02402 Statistik (Polyteknisk grundlag)

Uge 5: Hypotesetest

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

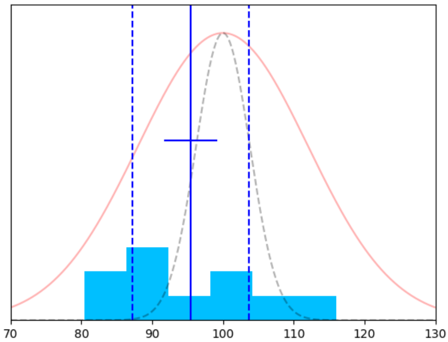
Dagsorden

1. Opsummering om konfidensintervaller
2. Hypotesetest
   - Nulhypotesen og den alternative hypotese
   - Teststørrelsen: t_{obs}
   - p-værdi
   - At påvise en *signifikant effekt*
   - Ét-sidet versus dobbelt-sidet test
3. Type I og Type II fejl
4. Modelkontrol
   - Q-Q plot
   - Transformation mod normalitet

Opsummering om konfidensintervaller

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Fra sidste uge: Statistisk inferens og konfidensintervaller



Den underliggende fordeling (rød linje) kender vi ikke, men vi kan beskrive den med en teoretisk model.

Stikprøve-gennemsnittet $\bar{X}$ følger en normalfordeling, $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ (sort stiplet linje), *hvis* den underliggende (røde) fordeling selv er en normalfordeling ELLER hvis n er stor nok (Central Limit Theorem) (så kan den røde fordeling se ud på alle mulige måder - den sorte vil stadig være normalfordelt).

Ud fra stikprøve data kan vi estimere $\hat{\mu} = \bar{x}$ og beregne et tilhørende konfidensinterval (de blå stiplede linjer).

Konfidensintervallet er et interval hvori vi tror på den "sande" værdi af μ befinder sig.

Bredden af konfidensintervallet afhænger af vores valg a signifikans niveau (α).

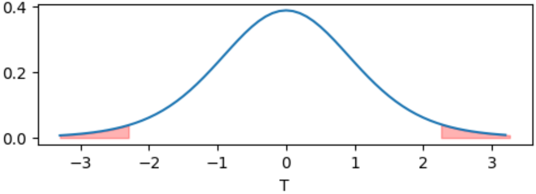
Fra sidste uge

Konfidensinterval og sigifikans niveau (1):

For at beregne konfidensinterval, betragtede vi størrelsen T:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

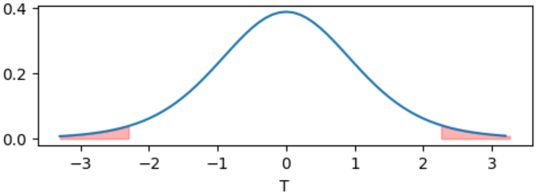
denne størrelse følger en t-fordeling: $T \sim t(n-1)$.



Det er mest sandsynligt at T antager værdier tæt på nul ($\bar{x}$ tæt på μ).
"Ekstreme" værdier (stor $|T|$) er mindre sandsynlige.

Fra sidste uge

Konfidensinterval og sigifikans niveau (2):



Værdierne $\pm t_{1-\alpha/2}$ er de værdier af T som afskærer de mest **ekstreme værdier** af T og dermed de værdier hvor $\bar{X}$ ligger længst væk fra μ.

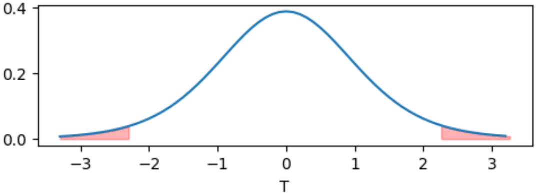
Hvis $\alpha = 0.05$, ligger de 95% *mindst ekstreme* værdier af T indenfor grænserne $\pm t_{1-\alpha/2}$, og dermed ligger de 5% *mest ekstreme* værdier af T udenfor $\pm t_{1-\alpha/2}$ (det røde område på plottet).

Kahoot!

(x2)

Fra sidste uge

Konfidensinterval og sigifikans niveau (3):



For at beregne konfidensintervallet skal vi vælge et **signifikansniveau** (α) og vi beregner da:

$$\left[\overline{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

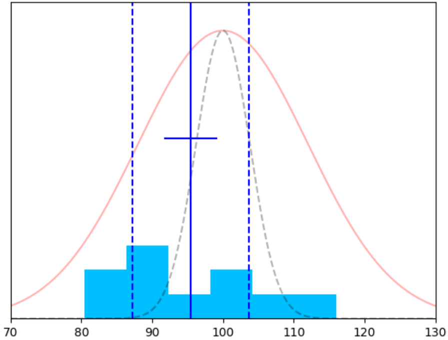
Større "konfidens" (mindre α) $\rightarrow$ bredere konfidensinterval

Hypotesetest

- Nulhypotesen og den alternative hypotese
- Teststørrelsen: t_{obs}
- p-værdi
- At påvise en *signifikant effekt*
- Ét-sidet versus dobbelt-sidet test

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Hypoteser og hypotesetests



Lad os nu sige vi tror på de nødvendige antagelser holder og vi har en stikprøve hvorfra vi estimerer $\hat{\mu}$ med tilhørende konfidensinterval.

På plottet er tegnet et 95% konfidensinterval (blå stiplede linjer).

Den underliggende fordeling (rød) kender vi ikke - den kunne ligge mere til højre/ventre og den kunne være bredere/smallere.

I dag skal vi tale om **hypoteser** og **tests**.

Man kunne fx have hypotesen $\mu = 120$. Lyder det som en god hypotese? Hvorfor/ikke?

Hvad med hypotesen $\mu = 90$?

Hvilke hypoteser vil vi "acceptere"? Og hvilke hypoteser vil vi "forkaste"?

Eksempel – letbane

Målinger af spændingsfald i letbane

I forbindelse med opførelsen af letbanen vil man undersøge spædningsfaldet et sted hvor man forventer at spændingsfaldet er nul (eller meget lille). Man foretager derfor 10 uafhængige målinger og får følgende resultater (i volt):

Stikprøve:

Måling	Spændingsfald
1	0.75
2	-0.85
3	4.23
4	2.12
5	3.04
6	0.53
7	-0.35
8	1.69
9	1.52
10	-0.42

Eksempel – letbane

Målinger af spændingsfald i letbane

I forbindelse med opførelsen af letbanen vil man undersøge spædningsfaldet et sted hvor man forventer at spændingsfaldet er nul (eller meget lille). Man foretager derfor 10 uafhængige målinger og får følgende resultater (i volt):

Stikprøve:

Måling	Spændingsfald
1	0.75
2	-0.85
3	4.23
4	2.12
5	3.04
6	0.53
7	-0.35
8	1.69
9	1.52
10	-0.42

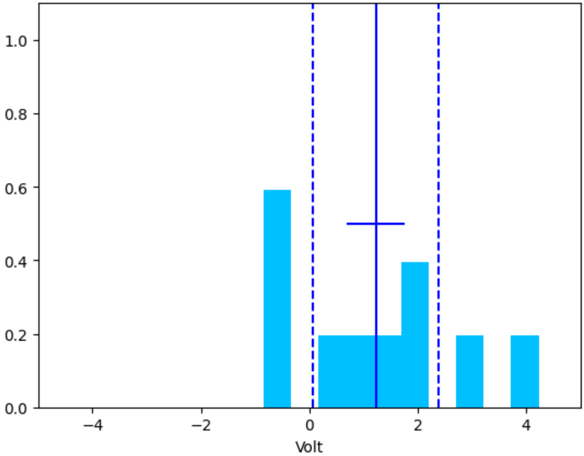
x_i = spændingsfald i måling nr. i
$n = 10$

$\bar{x} = 1.23$ (stikprøve-gennemsnit)
$s = 1.62$ (stikprøve-standardafvigelse)

$\hat{\mu} = 1.23$

$SE_{\hat{\mu}} = 1.62/\sqrt{10}$
95%-konfidensinterval for $\hat{\mu}$: $[0.07\,;2.4]$

Eksempel - letbane



Hvordan kunne den underliggende fordeling se ud?

Kunne den "sande" μ godt være 0?

Hypotesetest

Kunne den "sande" μ godt være 0?

Vi "tester" hypotesen ved at regne på *sandsynligheden* for at observere vores stikprøvedata(*) *hvis* nulhypotesen var sand.

(*) helt præcist udregner vi sandsynligheden for en tilsvarende *teststørrelse* (en stokastisk variabel).

02402 Statistik (Polyteknisk grundlag)

- Nulhypotesen og den alternative hypotese
- Teststørrelsen: t_{obs}
- p-værdi
- At påvise en *signifikant effekt*
- Ét-sidet versus dobbelt-sidet test

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Nulhypotesen og den alternative hypotese

Nulhypotesen:

Vi antager at μ har en bestemt værdi

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

hvor μ er det sande gennemsnit i populationen.

Nulhypotesen og den alternative hypotese

Nulhypotesen:

Vi antager at μ har en bestemt værdi

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

hvor μ er det sande gennemsnit i populationen.

Alternativ hypotese:

$$H_A: \ \mu \neq \mu_0$$

Nulhypotesen og den alternative hypotese

Nulhypotesen:

Vi antager at μ har en bestemt værdi

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

hvor μ er det sande gennemsnit i populationen.

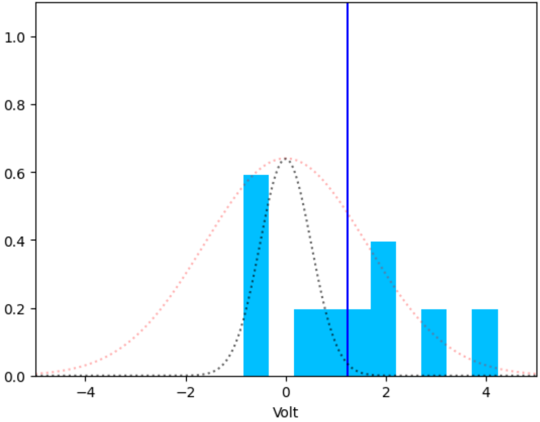
Alternativ hypotese:

$$H_A: \ \mu \neq \mu_0$$

Letbane-eksempel:

Fra data estimeres det gennemsnitlige spændingsfald: $\hat{\mu} = \bar{x} = 1.23$
$H_0 : \mu = 0$ (dvs spændingsfaldet svinger omkring gennemsnittet nul)
$H_A : \mu \neq 0$ (dvs gennemsnittet er ikke nul - og bør evt reguleres)

Eksempel - visualisering a nulhypotesen $\mu_0 = 0$



Her er tegnet en normalfordeling $N(0, s)$ (rød)
og en normalfordeling $N(0, \frac{s}{\sqrt{n}})$ (sort)
$(s = 1.62$ og $n = 10)$

02402 Statistik (Polyteknisk grundlag)

- Nulhypotesen og den alternative hypotese
- Teststørrelsen: t_{obs}
- p-værdi
- At påvise en *signifikant effekt*
- Ét-sidet versus dobbelt-sidet test

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Teststørrelse: t_{obs}

Nu vil vi finde en måde at "teste" vores nulhypotese:

Vi betragter størrelsen T under den **antagelse at nulhypotesen er sand**:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

denne størrelse følger en t-fordeling: $T \sim t(n-1)$.

Ud fra de observerede data beregner vi:

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

t_{obs} er vores **observation af den stokastiske variabel** T (under antagelse af at nulhypotesen er sand).

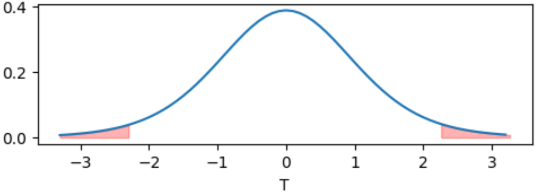
Spørgsmålet er om den observerede værdi t_{obs} er "for ekstrem"?

Student's t-test, eller bare t-test

Hypotesetest; t-test:

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

For at svare på om den observerede værdi t_{obs} er for "ekstrem", **tester** vi nu om t_{obs} ligger indenfor intervallet $[t_{\alpha/2}; t_{1-\alpha/2}]$ (hvilket afhænger af vores valg a signifikansniveau). Svaret på testen er Ja/Nej + angivelse af det valgte signifikansniveau α.



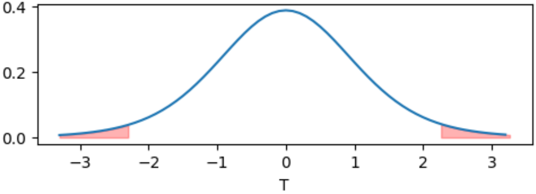
Hvis $|t_{obs}| > t_{1-\alpha/2}$ konkluderer vi at $|t_{obs}|$ er for stor til at vi tror på nulhypotesen - og vi siger at vi **forkaster nulhypotesen**.

Kristiske værdier for t_{obs}

Definition 3.31 The critical values

The $(1-\alpha)100\%$ critical values for the one-sample t-test are the $\alpha/2$- and $1-\alpha/2$-quantiles of the t-distribution with $n-1$ degrees of freedom

$$t_{\alpha/2} \text{ and } t_{1-\alpha/2}. \tag{3-23}$$



I ovenstående plot:

Hvad er de kritiske værdier $t_{\alpha/2}$ og $t_{1-\alpha/2}$? ____________________

Hvad visualiserer de røde arealer? ____________________

Eksempel – letbane

Hypotesen om spændingsfald $= 0$:

$$H_0: \; \mu = 0 \; (= \text{"}\mu_0\text{"})$$

hvor μ er det gennemsnitlige spændingsfald.

Eksempel – letbane

Hypotesen om spændingsfald $= 0$:

$$H_0: \ \mu = 0 \ \ (= "\mu_0")$$

hvor μ er det gennemsnitlige spændingsfald.

Udregn teststørrelsen

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.23 - 0}{1.62/\sqrt{10}} = 2.39$$

Eksempel – letbane

Hypotesen om spændingsfald $= 0$:

$$H_0: \ \mu = 0 \ \ (= \text{"}\mu_0\text{"})$$

hvor μ er det gennemsnitlige spændingsfald.

Udregn teststørrelsen

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.23 - 0}{1.62/\sqrt{10}} = 2.39$$

Sammenlign med kritiske værdier

(her vælges $\alpha = 0.05$):

$$[t_{\alpha/2}, t_{1-\alpha/2}] = [-2.26, 2.26]$$

Konklusion

$|t_{obs}| > t_{1-\alpha/2}$ og dermed *forkaster vi nulhypotesen*
(ved signifikansniveau 0.05).

t-test ud fra kristiske værdier

Method 3.32 The one-sample hypothesis test by the critical value

A null hypothesis is *rejected* if the observed test-statistic is more extreme than the critical values

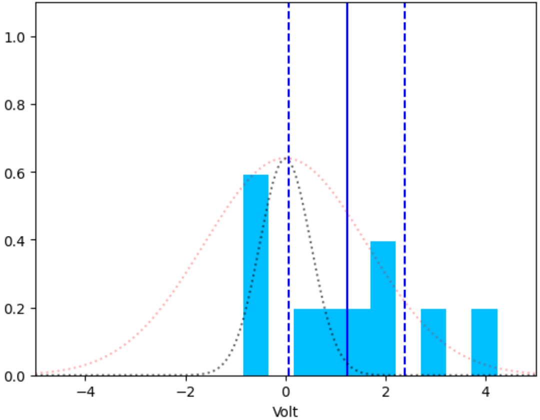
$$\text{If } |t_{\text{obs}}| > t_{1-\alpha/2} \text{ then } \textit{reject}, \tag{3-24}$$

otherwise *accept*.

Bemærk:

$|t_{obs}| > t_{1-\alpha/2}$, er det *samme* som at sige t_{obs} ligger udenfor intervallet $[t_{\alpha/2}; t_{1-\alpha/2}]$.

Eksempel - letbane (visualisering a nulhypotesen)



I dette tilfælde forkaster vi nulhypotesen.

Bemærk også at μ_0 ligger lige akurat udenfor konfidens-intervallet (95% konfidensintervallet for μ var $[0.07; 2.4]$)

Kahoot! (x2)

02402 Statistik (Polyteknisk grundlag)

- Nulhypotesen og den alternative hypotese
- Teststørrelsen: t_{obs}
- p-værdi
- At påvise en *signifikant effekt*
- Ét-sidet versus dobbelt-sidet test

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Eksempel - letbane

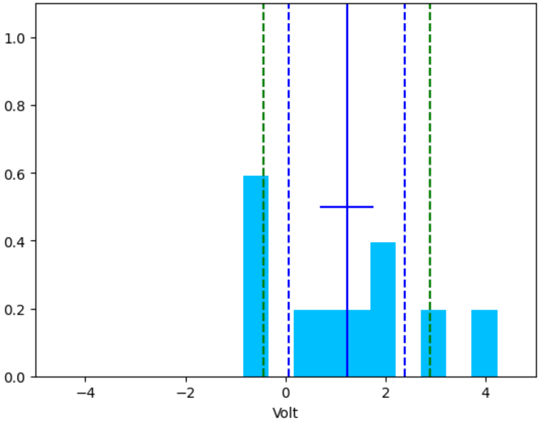
I vores eksempel med letbanen har vi $t_{obs} = 2.39$.

Men konklusionen på t-testen afhænger af valget af signifikansniveau (α):

For $\alpha = 0.05$:
Kritiske værdier for t_{obs}: $[-2.26, 2.26]$
95%-Konfidensinterval for μ: $[0.07; 2.4]$
Konklusion af t-test: nulhypotesen *forkastes*

For $\alpha = 0.01$:
Kritiske værdier for t_{obs}: $[-3.25, 3.25]$
99%-Konfidensinterval for μ: $[-0.43; 2.9]$
Konklusion af t-test: nulhypotesen *forkastes ikke*

Eksempel - letbane (visualisering a konfidensintervaller ved forskellig α



Vi kunne også spørge - hvor lille et **signifikansniveau (α)** skulle vi have valgt, *hvis* værdien $\mu = 0$ skulle ligge indenfor konfidensintervallet?

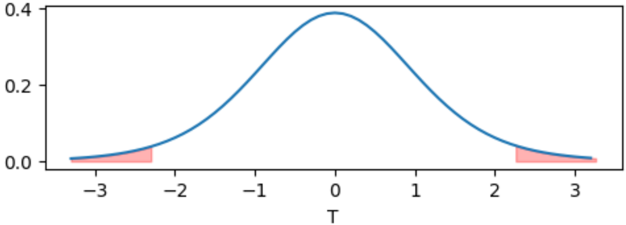
$\alpha = 0.05$ (blå), $\alpha = 0.01$ (grøn) ... ?

Sandsynligheden for at observere $T = t_{obs}$

Næste trin i t-testen er beregning af p-værdien:

Hvor lille signifikansniveau (α) skulle vi vælge for at t_{obs} ligger indenfor vores interval af accepterede T-værdier?

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



Vi skal dermed vælge grænserne præcis ved $T = \pm t_{obs}$

På den måde kommer t_{obs} lige akkurat med i intervallet, medens værdier af T der er "mere ekstreme" end t_{obs} ligger udenfor intervallet.

p-værdi for t-test med én (kvantitativ) stikprøve

Method 3.23 The one-sample t-test statistic and the p-value

For a (quantitative) one sample situation, the p-value is given by

$$p\text{-value} = 2 \cdot P(T > |t_{\text{obs}}|), \tag{3-20}$$

where T follows a t-distribution with $(n-1)$ degrees of freedom.
The observed value of the test statistics to be computed is

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \tag{3-21}$$

where μ_0 is the value of μ under the null hypothesis

$$H_0:\ \mu = \mu_0. \tag{3-22}$$

p-værdi sammenlignes med α

p-værdien skal sammenholdes med det ønskede signifikansniveau:

Det ønskede signifikansniveau udtrykkes ved værdien α
(fx $\alpha = 0.05$).

Hvis $p < \alpha$ *forkaster* vi nulhypotesen.

Hvis $p > \alpha$ *accepterer* vi nulhypotesen.

Eksempel – letbane

Hypotesen om intet spændingsfald:

$$H_0: \ \mu = 0 \ \ (= \text{"}\mu_0\text{"})$$

hvor μ er det gennemsnitlige spændingsfald.

Eksempel – letbane

Hypotesen om intet spændingsfald:

$$H_0: \ \mu = 0 \ \ (= "\mu_0")$$

hvor μ er det gennemsnitlige spændingsfald.

Udregn testværdien:

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.23 - 0}{1.62/\sqrt{10}} = 2.39$$

Eksempel – letbane

Hypotesen om intet spændingsfald:

$$H_0: \; \mu = 0 \; (= "\mu_0")$$

hvor μ er det gennemsnitlige spændingsfald.

Udregn testværdien:

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.23 - 0}{1.62/\sqrt{10}} = 2.39$$

Beregn p-værdien:

$$2P(T > 2.39) = 0.0404$$

Konklusion:

Hvis vi vælger $\alpha = 0.05$, får vi $p < \alpha \rightarrow$ Vi forkaster nulhypotesen.

Men havde vi valgt $\alpha = 0.04$ ville vi have accepteret nulhypotesen.

Definition og fortolkning af p-værdien (generelt)

Definition 3.22 The p-value

The p-value is the probability of obtaining a test statistic that is at least as extreme as the test statistic that was actually observed. This probability is calculated under the assumption that the null hypothesis is true.

"Evidens mod nulhypotesen":

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

Table 3.1: A way to interpret the evidence for a given p-value.

Trin i en t-test – Et overblik

Helt generelt består en t-test af følgende trin:

1. Formulér nulhypotesen (og modhypotesen) og vælg et signifikansniveau α (vælg "risikoniveauet").
2. Udregn værdien af teststørrelsen ud fra de observerede data.
3. Udregn p-værdien ud fra teststørrelsen holdt op imod den rette fordeling.
4. Sammenlign p-værdien med signifikansniveauet α og konkludér.

t-test i Python

Beregn t_{obs} "i hånden" og find p-værdien i den rette t-fordeling:

- `2*stats.t.cdf(-tobs, df=n-1))`

Man kan også bruge funktionen:

- `stats.ttest_1samp(...)` (fra `Scipy.stats` pakken)

Samme funktion kan desuden beregne et konfidensinterval direkte:

- `stats.ttest_1samp(...).confidence_interval(...)`

- Gå til Python notebook "voltage_drop.ipynb" i VS Code

Visual Studio Code

02402 Statistik (Polyteknisk grundlag)

- Nulhypotesen og den alternative hypotese
- Teststørrelsen: t_{obs}
- p-værdi
- At påvise en *signifikant effekt*
- Ét-sidet versus dobbelt-sidet test

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Sprogbrug om konklusion af en hypotesetest

Nulhypotesen

$H_0: \ \mu = \mu_0$

Den *alternative hypotese* (eller *Modhypotesen*)

$H_1: \ \mu \neq \mu_0$

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

En nulhypotese *afvises*, hvis p-værdi $< \alpha$ (α er valgt på forhånd).

Ellers siges nulhypotesen at være *'accepteret'*. Det er mere korrekt (langt at foretrække) at sige, at nulhypotesen ikke kan afvises.

Sprogbrug om konklusion af en hypotesetest

Definition 3.24 The hypothesis test

We say that we carry out a hypothesis test when we decide against a null hypothesis or not, using the data.

A null hypothesis is *rejected* if the p-value, calculated after the data has been observed, is less than some α, that is if the p-value $< \alpha$, where α is some pre-specified (so-called) *significance level*. And if not, then the null hypothesis is said to be *accepted*.

Remark 3.26

A note of caution in the use of the word *accepted* is in place: this should NOT be interpreted as having proved anything: *accepting* a null hypothesis in statistics simply means that we could not prove it wrong! And the reason for this could just potentially be that we did not collect sufficient amount of data, and *acceptance* hence proofs nothing at its own right.

En *"statistisk signifikant effekt"*

Ofte formuleres en nulhypotese så denne svarer til "ingen effekt".

Man udfører da en t-test og *hvis* nulhypotesen forkastes, konkluderer man omvendt at der er en *"statistisk signifikant effekt"*.

(derfor er de spændende resultater ofte når p-værdien er meget lille).

Definition 3.29 Significant effect

An *effect* is said to be *(statistically) significant* if the p-value is less than the significance level α. [a]

Lav din egen undersøgelse!

Nu er vi klædt på til at udtale os om "statistisk signifikante effekter"!

Prøv selv at finde på noget du vil undersøge:
Fomulér en nulhypotese, beregn t_{obs}, samt p-værdien og konkludér.

02402 Statistik (Polyteknisk grundlag)

- Nulhypotesen og den alternative hypotese
- Teststørrelsen: t_{obs}
- p-værdi
- At påvise en *signifikant effekt*
- Ét-sidet versus dobbelt-sidet test

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Ét-sidet versus dobbelt-sidet test

Indtil nu har det været underforstået, at testen er to-sidet (dobbeltsidet): (non-directional)

Alternativet til $H_0: \; \mu = \mu_0$ er $H_A: \; \mu \neq \mu_0$.

Ét-sidet versus dobbelt-sidet test

Indtil nu har det været underforstået, at testen er to-sidet (dobbeltsidet): (non-directional)

Alternativet til $H_0: \; \mu = \mu_0$ er $H_A: \; \mu \neq \mu_0$.

Der kan være andre situationer, f.eks. en-sidet (= directional) modhypoteser:

Alternativet til $H_0: \; \mu = \mu_0$ er $H_A: \; \mu < \mu_0$.

Ét-sidet versus dobbelt-sidet test

Indtil nu har det været underforstået, at testen er to-sidet (dobbeltsidet): (non-directional)

Alternativet til $H_0: \ \mu = \mu_0$ er $H_A: \ \mu \neq \mu_0$.

Der kan være andre situationer, f.eks. en-sidet (= directional) modhypoteser:

Alternativet til $H_0: \ \mu = \mu_0$ er $H_A: \ \mu < \mu_0$.

Vi holder os til den tosidet test (non-directional) i dette kursus!

Eksempel med t-test

Oftes bruges t-test til at teste om der er **forskel** på to ting.

Eksempel:
I et studie ønsker man at undersøge om der er signifikant forskel på indtag af kalorier i løbet af kvinders cyklus.

Man måler derfor **forskel i kalorieindtag** på 11 kvinder før og efter menstruation.

Målinger giver følgende data (forskel i kalorier): 1350, 1250, 1755, 1020, 745, 1835, 1540, 1540, 725, 1330, 1435

Hvad er en relevant nulhypotese?

Kahoot!

(x4)

Tror vi der er signifikant forskel?

Hvordan skal vi lave vores statistiske analyse af disse data?

Python:

- Gå til Python notebook "calorie_intake.ipynb" i VS Code

02402 Statistik (Polyteknisk grundlag)

Type I og Type II fejl

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Type I og Type II fejl

Der findes to slags fejl (dog kun én af gangen)

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

Type I fejl kaldes en falsk-positiv
Type II fejl kaldes en falsk-negativ
I dette sprogbrug er $H_0 =$ "negativ" ("ingen effekt")
og $H_1 =$ "positiv" (der er en "effekt")

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 person bliver stillet for en domstol under en specifik anklage.
Nul- og modhypotesen (den alternative hypotese) er:

H_0 : Personen er uskyldig.

H_1 : Personen er skyldig.

Retssalsanalogi

En person står stillet for en domstol:

En person bliver stillet for en domstol under en specifik anklage.
Nul- og modhypotesen (den alternative hypotese) er:

H_0 : Personen er uskyldig.

H_1 : Personen 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 hypotesetest

Theorem 3.39 Significance level and Type I error

The significance level α in hypothesis testing is the overall Type I risk

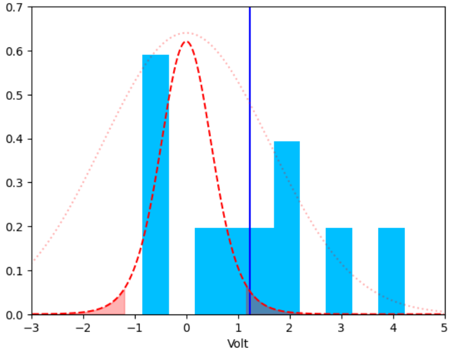
$$P(\text{"Type I error"}) = P(\text{"Rejection of } H_0 \text{ when } H_0 \text{ is true"}) = \alpha. \quad (3\text{-}41)$$

Fejlslutninger ved hypotesetest

Theorem 3.39 Significance level and Type I error

The significance level α in hypothesis testing is the overall Type I risk

$$P(\text{"Type I error"}) = P(\text{"Rejection of } H_0 \text{ when } H_0 \text{ is true"}) = \alpha. \qquad (3\text{-}41)$$

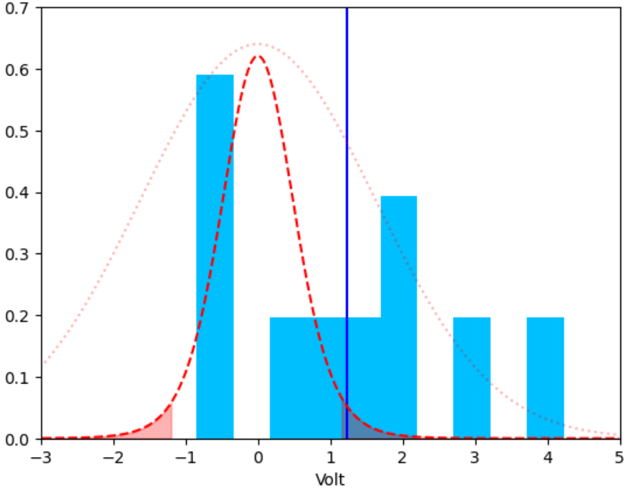


Fejlslutninger ved hypotesetest

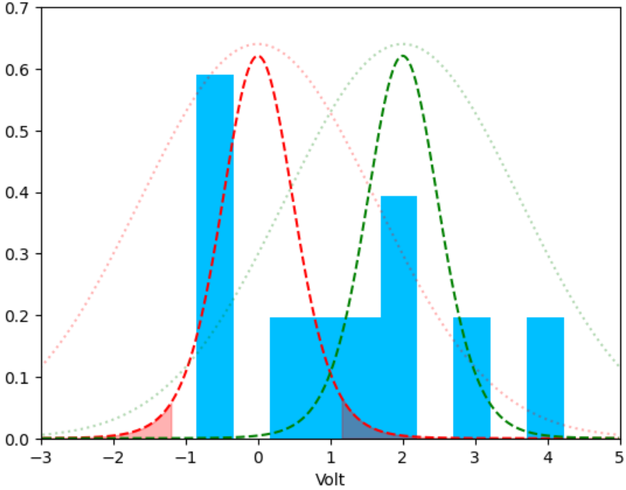
To mulige sandheder mod to mulige konklusioner:

	Afviser H_0	Afviser ikke H_0
H_0 er sand	Type I fejl (α)	Korrekt accept af H_0
H_0 er falsk	Korrekt afvisning af H_0	Type II fejl (β)

Type I fejl



Type I og Type II fejl



Mindre α = større β (og omvendt)

Modelkontrol

- Q-Q plot
- Transformation mod normalitet

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Modelkontrol

En statistisk model:

Et eksempel på en statistisk model kan være at data følger en normal fordeling med gennemsnit μ og varians σ^2:

$$X_i = \mu + \varepsilon_i \ , \quad \varepsilon_i \sim N(0, \sigma^2)$$

Men vi kan selvfølgelig også antage en anden fordeling.

Normalfordelings-antagelsen:

I statistik vil vi ofte antage at den underliggende fordeling er normalfordelt. I så fald bør vi tjekke om denne antagelse holder.

Vi kan få en ide om hvordan den underliggende fordeling ser ud - fx fra histogram af vores stikprøve data.

Python: Stikprøve fra underliggende normalfordeling

- Gå Python notebook "simulation_and_qqplot.ipynb" i VS Code

02402 Statistik (Polyteknisk grundlag)

- Q-Q plot
- Transformation mod normalitet

DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Modelkontrol

Det kan være svært at se fra et **histogram** om data er normalfordelt - især hvis stikprøven er lille. Derudover afhænger et histogram af valg af bin-size.

Man kan også vælge at plotte den **empiriske fordelingsfunktion** (ecdf).

Man kan også plotte et såkaldt **Q-Q plot** (quantile-quantile plot). Her plottes de sorterede observationer, $x_{(1)},\ldots,x_{(n)}$ mod de teoretiske fraktiler i normalfordelingen.
Der findes forskellige definitioner af fraktilerne:

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

(tilbage til simulering i Python)

Q-Q plot

Method 3.42 The Normal q-q plot

The ordered observations $x_{(1)}, \ldots, x_{(n)}$, called the sample quantiles, are plotted versus a set of expected normal quantiles $z_{p_1}, \ldots, z_{p_n}$. If the points are not systematically deviating from a line, we accept the normal distribution assumption. The evaluation of this can be based on some simulations of a sample of the same size.

The usual definition of $p_1, \ldots, p_n$ to be used for finding the expected normal quantiles is

$$p_i = \frac{i - 0.5}{n}, \; i = 1, \ldots, n. \tag{3-42}$$

Hence, simply the equally distanced points between $0.5/n$ and $1 - 0.5/n$. This formula is suitable for samples with $n > 10$ and can be used in Python by specifying `qqplot(...,a=1/2)`. For samples with $n \leq 10$, the formula

$$p_i = \frac{i - 3/8}{n + 1/4}, \; i = 1, \ldots, n, \tag{3-43}$$

which can be used in Python by specifying `qqplot(...,a=3/8)`, is preferred.

Q-Q plot

Q-Q plot er en visuel/grafisk inspektion af data.

Vi ahr ikke bevist at stikprøven kommer fra en normalfordeling.

Det kan være svært at afgøre om man synes punkterne i Q-Q plottet ligger tæt nok på den rette linje. For at have et sammenligningsgrundlag, kan det derfor være en god ide at sammenligne med simularet data (øvelse).

Kahoot!

02402 Statistik (Polyteknisk grundlag)

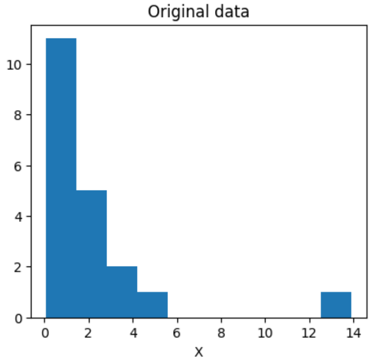
- Q-Q plot
- Transformation mod normalitet

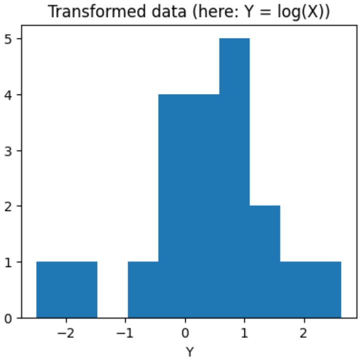
DTU Compute
Danmarks Tekniske Universitet
2800 Kgs. Lyngby

Transformation mod normalitet

Hvis data *ikke* er normalfordelt kan man ofte vælge at *transformere* sine data i håb om at de transformere data vil være mere normalfordelte.

I sådan et tilfælde vil man så udføre beregninger på transformerede data. Det kan blive relevant at transformere resultater tilbage til den oprindelige skala til sidst.





Transformation og fraktiler

Theorem 3.45 Transformations and quantiles

In general, the data transformations discussed in this section will preserve the quantiles of the data. Or more precisely, if f is a data transformation function (an increasing function), then

$$\text{The } p\text{th quantile of } f(Y) = f(\text{The } p\text{th quantile of } Y). \qquad (3\text{-}44)$$

Dagsorden

1. Opsummering om konfidensintervaller
2. Hypotesetest
   - Nulhypotesen og den alternative hypotese
   - Teststørrelsen: t_{obs}
   - p-værdi
   - At påvise en *signifikant effekt*
   - Ét-sidet versus dobbelt-sidet test
3. Type I og Type II fejl
4. Modelkontrol
   - Q-Q plot
   - Transformation mod normalitet

Tjekliste

Efter i dag skal du kunne:

- Forstå og beskrive konceptet med at vælge et signifikansniveau og udføre statistisk inferens (heruder beregne konfidensintervaller og udføre hypotesetest).
- DATA: Udføre en t-test, herunder beregne t_{obs} og p-værdi, og formulere en korrekt konklusion.
- PYTHON: Benyttet `stats.ttest_1samp(...)` (fra `Scipy.stats` pakken) til at udføre t-test og beregne konfidensintervaller.
- TEORI: Forstå baggrunden for udførelsen af en t-test, og dermed kunne forholde sig kritisk til dens validitet (og begrænsninger).
- Forstå og benytte almindelig statistisk sprogbrug, fx. termer som "statistisk signifikans", "effekt", "accept/forkastning af hypotese" og lign.
- Beskrive hvad der menes med Type 1 og Type 2 fejl.
- Forstå og vurdere et normal Q-Q plot.
- PYTHON: Tegne et normal Q-Q plot med `sm.qqplot(...)` (fra `Statsmodels` pakken))
- DATA: Vurdere om det er hensigtsmæssigt at *transformere* data, før der udføres en statistisk analyse

Øvelser Uge 5 (fortsættes næste slide)

3.4 **a)** "Find evidence against" dvs. beregn t_{obs} og tilsvarende p-værdi. Løs først opgaven baseret på oplysningerne om $\bar{x}$, s og n, og brug kun Python til at finde sandsynligheder og/eller fraktiler i t-fordelingen. Løs derefter opgaven igen, men denne gang baseret på værdierne i rå-data (stikprøven) og brug Pythons funktion `stats.ttest_1samp(...)`.

Brug til sidst tabel 3.1 til at "oversætte" resultaterne til en udtalelse om "evidence against H_0".

b) Brug Python til at finde sandsynligheder og/eller fraktiler i t-fordelingen. (supplér gerne med en tegning af t-fordeling og de kritiske værdier)

c) Løs først opgaven baseret på oplysningerne om $\bar{x}$, s og n, og brug kun Python til at finde sandsynligheder og/eller fraktiler i t-fordelingen. Løs derefter opgaven igen, men denne gang baseret på værdierne i rå-data (stikprøven) og brug Pythons funktion `stats.ttest_1samp(...).confidence_interval(...)`.

d) Brug gerne Python (hvis du har brug for det?).

Øvelser Uge 5 - fortsat

3.12 **a)** "Find evidence against" dvs. beregn t_{obs} og tilsvarende p-værdi. Løs først opgaven baseret på oplysningerne om $\bar{x}$, s og n (er beregnet sidste gang i Ex 3.11), og brug kun Python til at finde sandsynligheder og/eller fraktiler i t-fordelingen.
Løs derefter opgaven igen, men denne gang baseret på værdierne i rå-data (stikprøven) og brug Pythons funktion `stats.ttest_1samp(...)`.
Brug til sidst tabel 3.1 til at "oversætte" resultaterne til en udtalelse om "evidence against H_0".
b-c) Brug Python til at finde sandsynligheder og/eller fraktiler i t-fordelingen. (supplér gerne med en tegning af t-fordeling og de kritiske værdier)
d-g) Brug Python til at plotte.