02402: Introduktion til Statistik

Forelæsning 5: Hypotesetests

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Overview

Overview

Motiverende eksempel – sovemedicin

Forskel på sovemedicin?

I et studie er man interesseret i at sammenligne 2 sovemidler, A og B. For 10 testpersoner har man fået følgende resultater, der er givet i forlænget søvntid i timer (forskellen på effekten af de to midler er angivet):

Stikprøve, $n = 10$:

person	$x =$ B effect - A effect
1	1.2
2	2.4
3	1.3
4	1.3
5	0.9
6	1.0
7	1.8
8	0.8
9	4.6
10	1.4

$\bar{x} = 1.67$ (gennemsnit)
$s = 1.13$ (standardafvigelse)

Eksempel – sovemedicin

Hypotesen om ingen forskel:

$$H_0: \; \mu = 0$$

hvor μ er den gennemsnitlige forskel i søvnlængde ("effekten").

Stikprøvegennemsnit og -standardafvigelse:

$$\bar{x} = 1.670 = \hat{\mu}$$
$$s = 1.13 = \hat{\sigma}$$

Er data i overensstemmelse med nulhypotesen H_0?

Data: $\bar{x} = 1.67$, $H_0: \; \mu = 0$

NYT: p-værdi:

$$p\text{-værdi} = 0.00117$$

(Udregnet i scenariet at H_0 er sand).

NYT: Konklusion:

Eftersom data er usandsynligt under H_0, så **forkaster** vi H_0. Der er en **signifikant effekt** af middel B sammenlignet med middel A.

Overview

Method 3.23: One-sample t-test og p-værdi

Hvordan beregner vi p-værdien?

For en (kvantitativ) situation med **én stikprøve**, er p-værdien givet ved:

$$p\text{-værdi} = 2 \cdot P(T > |t_{\text{obs}}|)$$

hvor T følger en t-fordeling med $(n-1)$ frihedsgrader.
Den observerede værdi af teststørrelsen, som skal udregnes, er

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

hvor μ_0 er værdien af μ under nulhypotesen:

$$H_0: \; \mu = \mu_0$$

Definition og fortolkning af p-værdien (HELT generelt)

p-værdien udtrykker *evidens* imod nulhypotesen – Table 3.1:

$p < 0.001$	Very strong evidence against H_0
$0.001 \le p < 0.01$	Strong evidence against H_0
$0.01 \le p < 0.05$	Some evidence against H_0
$0.05 \le p < 0.1$	Weak evidence against H_0
$p \ge 0.1$	Little or no evidence against H_0

Definition 3.22 af p-værdien:

p-værdien er sandsynligheden for at observe en testsstørrelse som er **mindst lige så ekstrem** som den faktiske observerede testværdi. Denne sandsynligheden udregnes under antagelse om at nulhypotesen er sand.

Eksempel – sovemedicin

Hypotesen om ingen forskel:

$$H_0: \ \mu = 0$$

hvor μ er den gennemsnitlige forskel i søvnlængde.

Udregn testværdien:

$$t_{\text{obs}} = \frac{1.67 - 0}{1.13/\sqrt{10}} = 4.67$$

Beregn p-værdien:

$$2P(T > 4.67) = 0.00117$$

```
2 * (1 - pt(4.67, df = 9))
```

Fortolkningen af p-værdien ud fra Tabel 3.1:

Der er stærk evidens mod nulhypotesen.

Eksempel – sovemedicin – in R, manuelt

```
# Enter data
x <- c(1.2, 2.4, 1.3, 1.3, 0.9, 1.0, 1.8, 0.8, 4.6, 1.4)
n <- length(x) # sample size

# Compute 'tobs' - the observed test statistic
tobs <- (mean(x) - 0) / (sd(x) / sqrt(n))

# Compute the p-value as a tail-probability
# in the relevant t-distribution:
2 * (1 - pt(abs(tobs), df = n-1))

## [1] 0.001166
```


Eksempel – sovemedicin – in R, med indbygget funktion

```
t.test(x)

##
##  One Sample t-test
##
## data:  x
## t = 4.7, df = 9, p-value = 0.001
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.8613 2.4787
## sample estimates:
## mean of x
##      1.67
```


Definition af hypotesetest og signifikans (HELT generelt)

Definition 3.24. Hypotesetest:

Vi siger at vi *udfører et hypotesetest*, når vi vælger at afvise eller godkende en nulhypotese eller ej ud fra data.

En nulhypotese *afvises*, hvis p-værdien udregnet efter data er blevet observeret er mindre end α, dvs. hvis p-værdien $< \alpha$, hvor α er et på forhånd valgt *signifikansniveau*.

Ellers siges nulhypotesen at være *'godkendt'*.

Definition 3.29. Statistisk signifikans:

En *effekt* siges at være *(statistisk) signifikant* hvis p-værdien er mindre end signifikansniveauet α.

OFTEST bruges $\alpha = 0.05$.

Eksempel – sovemedicin

Med $\alpha = 0.05$ kan vi konkludere:

Idet p-værdien er mindre end α, så **forkaster** vi nulhypotesen.

Og dermed:

Vi har påvist en **signifikant effekt** af middel B sammenlignet med middel A. (Og dermed at B virker bedre end A).

Overview

Kritisk værdi

Definition 3.31 - kritiske værdier for t-testet:

$(1-\alpha)100\%$ kritiske værdier for det (tosidede) one-sample t-testet er $(\alpha/2)100\%$- og $(1-\alpha/2)100\%$ fraktilerne i t-fordelingen med $n-1$ frihedsgrader:

$$t_{\alpha/2} \text{ og } t_{1-\alpha/2}$$

Method 3.32: One-sample t-test ved kritiske værdier:

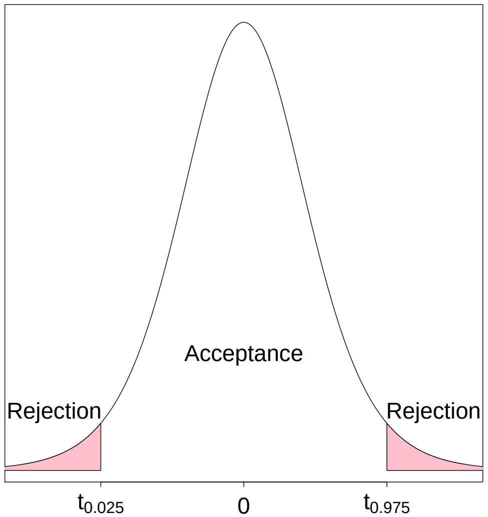
En nulhypotese *afvises* hvis den observerede teststørrelse er mere ekstrem end de kritiske værdier:

$$\text{Hvis } |t_{\text{obs}}| > t_{1-\alpha/2} \text{ så } \textit{forkast}$$

ellers *acceptér*.

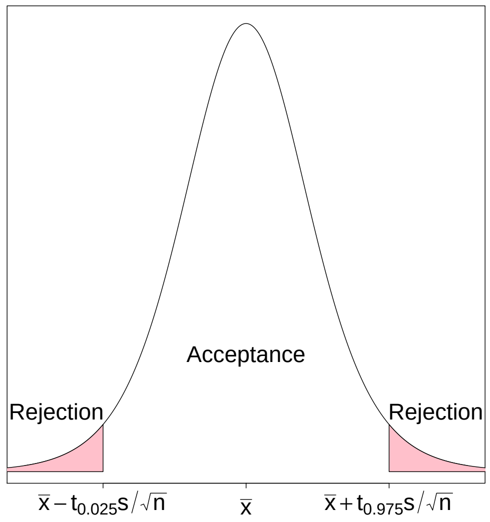
Kritisk værdi og hypotesetest

Acceptområdet består af de værdier af μ, som ikke er for langt væk fra stikprøvegennemsnittet – her på standardiseret skala:



Kritisk værdi og hypotesetest

Acceptområdet består af de værdier af μ, som ikke er for langt væk fra stikprøvegennemsnittet – nu på den oprindelige skala:



Kritisk værdi, konfidensinterval og hypotesetest

Theorem 3.33: Kritisk-værdi-metode = Konfidensinterval-metode

Vi betragter et $(1-\alpha)\cdot 100\%$ konfidensinterval for μ:

$$\bar{x} \pm t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Konfidensintervallet svarer til acceptområdet for H_0, når man tester (tosidede) hypotesen

$$H_0: \quad \mu = \mu_0$$

(Ny) fortolkning af konfidensintervallet:

Konfidensintervallet indeholder de værdier, som vi tror på i de givne data. (De værdier som ville accepteres ved det tilsvarende hypotesetest.)

Bevis:

Remark 3.34

Et μ_0 inden for konfidensintervallet opfylder at

$$|\bar{x} - \mu_0| < t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

hvilket er ækvivalent med:

$$\frac{|\bar{x} - \mu_0|}{\frac{s}{\sqrt{n}}} < t_{1-\alpha/2}$$

og igen med:

$$|t_{\text{obs}}| < t_{1-\alpha/2}$$

hvilket netop siger at μ_0 accepteres, idet t_{obs} er inden for de kritiske værdier.

Overview

Den alternative hypotese

Indtil nu – underforstået at testet er todsidet: (= non-directional)
Alternativet til $H_0: \ \mu = \mu_0$ er $H_1: \ \mu \neq \mu_0$.

MEN der kan være andre situationer, e.g. ensidet (= directional), "less":
Alternativet til $H_0: \ \mu = \mu_0$ er $H_1: \ \mu < \mu_0$.

Vi holder os til det tosidede test ('retningsløse') i dette kursus!

Steps ved hypotesetests – et overblik

Helt generelt består et hypotesetest af følgende trin:

1. Formulér hypotesen og vælg signifikansniveau α (vælg "risk-level").
2. Udregn, ud fra de observerede data, værdien af teststørrelsen.
3. Udregn p-værdien ud fra teststørrelsen holdt op imod den rette fordeling. Sammenlign p-værdien med signifikansniveauet α og konkludér.

   ELLER:

   Alternativt, konkludér ud fra de relevante kritiske værdier.

One-sample t-testet igen

Method 3.36 Et niveau α one-sample t-test:

1. Udregn t_{obs} som før.
2. Beregn evidensen mod *nulhypotesen* $H_0: \ \mu = \mu_0$ vs. den *alternative hypotese* $H_1: \ \mu \neq \mu_0$ ved

   $$p\text{–værdien} = 2 \cdot P(T > |t_{\text{obs}}|)\,,$$

   hvor t-fordelingen med $n-1$ frihedsgrader benyttes.
3. If p-værdien $< \alpha$, så afvises H_0. Ellers så godkendes H_0.

   ELLER: Alternativt, men ækvivalent, kan konklusionen (godkend/afvis) baseres på de kritiske værdier $\pm t_{1-\alpha/2}$:

   Hvis $|t_{\text{obs}}| > t_{1-\alpha/2}$ så afvises H_0, ellers accepteres H_0.

Fejlslutninger ved hypotesetests

Der findes to slags fejl (dog kun én af gangen!)
Type I: Afvisning af H_0 når H_0 er sand.
Type II: Ikke afvisning/godkendelse af H_0 når H_1 er sand.

Risikoen for de to typer fejl kaldes sædvanligvis:

$$P(\text{Type I fejl}) = \alpha$$

$$P(\text{Type II fejl}) = \beta$$

Retssalsanalogi

En person står stillet for en domstol:

En mand er stillet for en dommer, anklaget for en forbrydelse
Nulhypotesen og den alternative hypotese er:

H_0 : Den anklagede er uskyldig.

H_1 : Den anklagede er skyldig.

At man ikke kan bevises skyldig er ikke det samme som at man er bevist uskyldig:

Sagt på en anden måde:
Accept af en nulhypotese er IKKE et statistisk bevis for at nulhypotesen er sand!

Fejlslutninger ved hypotesetests

Theorem 3.39: Signifikansniveauet = Risikoen for Type I fejl

Signifikansniveauet α i hypotesetests er risikoen for en Type I fejl:

$$P(\text{Type I fejl}) = P(\text{Afvisning of } H_0 \text{ når } H_0 \text{ er sand}) = \alpha$$

To mulige sandheder vs. to mulige konklusioner:

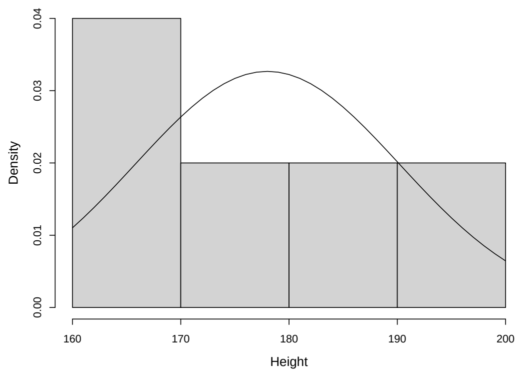
	Afviser H_0	Afviser ikke H_0
H_0 er sand	Type I fejl (α)	Korrekt godkendelse of H_0
H_0 er falsk	Korrekt afvisning of H_0 (Power)	Type II fejl (β)

$$\text{Mindre } \alpha = \text{ større } \beta \text{ (og omvendt)}$$

Overview

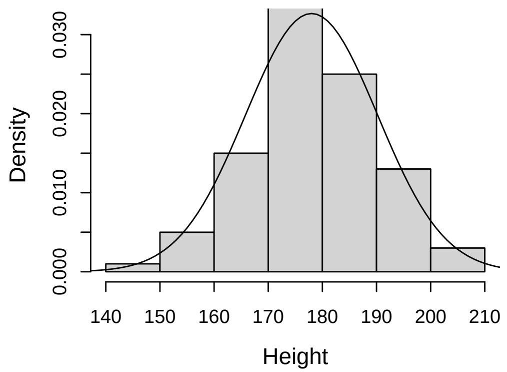
Eksempel – højde på studerende

```r
# Student heights data
x <- c(168, 161, 167, 179, 184, 166, 198, 187, 191, 179)

# Density histogram of student height data together with normal pdf
hist(x, xlab = "Height", main = "", freq = FALSE)
lines(seq(160, 200, 1), dnorm(seq(160, 200, 1), mean(x), sd(x)))
```



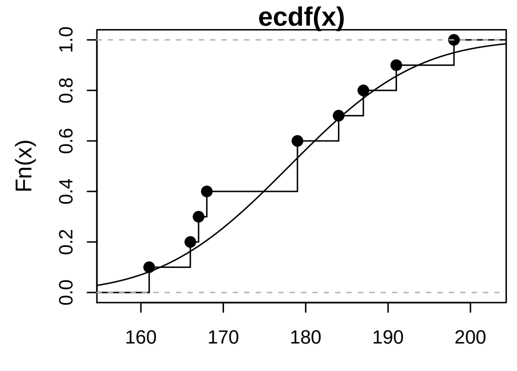
Eksempel – 100 observationer fra en normalfordeling

```
# Density histogram of simulated data from normal distribution
# (n = 100) together with normal pdf
xr <- rnorm(100, mean(x), sd(x))
hist(xr, xlab = "Height", main = "", freq = FALSE, ylim = c(0, 0.032))
lines(seq(130, 230, 1), dnorm(seq(130, 230, 1), mean(x), sd(x)))
```



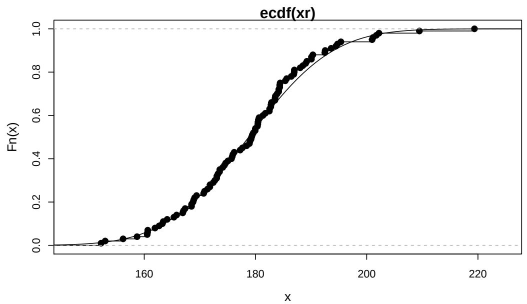
Eksempel – højder på studerende – ecdf

```
# Empirical cdf for student height data together
# with normal cdf
plot(ecdf(x), verticals = TRUE)
xp <- seq(0.9*min(x), 1.1*max(x), length.out = 100)
lines(xp, pnorm(xp, mean(x), sd(x)))
```



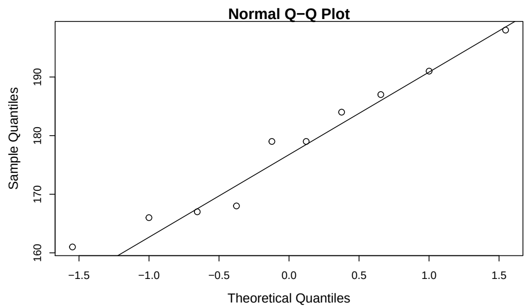
Eksempel – 100 observationer fra en normalfordeling – ecdf

```
# Empirical cdf of simulated data from normal distribution
# (n = 100) together with normal cdf
xr <- rnorm(100,  mean(x), sd(x))
plot(ecdf(xr), verticals = TRUE)
xp <- seq(0.9*min(xr), 1.1*max(xr), length.out = 100)
lines(xp, pnorm(xp, mean(xr), sd(xr)))
```



Eksempel – højder på studerende – q-q plot

```
# Normal q-q plot of student heights
qqnorm(x)
qqline(x)
```



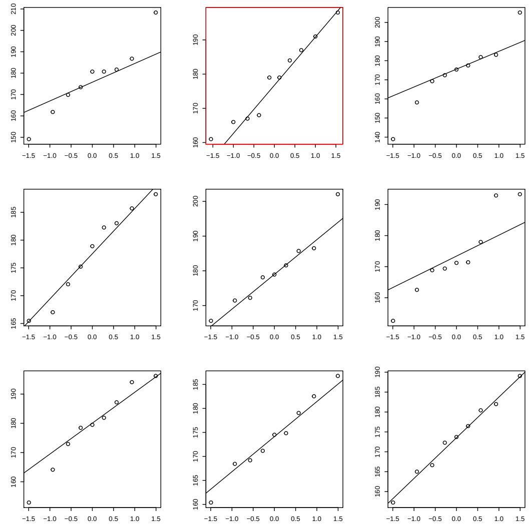
QQ-plot for normalfordeling

Metode 3.42- formel definition

De sorterede observationer, $x_{(1)},\ldots,x_{(n)}$ plottes mod en de teoretiske fraktiler i normalfordelingen. De findes forskellige definitioner af $p_1,\ldots,p_n$:

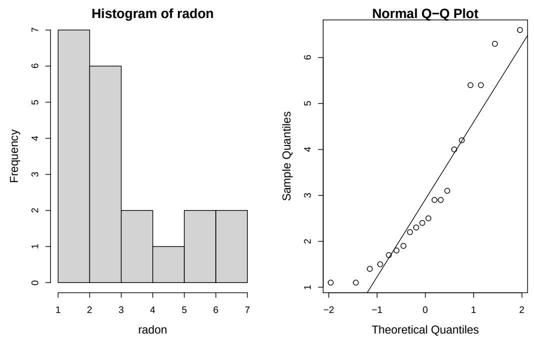
- I R, når $n > 10$:
$$p_i = \frac{i-0.5}{n},\ i = 1,\ldots,n$$
- I R, når $n \leq 10$:
$$p_i = \frac{i-3/8}{n+1/4},\ i = 1,\ldots,n$$

Eksempel – højder på studerende – sammenlign med andre simulerede data



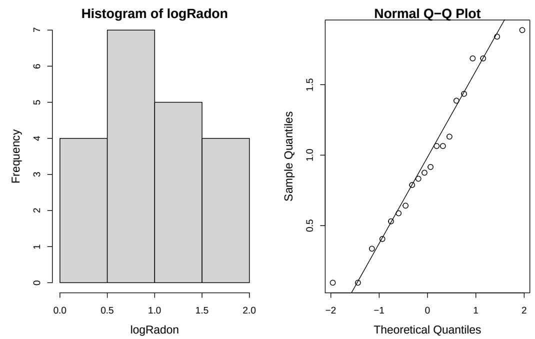
Eksempel – radon-data

```r
## Reading in the data
radon <- c(2.4, 4.2, 1.8, 2.5, 5.4, 2.2, 4.0, 1.1, 1.5, 5.4, 6.3,
       1.9, 1.7, 1.1, 6.6, 3.1, 2.3, 1.4, 2.9, 2.9)

## Histogram and q-q plot of data
par(mfrow = c(1,2))
hist(radon)
qqnorm(radon)
qqline(radon)
```



Eksempel – radon-data – log-transformede data er tættere på en normalfordeling

```r
# Transform data using the natural logarithm
logRadon<-log(radon)

## Histogram and q-q plot of transformed data
par(mfrow = c(1,2))
hist(logRadon)
qqnorm(logRadon)
qqline(logRadon)
```



Agenda

1. Motiverende eksempel – sovemedicin
2. One-sample t-test og p-værdi
3. Kritisk værdi og konfidensinterval
4. Hypotesetests – generelt
   - Den alternative hypotese
   - Den generelle metode
   - Fejlslutninger ved hypotesetests!
5. Modelkontrol: undersøge normalfordelingsantagelsen
   - Q-Q plot for normalfordeling
   - Transformation mod normalitet