2$

#En lenguaje R está implementada la prueba t, el siguiente código ejemplo la calcula para las muestras:

Primero digitamos las observaciones correspondientes a ambas muestras

```
IMC_Control <- c(23.6, 22.7, 21.2, 21.7, 20.7, 22.0, 21.8, 24.2, 20.1, 21.3, 20.5, 21.1, 20.4, 23.3, 24.8)
```

```
IMC_Pacientes <- c(25.6, 22.7, 25.9, 24.3, 25.2, 29.6, 21.3, 25.5, 27.4, 22.3, 24.4, 23.7,
```

Realizamos el contraste de igualdad de medias

```
t.test(IMC_Control, IMC_Pacientes, var.equal=TRUE, mu=0)
```

```
##
```

```
## Two Sample t-test
```

```
##
```

```
## data: IMC_Control and IMC_Pacientes
```

```
## t = -3.5785, df = 30, p-value = 0.001198
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -3.770935 -1.030653
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 21.97778 24.37857
```

#Se concluye entonces que existe diferencia significativa en el IMC para ambos grupos de pacientes, pues el p valor de la prueba resulta ser muy pequeño.

#PRUEBAS SOBRE DOS MUESTRAS PAREADAS

#Se cuenta con los datos simulados (con fines didácticos), de las observaciones de la presión arterial sistólica (PAS) en un grupo de 10 pacientes antes y después de un tratamiento con una dieta especial de bajosodio y medicamentos.

#Las hipótesis a contrastar son:

#H0: $\mu_1 = \mu_2$

#H1: $\mu_1 \neq \mu_2$

#El código en lenguaje R para calcular la prueba t para dos muestras apareadas es el siguiente

```
PAS.antes <- c(160,155,180,140,150,130,190,192,170,165)
```

```
PAS.despues <- c(139,135,175,120,145,140,170,180,149,146)
```

#verificando la normalidad

```
shapiro.test(PAS.antes)
```

```
##
```



```

## Shapiro-Wilk normality test
##
## data: PAS.antes
## W = 0.97021, p-value = 0.8928

shapiro.test(PAS.despues)

##
## Shapiro-Wilk normality test
##
## data: PAS.despues
## W = 0.92548, p-value = 0.4049

ks.test(PAS.antes, "pnorm", mean=mean(PAS.antes), sd=sd(PAS.antes))

##
## One-sample Kolmogorov-Smirnov test
##
## data: PAS.antes
## D = 0.10476, p-value = 0.9992
## alternative hypothesis: two-sided

ks.test(PAS.despues, "pnorm", mean=mean(PAS.despues), sd=sd(PAS.despues))

##
## One-sample Kolmogorov-Smirnov test
##
## data: PAS.despues
## D = 0.21871, p-value = 0.6495
## alternative hypothesis: two-sided

#realizando la prueba t
t.test(PAS.antes, PAS.despues, paired=TRUE, mu=0)

##
## Paired t-test
##
## data: PAS.antes and PAS.despues
## t = 4.0552, df = 9, p-value = 0.002862
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 5.880722 20.719278
## sample estimates:
## mean of the differences
## 13.3

#El valor del estadístico t es 4.0552, con gl = 9, P = 0.0029. Con estos resultados se rechaza
#tanto se concluye que la PAS antes y después del tratamiento es distinta, es decir, el tratamiento
#ha sido efectivo.

```

```

#PRUEBA DE HIPÓTESIS ACERCA DE LA VARIANZA DE DOS POBLACIONES
#El director de una sucursal de una compañía de seguros espera que dos de sus mejores agentes
#consigan formalizar por término medio el mismo número de pólizas mensuales.
#Los datos indican las pólizas formalizadas en los últimos 5 meses por ambos agentes

#introduciendo los datos
Agente_A <- c(12, 11, 18, 16, 13)
Agente_B <- c(14, 18, 18, 17, 16)
# realizando el contraste de igualdad de varianzas
var.test(Agente_A, Agente_B)

##
## F test to compare two variances
##
## data: Agente_A and Agente_B
## F = 3.0357, num df = 4, denom df = 4, p-value = 0.3075
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.3160711 29.1566086
## sample estimates:
## ratio of variances
## 3.035714

#UNIDAD 6: Práctica 24 - Análisis de Varianza (ANOVA).
#EJEMPLO 1.
#El Ministerio de Educación está interesado en implementar tres programas de estudio; con
#de medir la habilidad de lectura en los alumnos. Para ello, se eligen alumnos del sexto grado
#Colegio de San Salvador, 27 alumnos fueron asignados al azar, a cada uno de los tres grupos
#utilizó un programa diferente en cada grupo, se llevó a cabo un examen al inicio y al final de la
#implementación de los programas, los valores obtenidos representan la diferencia que hay entre la
#nota del examen que se hizo al inicio y al final de la implementación del programa.

#Contraste a un nivel de significancia del 5% de que los tres métodos de lectura producen
#efecto en la habilidad de lectura de los alumnos.

# Se digitan las observaciones
notas <- c(20,18,18,23,22,17,15,13,21,15,20,13,12,16,17,21,15,13,12,15,18,20,18,17,10,24,16)

# Se crea un vector de datos en el cual se diferencia cada uno de los programas de estudio
programas <- gl(n=3, k=9, labels=c("P1", "P2", "P3"))

#Crea la matriz de datos que contendrá la información del experimento
datos <- data.frame(notas = notas, programas = programas);datos

##      notas programas
## 1      20          P1
## 2      18          P1
## 3      18          P1
## 4      23          P1

```

```
## 5      22      P1
## 6      17      P1
## 7      15      P1
## 8      13      P1
## 9      21      P1
## 10     15      P2
## 11     20      P2
## 12     13      P2
## 13     12      P2
## 14     16      P2
## 15     17      P2
## 16     21      P2
## 17     15      P2
## 18     13      P2
## 19     12      P3
## 20     15      P3
## 21     18      P3
## 22     20      P3
## 23     18      P3
## 24     17      P3
## 25     10      P3
## 26     24      P3
## 27     16      P3

#Aplicando el análisis de varianza
mod1 <- aov(notas ~ programas, data = datos)

#Mostrando la tabla ANOVA
summary(mod1)

##              Df Sum Sq Mean Sq F value Pr(>F)
## programas      2  36.22   18.11    1.44  0.257
## Residuals     24 301.78   12.57

#UNIDAD 6: Práctica 25 - Diseños por bloques

#EJEMPLO 1.
#Se probaran 5 raciones respecto a sus diferencias en el engorde de novillos. Se dispone d
#para el experimento, que se distribuyen en 4 bloques (5 novillos por bloque) con base a s
#iniciar la prueba de engorde, los novillos más pesados se agruparon en un bloque, en otro
#agruparon los 5 siguientes más pesados y así sucesivamente. Los 5 tratamientos (raciones)
#asignaron al azar dentro de cada bloque.

#Utilizando un nivel de significancia del 5%, contraste la hipótesis de que las cinco
#raciones de comida producen el mismo efecto de engorde en los novillos.

# Definiendo el vector que contendrá el bloque al cual pertenecen los novillos.
bloques <- gl(n=4, k=1, length=20);bloques

## [1] 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
## Levels: 1 2 3 4
```

```

# Se crea el vector que contendrá los tratamientos de los novillos (raciones de alimento)
tratamientos <- gl(n=5, k=4);tratamientos

## [1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
## Levels: 1 2 3 4 5

# Se digitan los pesos de los novillos
peso <- c(0.9,1.4,1.4,2.3,3.6,3.2,4.5,4.1,0.5,0.9,0.5,0.9,3.6,3.6,3.2,3.6,1.8,1.8,0.9,1.4)

## [1] 0.9 1.4 1.4 2.3 3.6 3.2 4.5 4.1 0.5 0.9 0.5 0.9 3.6 3.6 3.2 3.6 1.8
## [18] 1.8 0.9 1.4

# Se registra en una hoja de datos los resultados del experimento
datos2 <- data.frame(bloques = bloques, tratamientos = tratamientos, peso = peso);datos2

##      bloques tratamientos peso
## 1          1             1  0.9
## 2          2             1  1.4
## 3          3             1  1.4
## 4          4             1  2.3
## 5          1             2  3.6
## 6          2             2  3.2
## 7          3             2  4.5
## 8          4             2  4.1
## 9          1             3  0.5
## 10         2             3  0.9
## 11         3             3  0.5
## 12         4             3  0.9
## 13         1             4  3.6
## 14         2             4  3.6
## 15         3             4  3.2
## 16         4             4  3.6
## 17         1             5  1.8
## 18         2             5  1.8
## 19         3             5  0.9
## 20         4             5  1.4

# Se aplica el análisis de varianza
mod2 <- aov(peso ~ tratamientos + bloques, data = datos2)

# Se muestra la tabla ANOVA del experimento
summary(mod2)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## tratamientos  4 30.712   7.678  39.107 8.6e-07 ***
## bloques       3  0.461   0.154   0.784  0.526
## Residuals    12  2.356   0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

#UNIDAD 6: Práctica 26 - Diseños bifactoriales

#EJEMPLO 1.

#Se llevó a cabo un estudio del efecto de la temperatura sobre el porcentaje de encogimiento de telas teñidas, con dos réplicas para cada uno de cuatro tipos de tela en un diseño totalmente aleatorizado. Los datos son el porcentaje de encogimiento de dos réplicas de tela secadas a cuatro temperaturas.

#Utilizando un nivel de significancia del 5%
Definiendo el vector que contendrá el factor A.

```
FactorA <- gl(n=4, k=8, length=32);FactorA
```

```
## [1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4
## Levels: 1 2 3 4
```

Se crea el vector que contendrá los tratamientos de los novillos (raciones de alimento).

```
FactorB<- gl(n=4, k=2,length=32);FactorB
```

```
## [1] 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4
## Levels: 1 2 3 4
```

Se digitan los pesos de los novillos

```
Porcentaje <- c(1.8, 2.1, 2.0, 2.1, 4.6, 5.0, 7.5, 7.9, 2.2, 2.4,4.2, 4.0, 5.4, 5.6,
9.8, 9.2, 2.8, 3.2, 4.4, 4.8, 8.7, 8.4, 13.2, 13.0, 3.2, 3.6, 3.3, 3.5, 5.7, 5.8,
10.9, 11.1);Porcentaje
```

```
## [1] 1.8 2.1 2.0 2.1 4.6 5.0 7.5 7.9 2.2 2.4 4.2 4.0 5.4 5.6
## [15] 9.8 9.2 2.8 3.2 4.4 4.8 8.7 8.4 13.2 13.0 3.2 3.6 3.3 3.5
## [29] 5.7 5.8 10.9 11.1
```

Se registra en una hoja de datos los resultados del experimento

```
datos3 <- data.frame(FactorA = FactorA, FactorB = FactorB, Porcentaje=Porcentaje);datos3
```

```
##      FactorA FactorB Porcentaje
## 1          1          1         1.8
## 2          1          1         2.1
## 3          1          2         2.0
## 4          1          2         2.1
## 5          1          3         4.6
## 6          1          3         5.0
## 7          1          4         7.5
## 8          1          4         7.9
## 9          2          1         2.2
## 10         2          1         2.4
## 11         2          2         4.2
## 12         2          2         4.0
## 13         2          3         5.4
## 14         2          3         5.6
## 15         2          4         9.8
## 16         2          4         9.2
## 17         3          1         2.8
```

```
## 18      3      1      3.2
## 19      3      2      4.4
## 20      3      2      4.8
## 21      3      3      8.7
## 22      3      3      8.4
## 23      3      4     13.2
## 24      3      4     13.0
## 25      4      1      3.2
## 26      4      1      3.6
## 27      4      2      3.3
## 28      4      2      3.5
## 29      4      3      5.7
## 30      4      3      5.8
## 31      4      4     10.9
## 32      4      4     11.1

# Se aplica el análisis de varianza
mod3 <- aov(Porcentaje ~ FactorA * FactorB, data = datos3)

# Se muestra la tabla ANOVA del experimento
summary(mod3)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## FactorA         3  41.88   13.96  279.18 5.05e-14 ***
## FactorB         3 283.94   94.65 1892.91 < 2e-16 ***
## FactorA:FactorB  9  15.86    1.76   35.24 7.09e-09 ***
## Residuals      16   0.80    0.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```