

Lijst met R commando's

Het statistisch pakket R kan gratis van het web gedownload worden vanop <http://cran.r-project.org>.

Observatievector invoeren

Het commando **c(...)** wordt gebruikt om vectoren in te geven, bijvoorbeeld om een observatievector in te voeren

```
x <- c(60, 72, 57, 90, 95, 72)
```

Opeenvolgende getallen bv. 1 2 3 4 ... 10:

```
x<- 1:10
```

Lengte van een vector

length(x)

Logische operatoren

Kleiner dan: <, groter dan: >, kleiner of gelijk aan: <=, groter of gelijk aan: >=, gelijk aan: ==, en: &, of: |

Tabel inlezen

tabel<-read.table('filenaam', header=T)

Het commando **attach(tabel)** kan gebruikt worden om de data van de tabel beschikbaar te maken.

Verwijderen van variabelen

Met behulp van het commando **rm()** kan je een bepaalde of alle variabelen uit de workspace in R verwijderen.

rm(variabele) verwijdert enkel de variabele met naam 'variabele'.

Om alle variabelen te verwijderen moet **rm(list = ls())** gebruikt worden. Dit kan ook door in het menu Misc 'Remove all objects' te selecteren én moet voor alle tabellen waarvoor attach(tabel) werd gebruikt, **detach(tabel)** worden uitgevoerd.

Deel van observaties in een kolom

Bijvoorbeeld de elementen weergeven van een kolom a waarvoor geldt dat de overeenkomstige elementen in kolom b kleiner zijn dan 8:

```
a[b<8]
```

Sommeren

Som van elementen in observatievector: **sum(x)**

Aantal elementen bv kleiner dan 8: **sum(x<8)**

Rangschikken

Het commando **sort(x)** rangschikt de waarnemingen x van klein naar groot.

Histogram

Het commando **hist(x)** genereert een histogram van de waarnemingen x.

Staafdiagram

Het commando **barplot(x)** genereert een staafdiagram van de frequenties of relatieve frequenties x. Het commando **barplot(x,names.arg=label)** plaatst onder elk staafje de naam corresponderend met het overeenkomstige element van de vector 'label'. Label is bijvoorbeeld gedefinieerd als **label<-c("label1","label2","label3",...)**

Stam- en bladdiagram

Het commando **stem(x)** genereert een stam- en bladdiagram van de gegevens beschreven door de vector x. Door een waarde op te geven voor 'scale', bijvoorbeeld **stem(x,scale=2)**, kan je de lengte van het stam-en bladdiagram aanpassen.

Puntenwolk of scatter plot

Het commando **plot(x,y)** genereert een puntenwolk/scatter plot van de waarnemingen y als functie van x. Het commando **plot(x,y,xlab="x-label",ylab="y-label",main="title")** genereert een puntenwolk van de waarnemingen y als functie van x waarbij de x-as wordt aangeduid met het label 'x-label', de y-as met het label 'y-label' en een titel met label 'title' bovenaan wordt geplaatst.

Box-plot

Het commando **boxplot(x)** genereert een box-plot van de gegevens beschreven door de vector x.

Het commando

boxplot(x1,x2,x3,ylab="y-label",xlab="x-label",names=c("name1","name2","name3"))

genereert drie box-plots van de gegevens beschreven door de vectoren x1, x2 en x3 respectievelijk. Hierbij wordt de y-as aangeduid met het label 'y-label' en de x-as met het label 'x-label'. Verder worden de drie box-plots op de x-as aangeduid met 'name1', 'name2' en 'name3'.

Het commando **fivenum(x)** geeft je de 5 kengetallen waarop een boxplot gebaseerd is: minimum, eerste kwartiel Q_1 , mediaan M_e , derde kwartiel Q_3 en maximum.

Kengetallen

Het rekenkundig gemiddelde, mediaan, variantie en standaarddeviatie van de waarnemingen x wordt berekend met de commando's **mean(x)**, **median(x)**, **var(x)** en **sd(x)**. De steekproefcorrelatiecoëfficiënt van de waarnemingen x en y wordt berekend met het commando **cor(x,y)**.

Sommatie uitrekenen

Het commando **sum(x*y)** met x en y twee vectoren van lengte n kan gebruikt worden om de som $\sum_{i=1}^n x_i y_i$ uit te rekenen.

Binomiale verdeling

Het commando **dbinom(x,n, π)** berekent de kans dat $X=x$ voor een Binomiaal verdeelde kansvariabele X met parameters n en π (**kansverdeling**).

Het commando **pbinom(x,n, π)** berekent de kans dat $X \leq x$ voor een Binomiaal verdeelde kansvariabele X met parameters n en π (**cumulatieve verdelingsfunctie**).

Het commando **qbinom(p, n, π)** berekent het (100p)de kwantiel y_p voor een Binomiaal verdeelde kansvariabele X met parameter n en π .

Poisson verdeling

Het commando **dpois(x,λ)** berekent de kans dat $X=x$ voor een Poisson verdeelde kansvariabele X met parameter λ (**kansverdeling**).

Het commando **ppois(x,λ)** berekent de kans dat $X \leq x$ voor een Poisson verdeelde kansvariabele X met parameter λ (**cumulatieve verdelingsfunctie**).

Het commando **qpois(p,λ)** berekent het (100xp)de kwantiel γ_p voor een Poisson verdeelde kansvariabele X met parameter λ .

Normale kansdichtheid

Het commando **dnorm(x,μ,σ)** berekent de kansdichtheid voor de waarde $X=x$ voor een normaal verdeelde kansvariabele X met verwachte waarde μ en variantie σ^2 (**kansdichtheid**).

Het commando **pnorm(x,μ,σ)** berekent de kans dat $X \leq x$ voor een normaal verdeelde kansvariabele X met verwachte waarde μ en variantie σ^2 (**cumulatieve verdelingsfunctie**).

Het commando **qnorm(p,μ,σ)** berekent het (100xp)de kwantiel γ_p voor een Binomiaal verdeelde kansvariabele X met parameter n en π ($P(X \leq \gamma_p) = p$).

Student's t-verdeling

Het commando **pt(x,n)** berekent de kans dat $X \leq x$ voor een t-verdeelde kansvariabele X met n het aantal vrijheidsgraden (**cumulatieve verdelingsfunctie**).

Het commando **qt(p,n)** berekent het (100xp)de kwantiel γ_p voor een t-verdeelde kansvariabele X met n het aantal vrijheidsgraden ($P(X \leq \gamma_p) = p$).

χ^2 -verdeling

Het commando **pchisq(x,n)** berekent de kans dat $X \leq x$ voor een χ^2 -verdeelde kansvariabele X met n het aantal vrijheidsgraden (**cumulatieve verdelingsfunctie**).

Het commando **qchisq(p,n)** berekent het (100xp)de kwantiel γ_p voor een χ^2 -verdeelde kansvariabele X met n het aantal vrijheidsgraden ($P(X \leq \gamma_p) = p$).

F-verdeling

Het commando **pf(x,n,m)** berekent de kans dat $X \leq x$ voor een F-verdeelde kansvariabele X met n vrijheidsgraden voor de teller en m vrijheidsgraden voor de noemer (**cumulatieve verdelingsfunctie**).

Het commando **qf(p,n,m)** berekent het (100xp)de kwantiel γ_p voor een F-verdeelde kansvariabele X met n vrijheidsgraden voor de teller en m vrijheidsgraden voor de noemer ($P(X \leq \gamma_p) = p$).

Kwantiendiagram

Het commando **qqnorm(x)** produceert een kwantiendiagram van de gegevens x .

Shapiro Wilk toets

Het commando **shapiro.test(x)** toetst of de gegevens x mogelijk afkomstig zijn uit een normaal verdeelde populatie door middel van een Shapiro Wilk test.

T-toets voor 1 gemiddelde

Het commando **t.test(x,mu=mu0,alternative="two.sided")** toetst de nulhypothese $\mu = \mu_0$ versus de alternatieve hypothese $\mu \neq \mu_0$.

Het commando **t.test(x,mu=mu0,alternative="less")** toetst de nulhypothese $\mu = \mu_0$ versus de alternatieve hypothese $\mu < \mu_0$.

Het commando **t.test(x,mu=mu0,alternative="greater")** toetst de nulhypothese $\mu=\mu_0$ versus de alternatieve hypothese $\mu>\mu_0$.

T-toets voor 2 gemiddeldes

Het commando **t.test(x,y,alternative="two.sided",paired=FALSE,var.equal=FALSE)** voert een T-toets uit voor het vergelijken van 2 gemiddeldes op basis van de vectoren x en y. **Alternative** specificeert de alternatieve hypothese. Gebruik "two.sided" voor een tweezijdige hypothesetoets, "greater" voor een rechts eenzijdig hypothesetoets, en "less" voor een links eenzijdig hypothesetoets. Met **paired** geef je aan of de steekproeven onafhankelijk (FALSE) of afhankelijk (TRUE) zijn. Met **var.equal** geef je aan of de varianties ongelijk (FALSE) of gelijk (TRUE) verondersteld mogen worden.

F-toets voor 2 varianties

Het commando **var.test(x,y,alternative="two.sided")** voert een F-toets uit voor het vergelijken van 2 varianties op basis van de vectoren x en y. **Alternative** specificeert de alternatieve hypothese. Gebruik "two.sided" voor een tweezijdige hypothesetoets, "greater" voor een rechts eenzijdig hypothesetoets, en "less" voor een links eenzijdig hypothesetoets.

Wilcoxon rangsomtoets en Wilcoxon rangtekentoets

Het commando **wilcox.test(x,y,paired=FALSE,alternative="greater")** voert de Wilcoxon rangsomtoets uit voor onafhankelijke steekproeven wanneer **paired** gelijk is aan FALSE op basis van de vectoren x en y. Wanneer **paired** gelijk is aan TRUE dan wordt de Wilcoxon rangtekentoets uitgevoerd. **Alternative** specificeert de alternatieve hypothese. Gebruik "two.sided" voor een tweezijdige hypothesetoets, "greater" voor een rechts eenzijdig hypothesetoets, en "less" voor een links eenzijdig hypothesetoets.

ANOVA

Beschouw een reeks van n waarnemingen, voorgesteld door een vector 'x' (lengte n). Met elk element wordt een specifieke behandeling (uit g mogelijke behandelingen) geassocieerd. Deze vector wordt voorgesteld door 'behandeling' (lengte n). Wanneer 'behandeling' numerieke waarden bevat (waarden 1, 2, 3,...) dan moet R eerst duidelijk gemaakt worden dat de nummering van de behandelingen op zich geen betekenis heeft:

behandeling<-as.factor(behandeling)

'behandeling' is nu een factor. Vervolgens kan in R een enkelvoudige variantie-analyse uitgevoerd worden met de volgende commando's:

lm1<-lm(x~behandeling)

anova(lm1)

Getest wordt of de gemiddelden voor elk van de g behandelingen gelijk zijn.

Tukey methode

Het commando **TukeyHSD(aov(x~behandeling))** kan gebruikt worden om simultaan 95% betrouwbaarheidsintervallen te construeren voor alle 2-aan-2 vergelijkingen van een reeks gemiddelden.

Bartlett toets voor het vergelijken van meer dan 2 varianties

Het commando **bartlett.test(x~behandeling)** kan gebruikt worden om de nulhypothese te toetsen dat de varianties (behorende bij de g behandelingen) gelijk zijn.

Shapiro Wilk test voor normaliteit van de residuele waarden

Het commando **shapiro.test(lm1\$residuals)** kan gebruikt worden om de nulhypothese te toetsen dat de residuele waarden, dit zijn de verschillen tussen de originele waarnemingen 'x' en het geschatte lineaire model 'lm1' normaal verdeeld zijn.

Lineaire regressie

In R kan een lineaire regressie worden uitgevoerd met de volgende commando's:

lm1<-lm(y~x)

met y de respons variabele en x de verklarende variabele.

Een scatterplot met bijbehorende regressierechte wordt verkregen met de commando's:

plot(x,y)

abline(lm1)

Betrouwbaarheidsintervallen voor schattingen van de regressiecoëfficiënten

95% betrouwbaarheidsintervallen voor de geschatte regressiecoëfficiënten worden verkregen met het commando

confint(lm1)

Significantie van het lineair verband

De hypothese $H_0: \beta_1=0$ versus de alternatieve hypothese $H_a: \beta_1 \neq 0$ volgt uit het 95% betrouwbaarheidsinterval voor de helling of uit de p-waarde van een T-toets zoals weergegeven in de tabel verkregen met het commando

summary(lm1)

Geobserveerde residu's

De geobserveerde residu's kunnen in R opgevraagd worden met het commando **lm1\$residuals**

Betrouwbaarheidsintervallen, predictie-intervallen en voorspellingen

Betrouwbaarheidsintervallen, predictie-intervallen, voorspellingen bij de waarden $x=x_0$ en grafieken van deze intervallen kunnen in R verkregen worden met volgende commando's waarbij x_0 een getalwaarde of een vector voorstelt.

grid<-data.frame(x=x0)

pc<-predict(lm1,interval="confidence",se=T,newdata=grid)

pp<-predict(lm1,interval="prediction",se=T,newdata=grid)

matlines(grid,pp\$fit,col="red")

matlines(grid,pc\$fit,col="blue")