

Tercera Semana

JMV

May 11, 2015

ejemplo t interval

usamos los datos de drogas de sueño (los originales de gooset)

```
data (sleep)
head(sleep)
```

```
##   extra group ID
## 1    0.7      1  1
## 2   -1.6      1  2
## 3   -0.2      1  3
## 4   -1.2      1  4
## 5   -0.1      1  5
## 6    3.4      1  6
```

son 10 tipos, y las id del tipo nm es el sujeto m en la medida n, son 2 medidas, una para cada medicacion

```
g1<-sleep$extra[1:10];g2<-sleep$extra[11:20]
difference<-g2-g1
mn <-mean(difference); s<- sd(difference); n<-10
```

El intervalo de confianza quedaria dado por

```
mn+c(-1,1)*qt(.975,n-1)*s/sqrt(n)
```

```
## [1] 0.7001142 2.4598858
```

Y aqui todas estas son equivalentes

```
t.test(difference)
```

```
##
##   One Sample t-test
##
## data:  difference
## t = 4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.7001142 2.4598858
## sample estimates:
## mean of x
##      1.58
```

```
t.test(g2,g1,paired=T)
```

```
##
## Paired t-test
##
## data: g2 and g1
## t = 4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.7001142 2.4598858
## sample estimates:
## mean of the differences
## 1.58
```

```
t.test(extra ~ I(relevel(group,2)),paired=T,data=sleep)
```

```
##
## Paired t-test
##
## data: extra by I(relevel(group, 2))
## t = 4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.7001142 2.4598858
## sample estimates:
## mean of the differences
## 1.58
```

Independent groups, por ejemplo separar un placebo de la medicina de verdad

Se supone 8 oc (anticonceptivos ya que nos ponemos) vs 21 controles, presión sanguínea

$X_{oc}=132.86$ $S_{oc}=15.34$ $X_c=127.44$ $S_c = 18.23$

pooled estimated variance

```
sp <- sqrt((7*15.34^2+20*18.23^2)/(21+8-2))
```

y el intervalo de confianza

```
132.86-127.44+c(-1,1)*qt(.975,27)*sp*sqrt(1/8+1/21)
```

```
## [1] -9.521097 20.361097
```

Vamos a ver que pasa si analizamos los datos de sueño en vez de como paired como agrupados (no se hacen las diferencias par a par sino a través de medias)

```
n1<- length(g1); n2<- length (g2)

sp <- sqrt( ( (n1-1)*sd(g1)^2+(n2-1)*sd(g2)^2)/(n1+n2-2))
md<- mean(g2)-mean(g1)
semd <-sp*sqrt(1/n1+1/n2)
```

El intervalo de confianza es

```
md + c(-1,1)*qt(.975,n1+n2-2)*semd
```

```
## [1] -0.203874  3.363874
```

otra forma de sacarlo

```
t.test(g2,g1,paired=F,var.equal=T)
```

```
##
## Two Sample t-test
##
## data:  g2 and g1
## t = 1.8608, df = 18, p-value = 0.07919
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.203874  3.363874
## sample estimates:
## mean of x mean of y
##      2.33      0.75
```

En vez de lo que teniamos antes

```
t.test(g2,g1,paired=T)
```

```
##
## Paired t-test
##
## data:  g2 and g1
## t = 4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.7001142 2.4598858
## sample estimates:
## mean of the differences
##                  1.58
```

y la putada es que ya no es estadisticamente significativo

Otro ejemplete

que lo hago mas que nada para seguir mamoneando con el r por que me aburro
datos de peso de pollo en diversos momentos de su vida

```
library(datasets)
data(ChickWeight)
library(reshape2)
```

Despues de esto lo casteamos para

```
wideCW<-dcast(ChickWeight,Diet+Chick~Time,value.var = "weight")
```

Así se dejan Diet y Chick como estaban (variables con valores, time se convierte para cada valor en una nueva columna, y se rellena la tabla con los valores de peso, así podremos sacar los valores de peso para cada fecha y sacar por dieta y pollo)

Le ponemos nombres cucos a las variables que antes tenían por nombre solo el número del día

```
names(wideCW)[- (1:2)]<-paste("time", names(wideCW[- (1:2)]),sep="")
```

Ahora para crear una columna con las diferencias entre dos días al nacer y el 21 (que me da que es cuando les dan matarile) usamos dplyr

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
wideCW<-mutate(wideCW,gain=time21-time0)
```

Vamos a comparar la dieta 1 con la 4 así que

```
wideCW14<- subset(wideCW,Diet %in% c(1,4))
```

Comparamos los intervalos suponiendo que las sigmas son iguales o distintas

```
rbind(
  t.test(gain~Diet, paired=F,var.equal = T,data=wideCW14)$conf,
  t.test(gain~Diet, paired=F,var.equal = F,data=wideCW14)$conf
)
```

```
##           [,1]      [,2]
## [1,] -108.1468 -14.81154
## [2,] -104.6590 -18.29932
```

T test

Cargamos los datos y hacemos un t test se hace una resta por que los datos son paired sería lo mismo que pasar los dos vectores y un paired = T

el t value es el punto en el que estaría de la distribución

```
library(UsingR)
```

```
## Loading required package: MASS
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
##
## Loading required package: HistData
## Loading required package: Hmisc
## Loading required package: grid
## Loading required package: lattice
## Loading required package: survival
## Loading required package: splines
## Loading required package: Formula

## Warning: package 'Formula' was built under R version 3.1.3

## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
##
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
##
## Attaching package: 'UsingR'
##
## The following object is masked from 'package:ggplot2':
##
##     movies
##
## The following object is masked from 'package:survival':
##
##     cancer
```

```
data(father.son)
t.test(father.son$sheight - father.son$fheight)
```

```
##
## One Sample t-test
##
## data: father.son$sheight - father.son$fheight
## t = 11.7885, df = 1077, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
```

```
## 95 percent confidence interval:
## 0.8310296 1.1629160
## sample estimates:
## mean of x
## 0.9969728
```

T test de dos grupos

recuperamos los datos de los pollos

```
library(datasets)
data(ChickWeight)
library(reshape2)

wideCW<-dcast(ChickWeight,Diet+Chick~Time,value.var = "weight")

names(wideCW)[-(1:2)]<-paste("time", names(wideCW[-(1:2)]),sep="")

library(dplyr)

wideCW<-mutate(wideCW,gain=time21-time0)
```

Nos quedamos de nuevo solo con las dietas 1 y 4

```
wideCW14<- subset(wideCW,Diet %in% c(1,4))
```

y hacemos el t test

```
t.test(gain~Diet, paired =F, var.equal = T, data=wideCW14)
```

```
##
## Two Sample t-test
##
## data: gain by Diet
## t = -2.7252, df = 23, p-value = 0.01207
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -108.14679 -14.81154
## sample estimates:
## mean in group 1 mean in group 4
## 136.1875 197.6667
```

P Values