# 02402 Statistik (Polyteknisk grundlag)

## Uge 6: Analyser med to stikprøver

Nicolai Siim Larsen
DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby



## Dagsorden





Hjælp os med at forbedre kurset:

Brug 5 minutter på at udfylde kursets midtvejsevaluering på DTU Inside!





## Overview

## Hypotesetest

Med udgangspunkt i et forskningsspørgsmål opstilles hypotesetestens grundrammer:

1. Definition og afgrænsning af populationen
2. Formulering af nulhypotesen (og modhypotesen)
3. Fastsættelse af signifikansniveauet (og teststyrken)

Baseret på kravene til hypotesetesten og en statistisk model opstilles et eksperiment (forsøg), hvorfra der udtages en eller flere repræsentative stikprøver:

- *Beregning af de nødvendige stikprøvestørrelser (Hvis muligt)*
- Undersøgelse af modellens antagelser

På baggrund af observationerne og den statistiske model udvælges en passende test. Testen kan evalueres på flere måder ved at sammenligne:

- Et konfidensinterval med parameterværdien under nulhypotesen
- En teststørrelse med kritiske værdier
- En $p$-værdi med signifikansniveauet





## Undersøgelse af modellens fordelingsantagelse

Man udleder teststørrelser på baggrund af nogle fordelingsantagelser. Man kan undersøge om fordelingsantagelserne er opfyldt på flere måder, bl.a. ved brug af:

- Histogrammer: Man grupperer observationerne og sammenligner gruppehyppighederne med en teoretisk tæthedsfunktion baseret på estimerede parameterværdier.
- Den empiriske fordelingsfunktion: Man tegner den empiriske fordelingsfunktion og sammenligner den med en teoretisk fordelingsfunktion baseret på estimerede parameterværdier.
- QQ-plot: Man sammenholder fraktiler fra den empiriske og en teoretisk fordeling.

Man kan sammenholde med plots baseret på simulerede data fra den teoretiske fordeling.





## QQ-plot

Man sammenligner fraktiler fra den empiriske fordeling med fraktiler fra den teoretiske fordeling. Hvis fordelingerne er ens, vil fraktilerne ligge på linjen $y = x$. Et QQ-plot kan bl.a. bruges til at vurdere:

- Fordelingens spredning (haler)
- Fordelingens skævhed (skewness)

Man kan teste mod alle fordelinger, men fraktilerne er kun unikt bestemt for absolut kontinuerte fordelinger.

Et normal QQ-plot er et QQ-plot, hvor den teoretiske fordeling er en normalfordeling. Normalfordelingen har nogle specielle egenskaber, som medfører, at man blot kan sammenligne med fraktiler fra en standardnormalfordeling. Få andre fordelinger, som f.eks. eksponentialfordelingen, har lignende egenskaber.





## QQ-plot

**Normalfordeling**
Lad $X \sim N(\mu, \sigma^2)$ med fraktiler $q_p$ for $p \in (0,1)$ og lad $Y \sim N(0,1^2)$ med fraktiler $z_p$ for $p \in (0,1)$. Så har man, at

$$p = \mathbb{P}(X \leq q_p) = \mathbb{P}(\sigma X^* + \mu \leq q_p) = \mathbb{P}\left(X^* \leq \frac{q_p - \mu}{\sigma}\right).$$

Da $X^*$ og $Y$ har samme fordeling, må $z_p = (q_p - \mu)/\sigma$.

**Eksponentialfordeling**
Lad $X \sim \text{Eks}(\lambda)$ med fraktiler $q_p$ for $p \in (0,1)$ og lad $Y \sim \text{Eks}(1)$ med fraktiler $f_p$ for $p \in (0,1)$. Så har man, at

$$p = \mathbb{P}(X \leq q_p) = \mathbb{P}(\lambda X \leq \lambda q_p).$$

Da $\lambda X$ og $Y$ har samme fordeling, må $f_p = \lambda q_p$.

## Links om QQ-plot

To gode links med diskussioner om normal QQ-plot:

https://stats.stackexchange.com/questions/101274/how-to-interpret-a-qq-plot

https://stats.stackexchange.com/questions/22258/what-is-the-use-of-the-line-produced-by-qqline-in-r

Hentet 2. oktober 2023.





## Overview







## Motiverende eksempel: Ernæringsstudie

**Forskel på energiforbrug?**

I et ernæringsstudie ønsker man at undersøge, om der er en forskel i energiforbruget for forskellige typer (moderat fysisk krævende) arbejde.

I studiet har 9 sygeplejersker fra hospital $A$ og 9 (andre) sygeplejersker fra hospital $B$ fået målt deres energiforbrug. Målingerne ses i følgende tabel (i enheden megajoule MJ):

Stikprøve fra hvert hospital: $n_1 = n_2 = 9$:

| Hospital A | Hospital B |
|---|---|
| 7.53 | 9.21 |
| 7.48 | 11.51 |
| 8.08 | 12.79 |
| 8.09 | 11.85 |
| 10.15 | 9.97 |
| 8.40 | 8.79 |
| 10.88 | 9.69 |
| 6.13 | 9.68 |
| 7.90 | 9.19 |





## Eksempel: Ernæringsstudie

**Hypotesen om ingen forskel (i det gennemsnitlige energiforbrug) ønskes undersøgt:**

$$H_0: \ \mu_A = \mu_B$$

**Stikprøvegennemsnit og -standardafvigelser:**

$$\hat{\mu}_A = \bar{x}_A = 8.293 \ (s_A = 1.428)$$
$$\hat{\mu}_B = \bar{x}_B = 10.298 \ (s_B = 1.398)$$

**Er data i overensstemmelse med nulhypotesen $H_0$?**

Data: $\bar{x}_B - \bar{x}_A = 2.005$

Nulhypotese: $H_0: \ \mu_B - \mu_A = 0$

NYT: **$p$-værdi for forskel:**

$$p = 0.0083$$

(Beregnet under antagelsen, at $H_0$ er sand.)

NYT: **Konfidensinterval for forskellen**:

$$2.005 \pm 1.412 = [0.59; \ 3.42]$$

## Overview







## Metode 3.49: Teststørrelsen i en (Welch) $t$-test med to uparrede stikprøver

**Forudsætninger:**

Testen gælder, når begge stikprøver er store, eller når begge stikprøver kommer fra normalfordelte populationer.

**Beregning af den observerede teststørrelse:**

Vi betragter følgende nulhypotese om forskellen i middelværdi mellem to *uafhængige* og *uparrede* stikprøver: (Bemærk fejl i bogen)

$$\delta = \mu_1 - \mu_2,$$

$$H_0: \ \delta = \delta_0,$$

så er den observerede teststørrelse:

$$t_{\text{obs}} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}.$$





## Sætning 3.50: Fordelingen af teststørrelsen

**Teststørrlsen er (tilnærmelsesvis) $t$-fordelt:**

Under nulhypotesen er teststørrlsen:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}.$$

Den følger tilnærmelsesvis en $t$-fordeling med $\nu$ frihedsgrader, hvor

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}},$$

hvis de to populationer er normalfordelte eller stikprøvestørrelserne er tilstrækkelig store.





## Eksempel: Ernæringsstudie

**Hypotesen om ingen forskel ønskes undersøgt:**

$$H_0: \ \delta = \mu_B - \mu_A = 0$$

mod det tosidede alternativ:

$$H_1: \ \delta = \mu_B - \mu_A \neq 0.$$

**Først beregnes $t_{\text{obs}}$ og $\nu$:**

$$t_{\text{obs}} = \frac{10.298 - 8.293}{\sqrt{2.0394/9 + 1.954/9}} = 3.01$$

og

$$\nu = \frac{\left(\frac{2.0394}{9} + \frac{1.954}{9}\right)^2}{\frac{(2.0394/9)^2}{8} + \frac{(1.954/9)^2}{8}} = 15.99$$

# Eksempel: Ernæringsstudie

Dernæst findes $p$-værdien:

$$p = 2 \cdot P(T > |t_{\text{obs}}|) = 2P(T > 3.01) = 2 \cdot 0.00415 = 0.0083$$

```
## Eksempel - Ernæringsstudie: P(T > 3.01)
1 - pt(3.01, df = 15.99)

## [1] 0.004154
```

Vurdér evidensen (Tabel 3.1):

Der er stærk evidens imod nulhypotesen.

Konklusion baseret på et 5%-signifikansniveau ($\alpha = 0.05$):

Vi forkaster nulhypotesen. Der er signifikant forskel på de to grupper – sygeplejersker på Hospital B kan siges at have et større (middel)energiforbrug end sygeplejesker på Hospital A.





# Overview







# Metode 3.47: Konfidensinterval for $\mu_1 - \mu_2$

Konfidensintervallet for forskellen i middelværdi:

For to stikprøver $(x_1, \ldots, x_{n_1})$ og $(y_1, \ldots, y_{n_2})$ er $(1-\alpha)$-konfidensintervallet for $\mu_1 - \mu_2$ givet ved:

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

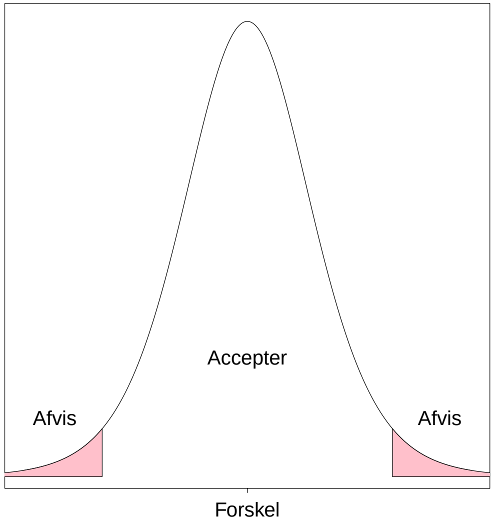
hvor $t_{1-\alpha/2}$ er $(1-\alpha/2)$-fraktilen i $t$-fordelingen med $\nu$ frihedsgrader (givet i sætning 3.50).





# Konfidensinterval og hypotesetest (Repetition)

Acceptområdet er de mulige værdier for $\mu_1 - \mu_2$, som ikke ligger for langt væk fra data:



## Eksempel: Ernæringsstudie

Lad os finde 95%-konfidensintervallet for $\mu_B - \mu_A$. Med $\nu = 15.99$ er den relevante $t$-fraktil givet ved

$$t_{0.975} = 2.120,$$

så konfidensintervallet bliver

$$10.298 - 8.293 \pm 2.120 \cdot \sqrt{\frac{2.0394}{9} + \frac{1.954}{9}}.$$

Udregnet giver dette:

$$[0.59; 3.42].$$





## Eksempel: Ernæringsstudie – Det hele i R:

```
# Indlæs data
xA = c(7.53, 7.48, 8.08, 8.09, 10.15, 8.4, 10.88, 6.13, 7.9)
xB = c(9.21, 11.51, 12.79, 11.85, 9.97, 8.79, 9.69, 9.68, 9.19)

# Udfør t-test med to uparrede stikprøver
t.test(xB, xA)

##
##  Welch Two Sample t-test
##
## data:  xB and xA
## t = 3, df = 16, p-value = 0.008
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5923 3.4166
## sample estimates:
## mean of x mean of y
##    10.298     8.293
```





## Overview



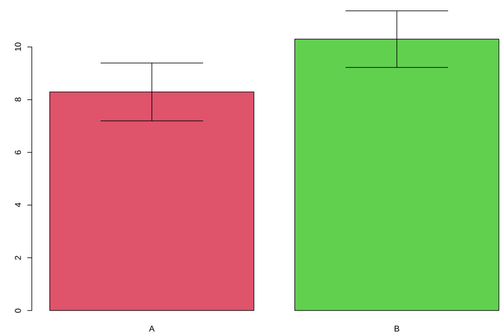




## Eksempel: Ernæringsstudie – Præsentation af resultater

Søjlediagrammer med *fejlbjælker* ses ofte:

Et søjlediagram med nogle fejlbjælker (error bars): Herunder vises 95%-konfidensintervallerne for hver gruppe:



## Vær varsom med at bruge "overlappende konfidensintervaller"

**Man bruger den forkerte variation til at vurdere forskellen:**

$$\sigma_{(\bar{X}_A - \bar{X}_B)} \neq \sigma_{\bar{X}_A} + \sigma_{\bar{X}_B}$$
$$\text{Var}(\bar{X}_A - \bar{X}_B) = \text{Var}(\bar{X}_A) + \text{Var}(\bar{X}_B)$$

Antag at de to standardafvigelser er 3 og 4: Summen er 7, men $\sqrt{3^2+4^2} = 5$.

**Det korrekte forhold mellem standardafvigelserne er således:**

$$\sigma_{(\bar{X}_A - \bar{X}_B)} < \sigma_{\bar{X}_A} + \sigma_{\bar{X}_B}.$$





## Vær varsom med at bruge "overlappende konfidensintervaller"

**Bemærkning 3.59 – Regel for brug af "overlappende konfidensintervaller":**

Når to konfidensintervaller IKKE overlapper: De to grupper er signifikant forskellige.

Når to konfidensintervaller overlapper: Ingen konklusion kan drages uden at undersøge konfidensintervallet for *forskellen* mellem grupperne.





## Overview







## Motiverende eksempel: Sovemedicin

**Forskel på sovemedicin?**

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

**Stikprøve med $n = 10$:**

| Person | $A$ | $B$ | $D = B - A$ |
|---|---|---|---|
| 1 | +0.7 | +1.9 | +1.2 |
| 2 | -1.6 | +0.8 | +2.4 |
| 3 | -0.2 | +1.1 | +1.3 |
| 4 | -1.2 | +0.1 | +1.3 |
| 5 | -1.0 | -0.1 | +0.9 |
| 6 | +3.4 | +4.4 | +1.0 |
| 7 | +3.7 | +5.5 | +1.8 |
| 8 | +0.8 | +1.6 | +0.8 |
| 9 | 0.0 | +4.6 | +4.6 |
| 10 | +2.0 | +3.4 | +1.4 |

$\bar{x} = 1.67$
$s = 1.13$

# Analyse af to parrede stikprøver: Parret $t$-test

```
# Indlæs data
x1 = c(.7,-1.6,-.2,-1.2,-1,3.4,3.7,.8,0,2)
x2 = c(1.9,.8,1.1,.1,-.1,4.4,5.5,1.6,4.6,3.4)

# Udregn forskellene
dif = x2 - x1

# Udfør en parret t-test
t.test(dif)

##
##  One Sample t-test
##
## data:  dif
## 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
```





# Analyse af to parrede stikprøver: Parret $t$-test

```
# En anden måde at udføre en parret t-test
t.test(x2, x1, paired = TRUE)

##
##  Paired t-test
##
## data:  x2 and x1
## t = 4.7, df = 9, p-value = 0.001
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.8613 2.4787
## sample estimates:
## mean of the differences
##                    1.67
```





# Forsøgsopsætning: Parrede og uafhængige stikprøver

**Fuldstændigt tilfældigt (uafhængige stikprøver):**

Vi har 20 patienter, som tilfældigt fordeles på to grupper (normalt lige mange i hver gruppe). Dvs. at der er forskellige (uafhængige) patienter i de to grupper.

**Parrede observationer (afhængige stikprøver):**

Vi har 10 patienter, som alle får begge behandlinger (typisk med noget tid imellem og med tilfældig rækkefølge af behandlingerne).
Dvs. de samme patienter fremgår i de to grupper.





# Eksempel: Sovemedicin – FORKERT analyse

```
# Forkert analyse (t-test for to uparrede stikprøver)
t.test(x1, x2)

##
##  Welch Two Sample t-test
##
## data:  x1 and x2
## t = -1.9, df = 18, p-value = 0.07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.4854  0.1454
## sample estimates:
## mean of x mean of y
##      0.66      2.33
```

## Overview

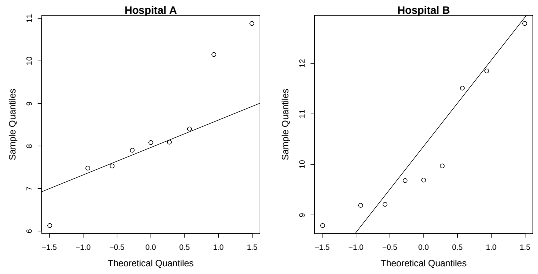
1. Opsummering
2. Motiverende eksempel: Ernæringsstudie
3. $t$-test med to uparrede stikprøver
4. Konfidensintervallet for forskellen i middelværdi
5. Overlappende konfidensintervaller?
6. $t$-test med to parrede stikprøver (parret $t$-test)
7. **Normalfordelingesantagelserne**
8. Styrke og stikprøvestørrelse – Forsøgsdesign
   - Krav til præcision
   - Styrke og stikprøvestørrelse – En stikprøve
   - Styrke og stikprøvestørrelse – To stikprøver
9. $t$-test med sammenvejet varians – Et alternativ





## Eksempel: QQ-plot for hver af stikprøverne:

```r
# QQ plot for hver stikprøve
qqnorm(xA, main = "Hospital A")
qqline(xA)
qqnorm(xB, main = "Hospital B")
qqline(xB)
```

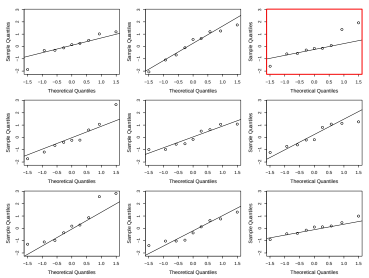







## Eksempel: Sammenligning med simulerede data, Hospital A

```r
# Flere simulerede QQ plot
require(MESS)
fitA <- lm(xA ~ 1)
qqnorm.wally <- function(x, y, ...) { qqnorm(y, ...); qqline(y, ...)}
wallyplot(fitA, FUN = qqnorm.wally, main = "")
```

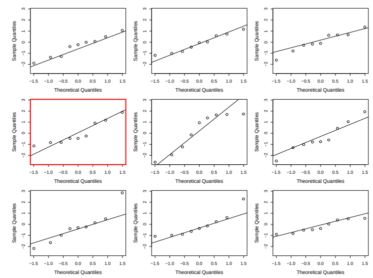






## Eksempel: Sammenligning med simulerede data, Hospital B

```r
# Flere simulerede QQ plot
fitB <- lm(xB ~ 1)
qqnorm.wally <- function(x, y, ...) { qqnorm(y, ...); qqline(y, ...)}
wallyplot(fitB, FUN = qqnorm.wally, main = "")
```



## Overview







## Forsøgsplanlægning med krav til præcisionen

*Fejlmarginen* (margin of error - ME) er defineret som

$$t_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

**Metode 3.63: Stikprøvestørrelse for konfidensintervallet baseret på en stikprøve**

Hvis $\sigma$ er kendt, eller vurderet til at være en bestemt værdi, så kan vi beregne den stikprøvestørrelse, som kræves for at opnå en given fejlmargin, med sandsynlighed $1-\alpha$.

$$n = \left(\frac{z_{1-\alpha/2}\cdot\sigma}{ME}\right)^2.$$





## Eksempel: Højde på studerende

**Stikprøvemiddelværdien og -standardafvigelsen:**

$$\bar{x} = 178$$
$$s = 12.21$$

**Estimater for populationens middelværdi og standardafvigelse:**

$$\hat{\mu} = 178$$
$$\hat{\sigma} = 12.21$$

**Hvis vi ønsker en fejlmargin på 3 cm på et 5%-signifikansniveau, hvor stor skal stikprøvestørrelsen $n$ så være?**

$$n = \left(\frac{1.96\cdot 12.21}{3}\right)^2 = 63.64 \approx 64.$$





## Forsøgsplanlægning: Styrke

**Hvad er styrken af et fremtidigt eksperiment:**

- Sandsynligheden for at detektere en (påstået) effekt.
- $P(H_0$ afvises) når $H_1$ er sand.
- Sandsynligheden for en korrekt afvisning of $H_0$.
- MEN: En nulhypotese kan være forkert på mange måder!
- I praksis: Brug en scenarie-baseret tilgang
  - F.eks. "Hvis $\mu = 86$, hvor sikkert vil mit forsøg være i stand til at detektere dette?"
  - F.eks. "Hvis $\mu = 84$, hvor sikkert vil mit forsøg være i stand til at detektere dette?"

## Forsøgsplanlægning: Styrke

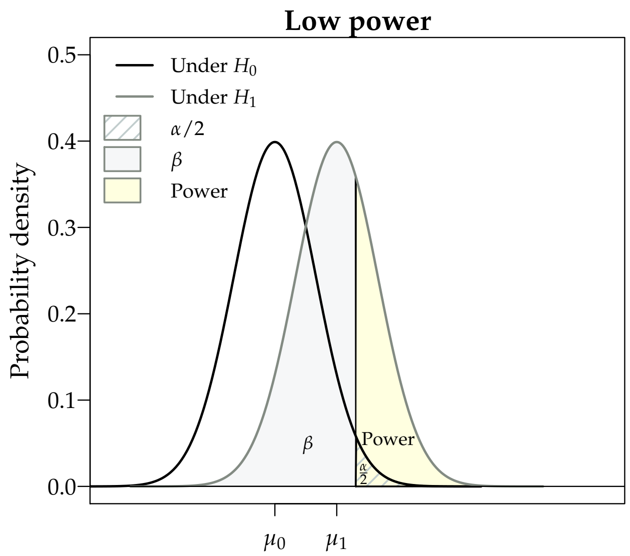
Hvis vi kender (eller antager) fire ud af de fem følgende størrelser, så kan vi finde den manglende:

- Stikprøvestørrelsen (sample size), $n$.
- Signifikansniveauet, $\alpha$, som vi tester på.
- Forskellen i middelværdi (effekt-størrelsen), $\mu_0 - \mu_1$.
- Populationsstandardafvigelsen, $\sigma$.
- Styrken (power), $1 - \beta$.





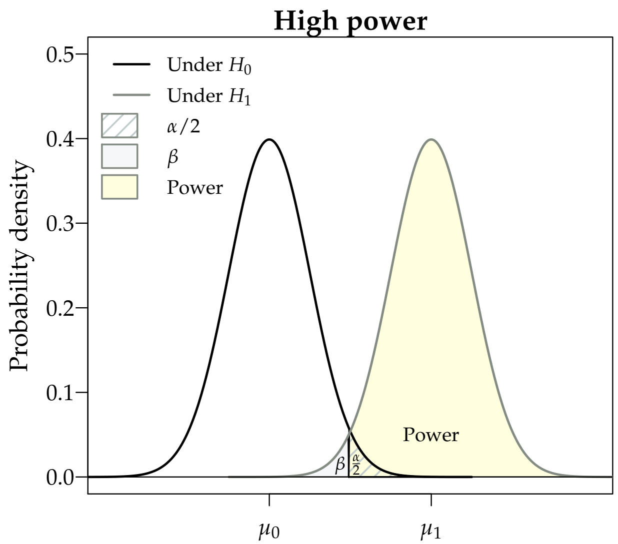
## Eksempel med lav styrke








## Eksempel med høj styrke








## Forsøgsplanlægning: Stikprøvestørrelsen $n$

**Det store spørgsmål: Hvor stort skal $n$ være?**

Vi skal have nok observationer til at kunne detektere en relevant effekt med høj styrke $1-\beta$ (typisk mindst 80%):

**Metode 3.65: Formel for stikprøvestørrelse med en stikprøve**

For en t-test med en stikprøve, hvor $\alpha$, $\beta$ og $\sigma$ er givet:

$$n = \left( \sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_0 - \mu_1} \right)^2 .$$

Her er $\mu_0 - \mu_1$ den forskel i middelværdi, som vi ønsker at måle, medens $z_{1-\beta}$ og $z_{1-\alpha/2}$ er fraktiler i standardnormalfordelingen.

# Eksempel: Styrken når $n = 40$

```
# Udregning af styrken (en stikprøve)
power.t.test(n = 40, delta = 4, sd = 12.21, type = "one.sample")

##
##      One-sample t test power calculation
##
##               n = 40
##           delta = 4
##              sd = 12.21
##       sig.level = 0.05
##           power = 0.5242
##     alternative = two.sided
```





# Eksempel: Stikprøvestørrelsen når styken skal være 80%

```
# Udregning af stikprøvestørrelsen (en stikprøve)
power.t.test(power = .80, delta = 4, sd = 12.21, type = "one.sample")

##
##      One-sample t test power calculation
##
##               n = 75.08
##           delta = 4
##              sd = 12.21
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
```





# Styrke og stikprøvestørrelse: To stikprøver

Find styrken af en test, som kan detektere en forskel/effektstørrelse på $2$ på et 5%-signifikansniveau, når $\sigma = 1$ og $n = 10$:

```
# Udregning af styrken (to stikprøver)
power.t.test(n = 10, delta = 2, sd = 1, sig.level = 0.05)

##
##      Two-sample t test power calculation
##
##               n = 10
##           delta = 2
##              sd = 1
##       sig.level = 0.05
##           power = 0.9882
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```





# Styrke og stikprøvestørrelse: To stikprøver

Find stikprøvestørrelsen, hvis en test med en styrke på 90% skal kunne detektere en forskel/effektstørrelse på $2$, når $\sigma = 1$:

```
# Udregning af stikprøvestørrelsen (to stikprøver)
power.t.test(power = 0.90, delta = 2, sd = 1, sig.level = 0.05)

##
##      Two-sample t test power calculation
##
##               n = 6.387
##           delta = 2
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

## Styrke og stikprøvestørrelse: To stikprøver

Hvilken effektstørrelse $\delta$ kan detekteres på et 5%-signifikansniveau i en test med en styrke på 90%, når $\sigma = 1$ og $n = 10$:

```
## Udregning af effektstørrelsen (to stikprøver)
power.t.test(power = 0.90, n = 10, sd = 1, sig.level = 0.05)

##
##      Two-sample t test power calculation
##
##               n = 10
##           delta = 1.534
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```





## Overview







## $t$-test med sammenvejet varians for to uparrede stikprøver

**Det *sammenvejede* (pooled) variansestimat (her antages $\sigma_1^2 = \sigma_2^2$)**

Metode 3.52

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}.$$

**Teststørrelsen i en $t$-test med sammenvejet varians, Metode 3.53**

Betragter vi nulhypotesen om forskellen i middelværdi mellem to *uafhængige* stikprøver:

$$\delta = \mu_1 - \mu_2,$$

$$H_0: \ \delta = \delta_0,$$

så er teststørrelsen i en $t$-test med den sammenvejede varians givet ved:

$$t_{\text{obs}} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{s_p^2/n_1 + s_p^2/n_2}}.$$





## Sætning 3.54: Fordelingen af teststørrelsen

**Resultatet**

Teststørrelsen i en $t$-test med sammenvejet varians:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{S_p^2/n_1 + S_p^2/n_2}}$$

følger under nulhypotesen (og under antagelsen $\sigma_1^2 = \sigma_2^2$) en $t$-fordeling med $n_1 + n_2 - 2$ frihedsgrader, hvis de to populationer er normalfordelte.

## Vi bruger altid "Welch"-versionen

Nogenlunde (idiot)sikkert at bruge Welch-versionen altid:

- Hvis $s_1^2 = s_2^2$, så er de to test ens. Hvis det er tilfældet, så foretrækker vi ikke nødvendigvis testen med den sammenvejede varians, da antagelsen om ens varianser kan være højst tvivlsom.
- Kun hvis de to varianser er meget forskellige, kan det ske, at de to test giver meget forskellige resultater. Hvis varianserne virker meget forskellige, brydes antagelsen om ens varianser formodentligt.
- I tilfælde med en lille stikprøvestørrelse i mindst en af grupperne, kan testen med sammenvejet varians give en højere styrke (under antagelse om ens varianser). I disse tilfælde er Welch-versionen en "forsigtig" tilgang.





## Dagsorden