

EXTRA

Twee plots naast elkaar	<code>par(mfrow=c(1,2))</code>
Plots naast elkaar uitschakelen	<code>dev.off()</code>

Z-toets

Standaardnormale verdeling	<code>dnorm(x (evt, mean=, sd=))</code>
Standaardnormale verdeling plot	<code>plot(x, dnorm(x))</code>
Kans berekenen	<code>pnorm(i, mean, sd)</code>
Voorspellingsinterval 95% berekenen	<code>qnorm(c(0.025, 0.975), mean, sd)</code>
Voorspellingsinterval 99% berekenen	<code>qnorm(c(0.05, 0.995), mean, sd)</code>
Voorspellingsinterval 95% van steekproef berekenen	<code>qnorm(c(0.025, 0.975), mean, sd/sqrt(j))</code>

i = variabel dat je wilt testen

j = steekproefgrootte (n)

T-test

Bij een gepaarde test, `t.test(na, voor, paired = T)`

Hypothesen

Shapiro-Wilk test	H0: De data is normaal verdeeld.	HA: De data is niet normaal verdeeld.
T-test	H0: Er is geen verschil tussen de verwachtingswaarde en het gemiddelde.	HA: Er is wel verschil tussen de verwachtingswaarde en het gemiddelde.
W-M-W-test	H0: De populatie distributie van X1 en X2 is hetzelfde.	HA: De populatie distributie van X1 en X2 zijn niet hetzelfde (verschillend).
Chi-kwadraat homogeniteit	H0: Er is geen onafhankelijkheid. (Geen verschil in verdeling)	HA: Er is geen onafhankelijkheid. (Wel verschil in verdeling.)
Chi-kwadraat goodness-of-fit	H0: Er is geen verband tussen X1 en X2,	HA: Er is een verband tussen X1 en X2.
ANOVA	H0: Er is geen verschil in de verwachtingswaarden.	HA: Er is minimaal één van de gemiddelde die significant verschillend is van de andere gemiddelden.
The Sign test	H0: De distributie van X1 is hetzelfde als X2.	HA: De distributie van X1 is niet hetzelfde als X2.



By **Kaji**

cheatography.com/kaji/

Not published yet.

Last updated 23rd April, 2019.

Page 1 of 4.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>

Hypothesen (cont)

WSR test H0: Er is geen verschil in Y tussen X1 en X2. HA: Er is wel verschil in Y tussen X1 en X2.

Shapiro Wilk test:

$P > 0.05$? Data is normaal verdeeld. H0 niet verwerpen.

$P < 0.05$? Data niet normaal verdeeld. H0 verwerpen.

- De gevonden verschillen berusten niet alleen op toeval.

Standaard berekeningen

Gemiddelde	<code>mean()</code>
Mediaan	<code>median()</code>
Variantie	<code>var()</code>
Standaarddeviatie	<code>sd()</code>
Kwantiel	<code>quantile()</code>
Gemiddelde, kwartielen en mediaan	<code>summary()</code>

Binomale testen

Binomale verdeling voor plot	<code>dbinom(x, aantal, kans)</code>
Kans berekenen	<code>pbinom(x, aantal, kans, lower.tail=TRUE)</code>
Kwantielen	<code>qbinom(x, aantal, kans, lower.tail=TRUE)</code>
x	<code>seq(startwaarde, stopwaarde, stapgrootte)</code>

Lineaire Regressie

Scatterplot maken	<code>plot(y~x)</code>
Lineaire regressielijn	<code>lm(y~x)</code>
Regressielijn toevoegen in plot	<code>lines(regressie\$fitted.values-data\$x, type="l")</code>
Correlatie coëfficiënt/bepalen	<code>cor(x, y)</code>
Sterkte van correlatie berekenen	<code>cor.test(x, y)</code>
Fitted waarden (verwachte y-waarde)	<code>fitted(regressie)</code>
Residue waarden (verschil tussen fitted en waargenomen)	<code>resid(regressie)</code>
Residue waarden in plot weergeven	<code>segments(x, y, x, fitted(regressie))</code>
Verschil fitted waarden tussen prediction interval	<code>predict.lm(regressie, int = 'prediction')</code>
Verschil fitted waarden tussen confidence interval	<code>predict.lm(regressie, int = 'confidence')</code>
Plot van opgestelde model	<code>abline(slope, interception)</code>

x <- variable met vaste waarden

y <- random

Y is somehow afhankelijk van X.

```
regressie <- lm(y~x)
```



By **Kaji**
cheatography.com/kaji/

Not published yet.
 Last updated 23rd April, 2019.
 Page 2 of 4.

Sponsored by **Readable.com**
 Measure your website readability!
<https://readable.com>

Chi-kwadraat test

Chi-kwadraat test (χ^2 berekenen)	<code>chisq.test(x)</code>
Voorspelde waarden bij geen verband	<code>chisq.test(x)\$expected</code>
Verskil tussen verwachtte en gevonden waarden weergeven	<code>chisq.test(x)\$residuals</code>
Kwantielen om te plotten/grenswaarde van de grootheid χ^2	<code>qchisq(0.99, df)</code>
Chi-kwadraat om te plotten	<code>dchisq()</code>
Conclusie	Er is wel/geen verband in verdeling tussen groepen.

Deze toets gebruiken voor data met een categorische respons: data die je in categorieën kan verdelen.

x = een matrix

df = degrees of freedom (aantal rijen - 1) * (aantal kolommen - 1) ! totaal niet meenemen

Chi-kwadraat Goodness of Fit

```
> rat5k = matrix(c(21,13,8,11,17),
+               dimnames=list(c("Teeg", "voer",
+                             "voer + drinken",
+                             "looprek", "drinken")))
> chisq.test(rat5k)

Chi-squared test for given probabilities

data:  rat5k
X-squared = 7.4286, df = 4, p-value = 0.1149
```

Komt een gevonden verdeling van waarden van één kwalitatief kenmerk overeen met verwachte verdeling?

Goodness of fit (aanpassing)

bv: dobbelsteen, wordt elke zijde 1/6 van alle keren geworpen?

Chi-kwadraat Homogeniteit

```
> ratLD = matrix(c(24,36,6,34),nrow=2,ncol=2, byrow=T,
+               dimnames=list(c("vrouw", "man"), c("donker", "licht")))
> ratLD
      donker licht
vrouw  24    36
man     6    34
> chisq.test(ratLD, correct = F)

Pearson's Chi-squared test

data:  ratLD
X-squared = 7.1429, df = 1, p-value = 0.007526
```

Is er een verband tussen kwalitatieve kenmerken?

independence and homogeneity (onafhankelijkheid en homogeniteit)

bv: mannen/vrouwen en rokers/niet-rokers

Niet-parametrische testen

Wilcoxon signed rank test	<code>wilcox.test(data1, data2, paired=TRUE, exact = TRUE, conf.level = (confidence level 0.95/0.99), conf.int = TRUE)</code>
Wilcoxon Mann-W-hitney test	<code>wilcox.test(data1, data2, alternative = 'greater/less, exact = FALSE, conf.int = TRUE, conf.level = 0.95/0.99)</code>

Verskil toetsen bij niet-normaal verdeelde data.

Bij onafhankelijke data: WMW.

Bij afhankelijke data: sign test of WSR.

Alternative alleen gebruiken bij eenzijdige toetsen.



ANOVA

1) Kijken of elke groep data normaal verdeeld is	<code>shapiro.test()</code>
2) 1 kolom met waarden, 1 kolom met indicator	<code>stack()</code>
3.1) One-way ANOVA	<code>anova(lm(waardes~indicator))</code>
3.2a) Two-way ANOVA per factor	<code>anova(lm(waardes~indicator1 + indicator2))</code>
3.2b) Two-way ANOVA interactie testen	<code>anova(lm(waardes~indicator1 : indicator2))</code>
3.2c) Two-way ANOVA factor + interactie	<code>anova(lm(waardes~indicator1 * indicator2))</code>
Boxplot maken van two way ANOVA	<code>boxplot(waardes ~ indicator1 + indicator2)</code>
Interactie plot	<code>interaction.plot(indicator1, indicator2, y)</code>
* Unstacken om de shapiro test te doen	<code>unstack(data, form = waardes ~ indicator)</code>
* Shapiro test voor alle groepen tegelijkertijd	<code>lapply(unstackdata, shapiro.test)</code>

Post-hoc

Bonferroni correctie	<code>paired.t.test(values, indicator, p.adjust.method="bonferroni")</code>
Tukey HSD	<code>TukeyHSD(aov(lm(values~indicator)))</code>

Gebruik dit na de ANOVA om te kijken welke categorieën precies afwijken.

values = y

indicators = x



By **Kaji**

cheatography.com/kaji/

Not published yet.

Last updated 23rd April, 2019.

Page 4 of 4.

Sponsored by **Readable.com**

Measure your website readability!

<https://readable.com>