

02402 Statistik (Polyteknisk grundlag)

Uge 6: Analyser med to stikprøver

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

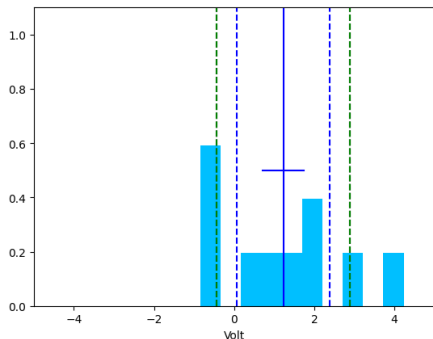
Dagsorden

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

Overview

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

Hypotesetest



t-test "cook-book":

- Nulhypotese $H_0 : \mu = \mu_0 (= 0?)$
- Beregn: $\bar{x}$, s , and $s.e.\bar{x} = s/\sqrt{n}$
- Beregn teststørrelse: $t_{obs} = \frac{\bar{x} - \mu_0}{s.e.\bar{x}}$
- Beregn p-værdi: $2 \cdot P(T > |t_{obs}|)$ (opslag i t-fordeling, lav tegning)
- Sammenhold p-værdi med ønsket α

- Ligger værdien μ_0 udenfor $(1 - \alpha) \cdot 100\%$ -konfidensintervallet? (fx 95%-CI)
 - Er den beregnede $|t_{obs}|$ større end $t_{1-\alpha/2}$? (fx $t_{0.975}$)
 - Er p-værdien mindre end α ? (fx 0.05)
- Ja? - så afviser vi nulhypotesen ved signifikansniveau α .

p-værdien svarer til det signifikansniveau (α) der lige netop ville indeholde nulhypotesen i konfidensintervallet.

I dag



I dag

- Vi vil ofte gerne vide om der er ("signifikant") forskel på to forskellige ting/grupper/interventioner.
- Vi har set eksempler hvor data afspejler målte forskelle - og vi tester om denne forskel i gennemsnit er nul.
- Men hvad hvis vi fx vil vide om der er forskel på to forskellige grupper, hvor vi ikke kan opsamle data for individuelle forskelle? (forskellen kan ikke måles på én statistisk enhed, grupperne kan også være forskellig størrelse)
- Så skal vi se på t-test for **to** stikprøver.

Overview

- 1 Opsummering fra sidste uge
- 2 **Motiverende eksempel: Ernæringsstudie**
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

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

Python

- Gå til dagens Python notebook i VS Code
 - "Example: Nutrition study 1"



Visual Studio Code

Kahoot!

Overview

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 **t -test med to stikprøver**
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

Forskel på to grupper



Hvad er nulhypotesen? Hvordan ville man formulere den? (diskutér 2 min)
(tavle)

Forskel på to grupper



Overvejelser til nulhypotesen:

- Der er ingen forskel på grupperne "A" og "B"
- De to stikprøver kommer fra *præcis samme* underliggende fordeling (?)
- **De underliggende fordelinger har samme gennemsnit:** $\mu_A = \mu_B$
- De underliggende fordelinger har samme *varsians*: $\sigma_A = \sigma_B$ (?)

Eksempel: Ernæringsstudie

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

$$H_0 : \mu_A = \mu_B \quad \text{eller} \quad H_0 : \mu_A - \mu_B = 0$$

Eksempel: Ernæringsstudie

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

$$H_0 : \mu_A = \mu_B \quad \text{eller} \quad H_0 : \mu_A - \mu_B = 0$$

Stikprøvegennemsnit og
-standardafvigelse:

$$\hat{\mu}_A = \bar{x}_A = 8.293 \quad (s_A = 1.428)$$

$$\hat{\mu}_B = \bar{x}_B = 10.298 \quad (s_B = 1.398)$$

Eksempel: Ernæringsstudie

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

$$H_0 : \mu_A = \mu_B \quad \text{eller} \quad H_0 : \mu_A - \mu_B = 0$$

Stikprøvegennemsnit og
-standardafvigelser:

$$\hat{\mu}_A = \bar{x}_A = 8.293 \quad (s_A = 1.428)$$

$$\hat{\mu}_B = \bar{x}_B = 10.298 \quad (s_B = 1.398)$$

Er data i overensstemmelse med
nulhypotesen H_0 ?

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

$$\text{Nulhypotese: } H_0 : \mu_B - \mu_A = 0$$

Eksempel: Ernæringsstudie

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

$$H_0 : \mu_A = \mu_B \quad \text{eller} \quad H_0 : \mu_A - \mu_B = 0$$

Stikprøvegennemsnit og
-standardafvigelser:

$$\hat{\mu}_A = \bar{x}_A = 8.293 \quad (s_A = 1.428)$$

$$\hat{\mu}_B = \bar{x}_B = 10.298 \quad (s_B = 1.398)$$

Er data i overensstemmelse med
nulhypotesen H_0 ?

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

$$\text{Nulhypotese: } H_0 : \mu_B - \mu_A = 0$$

NYT: p -værdi for forskel:

$$p = 0.0083$$

(Beregnet under antagelsen, at H_0 er sand.)

NYT: **Konfidensinterval for forskellen:**

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

Teori: Forskel mellem to gennemsnit

Hvad er vores nulhypotese?

$$H_0 : \delta = \mu_X - \mu_Y$$

Notation: δ er den "sande" forskel
(μ_X og μ_Y er de "sande" middelværdier)

Hvad er vores stokastiske variabel?

$$D = \bar{X} - \bar{Y}$$

Teori: Forskel mellem to gennemsnit

Middelværdi af D :

$$\mathbf{E}[D] = \mathbf{E}[\bar{X} - \bar{Y}] = \mathbf{E}[\bar{X}] - \mathbf{E}[\bar{Y}] = \mu_X - \mu_Y$$

Variansen af D :

$$\mathbf{V}[D] = \mathbf{V}[\bar{X} - \bar{Y}] = 1^2 \mathbf{V}[\bar{X}] + (-1)^2 \mathbf{V}[\bar{Y}] = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}$$

Standardafvigelse for D :

$$\sigma_D = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \quad \text{eller} \quad \text{”s.e.}_D\text{”} = \sqrt{(s.e._{\bar{X}})^2 + (s.e._{\bar{Y}})^2}$$

Teori: Notation (fra sidste uge: én stikprøve)

μ_X : det "sande" gennemsnit for X

σ_X^2 : variansen (den "sande" varians) for X

$\bar{X}$: stokastisk variabel der beskriver gennemsnit i en (teoretisk) stikprøve af størrelse n_X .

$\bar{x}$: stikprøve-gennemsnit beregnet ud fra rigtig data (stikprøven).

Vi udledte sidste gang teori om hvordan $\bar{X}$ fordeler sig:

$$\mathbf{E}[\bar{X}] = \mu_{\bar{X}} = \mu_X \quad \mathbf{V}[\bar{X}] = \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n_X}$$

Og så snakkede vi om "standard error"

$$s.e._{\bar{X}} = \sqrt{\frac{\sigma_X^2}{n_X}} \quad s.e._{\bar{x}} = \sqrt{\frac{s_X^2}{n_X}}$$

Kahoot!

Teori: Notation (to stikprøver)

δ : den "sande" forskel ("difference")

D : stokastisk variabel der beskriver forskellen mellem gennemsnit i to (teoretiske) stikprøver af størrelse n_X og n_Y .

d : observeret forskel beregnet ud fra rigtig data for to stikprøver.

Vi har netop udledt teori om hvordan D fordeler sig:

$$\mathbf{E}[D] = \text{"}\mu_D\text{"} = \mu_X - \mu_Y \quad \mathbf{V}[D] = \text{"}\sigma_D^2\text{"} = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}$$

Vi kan nu definere flere "standard errors":

$$s.e._D = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \quad s.e._d = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

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

Forudsætninger:

Testen gælder, når begge stikprøver er store (CLT), 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:

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

Sætning 3.50: Fordelingen af teststørrelsen

Teststørrelsen er (tilnærmelsesvis) t -fordelt:

Under nulhypotesen er teststørrelsen:

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

Den følger tilnærmelsesvis en t -fordeling med ν 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

- Hulhypotese $H_0 : \delta = \delta_0 (= 0?)$
 - Beregn: $d = \bar{x} - \bar{y}$ and $s.e._d$
 - Beregn teststørrelse: $t_{obs} = \frac{d - \delta_0}{s.e._d}$
 - Beregn p-værdi: $2 \cdot P(T > |t_{obs}|)$
(opslag i t-fordeling)
 - Sammenhold p-værdi med ønsket α
-
- Gå til dagens Python notebook i VS Code
 - "Example: Nutrition study 2"



Overview

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 **t -test med sammenvejede varians – Et alternativ**
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

t -test med sammenvejet varians for to uparrede stikprøver

Det *sammenvejede* (pooled) variansestimater (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

Betrægter 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, er Welch-versionen en mere "forsigtig" tilgang.

Overview

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 **Konfidensintervallet for differensen**
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

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

Konfidensintervallet for forskellen i middelværdi:

For to stikprøver $(x_1, \dots, x_{n_1})$ og $(y_1, \dots, y_{n_2})$ 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 v frihedsgrader (givet i sætning 3.50).

$$d \pm t_{1-\alpha/2} \cdot s.e.d$$

Det er samme "opskrift" som sidste gang.

Vær opmærksom på at beregne v til at indsætte som antal frihedsgrader i t -fordelingen.

Eksempel: Ernæringsstudie

- Gå til dagens Python notebook i VS Code
 - "Example: Nutrition study 3"



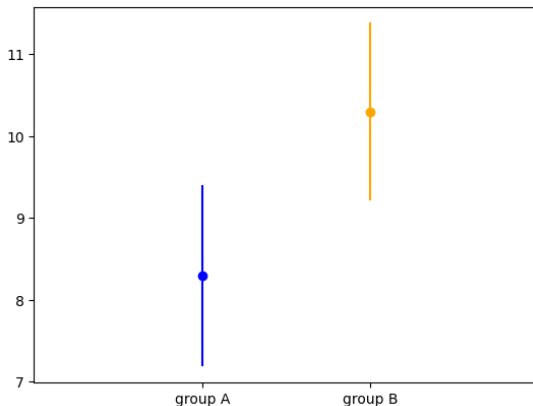
Visual Studio Code

Overview

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 **Overlappende konfidensintervaller?**
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

Overlappende konfidensintervaller

Kan I huske de overlappende konfidensintervaller?

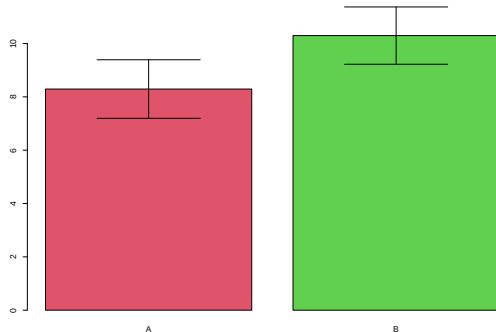


Selvom de to konfidensintervaller overlapper, endte vi med at konkludere at der ER signifikant forskel.

Alternativ præsentation fra bogen

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"

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

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 **t -test med to parrede stikprøver (parret t -test)**
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

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

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	$\bar{x} = 1.67$
6	+3.4	+4.4	+1.0	$s = 1.13$
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	

Eksempel: Sovemedicin

- Gå til dagens Python notebook i VS Code
 - "Example: Sleep medicine (Paired t-test)"



Visual Studio Code

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):

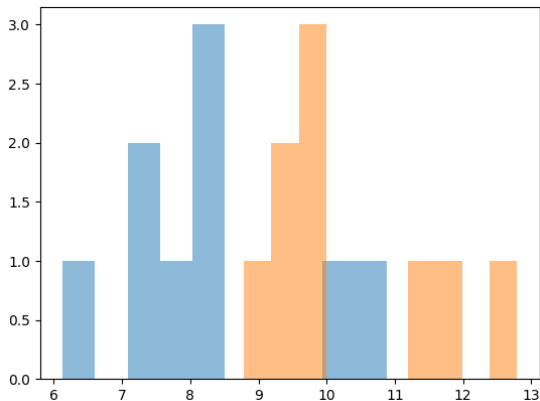
Vi har 10 patienter, som alle får begge behandlinger (typisk med noget tid imellem og med tilfældig rækkefølge af behandlingerne).

Dvs. de samme patienter fremgår i de to grupper.

Overview

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne**
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

Normalfordelingsantagelserne



Vi genbesøger data for energiforbrug hos sygeplejerskene på hospital A og hospital B.

QQ-plot

- Gå til dagens Python notebook i VS Code
 - "QQ-plots"



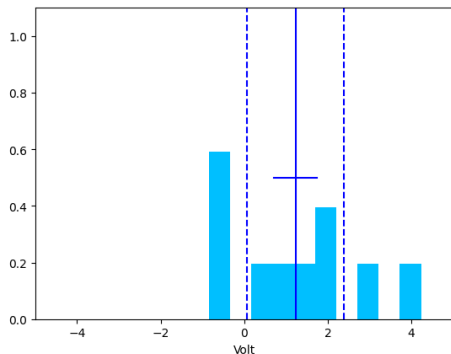
Visual Studio Code

Overview

- 1 Opsummering fra sidste uge
- 2 Motiverende eksempel: Ernæringsstudie
- 3 t -test med to stikprøver
- 4 t -test med sammenvejet varians – Et alternativ
- 5 Konfidensintervallet for differensen
- 6 Overlappende konfidensintervaller?
- 7 t -test med to parrede stikprøver (parret t -test)
- 8 Normalfordelingsantagelserne
- 9 Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

Eksempel - letbane

Lad os se på data fra letbane-eksemplet igen. Data måler spændingsfald henover et sted hvor spændingsfaldet bør være nul i gennemsnit.



Fra stikprøven ($n=10$):

$$\bar{x} = 1.23$$

$$s = 1.62$$

konfidensinterval:

$$\bar{x} \pm t_{1-\alpha/2} \cdot (s/\sqrt{n}) =$$
$$1.23 \pm 1.16$$
$$[0.07; 2.9]$$

I dette tilfælde har vi: $\hat{\mu} = 1.23 \pm 1.16$ volt ($\alpha = 0.05$).

Vi siger at "**Margin of Error**" (ME) er 1.16 (volt).

Eksempel - letbane

Konklusionen fra stikprøven blev at vi afviser nulhypotesen - dermed konkluderer vi at spændingsfaldet IKKE er nul (i gennemsnit).

Men chefen er ikke tilfreds og siger at vi må undersøge sagen bedre. Vores fejlmargen på ± 1.16 volt er for stor.

Vi planlægger derfor at lave en ny stikprøve.

Hvordan kan vi gøre det bedre end første gang?

Hvad kan vi gøre for at forbedre vores estimat af μ ?

Kahoot!

(x1)

Forsøgsplanlægning med krav til præcisionen

Fejlmarginen (margin of error - ME) er defineret som

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

For store nok stikprøver kan vi approximere $t_{1-\alpha/2}$ (fra t-fordeling) med $z_{1-\alpha/2}$ (fra normalfordeling).

Metode 3.63: Stikprøvestørrelse for konfidensintervallet baseret på en stikprøve

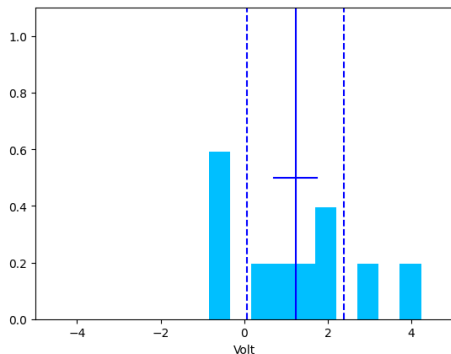
Hvis σ er kendt, eller vurderet til at være en bestemt værdi, så kan vi beregne den stikprøvestørrelse, som kræves for at opnå en given fejlmargin, med sandsynlighed $1 - \alpha$.

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

Eksempel - letbane

Lad os estimere størrelsen på stikprøven hvis vi ønsker ME på højst 0.5 volt.

Vi antager at $s = 1.62$ er et udemærket gæt på den sande σ .



$$\begin{aligned}
 n &= \left(\frac{z_{1-\alpha/2} \cdot \sigma}{ME} \right)^2 \\
 &= \left(\frac{1.96 \cdot 1.62}{0.5} \right)^2 \\
 &= 40.3
 \end{aligned}$$

Da vi ikke kan lave 40.3 målinger runder vi op til 41 (i den nye stikprøve er $n = 41$).

Planlægning (Styrke og stikprøvestørrelse)

Vi skal nu snakke om forsøgsplanlægning og "styrke".

Til dette skal vi overveje de to typer af fejl, som vi introducerede sidste gang:

Type I: Afvisning af H_0 , når H_0 er sand.

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

Type II: Accept (ikke afvisning) af H_0 , når H_1 er sand.

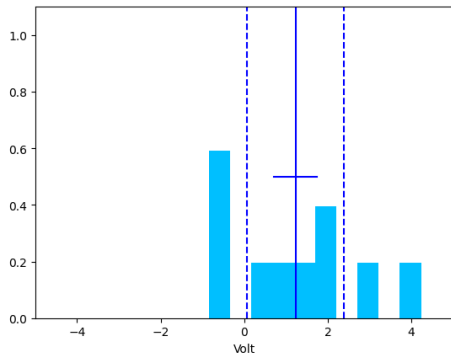
$$P(\text{Type II fejl}) = \beta$$

Eksempel: letbane

Chefen bliver nu lidt mere krævende. Hun siger at hvis det sande gennemsnitlige spændingsfald er ± 0.5 volt (eller mere), så vil vi gerne vide det.

Hvordan skal vi bære os ad?

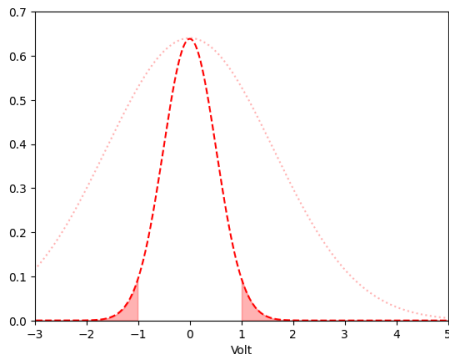
Vi gennemgår nu nogle forskellige **scenarier**.



Eksempel: Ietbane

Scenarie 1: Nulhypotesen er sand

- Hvis det sande gennemsnitlige spændingsfald er nul
- og vi laver en ny omgang målinger - denne gang med større n
- da vil risikoen for at forkaste nulhypotesen ved en fejl være α (fx 0.05, eller 5%)



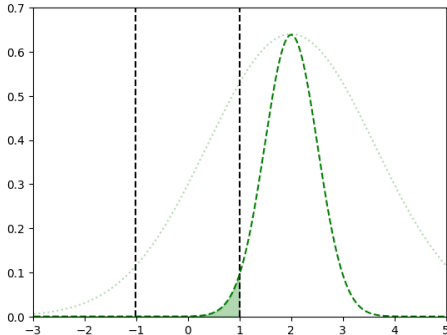
De værdier af $\bar{x}$ som vi potentielt accepterer som værende "ikke significant forskellige fra nul" ligger i området $0 \pm ME$. Bredden af dette område afhænger af størrelsen på stikprøven (n).

Kahoot!
(x2)

Eksempel: Ietbane

Scenarie 2: Modhypotesen er sand

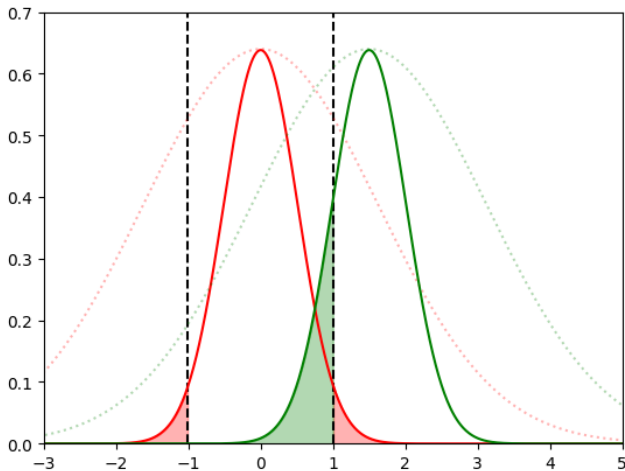
- Hvis det sande gennemsnitlige spændingsfald i virkeligheden er noget andet end nul - den sande værdi er μ_1
- og vi laver en ny omgang målinger - denne gang med større n
- Hvad er da risikoen (sandsynligheden β) for at acceptere (ikke afvise) nulhypotesen?



Ja, det kommer an på...
...hvor stor μ_1 egentlig er

Kahoot!
(x2)

Type I og Type II fejl



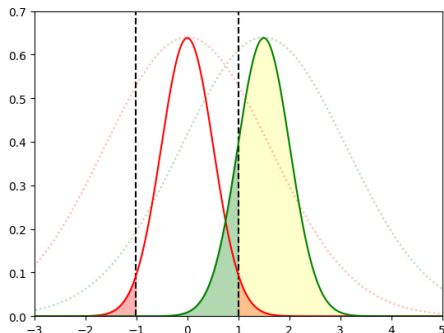
α : det røde areal

β : det grønne areal

$(1 - \beta)$ kaldes **styrken** (eng: **power**)

Forsøgsplanlægning: Styrke

- **Styrke** = $1 - \beta$
- $P(\text{Type II fejl}) = \beta$
- Styrke = "Sandsynligheden for en korrekt afvisning af H_0 ".
- Styrke = "Sandsynligheden for at detektere en (påstået) effekt".
- Typiske værdier for ønsket styrke er 80%, 90%, .. (noget relativt stort)
- I praksis: Brug en scenarie-baseret tilgang



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, α , som vi tester på.
- Forskellen i middelværdi (effekt-størrelsen), $\mu_0 - \mu_1$.
- Populationsstandardafvigelsen, σ .
- Styrken (power), $1 - \beta$.

(tavle)

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%):

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 α , β og σ er givet:

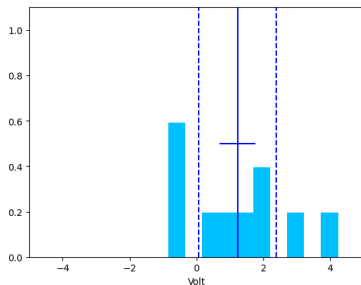
$$n = \left(\sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_0 - \mu_1} \right)^2.$$

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: Ietbane

Vi siger nu til chefen at vi kan designe den næste stikprøve således at vi vil have en 80% sandsynlighed for at "detektere" et spændingsfald på 0.5 volt (eller mere), ved et signifikansniveau $\alpha = 0.05$.

For at beregne den nødvendige stikprøvestørrelse, bruger vi (som informeret gæt): $\sigma = 1.62$, svarende til den spredning vi observerede i vores tidligere målinger.



Vi kommer da frem til at den næste stikprøve skal have 83 målinger:

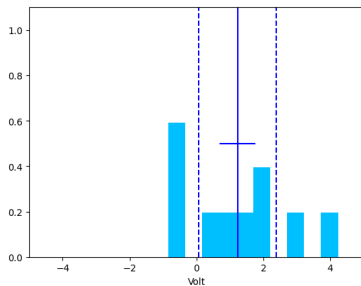
$$\begin{aligned}
 n &= \left(\sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_0 - \mu_1} \right)^2 \\
 &= \left(1.62 \frac{z_{0.80} + z_{0.975}}{0.5} \right)^2 \\
 &= 82.4
 \end{aligned}$$

Eksempel: Ietbane

"83 målinger bliver for dyrt", siger chefen.

"Kan vi nøjes med 60 målinger?"

Ja.. - men så får vores experiment en mindre styrke.



Ny styrkeberegning:

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

$\Rightarrow$

$$\begin{aligned} z_{1-\beta} &= \sqrt{60 \frac{0.5^2}{1.62^2}} - z_{0.975} \\ &= 0.43 \end{aligned}$$

svarende til en styrke på 0.67

Python

- Gå til dagens Python notebook i VS Code
 - "Example: Power calculations"



Visual Studio Code

Forsøgsplanlægning: To stikprøver

Det store spørgsmål: Hvor store skal n_1 og n_2 være?

Vi kan generalisere forsøgsplanlægningen til den situation hvor vi sammenligner to stikprøver. Stikprøvestørrelserne er nu givet ved n_1 og n_2 . Vi vil antage at variansen (σ^2) er ens i de to populationer.

Forsøgsplanlægning: To stikprøver

Det store spørgsmål: Hvor store skal n_1 og n_2 være?

Vi kan generalisere forsøgsplanlægningen til den situation hvor vi sammenligner to stikprøver. Stikprøvestørrelserne er nu givet ved n_1 og n_2 . Vi vil antage at variansen (σ^2) er ens i de to populationer.

Formel for stikprøvestørrelse med to stikprøver

For en t-test med to stikprøver, hvor α , β og σ er givet:

$$n_1 = (k + 1) \left(\sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_1 - \mu_2} \right)^2$$

Her er $\mu_1 - \mu_2$ den forskel i middelværdier, som vi ønsker at måle, medens $z_{1-\beta}$ og $z_{1-\alpha/2}$ er fraktiler i standardnormalfordelingen. Desuden er $k = n_1/n_2$ forholdet mellem det to stikprøvestørrelser, således at vi har $n_2 = n_1/k$.

Python

- Gå til dagens Python notebook i VS Code
 - "Example: Power calculations, 2 samples"



Visual Studio Code

Dagsorden

- ① Opsummering fra sidste uge
- ② Motiverende eksempel: Ernæringsstudie
- ③ t -test med to stikprøver
- ④ t -test med sammenvejet varians – Et alternativ
- ⑤ Konfidensintervallet for differensen
- ⑥ Overlappende konfidensintervaller?
- ⑦ t -test med to parrede stikprøver (parret t -test)
- ⑧ Normalfordelingsantagelserne
- ⑨ Styrke og stikprøvestørrelse – Forsøgsdesign
 - Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver