

# Chi-squared test in R

April 28, 2019

Applied STAtistics group at AAU

Department of Mathematical Sciences

Aalborg University



**AALBORG UNIVERSITY**  
DENMARK

# Introduction

Outline of session:

- ▶  $\chi^2$ -test in R.

---

Lecturer for this session is Ege Rubak, Dept. of Math. Sciences, AAU



# Raw data

```
popKids <- read.delim("https://asta.math.aau.dk/datasets?file=PopularKids.txt")
head(popKids, n = 15)
```

##	Gender	Grade	Age	Race	Urban.	Rural	School	Goals	Grades	Sports	Looks	Money
## 1	boy	5	11	White		Rural	Elm Sports		1	2	4	3
## 2	boy	5	10	White		Rural	Elm Popular		2	1	4	3
## 3	girl	5	11	White		Rural	Elm Popular		4	3	1	2
## 4	girl	5	11	White		Rural	Elm Popular		2	3	4	1
## 5	girl	5	10	White		Rural	Elm Popular		4	2	1	3
## 6	girl	5	11	White		Rural	Elm Popular		4	2	1	3
## 7	girl	5	10	White		Rural	Elm Popular		3	4	1	2
## 8	girl	5	10	White		Rural	Elm Grades		3	4	2	1
## 9	girl	5	10	White		Rural	Elm Sports		3	2	1	4
## 10	girl	5	10	White		Rural	Elm Sports		4	3	2	1
## 11	girl	5	11	White		Rural	Elm Sports		2	3	1	4
## 12	girl	4	10	White		Rural	Elm Grades		2	3	4	1
## 13	boy	4	9	White		Rural	Elm Popular		2	3	4	1
## 14	boy	4	9	White		Rural	Elm Popular		4	2	3	1
## 15	boy	4	9	Other		Rural	Elm Popular		4	3	2	1

## Contingency table

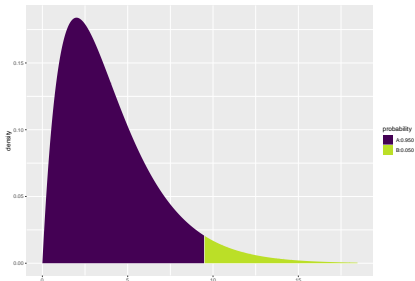
```
library(mosaic)
tab <- tally( ~ Urban.Rural + Goals, data = popKids)
tab
```

```
##           Goals
## Urban.Rural Grades Popular Sports
##   Rural      57      50      42
##   Suburban   87      42      22
##   Urban    103      49      26
```

# Chi-squared test in R

```
testStat <- chisq.test(tab, correct = FALSE)
testStat
```

```
##
## Pearson's Chi-squared test
##
## data:  tab
## X-squared = 18.828, df = 4, p-value = 0.0008497
qdist("chisq", 0.95, df = 4)
```



```
## [1] 9.487729
```

## Chi-squared test in R

### ► Expected table

```
testStat$expected
```

```
##           Goals
## Urban.Rural  Grades  Popular  Sports
##   Rural      76.99372 43.95188 28.05439
##   Suburban  78.02720 44.54184 28.43096
##   Urban     91.97908 52.50628 33.51464
```

### ► Std. residuals

```
testStat$stdres
```

```
##           Goals
## Urban.Rural  Grades  Popular  Sports
##   Rural      -3.9508449  1.3096235  3.5225004
##   Suburban    1.7666608 -0.5484075 -1.6185210
##   Urban       2.0865780 -0.7274327 -1.8186224
```

## Entering table data directly in R

```
data <- c(57, 87, 103, 50, 42, 49, 42, 22, 26)
tab <- matrix(data, nrow = 3, ncol = 3)
row.names(tab) <- c("Rural", "Suburban", "Urban")
colnames(tab) <- c("Grades", "Popular", "Sports")
tab
```

```
##           Grades Popular Sports
## Rural           57      50     42
## Suburban        87      42     22
## Urban          103      49     26
```

```
chisq.test(tab)
```

```
##
## Pearson's Chi-squared test
##
## data:  tab
## X-squared = 18.828, df = 4, p-value = 0.0008497
```