

02402 Statistik (Polyteknisk grundlag)

Uge 6: Analyser med to stikprøver

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

1 Sammenligning af to grupper

- Nulhypotese
- Differens mellem to stikprøvegegnemsnit
- Welch t -test for to stikprøver
- Pooled t -test for to stikprøver
- Konfidensintervallet for differensen
- t -test med to parrede stikprøver (parret t -test)

2 Styrke og stikprøvestørrelse – Forsøgsdesign

- Krav til præcision
- Styrke og stikprøvestørrelse – En stikprøve
- Styrke og stikprøvestørrelse – To stikprøver

02402 Statistik (Polyteknisk grundlag)

Sammenligning af to grupper

- Nulhypotese
- Differens mellem to stikprøvegennemsnit
- Welch t -test for to stikprøver
- Pooled t -test for to stikprøver
- Konfidensintervallet for differensen
- t -test med to parrede stikprøver (parret t -test)

Recap: Konfidensinterval og t -test for én stikprøve

t -test "cook-book":

- Nullhypotese $H_0 : \mu = \mu_0 (= 0?)$
- Beregn: $\bar{x}$, s , and $SE_{\bar{x}} = s/\sqrt{n}$
- Beregn CI: $\bar{x} \pm t_{1-\alpha/2} \cdot s/\sqrt{n}$
- Beregn teststørrelse: $t_{obs} = \frac{\bar{x} - \mu_0}{s.e.\bar{x}}$
- Beregn p-værdi: $2 \cdot P(T > |t_{obs}|)$

3 måder at nå frem til en konklusion:

- Ligger værdien μ_0 *udenfor* **konfidensintervallet**? (fx 95%-CI, dvs $\alpha = 0.05$)
- Er den beregnede $|t_{obs}|$ *større* end den **kritiske værdi** $t_{1-\alpha/2}$? (fx $t_{0.975}$)
- Er **p-værdien** *mindre* end α ? (fx $\alpha = 0.05$)

Ja? - så *afviser vi nulhypotesen* ved signifikansniveau α .

Sammenligning af 2 grupper



Er der *signifikant* forskel?

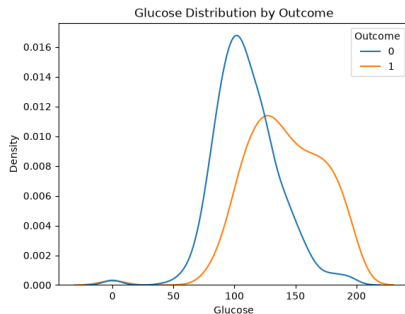
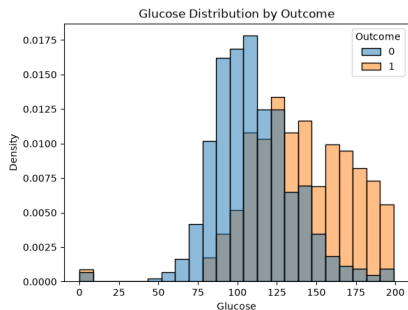
Er én ting bedre end en anden? Er der en *effekt*?

Kan vi regne p -værdier når vi har to uafhængige stikprøver?

Hvordan kan man opstille én test-størrelse for **to** stikprøver?

Eksempel med data

Gensyn med data for blodsukker-målinger lavet i forbindelse med en glucose tolerance test:



Blå: 500 raske personer

Gul: 268 personer med sukkersyge

Data fortæller os noget om blodsukker på "individ-niveau" og på "gruppe-niveau".

I dag skal vi tale om at sammenligne to grupper og beskrive forskelle.

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.

Eksempel: Energiforbrug

Forskel på energiforbrug?

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

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:

$$n_1 = 9, n_2 = 9:$$

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

Vi vil gerne vide om der er forskel på de to grupper.

Eksempel: Energiforbrug

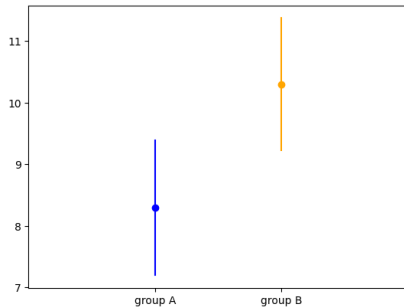
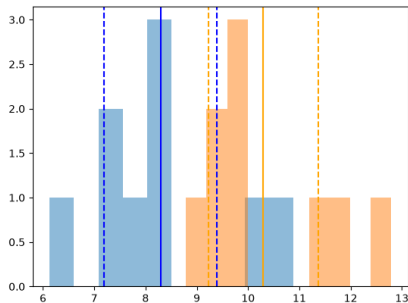
Hospital A	Hospital B
$\bar{x}_A = 8.29$	$\bar{x}_B = 10.3$
$s_A = 1.428$	$s_B = 1.398$
$\hat{\mu}_A = 8.29$	$\hat{\mu}_B = 10.3$
$SE_{\hat{\mu}_A} = 0.476$	$SE_{\hat{\mu}_B} = 0.466$
95%-CI: [7.196; 9.391]	95%-CI: [9.223, 11.37]

- Gå til dagens Python notebook i VS Code
 - "Energy consumption example"



Visual Studio Code

Eksempel: Energiforbrug



Kahoot!
(x3)

02402 Statistik (Polyteknisk grundlag)

- **Nulhypotese**
 - Differens mellem to stikprøvegegnemsnit
 - Welch t -test for to stikprøver
 - Pooled t -test for to stikprøver
 - Konfidensintervallet for differensen
 - t -test med to parrede stikprøver (parret t -test)

Forskel på to grupper



Hvordan formulerer vi en nulhypotese?

- Der er ingen forskel på grupperne "A" og "B" ...?
- De to stikprøver kommer fra samme underliggende sandsynlighedsfordeling ...?
- De underliggende fordelinger har samme gennemsnit: $\mu_A = \mu_B$...?
- De underliggende fordelinger har samme *varians*: $\sigma_A = \sigma_B$...?

Eksempel: Energiforbrug

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

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

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

Eksempel: Energiforbrug

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

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

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

Stikprøvegennemsnit:

$$\hat{\mu}_A = \bar{x}_A = 8.293$$

$$\hat{\mu}_B = \bar{x}_B = 10.298$$

Eksempel: Energiforbrug

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

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

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

Stikprøvegennemsnit:

$$\hat{\mu}_A = \bar{x}_A = 8.293$$

$$\hat{\mu}_B = \bar{x}_B = 10.298$$

Fra data får vi estimat af forskel:

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

$$\hat{\mu}_B - \hat{\mu}_A = 2.005$$

Eksempel: Energiforbrug

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

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

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

Stikprøvegennemsnit:

$$\hat{\mu}_A = \bar{x}_A = 8.293$$

$$\hat{\mu}_B = \bar{x}_B = 10.298$$

Fra data får vi estimat af forskel:

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

$$\hat{\mu}_B - \hat{\mu}_A = 2.005$$

NYT: p -værdi for forskel:

$$p = 0.0083$$

NYT: Konfidensinterval for forskellen:

$$95\text{-CI: } [0.59; 3.42]$$

Nulhypotese for en forskel mellem to grupper

Hvad er vores nulhypotese?

$$\begin{aligned} H_0 : \mu_X - \mu_Y &= \delta_0 \\ \delta &= \delta_0 \end{aligned}$$

Notation: δ er den "sande" forskel mellem populationernes middelværdier.

Oftentimes er nulhypotesen:

$$H_0 : \delta = 0$$

Der er "ingen forskel" (på μ_X og μ_Y).

Hvis man kan **afvise** denne nulhypotese siger man der **er** en "signifikant forskel" mellem grupperne.

02402 Statistik (Polyteknisk grundlag)

- Nulhypotese
- Differens mellem to stikprøvegentennsmit
- Welch t -test for to stikprøver
- Pooled t -test for to stikprøver
- Konfidensintervallet for differensen
- t -test med to parrede stikprøver (parret t -test)

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Differensen mellem to stikprøvegennemsnit

Vi beskriver **differensen** med den en stokastisk variabel:

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

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 :

$$\sigma_D = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

Standard error for en **differens**

Standard Error for differensen:

$$\hat{\sigma}_D = \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} = \sqrt{(\hat{\sigma}_{\bar{X}})^2 + (\hat{\sigma}_{\bar{Y}})^2}$$

eller

$$SE_{\hat{\delta}} = \sqrt{(SE_{\bar{x}})^2 + (SE_{\bar{y}})^2}$$

(Denne er ikke som sådan defineret i bogen, det er blot lidt ekstra forklaring på hvor størrelsen $\sqrt{s_1^2/n_1 + s_2^2/n_2}$ kommer fra)

02402 Statistik (Polyteknisk grundlag)

- Nulhypotese
- Differens mellem to stikprøvegegnemsnit
- **Welch t-test for to stikprøver**
- Pooled t -test for to stikprøver
- Konfidensintervallet for differensen
- t -test med to parrede stikprøver (parret t -test)

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Teststørrelsen t_{obs} for t-test med to stikprøver

For t-test med én stikprøve:

Vi betragter $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$, der følger en t -fordeling med $\nu = n - 1$ frihedsgrader.

Fra stikprøvedata beregnes $t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ og denne bruges til at udføre t -testen.

$$t_{obs} = \frac{\hat{\mu} - \mu_0}{SE_{\hat{\mu}}}$$

For t-test med to stikprøver:

Vi betragter $T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{S_X^2/n_X + S_Y^2/n_Y}}$, der følger en ??-fordeling (næste slide).

Fra stikprøvedata beregnes $t_{obs} = \frac{(\bar{x} - \bar{y}) - \delta_0}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$ og denne bruges til at udføre t -testen.

$$t_{obs} = \frac{\hat{\delta} - \delta_0}{SE_{\hat{\delta}}}$$

Fordeling for T ("Welch" test med to stikprøver)

||| Theorem 3.50 The distribution of the (Welch) two-sample statistic

The (Welch) two-sample statistic seen as a random variable

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}, \quad (3-49)$$

approximately, under the null hypothesis, follows a t -distribution with ν degrees of freedom, where

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}, \quad (3-50)$$

if the two population distributions are normal or if the two sample sizes are large enough.

"Welch" t -test for differens mellem to stikprøver

||| Method 3.51 The level α two-sample t -test

1. Compute the test statistic using Equation (3-48) and ν from Equation (3-50)

$$t_{\text{obs}} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \text{ and } \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

2. Compute the evidence against the *null hypothesis*^a

$$H_0 : \mu_1 - \mu_2 = \delta_0,$$

vs. the *alternative hypothesis*

$$H_1 : \mu_1 - \mu_2 \neq \delta_0,$$

by the

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

where the t -distribution with ν degrees of freedom is used

3. If $p\text{-value} < \alpha$: we reject H_0 , otherwise we accept H_0 ,

or

The rejection/acceptance conclusion can equivalently be based on the critical value(s) $\pm t_{1-\alpha/2}$:

if $|t_{\text{obs}}| > t_{1-\alpha/2}$ we reject H_0 , otherwise we accept H_0

$$\begin{aligned} t_{\text{obs}} &= \frac{\hat{\delta} - \delta_0}{SE_{\hat{\delta}}} \\ &= \frac{(\bar{x} - \bar{y}) - \delta_0}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}} \end{aligned}$$

Forudsætninger:

Testen gælder når **begge** stikprøver er store (CLT), eller når **begge** stikprøver kommer fra normalfordelte populationer.

Eksempel: Energiforbrug - Welch t -test

Udfør en Welch t -test:

- Opstil nulhypotese $H_0 : \delta = \delta_0$
- Beregn: $\hat{\delta} = \bar{x} - \bar{y}$ og $SE_{\hat{\delta}} = \sqrt{s_X^2/n_X + s_Y^2/n_Y}$
- Beregn teststørrelse: $t_{obs} = \frac{\hat{\delta} - \delta_0}{SE_{\hat{\delta}}}$
- Beregn p-værdi: $2 \cdot P(T > |t_{obs}|)$
(opslag i t-fordeling med korrekt v)
- Sammenhold p-værdi med ønsket α

Man kan bruge funktionen:

- `stats.ttest_ind(A, B, equal_var=False)` (fra `Scipy.stats` pakken)
- Gå til dagens Python notebook i VS Code
 - "T-test for two samples"



Visual Studio Code

02402 Statistik (Polyteknisk grundlag)

- Nulhypotese
- Differens mellem to stikprøvegegnemsnit
- Welch t -test for to stikprøver
- **Pooled t -test for to stikprøver**
- Konfidensintervallet for differensen
- t -test med to parrede stikprøver (parret t -test)

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Sammenvejet varians ("pooled" estimate of variance)

Forskel fra før: Nu laver vi den *ekstra antagelse* at variansen i de to underliggende fordelinger er ens: $\sigma_X = \sigma_Y = \sigma_p$

Vi estimerer variansen i de underliggende fordelinger med en "pooled variance" (sammenvejet varians), s_p^2 :

||| Method 3.52 The pooled two-sample estimate of variance

Under the assumption that $\sigma_1^2 = \sigma_2^2$ the *pooled* estimate of variance is the weighted average of the two sample variances

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \quad (3-51)$$

Fordeling for T ("pooled" test med to stikprøver)

Under antagelsen $\sigma_X^2 = \sigma_Y^2 = \sigma_p^2$ har vi $SE_{\hat{\delta}} = \sqrt{s_p^2/n_X + s_p^2/n_Y}$.

Nu betragter vi derfor $T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{S_X^2/n_X + S_Y^2/n_Y}} = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{S_p^2/n_X + S_p^2/n_Y}}$

|||| Theorem 3.54 The distribution of the pooled two-sample t -test statistic

The pooled two-sample statistic seen as a random variable:

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{S_p^2/n_1 + S_p^2/n_2}}. \quad (3-54)$$

follows, under the null hypothesis and under the assumption that $\sigma_1^2 = \sigma_2^2$, a t -distribution with $n_1 + n_2 - 2$ degrees of freedom if the two population distributions are normal.

Teststørrelsen t_{obs} for "pooled" test med to stikprøver

|||| Method 3.53 The pooled two-sample t -test statistic

When considering the null hypothesis about the difference between the means of two *independent* samples

$$\begin{aligned}\delta &= \mu_1 - \mu_2, \\ H_0 : \delta &= \delta_0.\end{aligned}\tag{3-52}$$

the pooled two-sample t -test statistic is

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{s_p^2/n_1 + s_p^2/n_2}}.\tag{3-53}$$

Eksempel: Energiforbrug - pooled t -test

Udfør en "pooled" t -test (under antagelsen $\sigma_X^2 = \sigma_Y^2$):

- Opstil nulhypotese $H_0 : \delta = \delta_0$
- Beregn: $\hat{\delta} = \bar{x} - \bar{y}$ og $SE_{\hat{\delta}} = \sqrt{s_p^2/n_X + s_p^2/n_Y}$ (kræver beregning af s_p^2)
- Beregn teststørrelse: $t_{obs} = \frac{\hat{\delta} - \delta_0}{SE_{\hat{\delta}}}$
- Beregn p-værdi: $2 \cdot P(T > |t_{obs}|)$
(opslag i t -fordeling med korrekt v . OBS: v er ikke det samme som for Welch testen!)
- Sammenhold p-værdi med ønsket α

Man kan bruge funktionen:

- `stats.ttest_ind(A, B, equal_var=True)` (fra `Scipy.stats` pakken)
- Gå til dagens Python notebook i VS Code
 - "T-test for two samples (pooled)"



Visual Studio Code

"Welch" eller "pooled" ?

- Jo færre antagelser jo bedre - vi foretrækker derfor "Welch" versionen.
- Hvis $s_1^2 = s_2^2$, så er de to test ens.
- 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.
- Senere i kurset, når vi skal sammenligne flere grupper, vil vi dog antage ens varians i alle grupperne.

Kahoot!
(x2)

02402 Statistik (Polyteknisk grundlag)

- Nulhypotese
- Differens mellem to stikprøvegegnemsnit
- Welch t -test for to stikprøver
- Pooled t -test for to stikprøver
- **Konfidensintervallet for differensen**
- t -test med to parrede stikprøver (parret t -test)

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Konfidensinterval for differens mellem to stikprøver

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

$$\hat{\delta} \pm t_{1-\alpha/2} \cdot SE_{\hat{\delta}}$$

Det er samme "opskrift" som sidste gang.

Vær opmærksom på at $t_{1-\alpha/2}$ er fraktil fra en t -fordeling med ν frihedsgrader - dvs man skal beregne ν (brug "Welch" versionen).

Konfidensinterval for differens mellem to stikprøver

||| Method 3.47 The two-sample confidence interval for $\mu_1 - \mu_2$

For two samples $x_1, \dots, x_n$ and $y_1, \dots, y_n$ the $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad (3-45)$$

where $t_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile from the t -distribution with ν degrees of freedom given from Equation (3-50)

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}. \quad (3-46)$$

Eksempel: Energiforbrug

Man kan også bruge:

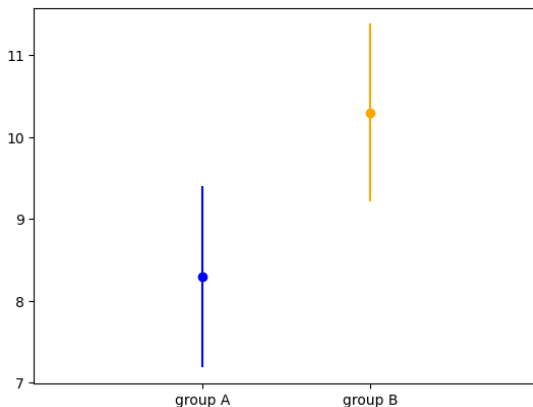
- `stats.ttest_ind(A, B, equal_var=False).confidence_interval(...)`
- Gå til dagens Python notebook i VS Code
 - "Confidence interval for difference between means"



Visual Studio Code

Overlappende konfidensintervaller

Kan I huske de overlappende konfidensintervaller? (for individuelle stikprøvegegnemsnit)

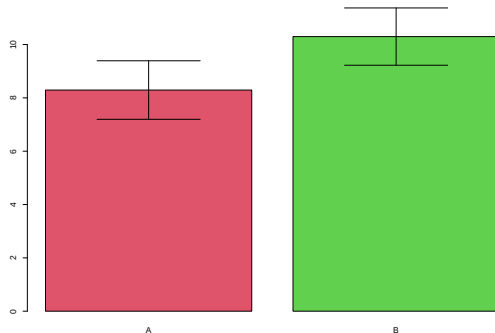


Selvom de to konfidensintervaller overlapper, endte vi med at konkludere at der ER signifikant forskel.

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

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.

|||| Remark 3.59

When interpreting two (and multi-) independent samples mean bar plots with added confidence intervals:

When two CIs do NOT overlap: The two groups are significantly different

When two CIs DO overlap: We do not know from this what the conclusion is (but then we can use the presented two-sample test method)

02402 Statistik (Polyteknisk grundlag)

- Nulhypotese
- Differens mellem to stikprøvegegnemsnit
- Welch t -test for to stikprøver
- Pooled t -test for to stikprøver
- Konfidensintervallet for differensen
- t -test med to parrede stikprøver (parret t -test)

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

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

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	

t -test for parret data

Når man har to "parrede" stikprøver, kan disse *oversættes* til én stikprøve med målte forskelle. Data analyseres derfor med en almindelig t -test for én stikprøve.

Man kan bruge:

- `stats.ttest_1samp(diff_data, pop_mean=0)`

Eller:

- `stats.ttest_rel(A, B)`

(*_rel* står for *relative*, dvs parret data)

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



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.

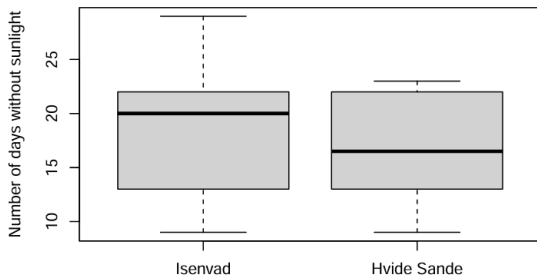
Parrede observationer (afhængige stikprøver):

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 stikprøver. Observationer fra samme patient er ikke uafhængige.

Eksempel: Eksamensopgave fra Maj 2021

Opgaven beskriver data for målinger af antal dage uden sollys, målt på to forskellige lokationer i Jylland (Kahoot efter næste slide):



The observations are sorted according to the year and for Isenvad stored in $\mathbf{x}$ and for Hvide Sande in $\mathbf{y}$:

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
$\mathbf{x}$	9	13	18	29	19	22	13	22	26	21
$\mathbf{y}$	9	12	13	23	17	22	15	17	22	16

Eksempel: Eksamensopgave fra Juni 2021

Exercise VI

A study was designed to investigate whether the use of recorded video lectures and group exercises affect the exam score of students. 9 students (denoted $\mathbf{x}$) did not use or attend either of the two activities. Other 9 students (denoted $\mathbf{y}$) used or attended both types of learning activities. The table below shows the final exam score of the students:

Students	1	2	3	4	5	6	7	8	9
$\mathbf{x}$	54	54	38	47	48	62	54	55	52
$\mathbf{y}$	67	70	71	60	68	65	68	67	64

Kahoot!
(x2)

02402 Statistik (Polyteknisk grundlag)

Styrke og stikprøvestørrelse – Forsøgsdesign

- Krav til præcision
- Styrke og stikprøvestørrelse – En stikprøve
- Styrke og stikprøvestørrelse – To stikprøver

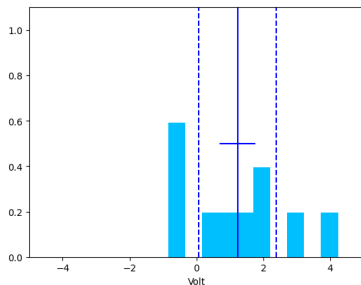
DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

02402 Statistik (Polyteknisk grundlag)

- Krav til præcision
 - Styrke og stikprøvestørrelse – En stikprøve
 - Styrke og stikprøvestørrelse – To stikprøver

Eksempel - Spænding i jordforbindelse

Lad os se på data fra eksemplet med spænding i jordforbindelse (fra sidste uge). Data måler spændingen mellem en jordforbindelse og referencejord og bør derfor være nul i gennemsnit.



Fra stikprøven ($n=10$):

$$\hat{\mu} = \bar{x} = 1.23$$

95% konfidensinterval:

$$[0.07; 2.9] = 1.23 \pm 1.16$$

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 - Spænding i jordforbindelse

Konklusionen fra stikprøven blev at vi afviser nulhypotesen - dermed konkluderer vi at spændingsfaldet IKKE er nul (i gennemsnit). Man beslutter derfor at jordforbindelsen i det elektriske anlæg bør udskiftes.

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") er defineret som

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

Vi approximerer $t_{1-\alpha/2}$ (fra t -fordeling med $\nu = n - 1$ frihedsgrader) med $z_{1-\alpha/2}$ (fra standard normalfordeling).

|||| Method 3.63 The one-sample CI sample size formula

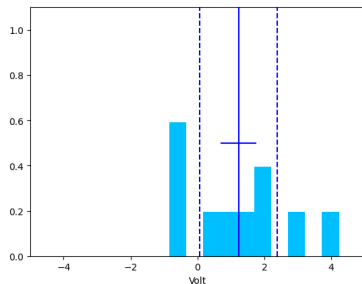
When σ is known or guessed at some value, we can calculate the sample size n needed to achieve a given margin of error, ME , with probability $1 - \alpha$ as

$$n = \left(\frac{z_{1-\alpha/2} \cdot \sigma}{ME} \right)^2. \quad (3-59)$$

Eksempel - Spænding i jordforbindelse

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 !



$$\begin{aligned} n &= \left(\frac{z_{1-\alpha/2} \cdot \sigma}{ME} \right)^2 \\ &= \left(\frac{1.96 \cdot 1.62}{0.5} \right)^2 \\ &= 40.33 \end{aligned}$$

Da vi ikke kan lave 40.33 målinger runder vi op til 41

Vi planlægger derfor en ny stikprøve med $n = 41$ målinger.

02402 Statistik (Polyteknisk grundlag)

- Krav til præcision
- Styrke og stikprøvestørrelse – En stikprøve
- Styrke og stikprøvestørrelse – To stikprøver

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

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.

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

Vi gennemgår nu de to forskellige **scenarier**

Nullhypotesen ER sand

Scenarie 1: Nulhypotesen er sand

- Hvis den sande middelværdi er $\mu = \mu_0$
- Og hvis den sande standardafvigelse er σ
- og vi har en stikprøve med n observationer
- og vi dermed forventer $ME = t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$
- da vil risikoen for at forkaste nulhypotesen ved en fejl være α (fx 0.05, eller 5%)

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

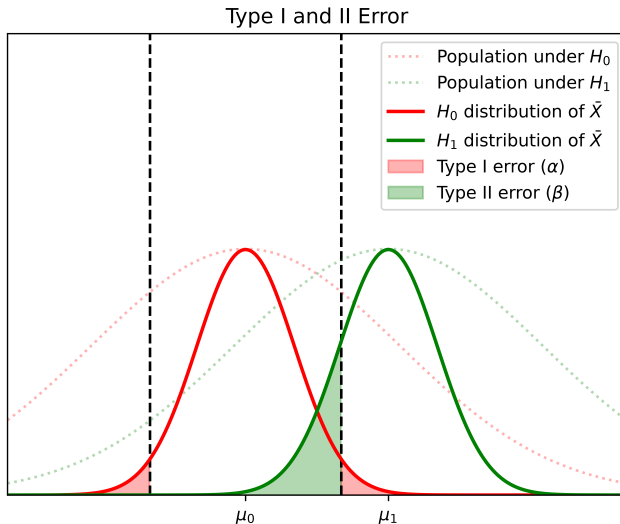
Nullhypotesen er IKKE sand

Scenarie 2: Nullhypotesen er IKKE sand

- Hvis den sande middelværdi er en **anden** værdi $\mu = \mu_1$
- Og hvis den sande standardafvigelse er σ (ligesom før)
- og vi har en stikprøve med n observationer (ligesom før)
- og vi dermed forventer $ME = t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ (ligesom før)
- Men hvad er risikoen (β) for at acceptere nulhypotesen ved en fejl?
 - Det kommer an på hvor stor afstanden $|\mu_0 - \mu_1|$ er
 - For hvis $|\mu_0 - \mu_1| \leq ME$ så accepterer vi nulhypotesen

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

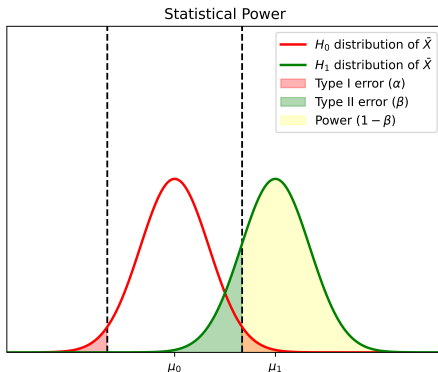
Type I og Type II fejl



α : det røde areal

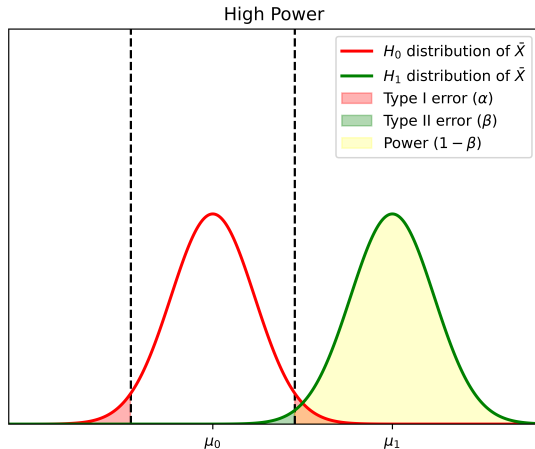
β : det grønne areal

Statistisk Styrke

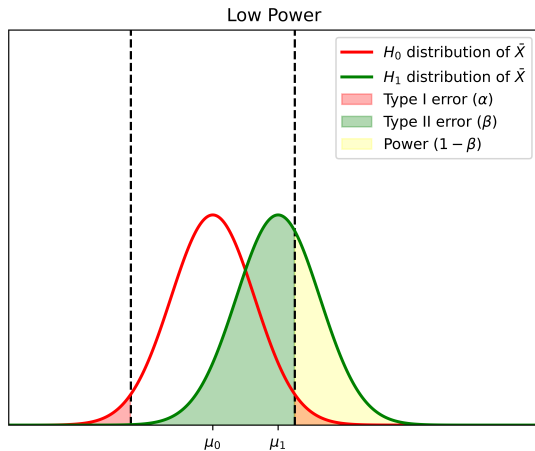


- **Styrke** = $1 - \beta$
- $P(\text{Type II fejl}) = \beta$
- Styrke = "Sandsynligheden for en korrekt afvisning af H_0 ".
- Styrke = "Sandsynligheden for at detektere en (påstået) effekt".
- Typiske værdier for ønsket styrke er 80%, 90%, .. (noget relativt stort)

Eksempel med høj styrke

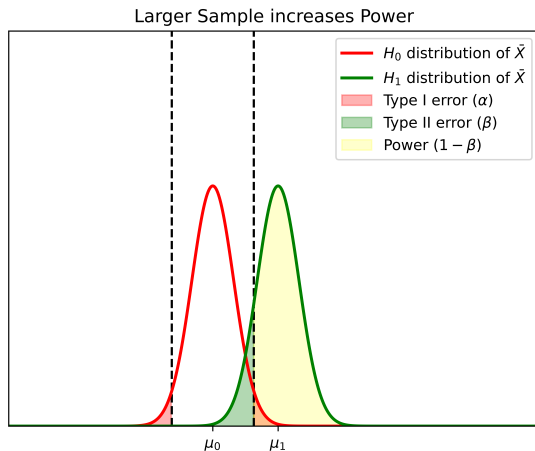


Eksempel med lav styrke



Hvad kan vi gøre?

Eksempel med større stikprøve



Større stikprøve!

Kahoot! (x4)

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%):

|||| Method 3.65 The one-sample sample size formula

For the one-sample t -test for given α , β and σ

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

where $\mu_0 - \mu_1$ is the difference in means that we would want to detect and $z_{1-\beta}$, $z_{1-\alpha/2}$ are quantiles of the standard normal distribution.

(Formlen i Method 3.65 kan udledes ud fra tegningen på forrige slide - du kan prøve selv hvis du har mod på en udfordring)

Forsøgsplanlægning: Styrke

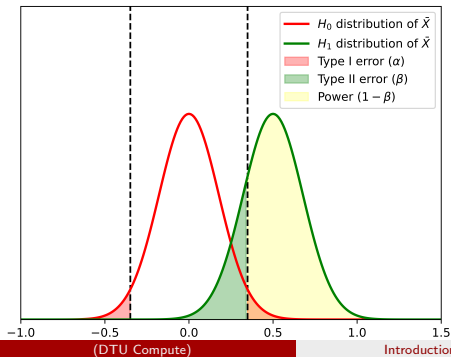
Hvis vi kender (eller antager) fire ud af de fem følgende størrelser, så kan vi finde den manglende:

- Stikprøvestørrelsen (sample size), n .
- Signifikansniveauet, α .
- Forskellen i middelværdi ("effekt-størrelsen"), $\mu_0 - \mu_1$.
- Populationsstandardafvigelsen, σ .
- Styrken (power), $1 - \beta$.

Eksempel - Spænding i jordforbindelse

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 82.4 målinger (rundes op til 83):

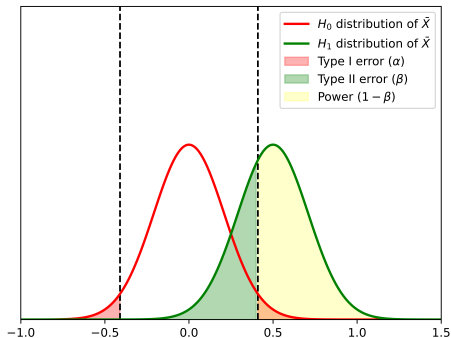
$$\begin{aligned}
 n &= \left(\sigma \frac{z_{1-\beta} + z_{1-\alpha/2}}{\mu_0 - \mu_1} \right)^2 \\
 &= \left(1.62 \frac{z_{0.80} + z_{0.975}}{0.5} \right)^2 \\
 &= 82.4
 \end{aligned}$$

Eksempel - Spænding i jordforbindelse

"83 målinger bliver for dyrt", siger chefen.

"Kan vi nøjes med 60 målinger?"

Ja.. - men så får vores eksperiment en mindre statistisk styrke.



Ny styrkeberegning:

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

$\Rightarrow$

$$\begin{aligned} z_{1-\beta} &= \sqrt{n \frac{(\mu_0 - \mu_1)^2}{\sigma^2}} - z_{0.975} \\ &= 0.43 \end{aligned}$$

svarende til en styrke på 0.67

Hvad er det der er 67% på plottet?

Python

Vi skal nu bruge nogle funktioner fra pakken `statsmodels.stats.power`

```
(import statsmodels.stats.power as smp)
```

Man kan bruge:

- `smp.TTestPower().solve_power(effect_size=delta/sd, alpha=0.05, power=0.80)`
- Gå til Python notebook
"power_calculations.ipynb" i VS Code



Visual Studio Code

02402 Statistik (Polyteknisk grundlag)

- Krav til præcision
- Styrke og stikprøvestørrelse – En stikprøve
- Styrke og stikprøvestørrelse – To stikprøver

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

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 $n_2 = n_1/k$.

(Denne formel er IKKE med i bogen, dvs det er ikke pensum)

Python

Vi skal bruge nogle funktioner fra pakken `statsmodels.stats.power`

```
(import statsmodels.stats.power as smp)
```

Til stikprøvestørrelser for test med to stikprøver kan man bruge:

- `smp.TTestIndPower().solve_power(effect_size=delta/sd, alpha=0.05, power=1-beta, ratio=k)`
- Gå til Python notebook
"power_calculations.ipynb" i VS Code



Visual Studio Code

Eksempel: Eksamensopgave fra Dec 2024

Exercise VII

Suppose we have collected exam scores from two groups:

Group 1: 82, 91, 85, 89, 88

Group 2: 76, 84, 80, 82, 83

We assume that the exam scores follow normal distributions with equal variances. Additionally, we assume that the exam scores can be considered independent and identically distributed (i.i.d.), within each group.

Question VII.1 (17)

What is the estimate of the pooled variance?

- 1 ☐ 9.00
- 2 ☐ 27.10
- 3 ☐ 11.25
- 4 ☐ 10.00
- 5 ☐ 8.00

Eksempel: Eksamensopgave fra Dec 2024

Question VII.2 (18)

What is the minimum number of observations required in each group (same number of observations in both groups) to achieve a power of 99% for detecting a difference in means of at least 4 between the two groups, assuming the variance is 20 (equal variances in both groups) and a significance level of 1%?

- 1 ☐ At least 56 (or 55, depending on calculation method)
- 2 ☐ At least 39 (or 38, depending on calculation method)
- 3 ☐ At least 62 (or 61, depending on calculation method)
- 4 ☐ At least 32 (or 31, depending on calculation method)
- 5 ☐ At least 82 (or 79, depending on calculation method)

Denne opgave krævede at man brugte Python (power beregning for 2 stikprøver). En relevant version for nuværende eksamen ville derfor fx være en opgave hvor noget Python output vises, eller en opgave hvor der regnes på kun én stikprøve (og man bruger Method 3.65, men får oplyst relevante fraktiler).

Hvilken Python funktion skal man bruge?

Hvad skal man angive `some effect_size`, `alpha`, `power` og `ratio`?

Tjekliste

Efter i dag skal du kunne:

- Beregne standard error for en differens
- Udføre t-test for to stikprøver - både Welch og pooled
- Redegøre for forskellen mellem Welch og pooled t-test
- Beregne konfidensinterval for en differens
- Redegøre for hvornår man bør udføre en parret t-test og kunne udføre denne
- Beskrive hvad der menes med Type 1 og Type 2 fejl.
- Beskrive hvad der menes med statistisk styrke (power)
- Beregne den nødvendige stikprøvestørrelse n , ud fra α , β , σ og $\mu_0 - \mu_A$
- Python: Benytte funktioner til at udføre diverse t-tests
- Python: Benytte funktioner til at beregne stikprøvestørrelse, power, mm.