ATBC BIOINFORMATICA

STATISTIEK IN R_

Week 5

HAN_UNIVERSITY
OF APPLIED SCIENCES

ONDERWERPEN

- Statistische toetsen
- Binomiale verdeling
  - dbinom()
  - binom.test
- Chi-kwadraattoets
  - chisq.test

STATS PACKAGE

- Bevat verschillende functies voor statistische toetsen, berekeningen en verdelingen.

- library(help = "stats")

BINOMIALE VERDELING

- De binomiale verdeling verteld de kans van het aantal successen *X* bij het aantal onafhankelijke herhalingen *n* met succeskans *p.*
- Dit word ook wel een Bernoulli experiment genoemd.

- dbinom(x, size, prob, log=FALSE)

DBINOM()

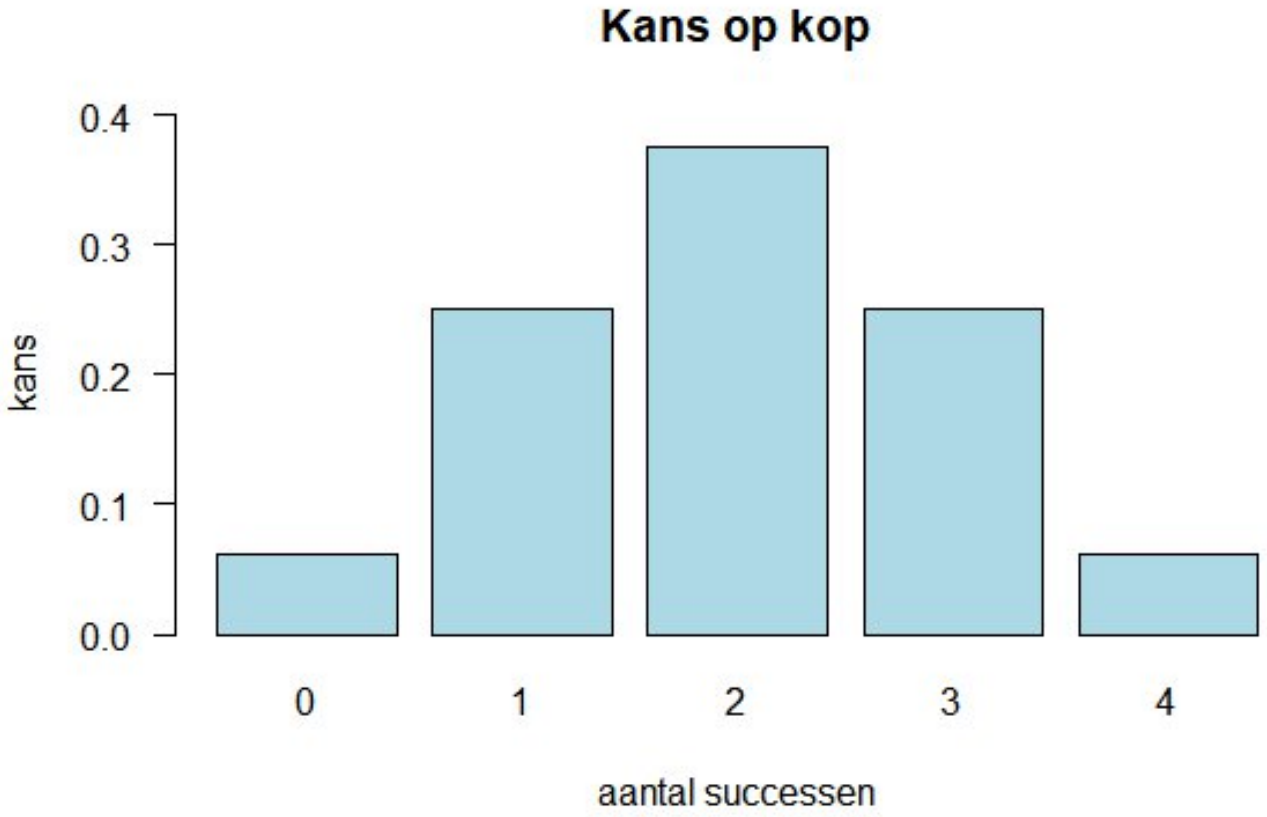
- De kans op het aantal keer “kop” bij het opgooien van een munt.
- Bij 4x gooien, hoe groot is de kans op 3x kop?

- dbinom(x=3, size=4, prob=1/2)
- dbinom(3, 4, 1/2)

DBINOM()

- Hoe groot is de kans op 0, 1, 2, 3 of 4x kop bij het 4x opgooien van een munt?
- dbinom(0:4, 4, 1/2)
- Opdracht: zet deze verdeling uit in een barplot, maak hier een volledige grafiek van.

BARPLOT




```
barplot(dbinom(0:4,4,1/2), ylim=c(0,0.40), names.arg=0:4, las=1, main = "Kans op kop",
xlab = "aantal successen", ylab = "kans", col = "lightblue")
```

DBINOM()

- Hoe groot is de kans op 3 of minder vaak kop bij het 4x opgooien van een munt?

- sum (dbinom(0:3, 4, 1/2))
- pbinom(q=3, size=4, p=1/2, lower.tail=T)

BINOMIAAL TESTEN

?binom.test

Exact Binomial Test

Description

Performs an exact test of a simple null hypothesis about the probability of success in a Bernoulli experiment.

Usage

```
binom.test(x, n, p = 0.5,
        alternative = c("two.sided", "less", "greater"),
        conf.level = 0.95)
```

Arguments

x	number of successes, or a vector of length 2 giving the numbers of successes and failures, respectively.
n	number of trials; ignored if x has length 2.
p	hypothesized probability of success.
alternative	indicates the alternative hypothesis and must be one of "two.sided", "greater" or "less". You can specify just the initial letter.
conf.level	confidence level for the returned confidence interval.

VOORBEELD

- Conover (1971), p. 97f.
- Under (the assumption of) simple Mendelian inheritance, a cross between plants of two particular genotypes produces progeny 1/4 of which are "dwarf" and 3/4 of which are "giant", respectively.
- In an experiment to determine if this assumption is reasonable, a cross results in progeny having 243 dwarf and 682 giant plants.
- If "giant" is taken as success, the null hypothesis is that p = 3/4 and the alternative that p != 3/4.

- binom.test(c(682, 243), p = 3/4)
- binom.test(682, 682 + 243, p = 3/4) # The same.

- => Data are in agreement with the null hypothesis.

OUTPUT & ATTRIBUTEN
BINOM.TEST(C(682, 243), P = 3/4)

Exact binomial test

data: c(682, 243)
number of successes = 682, number of trials = 925, p-value = 0.3825
alternative hypothesis: true probability of success is not equal to 0.75

95 percent confidence interval:
 0.7076683 0.7654066

sample estimates:
probability of success
            0.7372973

BINOM = binom.test(c(682, 243), p = 3/4)
attributes(BINOM)
$names
[1] "statistic"   "parameter"   "p.value"   "conf.int"   "estimate"
[6] "null.value" "alternative" "method"   "data.name"

$class
[1] "htest"

OPDRACHT 3.6.8

The sex ratio of newborn human infants is about 105 males : 100 females. If four infants are chosen at random, what is the probability that

(a) two are male and two are female?

(b) all four are male?

(c) all four are the same sex?

HAN_UNIVERSITY OF APPLIED SCIENCES

Bron: Samuels, M, Witmer, J.A., Schaffner, A.A.(2016), *Statistics for the Life Sciences*. Pearson.

OPDRACHT 3.6.12

Childhood lead poisoning is a public health concern in the United States. In a certain population, 1 child in 8 has a high blood lead level (defined as 30 mg/dl or more).
In a randomly chosen group of 16 children from the population, what is the probability that
(a) none has high blood lead?
(b) 1 has high blood lead?
(c) 2 have high blood lead?
(d) 3 or more have high blood lead?
[Hint: Use parts (a)–(c) to answer part (d).]

HAN_UNIVERSITY OF APPLIED SCIENCES

Bron: Samuels, M, Witmer, J.A., Schaffner, A.A.(2016), *Statistics for the Life Sciences*. Pearson.

CHI-KWADRAATTOETS

- Een chi-kwadraattoets is een toets om na te gaan of twee of meer verdelingen (populaties) van elkaar verschillen.

?CHISQ.TEST

Pearson's Chi-squared Test for Count Data
Description
chisq.test performs chi-squared contingency table tests and goodness-of-fit tests.

Usage
```
chisq.test(x, y = NULL, correct = TRUE,
        p = rep(1/length(x), length(x)), rescale.p = FALSE,
        simulate.p.value = FALSE, B = 2000)
```

Arguments
x a numeric vector or matrix. x and y can also both be factors.

y a numeric vector; ignored if x is a matrix. If x is a factor, y should be a factor of the same length.

correcta logical indicating whether to apply continuity correction when computing the test statistic for 2 by 2 tables: one half is subtracted from all |O - E||O−E| differences; however, the correction will not be bigger than the differences themselves. No correction is done if simulate.p.value = TRUE.

p a vector of probabilities of the same length as x. An error is given if any entry of p is negative.

rescale.p a logical scalar; if TRUE then p is rescaled (if necessary) to sum to 1. If rescale.p is FALSE, and p does not sum to 1, an error is given.

CHISQ.TEST VOORBEELD

- # goodness of fit
- x <- c(A = 20, B = 15, C = 25)
- chisq.test(x)

Chi-squared test for given probabilities

data: x
X-squared = 2.5, df = 2, p-value = 0.2865

CHISQ.TEST VOORBEELD

- # goodness of fit
- x <- c(89,37,30,28,2)
- p <- c(40,20,20,15,5)
- chisq.test(x, p = p, rescale.p = TRUE)

Chi-squared test for given probabilities

data: x
X-squared = 9.9901, df = 4, p-value = 0.04059

CHISQ.TEST VOORBEELD

```
# contigency table (example 10.5.1)

plover <- c(21,19,26,17,38,12,5,6,9)
plovmatrix <- matrix(plover, nrow=3, byrow=T,       dimnames = list(location=c("AF", "PD",
"G"),
                                year = c("2004","2005", "2006" )))
```

	year		
location	2004	2005	2006
AF	21	19	26
PD	17	38	12
G	5	6	9

```
chisq.test(plovmatrix)
```

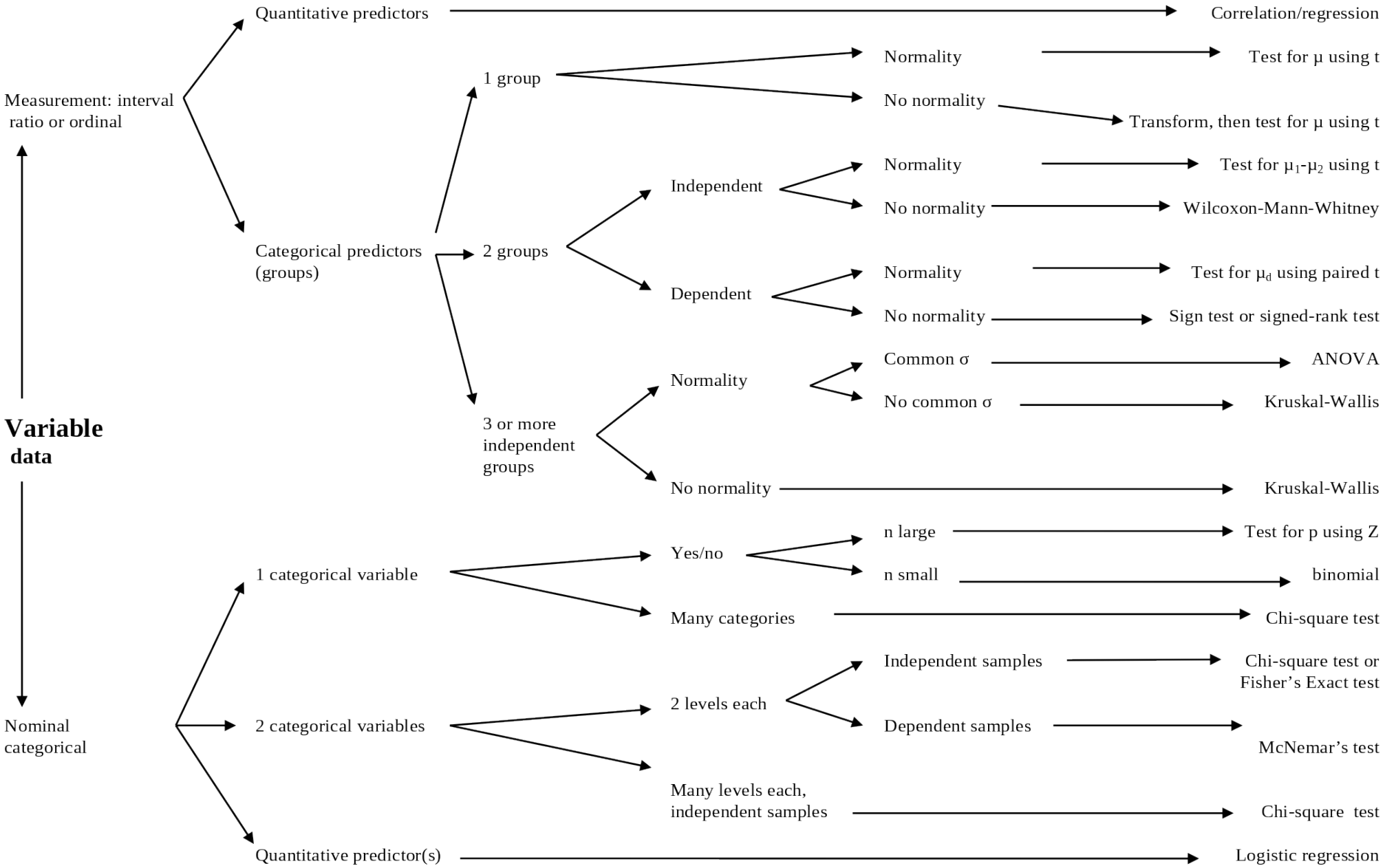
CHISQ.TEST(PLOVMATRIX)

- Pearson's Chi-squared test

- data: plovmatrix
- X-squared = 14.089, df = 4, p-value =0.007015

WANNEER WELKE TEST ?

Flowchart of inference methods



OPDRACHTEN

- Binom: Opgaven: 3.6: 1, 2, 4
- Chi g-o-f: Opgaven: 9.4: 1, 2, 4, 5, 10, 12

VERANTWOORDING

- In deze uitgave is géén auteursrechtelijk beschermd werk opgenomen
- Alle teksten © Gonny Henkes tenzij expliciet externe bronnen zijn aangegeven
- Screenshots op basis van eigen werk auteur en/of vernoemde websites
- Eventuele images zijn opgenomen met vermelding van bron