Hypotheses tests with continuous variables

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

Statistics and Bioinformatics Unit. Vall d'Hebron Institut de Recerca

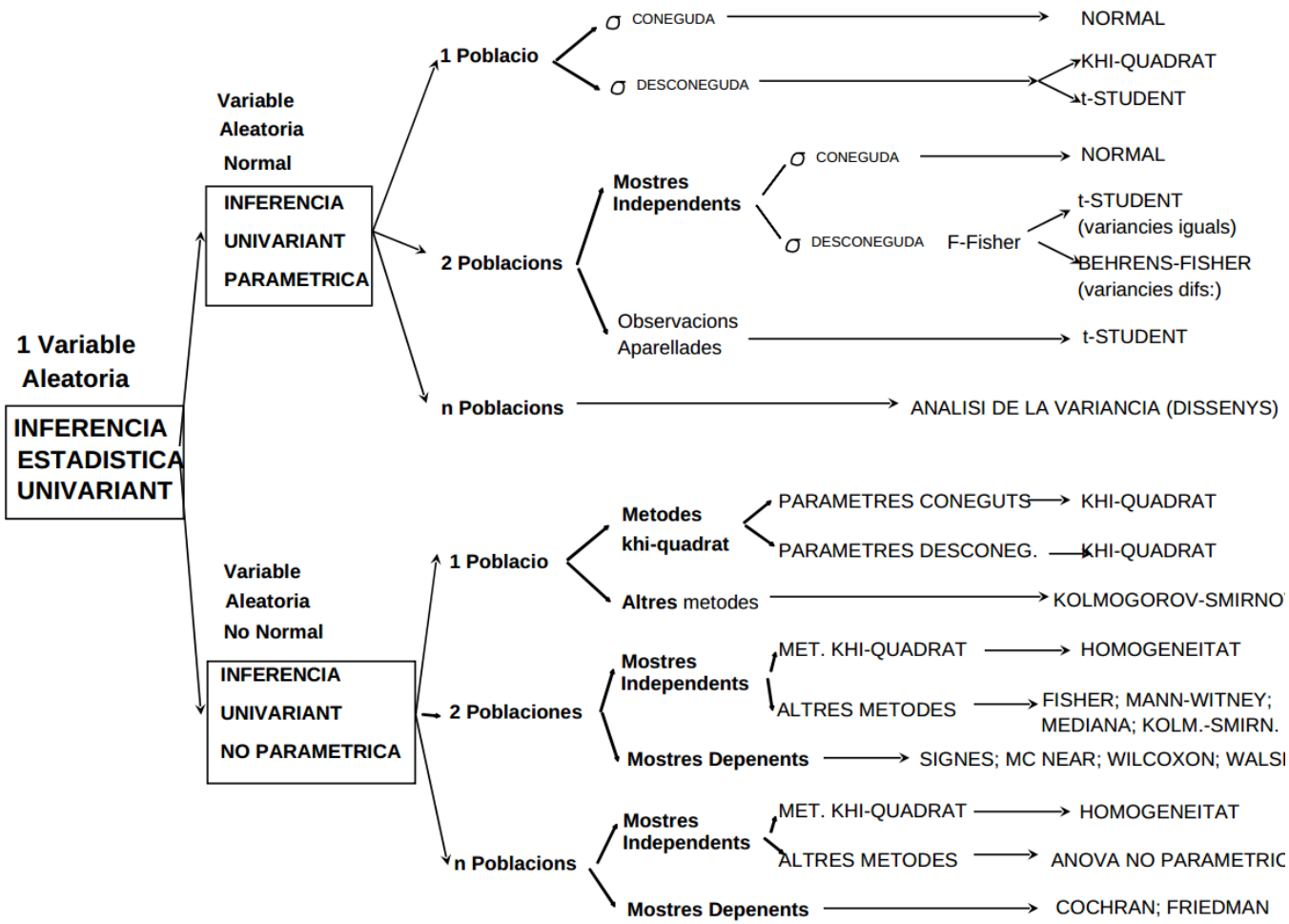
Versión 2023-05-30

Outline

Introduction

- Once the concept of hypothesis testing is established,
- Researchers face the problem of *which test should be applied at every possible situation*.
- For this, ideally, they should...
  - understand the problem and the questions addressed,
  - know available tests for each problem,
  - know (how to check) applicability assumptions of each test,
  - know how robust each test is to assumptions violation.
- Easier to say than to do.
  - Sometimes cheatsheets may be helpful, but be warned against a blind use, that is understand and be critic with the steps.

Which test is appropriate for which problem



Example situation (1): Introduction

- Many experimental questions may be answered through hypothesis testing.
- Imagine, for example, a study designed to compare two distinct hypertension control programs.
- 60 individuals with HTA were randomly assigned to either one or the other group (30 per group)
- Blood pressure was measured each month during a year
  - For simplicity we may keep only data at months 1 and 12

Example situation (2): Data collected

```
hta <- read_excel("datasets/hta.xls")
htaSimple <- hta %>% select(grupo, sexo, tas1, tad1, tas12,
head(as.data.frame(htaSimple))
```

```
##   grupo  sexo tas1 tad1 tas12 tad12
## 1     B VARON  150  100   150    80
## 2     B MUJER  160   90   150    90
## 3     B MUJER  150   90   110    85
## 4     A VARON  120   80   130    90
## 5     A MUJER  150   85   120    80
## 6     B MUJER  140   75   140    90
```

Example situation (3): Reasonable questions

- The goal of the study is to compare the treatment effect so a reasonable question is:
  - *Is the average decrease in "tad" the same in both groups A and B?*
- Or, if we are testing a new treatment "B", hat is intended to be batter than "A"
  - *Is the average decrease in "tad" greater in group B?*
- *Although they are not planned in this study* other relevant questions may lead to questions that need a test to be answered, such as:
  - *Is the average `tad` above 150?*
  - *Has the average `tad` (in group A) decreased in 12 months?*
  - *Is the average `tad` different in men and women at basal time?*

Types of tests (1): Confirmatory vs Independence

Distinct classifications can be found in textbooks

- Confirmatory
  - Is average HTA above 150?
  - Is the the tas1 variable normally distributed
- Independence
  - Is sex related to HTA (or is mean(HTA) the same in men or women)
  - Is average HTA decrease the same for both groups?
- *This classification is useful but artificial, not to say that the term "independence" is slightly abused*

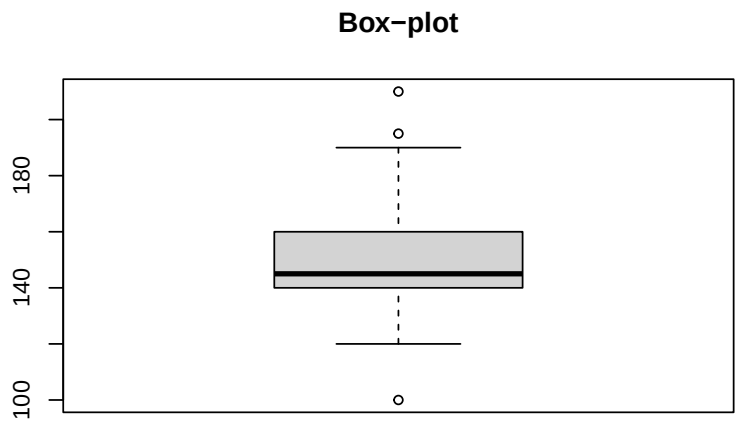
Types of tests (2): Parametric vs Non-parametric

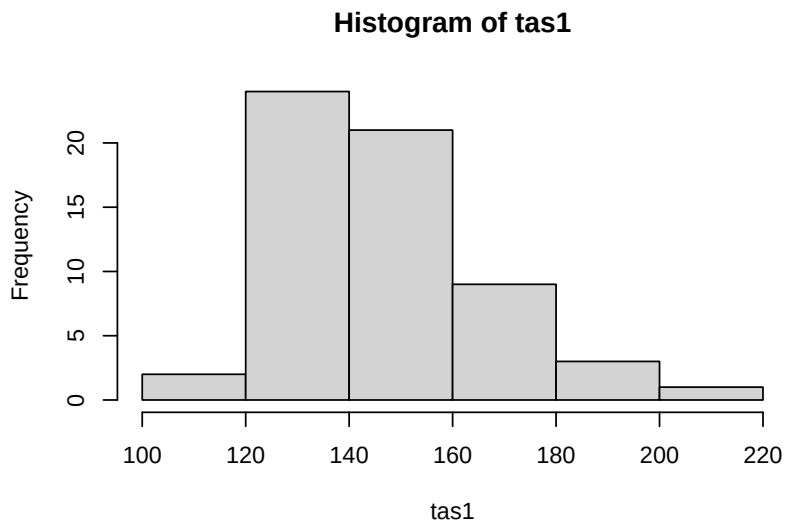
- Parametric tests
  - assume some underlying distribution for the data
  - pose the test in terms of the distribution's parameters
    - E.g. the t-test assumes normality and relies on the normal and t-distribution's parameters
- Non-parametric tests
  - Do not assume an underlying distribution, but they are not assumption-free!
  - Check: Distribution free is not assumption free
- Permutation tests
  - If sample size is not tiny *permutation tests* are a good alternative.

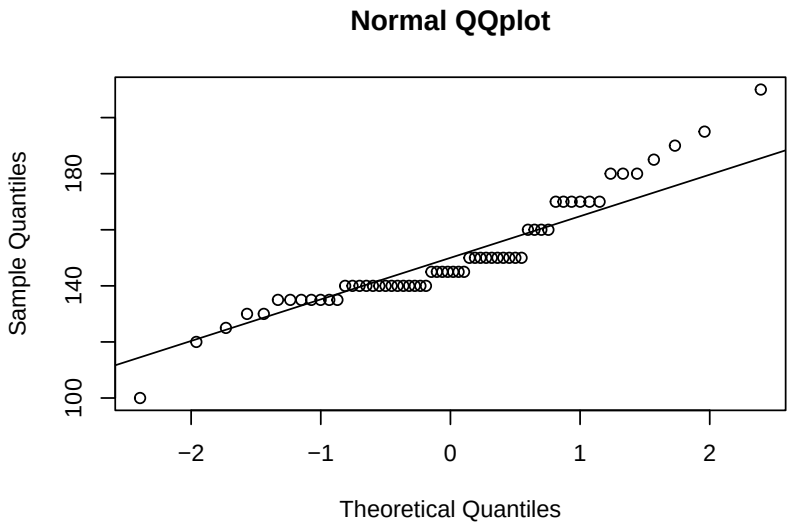
Hands on: Always start looking at the data

```
oldpar<-par(mfrow=c(1,1)) # Guarda los parámetros para el
par(mfrow=c(2,2)) # Dibuja cuatro gráficos por grafico
with(hta, boxplot(tas1, main="Box-plot") )
with(hta, hist(tas1) )
with(hta, qqnorm(tas1, main="Normal QQplot") )
with(hta, qqline(tas1) )
par(oldpar) # Vuelve a los parámetros de dibujo originales
```

Data visualization







Normality and tests

- The choice of test seems to pivot around the question of ¿is my data normally distributed?
- Leaving apart a tence to repeat what we have been taught, why is this so?
  - If the data is normally distributed there exist some *optimal tests* for some one or two sample problems.
  - Data show often a bell-shaped form that can be assimilated to have a gaussian distribution (it is *normal* to observe this).
  - Iven if data is not bell-shaped, if sample size is big enough th sample mean tends to be bell -shaped as sample size increase.
- In summary, normality is not only "practical" but common.
- As a consequence checking normality has become one of the first steps of any data analysis.
  - It doesn't hurt, but it is clearly over-rated.

Normality Test

- Normality tests can be used to decide if the data can be considered to follow a normal distribution.
- This is more a theoretical than practical issue because ...
  - If the sample size is too small, the test is not powerful enough.
  - If the sample size is too big, the test will almost always reject the normality htypothesis
- Normality can be tested
  - Graphically:
    - Check if it is a symmetric distribution
    - Probability graphs (QQ-plots)
  - Using Hypothesis test (Normality)
    - Kolmogorov tests
    - Shapiro-Wilks test

Normality test: Shapiro-Wilks

- In normality test, the null hypothesis is normality
- That is, rejecting the null suggests departure of normality.
  - H_0: The data follow a normal distribution
  - H_1: The data do not follow a normal distribution

```
with(hta,shapiro.test(tad1) ) # Shapiro Wilk test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  tad1
## W = 0.96622, p-value = 0.09512
```

One sample tests

- One sample tests refer to a single characteristic of the population such as:
  - *Is it true that the average tad in HTA population is above 150*?
- It is often said that they are less interesting because they are rarely used in most practical situations where the main goal is comparison
- *However*, noticing that a paired tests is equivalent to a one sample tests for the difference makes us realize their relevance.

One sample parametric vs non-parametric tests

- If we assume normality an appropriate test for a hypothesis about the mean is the *t-test*
  - $H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0, (\mu > \mu_0, \mu < \mu_0).$
- If we don't assume normality we can rely on
  - Wilcoxon rank test, if data re symmetrical
  - Sign test in other cases

Hands on one sample tests (1): t-test

- Is `tad` in HTA patients above 90?
- Notice that even if we are interested in "above 90" the null is "equality"
  - $H_0 : \mu = 90, \qquad H_1 : \mu > 90.$

```
with(hta,t.test(tad1,mu=90, alternative="greater") ) # One sampl
```

```
##
##  One Sample t-test
##
## data:  tad1
## t = -1.2137, df = 59, p-value = 0.8852
## alternative hypothesis: true mean is greater than 90
## 95 percent confidence interval:
##  86.23664      Inf
## sample estimates:
## mean of x
##  88.41667
```

Hands on one sample tests (2): Wilcoxon test

- In wilcoxon test the null hypothesis is about the median.

```
with(hta,wilcox.test(tad1,mu=90, alternative="greater") ) # One
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  tad1
## V = 429, p-value = 0.8934
## alternative hypothesis: true location is greater than 90
```

Homogeneity variance Test

```
library(car)
hta%>%
  group_by(sexo) %>%
  summarise(var = sd(tas1))
```

```
## # A tibble: 2 x 2
##   sexo    var
##   <chr> <dbl>
## 1 MUJER  17.6
## 2 VARON  22.1
```

```
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(tas1~factor(sexo),var.equal=TRUE ))
```

```
##
##  Two Sample t-test
##
## data:  tas1 by factor(sexo)
## t = -0.2471, df = 58, p-value = 0.8057
## alternative hypothesis: true difference in means between grou
## 95 percent confidence interval:
##  -11.603461   9.053519
## sample estimates:
## mean in group MUJER mean in group VARON
##            149.5946             150.8696
```

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

T test when variances are unequal

```
with(hta,t.test(tas1~factor(sexo),var.equal=FALSE ))
```

```
##
##  Welch Two Sample t-test
##
## data:  tas1 by factor(sexo)
## t = -0.23436, df = 39.098, p-value = 0.8159
## alternative hypothesis: true difference in means between grou
## 95 percent confidence interval:
##  -12.277927   9.727986
## sample estimates:
## mean in group MUJER mean in group VARON
##            149.5946             150.8696
```

- 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(tas1,tas12,paired=TRUE))
```

```
##
##  Paired t-test
##
## data:  tas1 and tas12
## t = 6.0672, df = 51, p-value = 1.609e-07
## alternative hypothesis: true mean difference is not equal to
## 95 percent confidence interval:
##   8.518285 16.943253
## sample estimates:
## mean difference
##        12.73077
```

```
summary(hta$tas1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   100.0   140.0   145.0   150.1   160.0   210.0
```

```
summary(hta$tas12)
```

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 0.016504 *
## Normal - Anormal == 0    -14.405      3.543  -4.065 0.000216 ***
## Normal - Frontera == 0    -3.310      2.405  -1.376 0.345695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```r
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 = 2.3459
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                          Estimate lwr      upr
## Frontera - Anormal == 0 -11.0943 -20.5007  -1.6879
## Normal - Anormal == 0   -14.4046 -22.7171  -6.0921
## Normal - Frontera == 0   -3.3103  -8.9533   2.3326
```

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
##
```