



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

# MÓDULO 6: Apartado 2

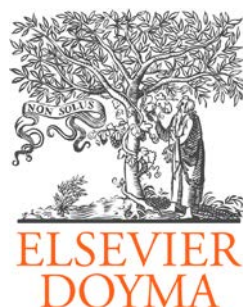
## Ejemplos

---

Héctor Rufino  
José Antonio González

Departament d'Estadística  
i Investigació Operativa  
UNIVERSITAT POLITÈCNICA DE CATALUNYA

 equator  
network



MEDICINA  
CLINICA

 TRIALS  
TRIALS

# 1. Generación de los datos basales

Generar una tabla sobre los datos basales para, al menos, una variable categórica, una ordinal, una numérica simétrica y una numérica asimétrica.

- **Datos:** IST (International Stroke Trial).

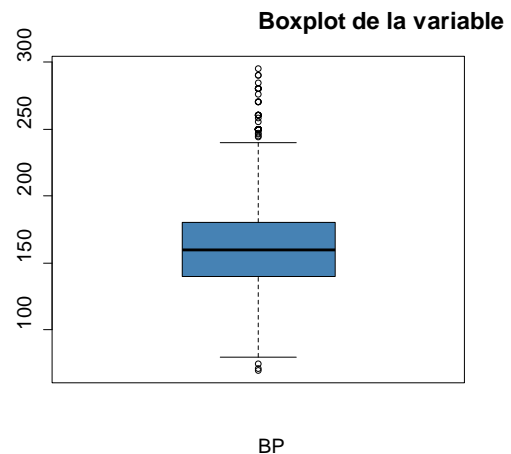
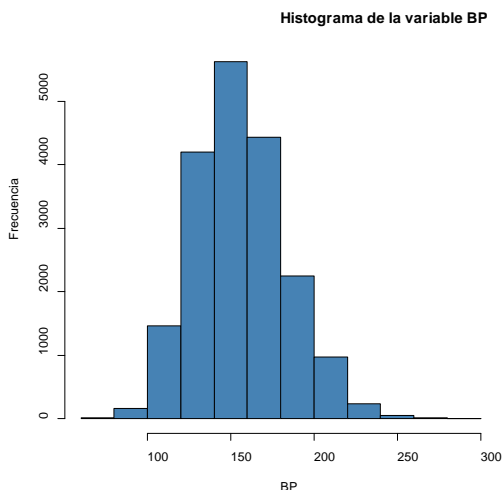
A la hora de mostrar el resumen de las diferentes variables para los datos basales se siguen las indicaciones de la guía CONSORT, que en el ítem 15 especifica cómo resumir una variable en función de cómo sea la misma. Entre las opciones disponibles, mostramos:

- **Variable simétrica:** aportar media y desviación estándar.
- **Variable asimétrica:** mostrar la mediana y el primer y tercer cuartil.
- **Variable categórica y variable ordinal:** resumir a través del número de observaciones en cada categoría y el porcentaje que representa.

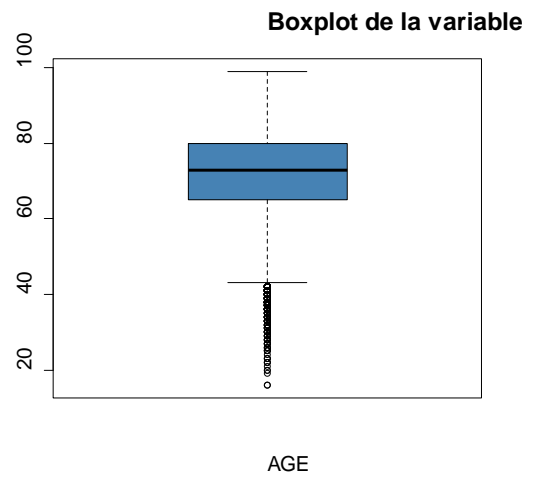
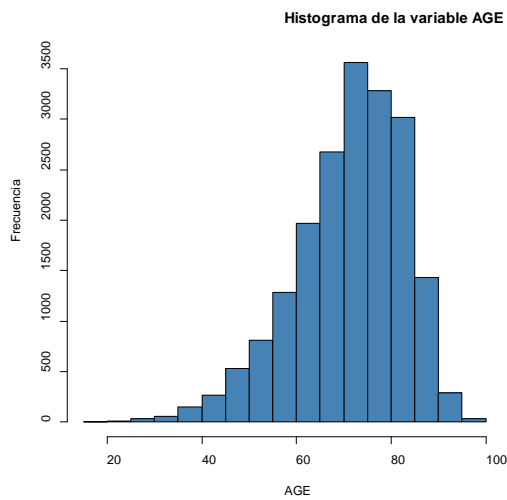
Para comprobar si una variable numérica es simétrica o asimétrica se pueden utilizar diferentes herramientas. Por ejemplo, para hacerlo de forma visual, se puede utilizar un histograma o Box-Plot.

Para este estudio, *The International Stroke Trial*, mostramos las siguientes variables que se resumirán en una tabla sobre datos basales:

- **Variable (razonablemente) simétrica:** presión sanguínea (BP)



- Variable continua asimétrica: edad (AGE)



- Variable categórica: género (SEX) y nivel de conciencia (RCONS)
- Variable ordinal: categorización de la variable numérica tiempo entre el AVC y la asignación al grupo (RDELAY)

La tabla resumen de los datos basales es:

VARIABLE	RESUMEN BASAL
<b>Systolic BP (mm Hg)</b>	160.16 (27.61)
<b>Age (yr)</b>	73 (65 / 80)
<b>Gender</b>	
Female	9028 (46.45%)
Male	10407 (53.55%)
<b>Conscious level</b>	
Unconscious	260 (1.33%)
Drowsy	4254 (21.89%)
Alert	14921 (76.77%)
<b>Delay (h)</b>	
0-3	843 (4.34%)
4-6	2322 (11.95%)
7-12	4114 (21.17%)
13-24	3180 (20.48%)
25-48	8176 (42.07%)

Se aporta, para cada variable, media (sd) o mediana (1r cuartil / 3r cuartil) o número (porcentaje)

## 2. Generación de los resultados principales

### 2.1. Respuesta dicotómica

- **Datos:** IST (International Stroke Trial)
- **Respuesta:** presencia (1) o ausencia (0) de evento en los primeros 14 días.
- **Intervenciones estudiadas:** (a) Heparina frente a No Heparina; y b) Aspirina frente a No Aspirina.

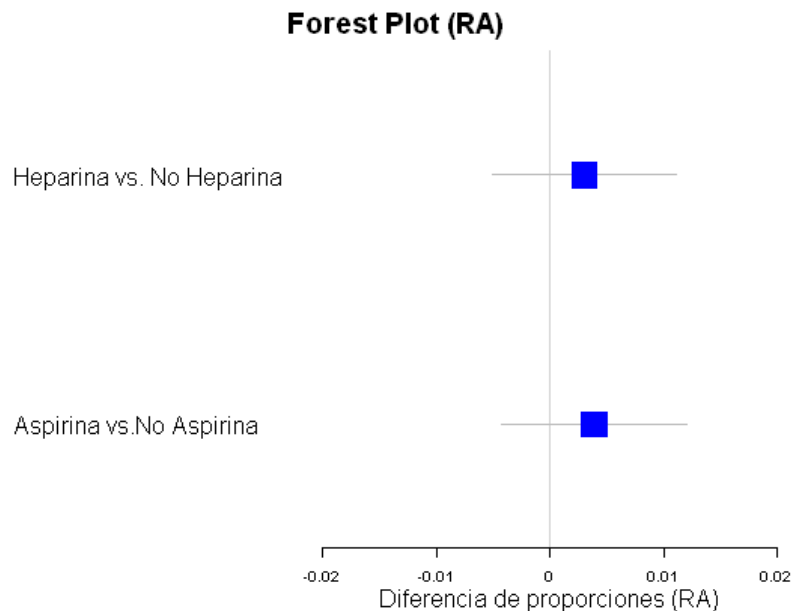
Para estudiar estas diferencias se puede utilizar la Diferencia de proporciones (RA), el Riesgo relativo (RR), o el Odds Ratio (OR). En este informe se realizará el análisis para las tres. Repase su definición y cálculo en el capítulo 8 del módulo 3.

- Diferencia de proporciones (RA)

A continuación se muestra una tabla con el cálculo de la estimación del RA para los dos análisis que se quieren realizar y del intervalo de confianza del RA(95%).

Events	Heparin vs. No Heparin			Aspirin vs. No aspirin		
	Heparin	No Heparin		Aspirin	No Aspirin	
	876 (0.0901)	905 (0.0931)	0.00298 (-0.00514, 0.01109)	872 (0.0897)	909 (0.0935)	0.00386 (-0.00426, 0.01197)

Y gráficamente, se puede representar de la siguiente forma:



A la vista de los resultados, como el intervalo de confianza del 95% de a diferencia de riesgos incluye el 0 en los dos análisis, no hay diferencias significativas en la proporción

de pacientes con eventos durante los primeros 14 días ni para el tratamiento de Heparina, ni para el de Aspirina.

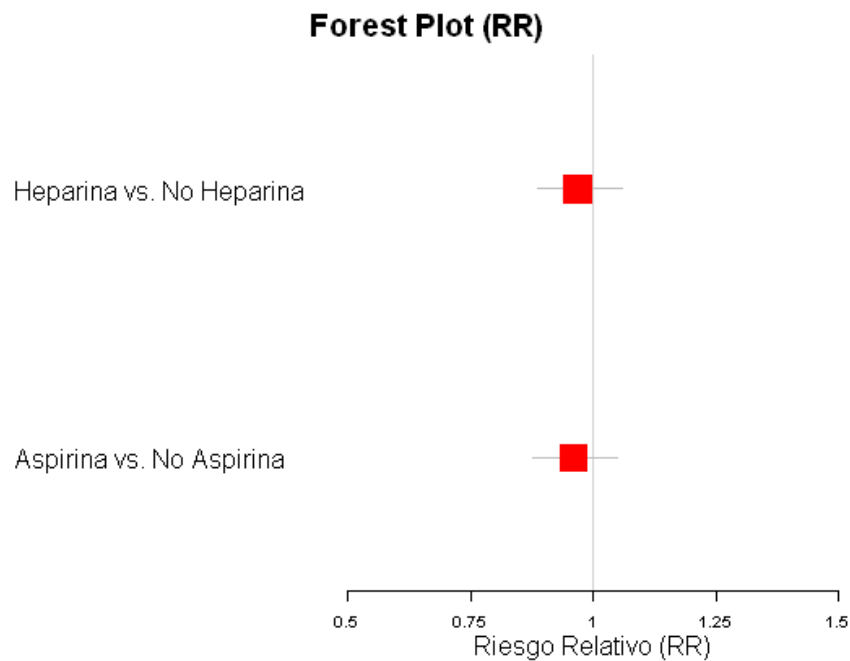
- Riesgo relativo (RR)

En este caso, al ser una medida basada en cocientes, se obtiene la asimetría en la escala logarítmica, por lo que el intervalo de confianza se calcula sobre el logaritmo del RR, con el objetivo de proporcionar un intervalo simétrico alrededor de la estimación puntual. Si no hay diferencias entre los dos grupos, el RR valdrá 1 y su logaritmo, 0 ( $\log 1 = 0$ ).

En la siguiente tabla se muestran los resultados obtenidos para el RR:

	Heparin vs. No Heparin				Aspirin vs. No aspirin			
	Heparin	No Heparin			Aspirin	No Aspirin		
Events	876 (0.0901)	905 (0.0931)	-0.0325 (-0.121, 0.056)	0.968 (0.886, 1.057)	872 (0.0897)	909 (0.0935)	-0.0421 (-0.131, 0.046)	0.9588 (0.877, 1.047)

Gráficamente:



Como los intervalos de confianza del logaritmo del RR incluyen el valor 0 (y el intervalo de confianza del RR incluye el valor 1), tampoco con el RR hay diferencias significativas ni para Heparina, ni para Aspirina.

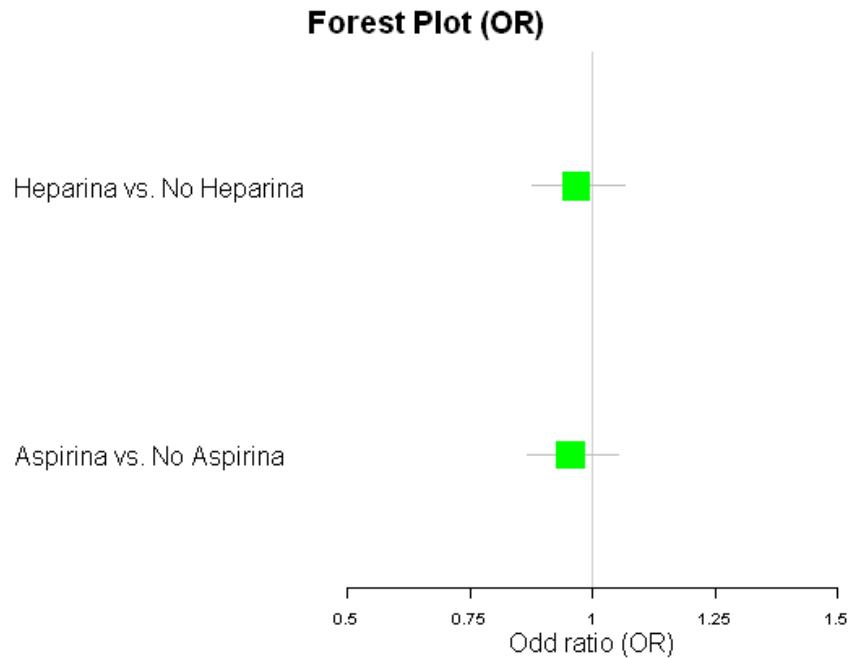
- Odds ratio (OR)

También ahora recurrimos a los logaritmos.

Para los datos que se están tratando se tienen los siguientes resultados:

	Heparin vs. No Heparin				Aspirin vs. No aspirin			
	Heparin	No Heparin			Aspirin	No Aspirin		
Events	876 (0.0901)	905 (0.0931)	-0.0357 (-0.133, 0.062)	0.9649 (0.875, 1.064)	872 (0.0897)	909 (0.0935)	-0.0463 (-0.144, 0.051)	0.9548 (0.866, 1.052)

Y, gráficamente:



Y se llega a la misma conclusión que con las medidas de riesgo anteriores ya que ambos IC<sub>95%</sub> cubren los valores de no efecto.

## 2.2. Respuesta numérica

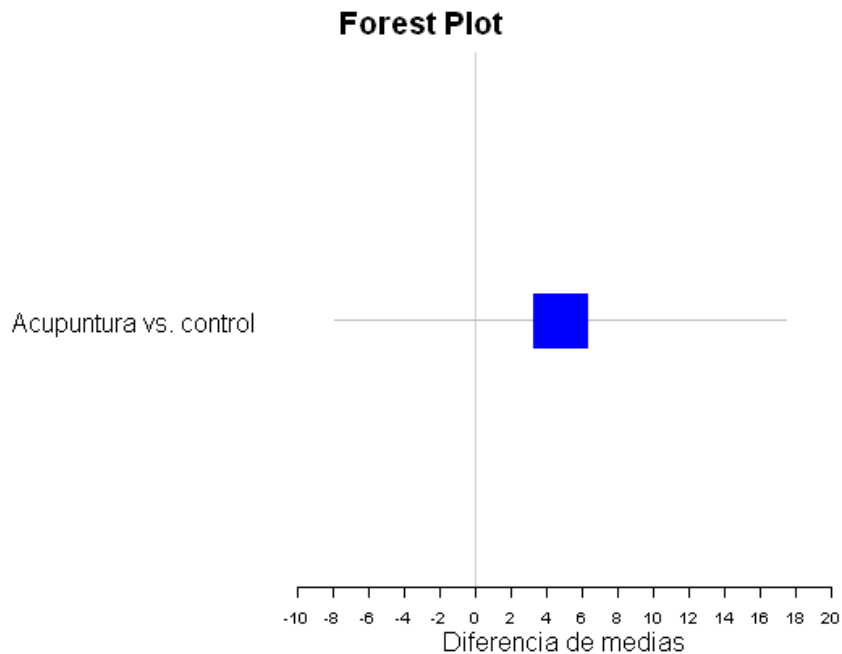
- **Datos:** Study of acupuncture for chronic headache.
- **Respuesta:** diferencia en el grado de dolor de cabeza, según un indicador que toma valores del 0 (ausencia) al 100 (máximo dolor), entre el valor inicial y la respuesta después del tratamiento.
- **Intervenciones estudiadas:** acupuntura frente a control.

Para estudiar si existen o no diferencias entre los pacientes asignados a la acupuntura y los pacientes asignados al grupo control, utiliza la diferencia entre medias para la variable cambio o diferencia entre la respuesta final y el valor inicial (valores negativos indican menor dolor final). Valore repasar los apartados 2 del capítulo 13 y 4 del 8, para el cálculo de intervalos de confianza para la diferencia de medias.

En la siguiente tabla se muestra, para ambos grupos, la media y la desviación estándar entre paréntesis. Además, se incluye la estimación puntual y por intervalo de la diferencia entre medias que estima la magnitud del efecto:

Acupuntura	Control	Diferencia medias (desviación típica)	
3.92 (63.08)	8.68 (53.13)	4.76 (58.58)	(-7.91, 17.42)

Los resultados anteriores se pueden representar con un Forest Plot:



A partir de las estimaciones puntuales de la media de la diferencia en el dolor de cabeza antes y después del tratamiento, se observa que, en media, el dolor de cabeza aumenta en ambos grupos.

También se observa que el grupo que ha recibido acupuntura aumenta menos el dolor, pero esta diferencia no es significativa, ya que su IC<sub>95%</sub> incluye el valor 0. En resumen, no se ha logrado establecer un efecto significativo de la acupuntura.

### 2.3. Respuesta tiempo de supervivencia

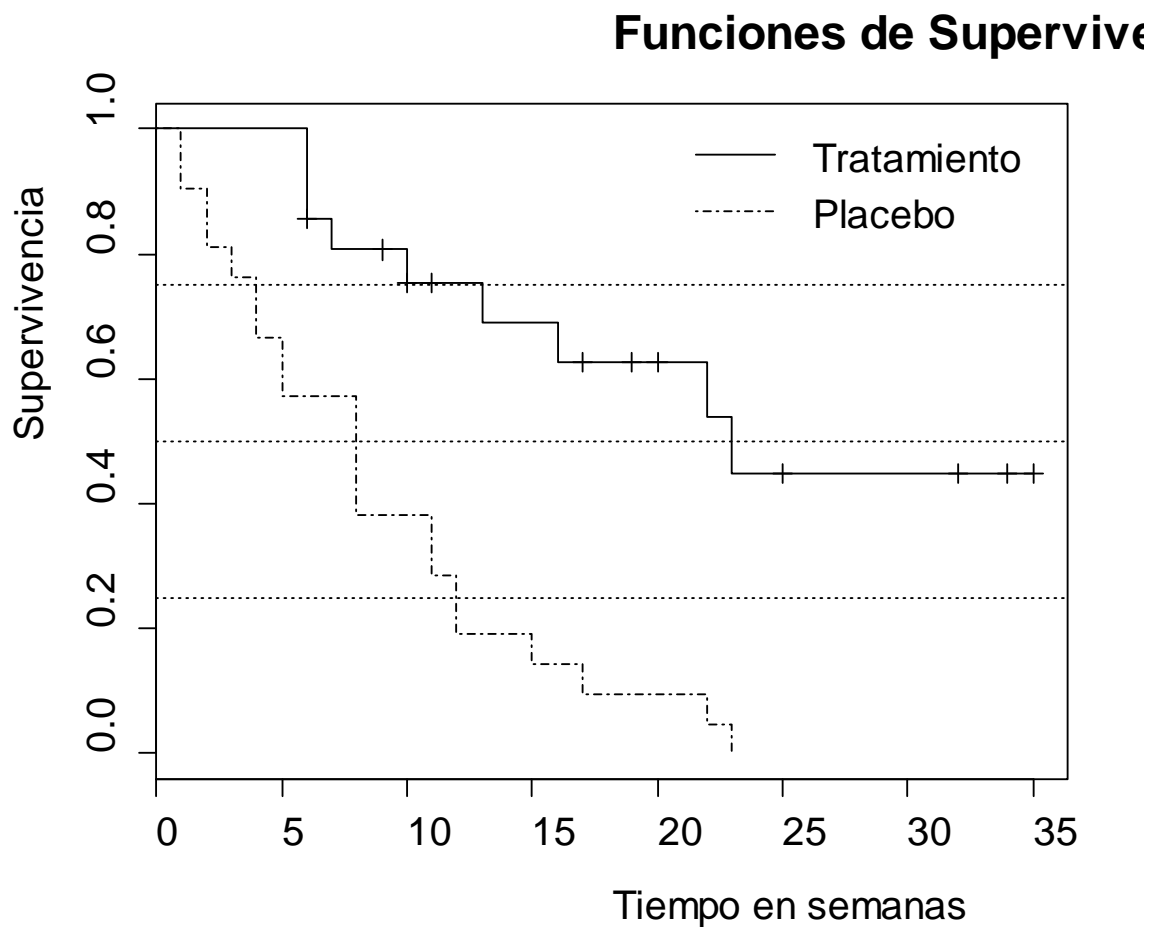
- **Datos:** Anderson
- **Respuesta:** tiempo (en semanas) hasta la recaída en pacientes con Leucemia.
- **Intervenciones estudiadas:** Tratamiento vs. placebo.

Se quiere analizar si el tratamiento (respecto a su ausencia, placebo) cambia el tiempo hasta la recaída.

Una de las vías para ello es comparar la función de supervivencia de ambos grupos. El método de Kaplan-Meier, para comparar la frecuencia de eventos a lo largo del tiempo, estima una única razón de tasas común para todos los tiempos de seguimiento.

Repase el apartado 3 del capítulo 11 sobre Análisis de Supervivencia.

El primer paso consiste en representar las funciones del tiempo hasta la recaída en ambos grupos en un mismo gráfico.



El gráfico muestra un mayor tiempo hasta la recaída. Un paciente prototípico, representado por la mediana, alargaría su tiempo hasta la recaída de unos 8 a 23 semanas; mientras que uno



grave, representado por el percentil 75% de supervivencia (0,25 del tiempo hasta la recaída), lo alargaría de unos 4 a 11 semanas. Esta recaída alcanza al 75% de los pacientes con placebo a las 12 semanas, pero la longitud del seguimiento no permite conocer este tiempo en los tratados, que sería, por tanto, superior a 36 semanas.

Para analizar si existen diferencias estadísticamente significativas entre ambos tratamientos, se puede utilizar el modelo de Cox, que supone riesgos proporcionales.

La salida que proporciona el R de la regresión de Cox es la siguiente:

```
Call:
coxph(formula = leuk ~ g)

n= 42, number of events= 30

              coef exp(coef) se(coef)      z Pr(> |z|)
gtest      -1.5721   0.2076   0.4124  -3.812 0.000138 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
gtest      0.2076      4.817   0.09251   0.4659
```

De la salida de R se observa que el HRR del tratamiento respecto al placebo vale 0.2076, con un Intervalo de confianza del 95%= (0.09251, 0.4659), de donde se desprende que el tratamiento reduce la frecuencia de eventos en un 79,2% (53.4 a 90.7%).

El intervalo de confianza del HRR no incluye el valor 1, por lo que el tratamiento reduce el riesgo de recaída de forma significativa.

Parece razonable representar la reducción del riesgo con un HRR constante ya que la inspección visual apoya la premisa de proporcionalidad de los riesgos entre estas 2 intervenciones: a lo largo de todo el periodo de seguimiento considerado, administrar el tratamiento reduce a una quinta parte la frecuencia de eventos.

Nota: El curso de observacionales enseña a reproducir estos resultados ajustando por terceras variables.

Nota técnica: la aleatorización hace que este ajuste no sea necesario en los ensayos clínicos bien ejecutados (p.e., sin pérdidas). Pero como puede tener otras ventajas (p.e., en eficiencia), algunos autores especifican como análisis principales o análisis ajustado.

## 3. Scripts R

### 3.1. Generación de los datos basales

```
#Lectura de los datos
datos      <-      read.csv('1745-6215-12-101-s1.csv',dec=".",sep="," ,
                           header=TRUE)

##VARIABLE CONTINUA
###Simétrica: BP
#Histograma de la variable
hist(datos$RSBP,col="steelblue",xlab="BP",ylab="Frecuencia",
      main="Histograma de la variable BP")
#Media
mediaBP<-mean(datos$RSBP)
#Desviación estándar
sdBP<-sd(datos$RSBP)

###Asimétrica: AGE
#Histograma de la variable
hist(datos$AGE,col="steelblue",xlab="AGE",ylab="Frecuencia",
      main="Histograma de la variable AGE")
#Mediana
medianaAGE<-median(datos$AGE)
#Primer y tercer cuartil
percentilesaAGE<-quantile(datos$AGE,probs=c(0.25,0.75))

#VARIABLE CATEGÓRICA
#Conscious level
##Recuentos
tablaCL <- table(datos$RCONS)
tablaCL
##Porcentajes
prop.table(tablaCL)*100

#Sex
##Recuentos
tablaSEX <- table(datos$SEX)
tablaSEX
#Porcentajes
prop.table(tablaSEX)*100

#VARIABLE ORDINAL
#Categorización la variable continua: Delay
datos$Delay_cat      <-      cut(datos$RDELAY,breaks=c(0,4,7,13,24,49),
                                right=FALSE)
#La función cut permite categorizar una variable numérica.
#El parámetro breaks especifica los cortes que quieres
#realizar. El parámetro right=FALSE indica que el intervalo
#lo coge cerrado por la izquierda y abierto por la derecha
#Construimos la tabla con los recuentos de las categorías
tabla <- table(datos$Delay_cat)
tabla
#Construimos la tabla con los porcentajes de las categorías
prop.table(tabla)*100
```

## 3.2. Generación de los resultados principales

### 3.2.1. Respuesta dicotómica

```
####Análisis para la variable que mide el número de eventos en los
primeros 14 días

##HERAPIN vs. NO HERAPIN
###Grupo Herapin
#Recuentos
THEP <- sum(datos$ID[datos$RXHEP!="N"])
#Proporciones
PTHEP <- THEP/length(datos$RXHEP[datos$RXHEP!="N"])

###Grupo No Herapin
#Recuentos
TNHEP <- sum(datos$ID[datos$RXHEP=="N"])
#Proporciones
PTNHEP <- TNHEP/length(datos$RXHEP[datos$RXHEP=="N"])

###Análisis

##A continuación se muestran los cálculos para obtener las
estimaciones puntuales y por intervalo de las medidas de riesgo RA,
RR, OR. Si usted dispone de una base de datos no muy grande, puede
utilizar la función epi2x2 (de la librería 'epibasix') explicada en la
página 35 del capítulo 8.

# Estimación puntual e IC para el RA
nNHEP<-length(datos$RXHEP[datos$RXHEP=="N"])
nHEP<-length(datos$RXHEP[datos$RXHEP!="N"])
RAhep<-PTNHEP-PTHEP
IC.LRAhep<- RAhep-
  (1.96 * sqrt((PTHEP*(1-PTHEP))/nHEP+
               (PTNHEP*(1-PTNHEP))/nNHEP))
IC.URAhep<- RAhep +
  (1.96 * sqrt((PTHEP*(1-PTHEP))/nHEP+
               (PTNHEP*(1-PTNHEP))/nNHEP))

# Estimación puntual e IC para el RR
logRRhep<-log(PTHEP/PTNHEP)
IC.LRRhep<- logRRhep-
  1.96 * sqrt((1-PTHEP)/(nHEP*PTHEP)+
              (1-PTNHEP)/(PTNHEP*nNHEP))
IC.URRhep<-logRRhep +
  1.96 * sqrt((1-PTHEP)/(nHEP*PTHEP)+
              (1-PTNHEP)/(PTNHEP*nNHEP))

# Estimación puntual e IC para el OR
logORhep<-log(((THEP)/(nHEP-THEP))/((TNHEP)/(nNHEP-TNHEP)))
IC.LORhep<- logORhep-
  1.96 * sqrt((1/THEP)+(1/(nHEP-THEP))+
              (1/TNHEP)+(1/(nNHEP-TNHEP)))
IC.UORhep<-logORhep +
  1.96 * sqrt((1/THEP)+(1/(nHEP-THEP))+
              (1/TNHEP)+(1/(nNHEP-TNHEP)))

##ASPIRIN vs. NO ASPIRIN
```

```

###Grupo Aspirin
#Recuentos
TASP<-sum(datos$ID[datos$RXASP!="N"])
#Proporciones
PTASP<-TASP/length(datos$RXASP[datos$RXASP!="N"])

###Grupo No Aspirin
#Recuentos
TNASP<-sum(datos$ID[datos$RXASP=="N"])
#Proporciones
PTNASP<-TNASP/length(datos$RXASP[datos$RXASP=="N"])

###Análisis
# Estimación puntual e IC para el RA
nNASP<-length(datos$RXASP[datos$RXASP=="N"])
nASP<-length(datos$RXASP[datos$RXASP!="N"])
RAasp<-PTNASP-PTASP
IC.Lasp<- RAasp-
  (1.96 * sqrt((PTASP*(1-PTASP))/nASP+
               (PTNASP*(1-PTNASP))/nNASP))
IC.Uasp<- RAasp +
  (1.96 * sqrt((PTASP*(1-PTASP))/nASP+
               (PTNASP*(1-PTNASP))/nNASP))

# Estimación puntual e IC para el RR
logRRasp<-log(PTASP/PTNASP)
IC.LRRasp<- logRRasp-
  1.96 * sqrt((1-PTASP)/(nASP*PTASP)+
              (1-PTNASP)/(PTNASP*nNASP))
IC.URRasp<-logRRasp +
  1.96 * sqrt((1-PTASP)/(nASP*PTASP)+
              (1-PTNASP)/(PTNASP*nNASP))

# Estimación puntual e IC para el OR
logORasp<-log(((TASP)/(nASP-TASP))/((TNASP)/(nNASP-TNASP)))
IC.LORasp<- logORasp-
  1.96 * sqrt((1/TASP)+
              (1/(nASP-TASP))+(1/TNASP)+(1/(nNASP-TNASP)))
IC.UORasp<-logORasp +
  1.96 * sqrt((1/TASP)+(1/(nASP-TASP))+
              (1/TNASP)+(1/(nNASP-TNASP)))

#Obtención de los forest plots
##Paquete necesario
install.packages("Gmisc")
library(Gmisc)

##RA
forestplot2(mean=c(RAhep, RAasp),
            lower=c(IC.LRAhep,IC.Lasp),
            upper=c(IC.URAhep,IC.Uasp),
            labeltext=c("Heparina vs. No Heparina", "Aspirina vs.No
Aspirina"),
            clip=c(-1,1.5),
            # Getting the ticks auto-generate is
            # a nightmare - it is usually better to
            # specify them on your own
            xticks=seq(from=-0.02,to=0.02,by=0.01),
            boxsize=0.1,
            col=fpColors(box=c("blue", "darkred")),
            xlab="Diferencia de proporciones (RA)",

```

```

        main="Forest Plot (RA)",
        new_page=TRUE)

##RR
forestplot2(mean=c(exp(logRRhep), exp(logRRasp)),
            lower=c(exp(IC.LRRhep),exp(IC.LRRasp)),
            upper=c(exp(IC.URRhep),exp(IC.URRasp)),
            labeltext=c("Heparina vs. No Heparina", "Aspirina vs. No
Aspirina"),
            clip=c(-1,1.5),
            # Getting the ticks auto-generate is
            # a nightmare - it is usually better to
            # specify them on your own
            xticks=seq(from=0.5,to=1.5,by=0.25),
            boxsize=0.1,
            col=fpColors(box="red"),
            xlab="Riesgo Relativo (RR)",
            main="Forest Plot (RR)",
            new_page=TRUE,zero=1)

##OR
forestplot2(mean=c(exp(logORhep), exp(logORasp)),
            lower=c(exp(IC.LORhep),exp(IC.LORasp)),
            upper=c(exp(IC.UORhep),exp(IC.UORasp)),
            labeltext=c("Heparina vs. No Heparina", "Aspirina vs. No
Aspirina"),
            #clip=c(-1,1.5),
            # Getting the ticks auto-generate is
            # a nightmare - it is usually better to
            # specify them on your own
            xticks=seq(from=0.5,to=1.5,by=0.25),
            boxsize=0.1,
            col=fpColors(box="green"),
            xlab="Odd ratio (OR)",
            main="Forest Plot (OR)",
            new_page=TRUE,zero = 1)

```

### 3.2.2. Respuesta numérica

```

#Leemos los datos
datos<- read.table("clipboard",header=T,na.strings=999)

#Calculamos la diferencia de la variable respuesta antes y después del
tratamiento
datos$difff<- as.integer(datos$pk2)-as.integer(datos$pk1)

#Group=0 -> grupo control
#Group=1 -> grupo acupuntura

#Media y desviación típica para cada grupo
m0 <- mean(datos$difff[datos$group==0],na.rm=T)
sd0 <- sd(datos$difff[datos$group==0],na.rm=T)
m1 <- mean(datos$difff[datos$group==1],na.rm=T)
sd1 <- sd(datos$difff[datos$group==1],na.rm=T)

#diferencia de medias y desviación típica
dm <- mean(datos$difff[datos$group==1],na.rm=T)-
mean(datos$difff[datos$group==0],na.rm=T)
sd(datos$difff,na.rm=T)

#Calculamos el intervalo de confianza de la diferencia de medias
test <- t.test(datos$difff~datos$group)

```

```

#Obtención de los forest plots
##Paquete necesario
install.packages("Gmisc")

##Diferencia de medias
forestplot2(mean=-dm,
            lower=test$conf.int[1],
            upper=test$conf.int[2],
            labeltext=c("Acupuntura vs. control"),
            clip=c(-10,20),
            # Getting the ticks auto-generate is
            # a nightmare - it is usually better to
            # specify them on your own
            xticks=seq(from=-10,to=20,by=2),
            boxsize=0.1,
            col=fpColors(box="blue"),
            xlab="Diferencia de medias",
            main="Forest Plot",
            new_page=TRUE)

```

### 3.2.3. Respuesta tiempo de supervivencia

```

##Cargamos el paquete necesario
library(survival)

#Leemos la base de datos
anderson<-read.table("anderson.txt",head=TRUE)
attach(anderson)

#Calculamos la variable tiempo con los tiempos censurados indicados
leuk<-Surv(t,d)

#Gráfico de las funciones de supervivencia
leuk.grup<-survfit(leuk~g)
plot(leuk.grup,xlab="Tiempo en
semanas",ylab="Supervivencia",lty=c(1,4))
legend(20,1.05,c("Tratamiento","Placebo"),lty=c(1,4),bty="n")
title("Funciones de Supervivencia por tratamiento")
abline(h=c(0.25,0.5,0.75),lty=3)

#Modelo de Cox
mod.cox <- coxph(leuk ~ g)
summary(mod.cox)

```