

Sessie 2: parametrische testen

1variabele

Doe t.test of prop.test

Meerdere variabelen

a) 2 categorieën bv ja, nee met cijfers

stap 1: tapply om te gemiddelde te vergelijken

stap 2: boxplot om gemiddelde te vergelijken

stap 3: qqnorm om te testen of normaal verdeeld

stap 4: var.test om te testen of varianties gelijk zijn

stap 5: t.test met gelijke of ongelijke varianties

b) 2x2 categorieën bv airco ja/nee, verkoper part/prof met proporties

stap1: zijn zowel het #successen als #mislukkingen groter dan 5?

Stap 2: hoeveel % liggen de 2 proporties uit elkaar

Stap 3: doe prop.test met (c(k1,k2),c(n1,n2))

c) 3categorieën bv merk 1, merk2, merk3 met cijfers

stap 1: tapply om gemiddelde te vergelijken

stap 2: boxplot om gemiddelde te vergelijken

stap 3: qqnorm om te testen of normaal verdeeld

stap 4: testobs<-aov(), summary(testobj) om variantie te testen want dit kan niet met var.test door proporties

stap 5:post hoc analyse met tukey HSD

Sessie 3: regressie

Lineaire regressie

Stap 1: plot om te zien of model lineair is, zelf als het er niet zo uit ziet veronderstellen we het voor deze oef toch lineair.

Stap 2: maak model aan en definieer het als lineair. Model<-lm()

Stap 3: summary(model)

Stap 4: voorwaarden checken: qqplot, plot(model), is normaal verdeeld?, homoscedastisch?, uitschieters?

Stap 5: predict om waarden te schatten

Exponentieel regressiemodel

Doe exact hetzelfde maar vergeet niet alles log() te doen

Sessie 4: niet-parametrische testen

Testen voor verdeling

Chisq.test

Vragen zoals is deze steekproef representatief?

Je hebt een steekproef en de kansen, dus welke waarden er verwacht worden

Stap 1: test of iedere klasse min verwachte waarde 5 heeft

Stap2: `chisq.test(steekproef,kansen)`

Kolmogorov-Smirnov test

Volledig gespecificeerde continue verdeling

`Ks.test(tijden,"pexp",1/25)`

Simulatie doen: `set.seed(startwaarde)`

`X<-runif` en dan `ks.test` doen

Shapiro-wilkes test

Test voor normale verdeling

`Shapiro.test()`

Testen voor centrale waarden

1 variabele

Als we geen enkele andere test mogen doen, doen we deze

Stap 1: doe `qqnorm` en histogram om te checken of normaal verdeeld en symmetrisch

Stap 2: als niet norm verdeeld en niet symmetrisch tekentest `binom.test(wat aan test voldoet, alles,p=0.5)`

Stap 3: als symmetrisch, wilcoxon test, 1 variabele.

Meerdere variabelen

Stap 1: Vergelijk de histogrammen en qqplots om te weten of ongeveer gelijk verdeeld

Stap2 : `tapply` om medianen te vergelijken

Stap 3: `wilcox.test` als 2 variabelen, Kruskal-Wallis als meer dan 2

Testen voor onafhankelijkheid

Chisq test

Stap 1: test of de verwachte waarden minimum 5 zijn.

Stap2 : zo ja, doe `chisq.test`, zo nee, als tabel 2x2 is kan je fishertest doen, als niet 2x2 doe je toch `chisq` maar hou je rekening dat het niet helemaal mag.

Correlatietesten spearman of Kendall, afhankelijk van wat ze vragen.