Application of Hypothesis Tests

Alex Sanchez, Miriam Mota and Santiago Perez-Hoyos

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

Versión 2022-02-14

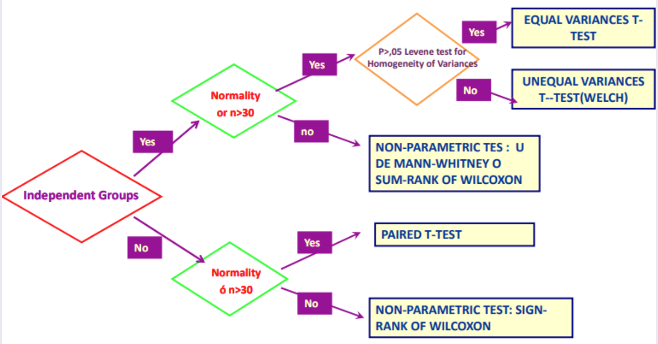
Outline

Introduction

- Once the concept of hypothesis testing is established,
- Researchers face the problem of *which test should be applied at every possible situation*.
- Best solution:
  - understand the problem and the questions addressed
  - know available tests for each problem
  - be aware of applicability assumptions of each test and how to check them.
- 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.

Select the test according to each situation (1)

Some test options when variables are numerical



Introductory Example

- A study was 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

B	C	D	E	F	G	H	I
sexo	grupo	tas1	tad1	tas2	tad2	tas3	tad3
VARON	B	150	100	150	90	170	
MUJER	B	160	90	170	90	160	
MUJER	B	150	90	110	90	115	
VARON	A	120	80	140	90	140	
MUJER	A	150	85	145	85	160	
MUJER	B	140	75	160	70	135	
MUJER	A	150	100	140	90	130	
VARON	A	160	90	170	90	170	
MUJER	A	145	105	170	95	140	
MUJER	A	210	110				
MUJER	A	170	100	170	90	170	
MUJER	B	140	90	140	90	100	

Introductory Example (2)

```
hta <- read_excel("datasets/hta.xls")
print(head(hta))
```

```
## # A tibble: 6 x 27
##   numero sexo  grupo  tas1  tad1  tas2  tad2  tas3  tad3
##    <dbl> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl
## 1      1 VARON B       150   100   150    90   170    90
## 2      2 MUJER B       160    90   170    90   160    80
## 3      3 MUJER B       150    90   110    90   115    90
## 4      4 VARON A       120    80   140    90   140    90
## 5      5 MUJER A       150    85   145    85   160    90
## 6      6 MUJER B       140    75   160    70   135    75
## # ... with 14 more variables: tas6 <dbl>, tad6 <dbl>, ta
## #   tas8 <dbl>, tad8 <dbl>, tas9 <dbl>, tad9 <dbl>, tas1
## #   tas11 <dbl>, tad11 <dbl>, tas12 <dbl>, tad12 <dbl>
```

Some questions to be answered

- Is diastolic (min) tension above 90, “on average”, at the beginning of the study?
- Are samples at baseline equivalent? – In Age? Sex%? Sist? Diast?
- Has there been a change in BP between month 1 (first measure) and month 12?
- Has this change been different between treatment groups?

What type of test for each question?

- Is diastolic (min) tension above 90, "on average", at the beginning of the study?

  – *One variable. Test about the mean*

- Are samples at baseline equivalent?

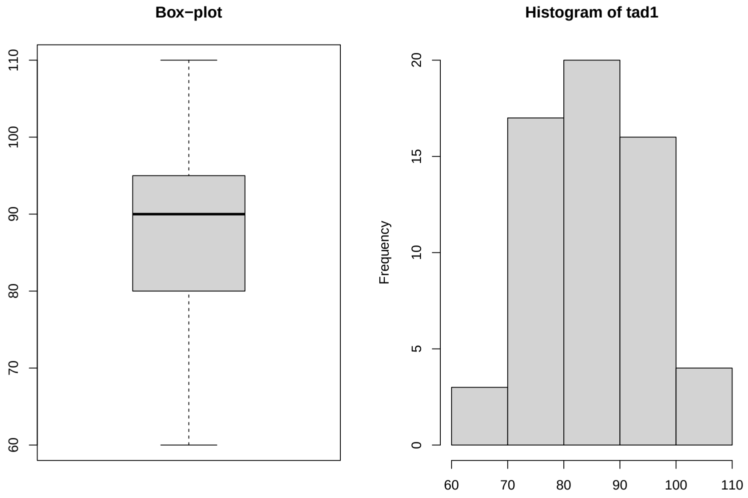
  – *Comparison between distinct groups of individuals in two groups (A/B, Male/Fem)*

- Has there been a change in BP between month 1 (first measure) and month 12?

  – *Comparison between same individuals at different time points*

Always start looking at the data

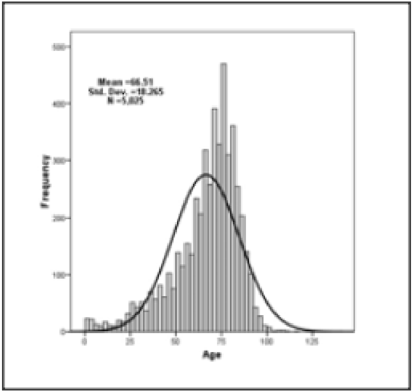
```
par(mfrow=c(1,2)) # Draw four plots in one panel
with(hta, boxplot(tad1, main="Box-plot") )
with(hta, hist(tad1) )
```



General approach

- Hypothesis tesing is useful to answer some questions in problems.
- These can be applied in distinct scenarios:
  - With numerical/continuos or categorical variables.
  - With data distribued as a gaussian population or not
  - With one or more groups of data.
  - With dependent or independent data.
- Try not to blindly apply recipes
  - Think about the situation that generated the data
  - Think about how your question becomes a test
  - Think about the assumptions of the test, how to check them and how robust may the test be to assumption violations.
  - Know how to apply the test
  - Do not blindly and dumby rely on cutoffs!

Normality Test




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

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

One sample Test

- We do not use it very often.
- Very similar to estimation questions. It can be solved calculating a confidence interval
- Idea: We want to verify from a sample a previous hypothesis about the mean in a population
- Example: *Can it be accepted that the initial TAD is 90 or greater in hypertensive patients?*
- If data is assumed to follow a normal distribution: *t-test*
- If data is **not** assumed to follow a normal distribution: *wilcoxon-tets*

Example of one-sample test

Assuming data does not follow a normal distribution ...

```
with(hta,wilcox.test(tad1,mu=90) ) # One sample wilcoxon.test
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  tad1
## V = 429, p-value = 0.2173
## alternative hypothesis: true location is not equal to 90
```

The t-test is robust to small departures of normality ...

```
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
```

Exercise 1

1. Check the normality of tas1 variable, call it “TAS” in hta dataset
2. Can it be accepted that the initial TAS is 120 in Hipertensive patients?
3. Find the 95% confidence interval for the mean of tas1 variable
4. Extra: Can it be accepted that the initial TAS is higher than 120 in Hipertensive women?

Comparison between two groups ("two-sample problems")

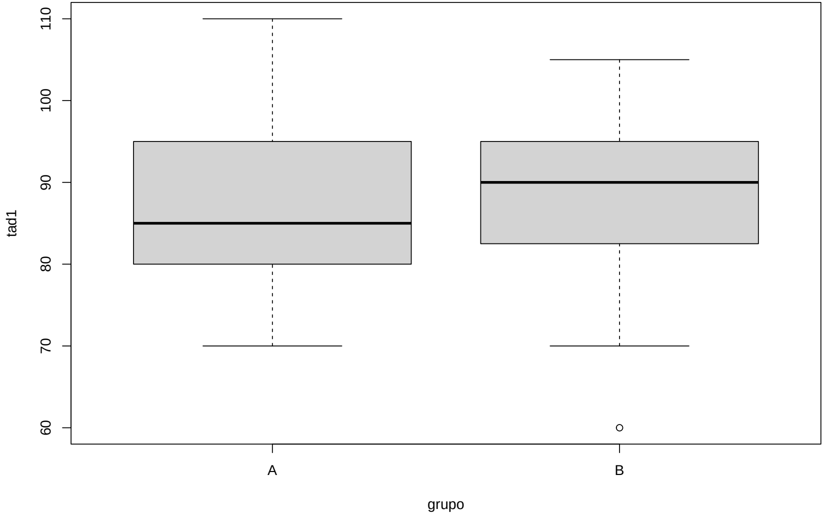
- Most of the times, tests are associated with comparison between two or more groups.
  - Are the two groups A and B comparable?, that is, given a certain variable, does it take on average, the same value, at baseline time?
  - Is blood pressure comparable between first and 12th measures?

Notice that:

- the first question implies comparison between distinct groups of individuals
- while the first one assumes comparison between the same individuals at two time points.

Start looking at the data

```
with(hta, boxplot(tad1~grupo))
```



Compare two groups assuming normality:

two sample t-test for independent groups

- The two-sample t-test is used to compàre two groups assuming normality.
- The test changes depending on if the variances of the two groups can be considered equal or not.
  - This preliminary comparison is omitted here.

```
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 between group MUJER and group VARON is not equal to 0
## 95 percent confidence interval:
##  -4.852834  6.768228
## sample estimates:
## mean in group MUJER mean in group VARON
##            88.78378            87.82609
```

Compare two groups without assuming normality

U Mann-Whitney or Sum Rank non parametric test

```
hta%>%
  group_by(grupo) %>%
  summarise(median = median(tad1))
```

```
## # A tibble: 2 x 2
##   grupo median
##   <chr>  <dbl>
## 1 A         85
## 2 B         90
```

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

```
##
##  Wilcoxon rank sum test
##
## data:  tad1 by factor(grupo)
## W = 432, p-value = 0.7926
## alternative hypothesis: true location shift is not equal to 0
```

- Null Hypothesis cannot be rejected

Exercise 2

1. Is Diastolic pressure ("TDA") comparable at baseline time between Men and women?
2. What is the Hypothesis that we want to test? Describe the null hypothesis and the alternative hypothesis.
3. What test would be appropiate to answer the question?
4. Compute and decide
5. Apply a non-parametric test and compare the results

Comparisons with dependent (paired) data

- If we consider two groups of dependent data we are in a prticular situation.
  - Apparently two groups
  - Only one group of individuals
- Computer programs usually provide tests for "paired data", but they are essentially one-sample tests fro the difference between the values of the same individuals.
- Again, depending on, if normality is assumed or not, we rely on *paired-t-test* or *wilcoxon test for paired data*.

Exercise 3

- Can we consider that systolic and diastolic pressure have changed between baseline and month 12?
- Choose the appropriate test and apply it to yioeld a decision.
- Think carefully about the hypothesis being tested.
  - What is a reasonable option for the alternative hypothesis?

Categorical Variables

- Categorical variables represent facts that can be better described with *labels* than with numbers.
  - Example: `Sex`, better choose from {`Male` , `Female`} than from: {1,2}.
- Sometimes ordering of labels makes sense, although *it is not reasonable to assign numbers to categories*:
  - Example: `Tumor stage`: {1,2,3,4}, but $1 + 2 \neq 3$!!!
  - `Sex` is an example of a categorical variable in `nominal` scale
  - `Stage` is an example of a categorical variable in `ordinal` scale

Representing categorical variables in R

- Categorical variables are well represented with *factors*

```
sex <- factor(c("Female", "Male"))
blood_group <- factor(c("A", "B", "AB", "O"))
```

- Be careful with the names of factors, by default, *levels* assigned in alphabetical order.

```
levels(blood_group)
```

```
## [1] "A"  "AB" "B"  "O"
```

- To verify class of a variable

```
class(sex)
```

```
## [1] "factor"
```

Creating factors

- Factors can be created ...
  - automatically, when reading a file or
    - Not all functions for reading data from file will create a factor!!!
    - Usually levels will be defined from alphabetic order
  - using the `factor` or the `as.factor` commands.
    - more flexible

Create factors automatically

- This is achieved by
  - Using the `read.table` or `read.delim` functions for reading
    - Setting the "character variables as.factors" to TRUE
- Example
  - Load the `diabetes` dataset using the `Import Dataset` feature of Rstudio
    - From text (base) (use the file `osteoporosis.csv`)
    - From text (readr) (use the file `osteoporosis.csv`)
  - What is the class of the variable `menop`

Create factors automatically

```
osteo1 <- read.csv("datasets/osteoporosis.csv",sep = "\t",
                   stringsAsFactors=TRUE)
class(osteo1$menop)
summary(osteo1$menop)
str(osteo1)

library(readr)
osteo2 <- read_delim("datasets/osteoporosis.csv", "\t",
                     escape_double = FALSE)
class(osteo2$menop)
osteo2$menop <- as.factor(osteo2$menop)
class(osteo2$menop)
summary(osteo2$menop)
str(osteo2)
```

Exercise 4

- Select `diabetes.xls` datasets
- Read the dataset into R and check that the categorical variables you are interested (mort, tabac, ecg) in are converted into factors.
- Confirm the conversion by summarizing the variables

The analysis of categorical variables

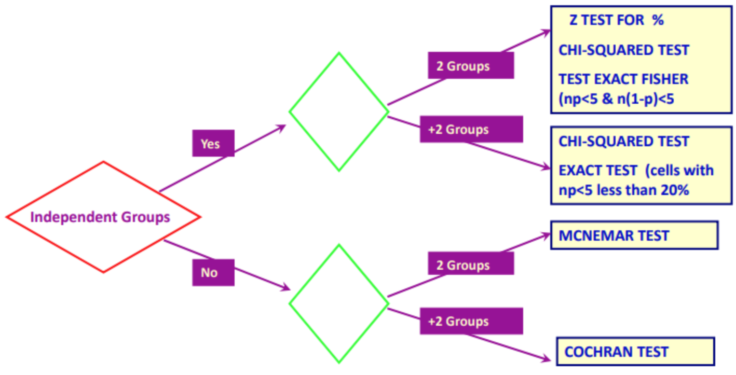
- The analysis of categorical data proceeds as usual:
- Start exploring the data with the tables and graphics
- Proceed to estimation and/or testing if *appropriate*
- Estimation
  - Proportions: Point estimates, confidence intervals
- Testing
  - One variable (tests with proportions)
  - With two variables (chi-square and related)

Types of test with categorical variables

- One variable (tests with proportions)
  - Does the proportion (% affected) match a given value?
  - Is the proportion (% affected) the same in two populations?
- With two variables (chi-square and related)
  - Is there an association between two categorical variables?
  - Is there a relationship between the values of a categorical variable before and after treatment?

Select the test according to each situation (3)

Some test options when variables are categorical

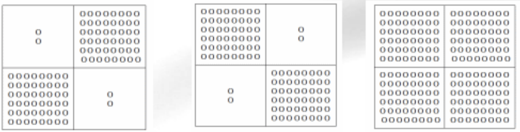


Example

Consider the following study relating smoking and cancer.

Load data: "dadescancer.csv"

	Smoking X=1	Non smoking X=0	TOTAL
CANCER Y=1	190	87	277
NO CANCER Y=0	60	163	223
TOTAL	250	250	500



Our goal here would be to determine if there is an association between smoking and cancer.

Crosstabulating a dataset

- Data may come from a table (aggregated) or disagregated in a data file.
- In this case we need to build the table applying "cross-tabulation"

```
dadescancer <- read.csv("datasets/dadescancer.csv",
                        stringsAsFactors = TRUE)
```

```
#attach(dadescancer)
mytable <-table(dadescancer$cancer, dadescancer$fumar)
mytable
```

```
##
##              Fuma No fuma
##   Cancer      190      87
##   No cancer    60     163
```

Crosstabulation (2): Marginal tables

Marginal values are important to understand the structure of the data:

```
mytable<- addmargins(mytable)
mytable
```

```
##
##              Fuma No fuma Sum
##   Cancer      190      87 277
##   No cancer    60     163 223
##   Sum         250     250 500
```

Crosstabulation (3): In percentages

Showing tables as percentages is useful for comparisons

```r
prop.table(mytable) # cell percentages
```

```
##
##                Fuma No fuma    Sum
##   Cancer     0.0950  0.0435 0.1385
##   No cancer  0.0300  0.0815 0.1115
##   Sum        0.1250  0.1250 0.2500
```

```r
prop.table(mytable, 1) # row percentages
```

```
##
##                   Fuma   No fuma       Sum
##   Cancer     0.3429603 0.1570397 0.5000000
##   No cancer  0.1345291 0.3654709 0.5000000
##   Sum        0.2500000 0.2500000 0.5000000
```

```r
# prop.table(mytable, 2) # column percentages
```

Exercise 5

- With the diabetes dataset repeat the crosstabulation done above using
  - Two categorical variables
  - Variable "mort" and a newly created variable "bmi30" created by properly categorizing variable bmi.

One variable: Proportion tests

- According to medical literature, in the period 1950-1980, the proportion of obese individuals (defined as: BMI ≥ 30) was 15% in the population of men over 55 years old.
- A random sample obtained from the same population between 2000 and 2003 showed that, over a total of 723 men older than 55, 142 were obese.
- With a significance level of 5%, can we say that the population of men older than 55 in 2000-2003 had the same proportion of obese cases than that population had in 50'-80'?

Proportion tests with R

Alternative "NOT EQUAL". This is set by default.

```
prop.test(x=142, n=723, p=0.15)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  142 out of 723, null probability 0.15
## X-squared = 11.849, df = 1, p-value = 0.0005768
## alternative hypothesis: true p is not equal to 0.15
## 95 percent confidence interval:
##  0.1684325 0.2276606
## sample estimates:
##         p
## 0.1964039
```

Alternative "GREATER"

```r
prop.test(x=142, n=723, p=0.15, alternative="g")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  142 out of 723, null probability 0.15
## X-squared = 11.849, df = 1, p-value = 0.0002884
## alternative hypothesis: true p is greater than 0.15
## 95 percent confidence interval:
##  0.1725953 1.0000000
## sample estimates:
##         p
## 0.1964039
```

Alternative "LESS THAN"

```
prop.test(x=142, n=723, p=0.15, alternative="l")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  142 out of 723, null probability 0.15
## X-squared = 11.849, df = 1, p-value = 0.9997
## alternative hypothesis: true p is less than 0.15
## 95 percent confidence interval:
##  0.0000000 0.2225404
## sample estimates:
##         p
## 0.1964039
```

Notice that *choosing the wrong alternative may yield unreasonable conclusions*.

Estimation comes with proportion test

- `prop.test` does **three** distinct calculations
  - A test for the hypothesis $H_0 : p = p_0$ is performed
  - A confidence interval for p is built based on the sample
  - A point estimate for p is also provided.

```
> prop.test(x=142, n=723, p=0.15)

        1-sample proportions test with continuity correction

data:  142 out of 723, null probability 0.15
X-squared = 11.849, df = 1, p-value = 0.0005768
alternative hypothesis: true p is not equal to 0.15
95 percent confidence interval:
 0.1684325 0.2276606
sample estimates:
        p
0.1964039
```

Hypothesis Test

Confidence interval

Point estimate

Exercise 3

- In the diabetes dataset.
  - Test the hypothesis that the proportion of patients with `bmi30` is higher than 40%
    - In the global population of the study
    - Only in patients with ‘mort’ equal “Muerto”

Contingency tables

- A contingency table (a.k.a cross tabulation or cross tab) is a matrix-like table that displays the (multivariate) frequency distribution of the variables.
- It is bidimensional, and classifies all observations according to two categorical variables (A and B, rows and columns).

Clasif	B_1	B_2	...	B_s	Total
A_1	n_{11}	n_{12}	...	n_{1s}	$n_{1\bullet}$
A_2	n_{21}	n_{22}	...	n_{2s}	$n_{2\bullet}$
...	...	...	...	...	
A_r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$		$n_{\bullet s}$	N

Chi-squared test

- A *familiy* of tests receiving its name because they all rely on the *Chi-Squared distribution* to compute the test probabilities.

Chi squared independence test

- When the sample comes from a single population with 2 categorical variables, the aim is to determine if there is relationship between them.

Chi squared homogeneity test

- When each row is a sample from distinct populations (groups, subgroups. . .), the aim is to determine if both groups have significative differences in that variable

Chi-squared tests

- When we have:
  - quantitative data,
  - two or more categories,
  - independent observations,
  - adequate sample size ($>$10)
- and our questions are like. . .
  - *Do the number of individuals or objects that fall in each pair of categories differ significantly from the number you would expect if there was no association?*
  - *Is this difference between the expected and observed due to chance ("sampling variation"), or is it a real difference?*

Chi squared.test: Observed vs expected

Observades	Braf -	Braf +
Grau 1	97	5
Grau 2	81	7
Grau 3	32	18

Esperades	Braf -	Braf +
Grau 1	89.25	12.75
Grau 2	77.00	11.00
Grau 3	43.75	6.25

Chi squared tests: Observed vs expected with R

```
require(gmodels)
mytable <- table(dadescancer$cancer, dadescancer$fumar)
CrossTable(mytable,expected = T,prop.chisq = F,prop.c = F,p
```

```
   Cell Contents
|-------------------------|
|                       N |
|              Expected N |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  500


             |
             |      Fuma |   No fuma | Row Total |
-------------|-----------|-----------|-----------|
      Cancer |       190 |        87 |       277 |
             |   138.500 |   138.500 |           |
             |     0.380 |     0.174 |           |
-------------|-----------|-----------|-----------|
   No cancer |        60 |       163 |       223 |
             |   111.500 |   111.500 |           |
             |     0.120 |     0.326 |           |
-------------|-----------|-----------|-----------|
Column Total |       250 |       250 |       500 |
-------------|-----------|-----------|-----------|
```

Chi squared tests with R

```
chisq.test (mytable)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity corre
##
## data:  mytable
## X-squared = 84.214, df = 1, p-value < 2.2e-16
```

Fisher test. an assumptions-free alternative

Chi-squared test require that sample sizes are "big" and expected frequencies are, at least greater than 5.

Fisher test can be an alternative if these assumptions are not met, especially for two times two tables.

```
fisher.test(mytable)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  mytable
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to
## 95 percent confidence interval:
##  3.945907 8.936465
## sample estimates:
## odds ratio
##   5.909114
```

Exercise 6

- Use the diabetes dataset to study if it can be detected an association between the variables `mort` and `tabac` in the diabetes dataset.
- Do not start with a test but with an appropriate summarization and visualization!

Mcnemar test

Mcnemar test is used to compare the frequencies of paired samples of dichotomous data

- Ho: There is no significant change in individuals after the treatment
- H1: There is a significant change in individuals after the treatment

Mcnemar test. Example

SELF * SURGICAL Crosstabulation

Count

		SURGICAL		Total
		Rupture	No Rupture	
SELF	Rupture	69	28	97
	No Rupture	5	63	68
Total		74	91	165

```r
.Table <- matrix(c(69,28,5,63), 2, 2, byrow=TRUE)
mcnemar.test(.Table)
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  .Table
## McNemar's chi-squared = 14.667, df = 1, p-value = 0.0001
```