Hvad skal vi med t-testet

Søren Højsgaard[*]

June 12, 2018

Contents

Hvad skal vi med t-testet?

Betragt en lineær regressionsmodel

$$y_i = \alpha + \beta x_i + e_i$$

Variansen på estimatet $\hat{\beta}$ er $\sigma^2_\beta = \sigma^2 c$, hvor σ^2 er residualvariansen, dvs. variansen på y_i'erne (eller på e_i'erne om man vil, det er det samme) og c er en konstant hvis værdi er let at regne ud, men vi skal ikke gøre dette her.

t–testet og z–testet

I introducerende statistikkurser lærer man, at når man skal teste hypotesen at $\beta = \beta_0$ (hvor β_0 er et givet tal) så skal man skelne mellem situationerne hvor σ^2 er kendt og ukendt (i praksis er σ^2 næsten aldrig kendt).

Hvis σ^2 er kendt, så ser vi på *teststørrelsen*

$$z = \frac{\hat{\beta} - \beta_0}{\sqrt{\sigma^2_\beta}} = \frac{\hat{\beta} - \beta_0}{\sigma_\beta}$$

Dvs. z "måler" hvor mange standardafvigelser estimatet $\hat{\beta}$ ligger fra β_0. Numerisk store værdier af z bevirker, at man tvivler på hypotesen.

så kan man se på en t–test størrelse

$$t = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{\sigma}^2_\beta}}$$

Dvs. t "måler" hvor mange standardafvigelser estimatet $\hat{\beta}$ ligger fra β_0. Numerisk store værdier af t bevirker, at man tvivler på hypotesen.

Hvis hypotesen er sand, så skal t vurderes i en t–fordeling med $N - 1$ frihedsgrader, hvor N er antal observationer. Numerisk store værdier af t er får os til at tvivle på hypotesen.

Hvis σ^2 er kendt (det er σ^2 næsten aldrig i praksis, men lad os lige lade som om) så er der ingen grund til at bruge et estimat for σ^2 ovenfor. I så fald bliver variansen på $\hat{\beta}$ givet ved $\sigma^2_\beta = \sigma^2 c$. I dette tilfælde bliver t–test størrelsens pendant

[*]University of Aalborg, Denmark

Hvis hypotesen er sand, så skal u vurderes i en $N(0,1)$–fordeling. Numerisk store værdier af t er får os til at tvivle på hypotesen.

I et introducerende statistikkursus vil man ofte lære, at hvis antallet af frihedsgrader f er stort så vil en t_f-fordeling (en t fordeling med f frihedsgrader) ligne en standard normal fordeling (en $N(0,1)$ fordeling) så meget, at man simpelthen kan vurdere t i en $N(0,1)$ fordeling. Dette svarer præcist til at lade som om at $\hat{\sigma}^2_\beta$ er den sand værdi for spredningen på $\hat{\beta}$.

Hvis man vurderer t i en $N(0,1)$ fordeling når der er ganske få frihedsgrader, så kan resultaterne derimod blive forkerte - meget forkert endda.

For at illustrere dette laver vi følgende tankeeksperiement: Vi skal sammenligne en behandling for en given sygdom med en placebo, så vi har to grupper med M patienter i hver. Dette kan håndteres med en lineær regressionsmodel: Lad x være en variabel der er 1 for de patienter, der har fået behandling og 0 for dem, der har fået placebo. Så vil middelværdien for patienter med placebo være α mens middelværdien for patienter med behandlingen være $\alpha + \beta$ så behandlingseffekten vil være β. Da vi laver eksperimentet selv, så står det os frit for at vælge β og vi sætter $\beta = 0$, således der ikke er en behandlingseffekt. Hvis vi tester hypotesen $\beta = 0$ på niveau 5% så vil vi med 5% sandsynlighed forkaste hypotesen (der jo er sand, for sådan er eksperimentet lavet). Lad os gentage eksperimentet 1000 gange. Så ville vi skulle forkaste hypotesen omkring 50 gange - ellers er der noget helt galt.

Hvis M (antal patienter per gruppe) er lille og vi laver et normalfordelingstest, så vil vi ikke få forkastet hypotesen i 5% af tilfældene men måske i 10% af tilfældene.

```
M <- 3
beta <-0
mu <- c(rep(0, M), rep(beta, M))
grp <- factor(c(rep("placebo", M), rep("behandling", M)))

y <- rnorm(2 * M, mean=mu)
tb <- summary(lm(y ~ grp))$coef
tb <- as.data.frame(tb)
prt <-2 * (1 - pt(abs(tb$`t value`), df=2 * M - 2))
prn <- 2 * (1 - pnorm(abs(tb$`t value`)))
c(prt[2], prn[2])
```

```
## [1] 0.1219 0.0503
```

```
do_sim <- function(){
  y <- rnorm(2 * M, mean=mu)
  tb <- summary(lm(y ~ grp))$coef
  tb <- as.data.frame(tb)
  prt <-2 * (1 - pt(abs(tb$`t value`), df=2 * M - 2))
  prn <- 2 * (1 - pnorm(abs(tb$`t value`)))
  c(prt[2], prn[2])
}

do_sim()
```

```
## [1] 0.6531 0.6278
```

```
Nsim <- 1000
sim <- replicate(Nsim, do_sim())

sum(sim[1,] <= 0.05) / Nsim
```

```
## [1] 0.055
```

```
sum(sim[2,] <= 0.05) / Nsim
```

```
## [1] 0.129
```

F-test og Wald-test

Et alternativt men ækvivalent test (giver samme resultat) er et F–test:

$$F = t^2 = \frac{(\hat{\beta} - \beta_0)^2}{\hat{\sigma}^2_\beta}$$

er under hypotesen F-fordelt med een tællerfrihedsgrad og $N-1$ nævnerfrihedsgrader.

$$W = u^2 = \frac{(\hat{\beta} - \beta_0)^2}{\sigma_\beta^2}$$

Dette kaldes et Wald–test. Forskellen mellem et Wald-test og et F-test er altså, at i det første er variansen kendt; i F-test tages der højde for at variansen er estimeret fra data og den usikkerhed i estimatet, der følger dermed.

–> –> –> –> –>

–>

–>

–> –> –> –>

–> –> –> –> –> –>

–>

–>

–> –> –> –>

–> –>

–> –> –> –> –> –>

–>

–> –> –>

–>

–> –>

–>

–> –> –> –>

–> –> –> –> –> –> –> –> –> –>

–> –> –> –> –> –> –> –> –> –>