

Quantitative Hypothesis Test

Alex Sanchez, Miriam Mota, Ricardo Gonzalo and
Santiago Perez-Hoyos

Statistics and Bioinformatics Unit. Vall d'Hebron Institut de
Recerca 14/10/2020

```

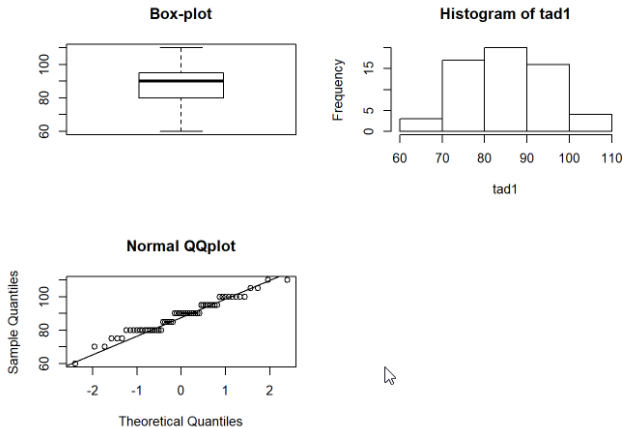
oldpar<-par(mfrow=c(1,1)) # Guarda los parámetros para el dibujo
par(mfrow=c(2,2)) # Dibuja cuatro gráficos por grafico
with(hta, boxplot(tad1, main="Box-plot" )

with(hta, hist(tad1) )

with(hta, qqnorm(tad1, main="Normal QQplot" )
;with(hta, qqline(tad1) )

par(oldpar) # Vuelve a los parámetros de dibujo originales

```



Normality Test

```
with(hta,shapiro.test(tas1) ) # Shapiro Wilk test
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  tas1
```

```
## W = 0.91869, p-value = 0.0006796
```

One sample Test

```
with(hta,t.test(tad1,mu=90) ) # One sample T.test
```

```
##  
##  One Sample t-test  
##  
## data:  tad1  
## t = -1.2137, df = 59, p-value = 0.2297  
## alternative hypothesis: true mean is not equal to 90  
## 95 percent confidence interval:  
##  85.80626 91.02707  
## sample estimates:  
## mean of x  
##  88.41667
```

Homogeneity variance Test

```
library(car)
hta%>%
  group_by(sexo) %>%
  summarise(var = sd(tad1))
```

```
## # A tibble: 2 x 2
##   sexo      var
##   <chr> <dbl>
## 1 MUJER  9.08
## 2 VARON 11.8
```

```
with(hta, leveneTest(tad1~factor(sexo), center="median"))
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##           Df F value Pr(>F)
## group    1  1.3506 0.2499
##           58
```

- p value is over 0.05
- We can assume homogeneity of variances

T test when variances are equal

```
with(hta,t.test(tad1~factor(sexo),var.equal=TRUE ))
```

```
##
```

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

```
##
```

```
## data: tad1 by factor(sexo)
```

```
## t = 0.35427, df = 58, p-value = 0.7244
```

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

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

```
## -4.453505 6.368899
```

```
## sample estimates:
```

```
## mean in group MUJER mean in group VARON
```

```
## 88.78378 87.82609
```

- Type I Error is over than 0.05
- We cannot reject mean equality

T test when variances are unequal

```
with(hta,t.test(tad1~factor(sexo),var.equal=FALSE ))
```

```
##
```

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

```
##
```

```
## data: tad1 by factor(sexo)
```

```
## t = 0.33362, df = 38.144, p-value = 0.7405
```

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

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

```
## -4.852834 6.768228
```

```
## sample estimates:
```

```
## mean in group MUJER mean in group VARON
```

```
## 88.78378 87.82609
```

- Same conclusions as before
- Test is also known as Welch test

U Mann-Whitney or Sum Rank non parametric test

```
with(hta,wilcox.test(tad1~factor(sexo)  
  ,alternative='two.sided',exact=TRUE, correct=FALSE))
```

```
##
```

```
## Wilcoxon rank sum test
```

```
##
```

```
## data: tad1 by factor(sexo)
```

```
## W = 434, p-value = 0.8955
```

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

```
hta%>%
```

```
  group_by(sexo) %>%
```

```
  summarise(median = median(tad1))
```

```
## # A tibble: 2 x 2
```

```
##   sexo median
```

```
##   <chr>   <dbl>
```

```
## 1 MUJER      90
```

```
## 2 VARON      90
```

- Null Hypothesis cannot be rejected

Paired T-test

```
with(hta,t.test(tad1,tad12,paired=TRUE))
```

```
##
```

```
## Paired t-test
```

```
##
```

```
## data: tad1 and tad12
```

```
## t = 1.8507, df = 51, p-value = 0.07001
```

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

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

```
## -0.2364274 5.8133505
```

```
## sample estimates:
```

```
## mean of the differences
```

```
## 2.788462
```

- P value is over 0.05

Paired Sign-Rank Wilcoxon Test

```
with(hta,wilcox.test(tad1,tad12,  
    exact=TRUE, paired=TRUE))
```

```
##
```

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

```
##
```

```
## data: tad1 and tad12
```

```
## V = 478.5, p-value = 0.05333
```

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

Read diabetes data

```
library(readxl)
library(dplyr)
library(magrittr)
diabetes <- read_excel("datasets/diabetes.xls")
sapply(diabetes, class)
```

```
##      numpacie      mort      tempsviu      edat      bmi      edatdiag
## "numeric" "character" "numeric" "numeric" "numeric" "numeric"
##      tabac      sbp      dbp      ecg      chd
## "character" "numeric" "numeric" "character" "character"
```

```
diabetes_factor <- diabetes %>%
  mutate_if(sapply(diabetes, is.character), as.factor) %>%
  select (-numpacie)
```

```
diabetes%>%
  group_by(ecg) %>%
  summarise( n=n(),
            mean = mean(edat),
            sd=sd(edat))
```

```
## # A tibble: 3 x 4
##   ecg      n mean   sd
##   <chr> <int> <dbl> <dbl>
## 1 Anormal    11  64.9  6.76
## 2 Frontera   27  53.8 11.4
## 3 Normal   111  50.5 11.5
```

ANOVA

```
anova<-aov(edat~ecg,data=diabetes_factor)
summary(anova)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## ecg           2    2166   1083.0     8.619 0.00029 ***
## Residuals    146   18347    125.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Multicomparison

```
library(multcomp)
tuk <- glht(anova, linfct = mcp(ecg = "Tukey"))

print(summary(tuk)) # pairwise tests

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = edat ~ ecg, data = diabetes_factor)
##
## Linear Hypotheses:
##
##           Estimate Std. Error t value Pr(>|t|)
## Frontera - Anormal == 0 -11.094      4.010  -2.767 0.016516 *
## Normal - Anormal == 0  -14.405      3.543  -4.065 0.000224 ***
## Normal - Frontera == 0   -3.310      2.405  -1.376 0.345714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
print(confint(tuk, level=0.95)) # confidence intervals
```

```
##  
## Simultaneous Confidence Intervals  
##  
## Multiple Comparisons of Means: Tukey Contrasts  
##  
##  
## Fit: aov(formula = edat ~ ecg, data = diabetes_factor)  
##  
## Quantile = 2.3458  
## 95% family-wise confidence level  
##  
##  
## Linear Hypotheses:  
##  
##           Estimate lwr      upr  
## Frontera - Anormal == 0 -11.0943 -20.5005 -1.6881  
## Normal - Anormal == 0 -14.4046 -22.7169 -6.0922  
## Normal - Frontera == 0 -3.3103 -8.9531 2.3325
```

Multicomparison plot

```
plot(confint(tuk))
```

Kruskal-Wallis Test

```
diabetes_factor%>%  
  group_by(ecg) %>%  
  summarise(median = median(edat))
```

```
## # A tibble: 3 x 2  
##   ecg      median  
##   <fct>    <dbl>  
## 1 Anormal      64  
## 2 Frontera     53  
## 3 Normal       49
```

```
kruskal.test(edat~ecg,data=diabetes_factor)
```

```
##
```

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

```
##
```

```
## data: edat by ecg
```

```
## Kruskal-Wallis chi-squared = 17.483, df = 2, p-value = 0.0001
```


Dunn Test for multiple comparison

```
library(dunn.test)
with(diabetes_factor,dunn.test(edat,ecg,method="bonferroni"))
```

```
##    Kruskal-Wallis rank sum test
##
## data: edat and ecg
## Kruskal-Wallis chi-squared = 17.4826, df = 2, p-value = 0
##
##
##                                     Comparison of edat by ecg
##                                     (Bonferroni)
## Col Mean-|
## Row Mean |      Anormal    Frontera
## -----+-----
## Frontera |      2.721182
##           |      0.0098*
##           |
## Normal   |      4.075469    1.467464
##           |      0.0001*      0.2134
##
```