

Análisis Estadístico con R

Análisis de comparación

CNE/ISCIH

Estructura del curso

- 1 Inferencia
- 2 Comparacion medias
- 3 Comparación proporciones

Section 1

1. Inferencia

Carga de datos

```
rm(list=ls())  
  
setwd("/Users/pfernandezn/Desktop/datos")  
  
load("datos.curso1.RData")
```

Estimador muestral vs Parámetro poblacional

```
mujeres<-datos[datos$sexo%in%"Mujer",]  
mean(mujeres$peso)
```

```
## [1] 59.90394
```

```
sd(mujeres$peso)
```

```
## [1] 1.069147
```

```
set.seed(20)  
teorico_mujeres<- rnorm(n=100000, mean = mean(mujeres$peso), sd = sd(mujeres$peso))
```

Estimador muestral vs Parámetro poblacional

```
hombres<-datos[datos$sexo%in%"Hombre",]  
mean(hombres$peso)
```

```
## [1] 79.991
```

```
sd(hombres$peso)
```

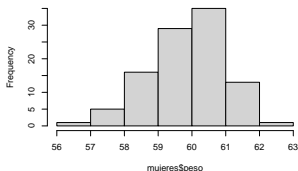
```
## [1] 0.9641111
```

```
set.seed(20)  
teorico_hombres<- rnorm(n=100000, mean = mean(hombres$peso), sd = sd(hombres$peso))
```

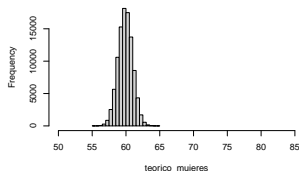
Estimador muestral vs Parámetro poblacional

```
par(mfrow=c(2,2))  
hist(mujeres$peso,main="Muestral mujeres")  
hist(teorico_mujeres,main="Teorico mujeres",xlim=c(50,85))  
hist(hombres$peso,main="Muestral hombres")  
hist(teorico_hombres,main="Teorico hombres",xlim=c(50,85))
```

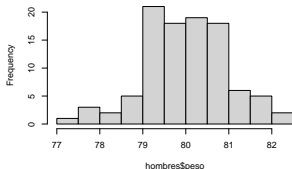
Muestral mujeres



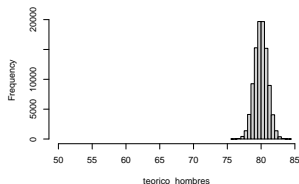
Teorico mujeres



Muestral hombres



Teorico hombres



Intervalo de confianza media

```
res<-t.test(hombres$"peso")
res
```

```
##
##  One Sample t-test
##
## data:  hombres$peso
## t = 829.69, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  79.7997 80.1823
## sample estimates:
## mean of x
##      79.991
```

```
mean_ic95<-paste(round(res$"estimate",2)," (",round(res$"conf.int"[1],2),"-
round(res$"conf.int"[2],2),")",sep="")
mean_ic95
```

```
## [1] "79.99 (79.8-80.18)"
```


Intervalo de confianza media

```
res<-t.test(hombres$"peso", conf.level = 0.99)
res
```

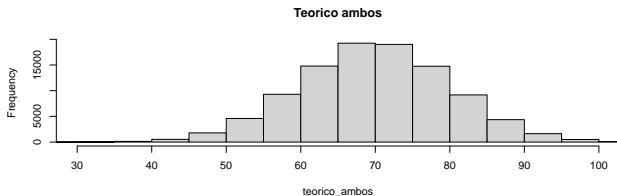
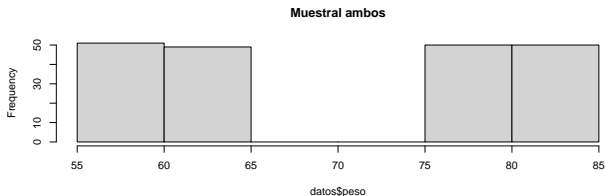
```
##
##  One Sample t-test
##
## data:  hombres$peso
## t = 829.69, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  79.73779 80.24422
## sample estimates:
## mean of x
##      79.991
```

```
mean_ic99<-paste(round(res$"estimate",2)," (",round(res$"conf.int"[1],2),"-
round(res$"conf.int"[2],2),")",sep="")
mean_ic99
```

```
## [1] "79.99 (79.74-80.24)"
```

Intervalo de confianza media

```
teorico_ambos<- rnorm(n=100000, mean = mean(datos$peso), sd = sd(datos$peso))
par(mfrow=c(2,1))
hist(datos$peso,main="Muestral ambos")
hist(teorico_ambos,main="Teorico ambos",xlim=c(30,100))
```



Intervalo de confianza media

```
res<-t.test(datos$"peso")
res
```

```
##
##  One Sample t-test
##
## data:  datos$peso
## t = 97.75, df = 199, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  68.53638 71.35856
## sample estimates:
## mean of x
##  69.94747
```

```
mean_ic95<-paste(round(res$"estimate",2)," (",round(res$"conf.int"[1],2),"-
round(res$"conf.int"[2],2),")",sep="")
mean_ic95
```

```
## [1] "69.95 (68.54-71.36)"
```

Section 2

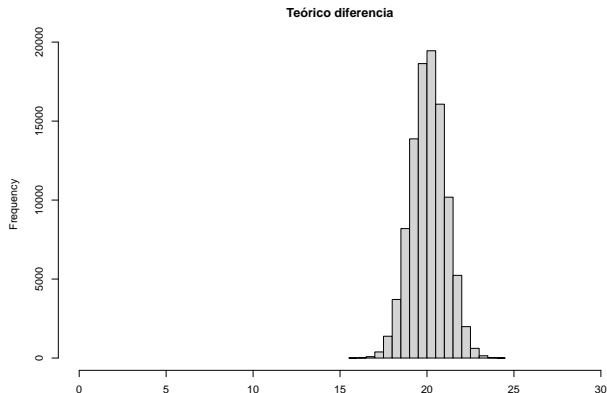
2. Comparación medias

Comparación peso hombres y mujeres

```
diferencia<-mean(hombres$"peso")-mean(mujeres$"peso")  
diferencia
```

```
## [1] 20.08706
```

```
teorico_diferencia<-rnorm(n=100000, mean = diferencia, sd = 1)  
hist(teorico_diferencia,main="Teórico diferencia",xlim=c(0,30))
```



Comparación peso hombres y mujeres

```
res<-t.test(hombres$"peso",mujeres$"peso", conf.level = 0.95)
```

```
res
```

```
##
```

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

```
##
```

```
## data: hombres$peso and mujeres$peso
```

```
## t = 139.53, df = 195.92, p-value < 2.2e-16
```

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

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

```
## 19.80315 20.37098
```

```
## sample estimates:
```

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

```
## 79.99100 59.90394
```

```
res$"p.value"
```

```
## [1] 4.807316e-198
```

Comparación peso hombres y mujeres

```
res<-t.test(hombres$"peso",mujeres$"peso", conf.level = 0.95)
```

```
res
```

```
##
```

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

```
##
```

```
## data: hombres$peso and mujeres$peso
```

```
## t = 139.53, df = 195.92, p-value < 2.2e-16
```

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

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

```
## 19.80315 20.37098
```

```
## sample estimates:
```

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

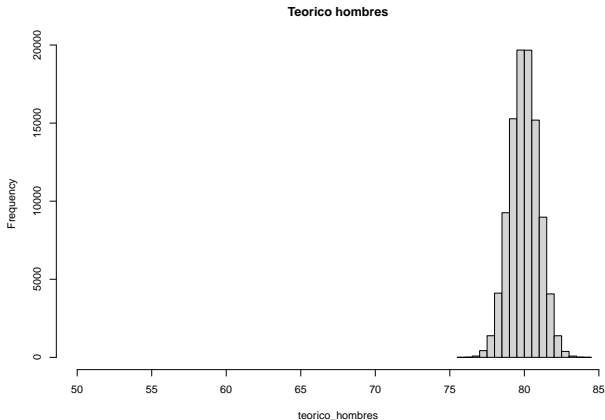
```
## 79.99100 59.90394
```

```
res$"p.value"
```

```
## [1] 4.807316e-198
```

Comparamos con valor de la OMS para peso hombres (90 sd=1)

```
hist(teorico_hombres,main="Teorico hombres",xlim=c(50,85))
```



Comparamos con valor de la OMS para peso hombres (90 sd=1)

```
res<-t.test(hombres$"peso",mu = 90,sd=1, conf.level = 0.95)
```

```
res
```

```
##  
## One Sample t-test  
##  
## data:  hombres$peso  
## t = -103.82, df = 99, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 90  
## 95 percent confidence interval:  
## 79.7997 80.1823  
## sample estimates:  
## mean of x  
## 79.991
```

```
res$"p.value"
```

```
## [1] 7.625474e-103
```

Contraste unilateral o bilateral (una cola o dos colas)

Comparamos con valor de la OMS para peso hombres (79.50 sd=1)

```
res_dos_colas<-t.test(hombres$"peso",mu = 79.80,sd=1,  
conf.level = 0.95,alternative = c("two.sided"))  
res_dos_colas$"p.value"
```

```
## [1] 0.05034895
```

Contraste unilateral o bilateral (una cola o dos colas)

Comparamos con valor de la OMS para peso hombres (79.50 sd=1)

```
res_less<-t.test(hombres$"peso",mu = 79.80 ,sd=1,  
conf.level = 0.95,alternative = c("less"))  
res_less$"p.value"
```

```
## [1] 0.9748255
```

Contraste unilateral o bilateral (una cola o dos colas)

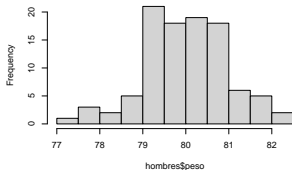
Comparamos con valor de la OMS para peso hombres (79.50 sd=1)

```
res_greater<-t.test(hombres$"peso",mu = 79.80,sd=1,  
conf.level = 0.95,alternative = c("greater"))  
res_greater$"p.value"  
  
## [1] 0.02517448
```

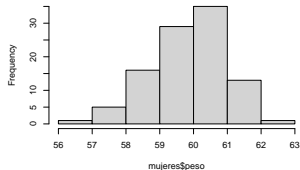
Asunciones (normalidad)

```
par(mfrow=c(2,2))  
hist(hombres$"peso")  
hist(mujeres$"peso")  
qqnorm(hombres$"peso")  
qqnorm(mujeres$"peso")
```

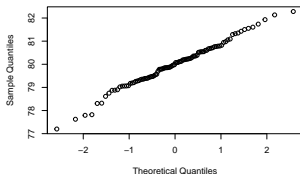
Histogram of hombres\$peso



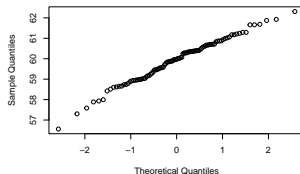
Histogram of mujeres\$peso



Normal Q-Q Plot



Normal Q-Q Plot



Asunciones (normalidad)

```
shapiro.test(hombres$"peso")
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  hombres$peso  
## W = 0.98918, p-value = 0.5985
```

```
shapiro.test(mujeres$"peso")
```

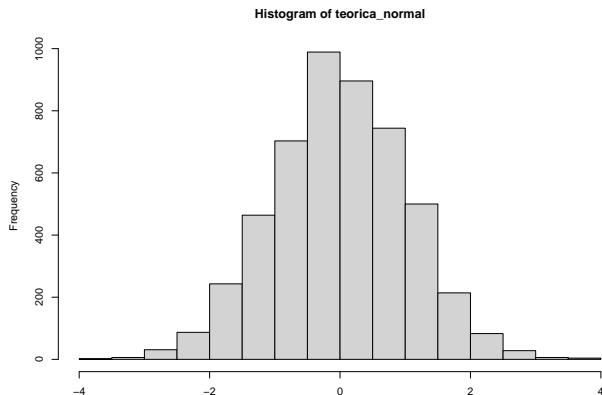
```
##  
##  Shapiro-Wilk normality test  
##  
## data:  mujeres$peso  
## W = 0.98709, p-value = 0.4439
```

Asunciones (normalidad)

```
set.seed(10)
teorica_normal<-rnorm(5000)
shapiro.test(teorica_normal)$p.value
```

```
## [1] 0.6136662
```

```
hist(teorica_normal)
```



Asunciones (homogeneidad de varianzas)

```
var.test(x=hombres$"peso",y=mujeres$"peso")
```

```
##  
## F test to compare two variances  
##  
## data:  hombres$peso and mujeres$peso  
## F = 0.81317, num df = 99, denom df = 99, p-value = 0.3052  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
##  0.5471318 1.2085544  
## sample estimates:  
## ratio of variances  
##           0.8131657
```


Tener en cuenta que las muestras no tienen varianzas homogéneas

```
res<-t.test(hombres$"peso",mujeres$"peso", conf.level = 0.95)
res
```

```
##
## Welch Two Sample t-test
##
## data:  hombres$peso and mujeres$peso
## t = 139.53, df = 195.92, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  19.80315 20.37098
## sample estimates:
## mean of x mean of y
##  79.99100  59.90394
```

Tener en cuenta que las muestras no tienen varianzas homogéneas

```
res<-t.test(hombres$"peso",mujeres$"peso", conf.level = 0.95,var.equal = TR  
res
```

```
##  
## Two Sample t-test  
##  
## data:  hombres$peso and mujeres$peso  
## t = 139.53, df = 198, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 19.80316 20.37097  
## sample estimates:  
## mean of x mean of y  
## 79.99100 59.90394
```

Comparacion de media

Test no parametricos

```
res<-wilcox.test(hombres$"peso",mujeres$"peso")
```

```
res
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data:  hombres$peso and mujeres$peso
```

```
## W = 10000, p-value < 2.2e-16
```

```
## alternative hypothesis: true location shift is not equal to 0
```

```
res<-wilcox.test(hombres$"peso",mu=79.80)
```

```
res
```

```
##
```

```
## Wilcoxon signed rank test with continuity correction
```

```
##
```

```
## data:  hombres$peso
```

```
## V = 3142, p-value = 0.03403
```

```
## alternative hypothesis: true location is not equal to 79.8
```

```
res<-t.test(hombres$"peso",mu=79.80)
```

Comparacion de media

Test no parametricos

```
res<-wilcox.test(hombres$"peso",mujeres$"peso")  
res$p.value
```

```
## [1] 2.562144e-34
```

```
res<-t.test(hombres$"peso",mujeres$"peso")  
res$p.value
```

```
## [1] 4.807316e-198
```

Comparacion de media

Test no parametricos

```
res<-wilcox.test(hombres$"peso",mu=79.80)  
res$p.value
```

```
## [1] 0.03402905
```

```
res<-t.test(hombres$"peso",mu=79.80)  
res$p.value
```

```
## [1] 0.05034895
```

Comparacion de más de dos grupos (en hombres: peso por nivel de estudios)

Paramétrico

```
table(hombres$"nivel.estudios")
```

```
##  
## Alto Bajo Medio  
## 44 33 23
```

```
tapply(hombres$"peso",hombres$"nivel.estudios",mean)
```

```
## Alto Bajo Medio  
## 80.13396 79.92041 79.81881
```

```
anova(lm(hombres$"peso"~hombres$"nivel.estudios"))
```

```
## Analysis of Variance Table  
##  
## Response: hombres$peso  
##  
## Df Sum Sq Mean Sq F value Pr(>F)  
## hombres$nivel.estudios 2 1.746 0.87277 0.9378 0.395  
## Residuals 97 90.276 0.93068
```

Comparacion de más de dos grupos (comparaciones dos a dos)

```
pairwise.t.test(x=hombres$"peso",g=as.factor(hombres$"nivel.estudios"),p.ad
```

```
##  
## Pairwise comparisons using t tests with pooled SD  
##  
## data:  hombres$peso and as.factor(hombres$nivel.estudios)  
##  
##      Alto Bajo  
## Bajo  1.00 -  
## Medio 0.62 1.00  
##  
## P value adjustment method: bonferroni
```

Comparacion de más de dos grupos (comparaciones dos a dos)

```
bartlett.test(hombres$"peso"~hombres$"nivel.estudios")
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data:  hombres$peso by hombres$nivel.estudios  
## Bartlett's K-squared = 4.6781, df = 2, p-value = 0.09642
```

```
oneway.test(hombres$"peso"~hombres$"nivel.estudios")
```

```
##  
## One-way analysis of means (not assuming equal variances)  
##  
## data:  hombres$peso and hombres$nivel.estudios  
## F = 1.1875, num df = 2.000, denom df = 58.221, p-value = 0.3123
```


Correccion de p-valores por comparaciones multiples

```
pvalores<-c(0.001,0.43,0.54300)  
p.adjust(pvalores,method="BH")
```

```
## [1] 0.003 0.543 0.543
```

Comparacion de más de dos grupos (en hombres: peso por nivel de estudios)

No parametrico

```
kruskal.test(hombres$"peso"~hombres$"nivel.estudios")
```

```
##
```

```
##  Kruskal-Wallis rank sum test
```

```
##
```

```
## data:  hombres$peso by hombres$nivel.estudios
```

```
## Kruskal-Wallis chi-squared = 2.2438, df = 2, p-value = 0.3257
```

Section 3

3. Comparacion de proporciones

Proporciones (Intervalo de confianza)

```
fumadores <- c( 83, 70 )
pacientes <- c( 86, 82 )

prop1<-prop.test(fumadores[1],pacientes[1])
paste(round(prop1$"estimate",2)," (",round(prop1$"conf.int"[1],2),"-",
round(prop1$"conf.int"[2],2),")",sep="")
```

```
## [1] "0.97 (0.89-0.99)"
```

```
prop2<-prop.test(fumadores[2],pacientes[2])
paste(round(prop2$"estimate",2)," (",round(prop2$"conf.int"[1],2),"-",
round(prop2$"conf.int"[2],2),")",sep="")
```

```
## [1] "0.85 (0.75-0.92)"
```

Comparacion proporciones

```
prop.test(fumadores, pacientes)
```

```
##  
## 2-sample test for equality of proportions with continuity correction  
##  
## data: fumadores out of pacientes  
## X-squared = 5.1155, df = 1, p-value = 0.02371  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## 0.01377774 0.20913775  
## sample estimates:  
##      prop 1      prop 2  
## 0.9651163 0.8536585
```

Comparacion proporciones

```
chisq.test(matrix(c(83,70,3,8),2))
```

```
##  
##  Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  matrix(c(83, 70, 3, 8), 2)  
## X-squared = 2.0103, df = 1, p-value = 0.1562
```

Comparacion proporciones

```
fisher.test(matrix(c(83,70,3,8),2))
```

```
##  
## Fisher's Exact Test for Count Data  
##  
## data:  matrix(c(83, 70, 3, 8), 2)  
## p-value = 0.1185  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
##  0.7190756 19.0735432  
## sample estimates:  
## odds ratio  
##  3.140762
```

Comparacion proporciones (más de dos grupos)

```
fumadores <- c( 83, 90, 129, 70 )  
pacientes <- c( 86, 93, 136, 82 )  
prop.test(fumadores,pacientes)
```

```
##  
## 4-sample test for equality of proportions without continuity  
## correction  
##  
## data: fumadores out of pacientes  
## X-squared = 12.6, df = 3, p-value = 0.005585  
## alternative hypothesis: two.sided  
## sample estimates:  
##      prop 1      prop 2      prop 3      prop 4  
## 0.9651163 0.9677419 0.9485294 0.8536585
```


Comparacion proporciones (más de dos grupos)

```
fumadores <- c( 83, 90, 129, 70 )  
pacientes <- c( 86, 93, 136, 82 )  
prop.trend.test(fumadores,pacientes)
```

```
##  
##  Chi-squared Test for Trend in Proportions  
##  
## data:  fumadores out of pacientes ,  
##  using scores: 1 2 3 4  
## X-squared = 8.2249, df = 1, p-value = 0.004132
```