Introduktion til Statistik

Forelæsning 5: Hypotesetest og modelkontrol - one sample

Peder Bacher

DTU Compute, Dynamiske Systemer
Bygning 303B, Rum 010
Danmarks Tekniske Universitet
2800 Lyngby – Danmark
e-mail: pbac@dtu.dk

Forår 2021

Kapitel 3: Hypotesetests for én gruppe/stikprøve

Grundlæggende koncepter:

- Hypoteser *(H_0 vs. H_1)*
- p-værdi *(Sandsynlighed for observeret eller mere ekstrem værdi af teststørrelsen, hvis H_0 er sand, e.g. $P(T > t_{\text{obs}})$)*
- Type I fejl *(I virkeligheden ingen effekt, men H_0 afvises)*
    - $P(\text{Type I}) = \alpha$ *(Sandsynligheden for at begå type I fejl)*
- Type II fejl *(I virkeligheden effekt, men H_0 afvises ikke)*
    - $P(\text{Type II}) = \beta$ *(Sandsynligheden for type II fejl)*
- Modelkontrol

Specifikke metoder, én gruppe:

- t-test for middelværdiniveau
- Modelkontrol med normal qq-plot

Chapter 3: One sample hypothesis testing

General concepts:

- Hypotheses *(H_0 vs. H_1)*
- p-value *(Probability for observing the test value or more extreme, if H_0 is true, e.g. $P(T > t_{\text{obs}})$)*
- Type I error *(No effect in reality, but H_0 is rejected)*
  - $P(\text{Type I}) = \alpha$ *(The probability for a Type I error)*
- Type II error: *(In reality an effect, but H_0 is not rejected)*
  - $P(\text{Type II}) = \beta$ *(The probability for a Type II error)*
- Model validation

Specific methods, one sample:

- t-test for the mean
- Model validation with normal q-q plot

Oversigt

1. One-sample t-test og p-værdi
2. p-værdier og hypotesetest
3. Kritisk værdi og konfidensinterval
4. Hypotesetests (helt generelt)
    - Hypotesetest med alternativer
    - Den generelle metode
    - Type I og type II fejl
5. Check af normalfordelingsantagelse
    - The normal q-q plot
    - Transformation towards normality

Spørgsmål om fordelingen af stikprøvegennemsnittet og standardisering (socrative.com - ROOM:PBAC)

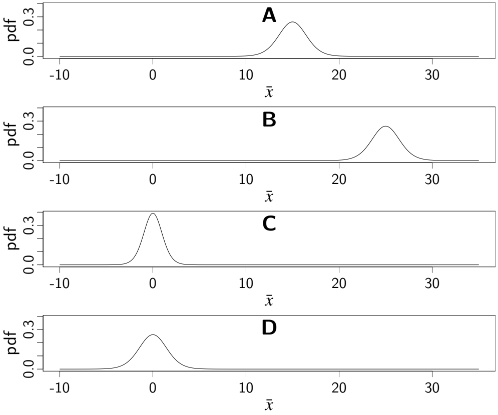
Hvilken pdf repræsenterer fordelingen af stikprøvegennemsnittet

$$\bar{X}$$

for

$$\mu = 15$$

(stikprøvestørrelse $n = 16$)
(stikprøvestandardafvigelse $s = 8$)



A B C eller D? Svar: A

Spørgsmål om fordelingen af stikprøvegennemsnittet og standardisering (socrative.com - ROOM:PBAC)

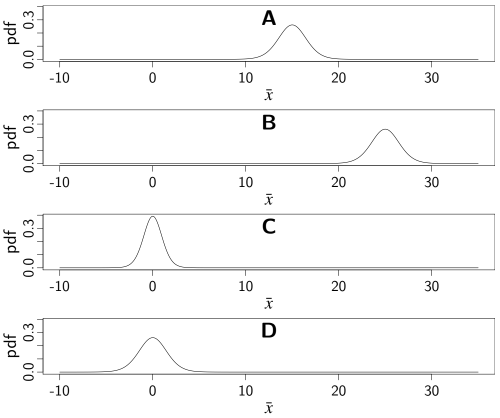
Hvilken pdf repræsenterer fordelingen af

$$\bar{X} - \mu$$

for

$$\mu = 15$$

(stikprøvestørrelse $n = 16$)
(stikprøvestandardafvigelse $s = 8$)



A B C eller D? Svar: D

Spørgsmål om fordelingen af stikprøvegennemsnittet og standardisering (socrative.com - ROOM:PBAC)

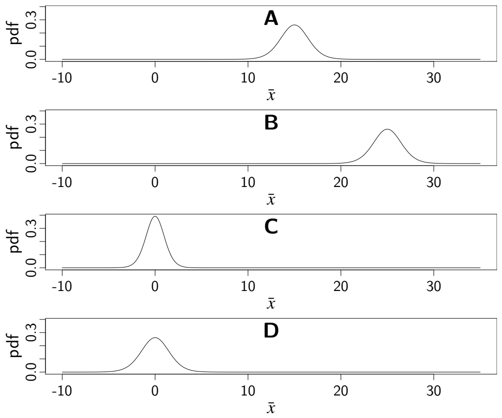
Hvilken pdf representerer fordelingen af

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

for

$$\mu = 15$$

(stikprøvestørrelse $n = 16$)
(stikprøvestandardafvigelse $s = 8$)



A B C eller D? Svar: C

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

Hvad er p-værdien og hvordan beregnes den?

Man fremsætter *nulhypotesen*

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

under hvilken man beregner *teststørrelsen*

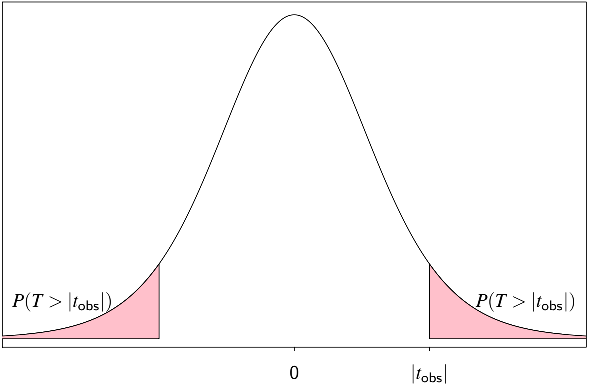
$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

som man derefter bruger til at beregne p-værdien

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

- p-værdien er altså: *Hvis nulhypotesen er sand, hvor sandsynligt er det da at få den observerede værdi af teststørrelsen (t_{obs}) eller mere ekstremt?*

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



Fortæller noget om: "*hvor sandsynligt er det at få det observerede data under H_0*" (dvs. hvis H_0 er sand)

Definition og fortolkning af p-værdien (HELT generelt)

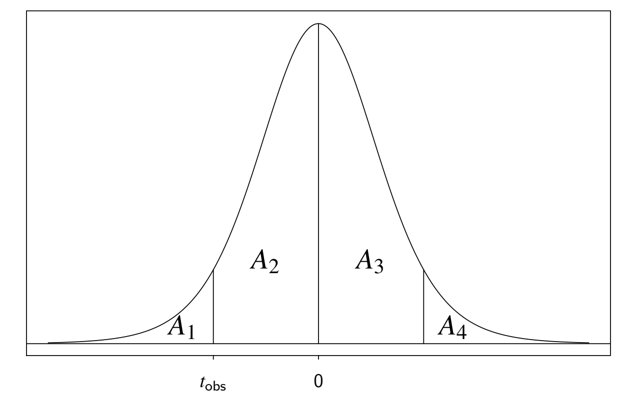
Definition 3.22 af p-værdien:

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.

p-værdien udtrykker *evidence* imod nulhypotesen – Tabel 3.1:

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

Spørgsmål om p-værdi (socrative.com - ROOM:PBAC)



Hvad er $2 \cdot P(T > |t_{\text{obs}}|)$?

A: $A_1 + A_2$ B: $A_3 + A_4$ C: $A_1 + A_4$ D: $A_2 + A_3$

Svar: C, husk p-værdi $= 2 \cdot P(T > |t_{\text{obs}}|)$ altså som $2 \cdot A_4$

Motiverende eksempel - sovemedicin

Forskel på sovemedicin?

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

Stikprøve, $n = 10$:

Person	$x =$ Beffekt - Aeffekt
1	1.2
2	2.4
3	1.3
4	1.3
5	0.9
6	1.0
7	1.8
8	0.8
9	4.6
10	1.4

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

Eksempel - sovemedicin

Hypotesen om ingen forskel på sovemedicin A og B ønskes undersøgt:

$$H_0: \ \mu = 0$$

Er data i overenstemmelse med nulhypotesen H_0?

Hvor "sandsynligt" er $\bar{x} = 1.67$ hvis $H_0 : \mu = 0$ er sand?

Beregne p-værdien:

Sandsynlighed for mere ekstremt data hvis H_0 er sand:

$$2 \cdot P(T > |t_{\text{obs}}|) = 2 \cdot P(T > 4.67) = 0.00117$$

Beregne teststørrelsen:

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.67 - 0}{1.13/\sqrt{10}} = 4.67$$

NYT: **Konklusion**:

Idet data er usandsynligt under H_0, så **forkaster** vi H_0 - vi har påvist en **signifikant effekt** af middel B ift. middel A.

Eksempel - sovemedicin manuelt i R

```
## Angiv data
x <- c(1.2, 2.4, 1.3, 1.3, 0.9, 1.0, 1.8, 0.8, 4.6, 1.4)
n <- length(x)
## Beregn den observerede t værdi - den observerede test statistik
tobs <- (mean(x) - 0) / (sd(x) / sqrt(n))
## Beregn p-værdien, som sandsynligheden for at få tobs eller mere ekstremt
pvalue <- 2 * (1-pt(abs(tobs), df=n-1))
pvalue

## [1] 0.00117
```

Eksempel - sovemedicin med indbygget funktion i R

```r
## Kald funktionen med data x
t.test(x)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 4.67, df = 9, p-value = 0.00117
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.861 2.479
## sample estimates:
## mean of x
##      1.67
```

Definition af hypotesetest og signifikans

Definition 3.24 Hypotesetest:

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 (i.e. p-value $< \alpha$), where α is some pre-specifed (so-called) *significance level*. And if not, then the null hypothesis is said to be *accepted*.

Definition 3.29 Statistisk signifikans:

An *effect* is said to be *(statistically) significant* if the p-value is less than the significance level α.
(OFTE bruges $\alpha = 0.05$)

Eksempel - sovemedicin

Konklusion for test af sovemedicin

Fortolkning med p-værdien.

Med $\alpha = 0.05$ kan vi konkludere:

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

Og dermed:

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

Spørgsmål om p-værdi (socrative.com - ROOM:PBAC)

```
## Kald funktionen med nul-hypotesen H_0: mu=1
t.test(x, mu=1)

##
##  One Sample t-test
##
## data:  x
## t = 1.87, df = 9, p-value = 0.0937
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##  0.861 2.479
## sample estimates:
## mean of x
##      1.67
```

Signifikansniveau $\alpha = 0.05$. Bliver H_0 afvist?

- A) $H_0 : \mu = 1$ afvises ikke og må accepteres
- B) $H_0 : \mu = 1$ afvises
- C) Ved ikke Svar A) den accepteres da p-værdien er 0.09 er større end $\alpha = 0.05$

Kritisk værdi

Definition 3.31 - de kritiske værdier for t-testet:

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

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

Metode 3.32: One-sample t-test vha. kritisk værdi:

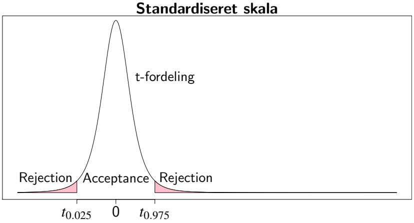
A null hypothesis is *rejected* if the observed test-statistic is more extreme than the critical values:

$$\text{If } |t_{\text{obs}}| > t_{1-\alpha/2} \text{ then } \textit{reject } H_0$$

otherwise *accept*.

Hypotesetests

Hvis t_{obs} er i acceptområdet, så accepteres $H_0 : \mu = \mu_0$



Kritisk værdi, konfidensinterval og hypotesetest

Theorem 3.33:
Kritisk-værdi-metode ækvivalent med konfidensinterval-metode

We consider a $(1-\alpha)\cdot 100\%$ confidence interval for μ

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

The confidence interval corresponds to the acceptance region for H_0 when testing the (non-directional) hypothesis

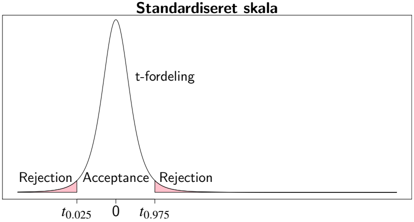
$$H_0: \quad \mu = \mu_0$$

(Ny) fortolkning af konfidensintervallet:

Nulhypoteser hvor μ_0 er udenfor konfidensintervallet ville være blevet afvist

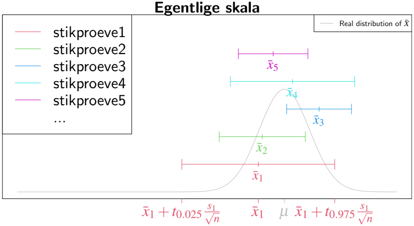
Hypotesetests

Hvis t_{obs} er ude af acceptområdet, så afvises $H_0 : \mu = \mu_0$



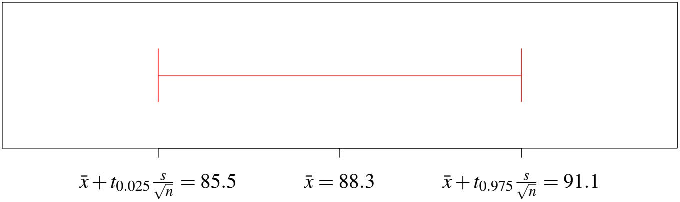
Konfidensintervallet

Nulhypoteser med μ_0 udenfor konfidensintervallet ville være blevet afvist



Spørgsmål om konfidensinterval (socrative.com - ROOM:PBAC)

Afgør på *signifikansniveau* $\alpha = 5\%$ om en type PC skærm lever op til specifikationen af et effektforbrug på $\mu = 83$ W. Der er taget en stikprøve af denne type skærm og et 95% konfidensinterval for middelværdien af effektforbruget μ er beregnet til:



Hvilken af følgende hypoteser skal testes og hvilken konklusion er korrekt?

- A) $H_0 : \mu = 0$ accepteres og signifikant højere effektforbrug er påvist
- B) $H_0 : \mu = 0$ afvises og signifikant højere effektforbrug er påvist
- C) $H_0 : \mu = 83$ accepteres og signifikant højere effektforbrug er ikke påvist
- D) $H_0 : \mu = 83$ afvises og signifikant højere effektforbrug er påvist
- E) Ved ikke Svar D) da $\mu = 83$ ligger udenfor og under konfidensintervallet

Den alternative hypotese

Den alternative hypotese H_1 er negationen af nulhypotesen H_0

Indtil nu - underforstået: (= non-directional)

Alternativet til $H_0: \ \mu = \mu_0$ er : $H_1: \ \mu \neq \mu_0$

MEN der kan være andre settings, f.eks. one-sided (=directional), less:

Alternativet til $H_0: \ \mu \geq \mu_0$ er $H_1: \ \mu < \mu_0$

I kurset er kun inkluderet opgaver med "non-directional"

Steps ved hypotesetests - et overblik

Helt generelt består et hypotesetest af følgende trin:

1. Formuler hypoteserne (H_0 og H_1) og vælg signifikansniveau α (choose the "risk-level")
2. Beregn med data værdien af teststatistikken
3. Beregn p-værdien med teststatistikken og den relevante fordeling, og sammenlign p-værdien med signifikansniveauet og drag en konklusion

   eller

   Lav konklusionen ved de relevante kristiske værdier

Metode 3.36: The level α one-sample t-test

1. Compute t_{obs} using Equation (3-21): $t_{\text{obs}} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
2. Compute the evidence against the *null hypothesis*

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

   vs. the *alternative hypothesis*

   $$H_1: \quad \mu \neq \mu_0,$$

   by the

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

   where the t-distribution with $n-1$ degrees of freedom is used
3. If p-value $< \alpha$: We reject H_0, otherwise we accept H_0,

   or

   The rejection/acceptance conclusion could alternatively, but equivalently, be made 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

Mulige fejl ved hypotesetests

To mulige sandheder vs. to mulige konklusioner:

	Reject H_0	Fail to reject H_0
H_0 is true	Type I error (α)	Correct acceptance of H_0
H_0 is false	Correct rejection of H_0	Type II error (β)

Eksempel - sovemedicin

To mulige sandheder vs. to mulige konklusioner:

	Reject H_0	Fail to reject H_0
Sand H_0: Ingen forskel på A og B	Type I fejl (α)	Korrekt accept af H_0
Falsk H_0: Forskel på A og B	Korrekt afvisning af H_0	Type II fejl (β)

Mulige fejl ved hypotesetests

Der findes to slags fejl (dog kun een af gangen!)

Type I: Rejection of H_0 when H_0 is true
Type II: Non-rejection of H_0 when H_1 is true

Risikoen for de to typer fejl kaldes sædvanligvis:

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

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

Theorem 3.39: "Signifikansniveauet" = "Risikoen for Type I fejl":

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

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

Eksempel: Retsalsanalogi

En person står stillet for en domstol:

A man is standing in a court of law accused of criminal activity.
The null- and the the alternative hypotheses are:

H_0 : The man is not guilty

H_1 : The man is guilty

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

Absence of evidence is NOT evidence of absence! Or differently put:

Accepting a null hypothesis is NOT a statistical proof of the null hypothesis being true!

Transport tid

Hypotesetests om studerendes transport tid til DTU fredag morgen

Tag link i meddelelse og indtast din transporttid i dag.

Kan det påvises at transporttiden for studerende på cykel er forskellig fra 20 minutter?

Kan det påvises at transporttiden for studerende på cykel er mere end 20 minutter?

Kan det påvises at tage mere end 20 minutter for studerende i bil?

Find de kritiske værdier for studerende i tog og bus.

Normaltfordelt data?

Teknikker til at undersøge om data kommer fra en normalfordeling:

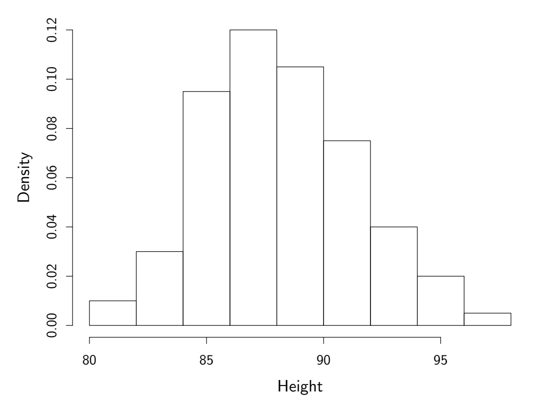
- Empirisk fordelings funktion (ecdf)
- Normal q-q plot

Transformer for at for mere normalfordelt data:

- Brug log-transformation til at få mere normalfordelte observationer

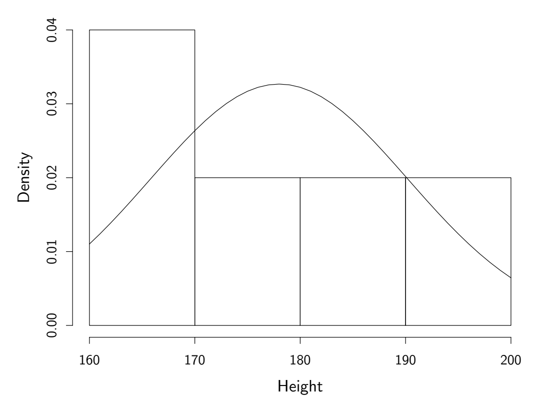
Se på 100 sim. observationer fra en normal fordeling

```
## 100 simulerede observationer fra normalfordeling
xr <- rnorm(100, mean(x), sd(x))
hist(xr, xlab="Height", main="", prob=TRUE)
lines(seq(130, 230, 1), dnorm(seq(130, 230, 1), mean(x), sd(x)))
```



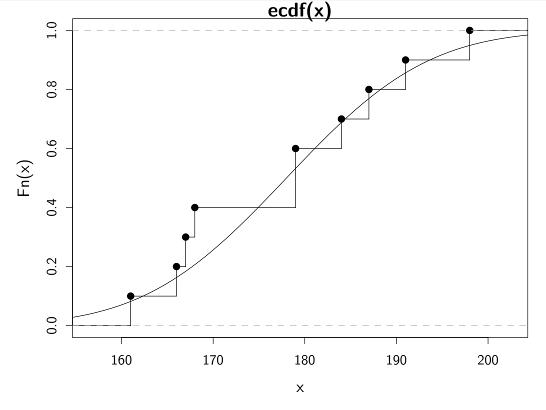
Højde af studerende - er de normalfordelt?

```
## Empirisk og teoretisk pdf af højdeeksempel
x <- c(168,161,167,179,184,166,198,187,191,179)
hist(x, xlab="Height", main="", prob=TRUE)
lines(seq(160, 200, 1), dnorm(seq(160, 200, 1), mean(x), sd(x)))
```



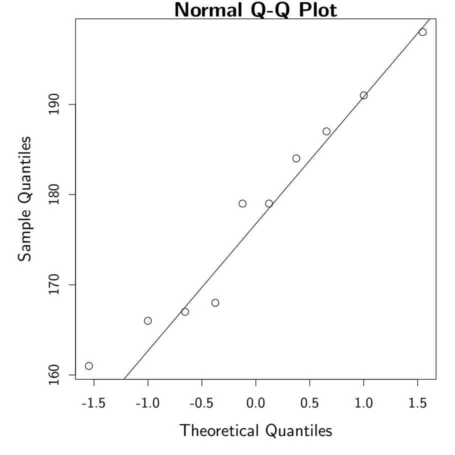
Højde af studerende - ecdf

```
## Empirisk og teoretisk fordelingsfunktion (ecdf og cdf)
plot(ecdf(x), verticals = TRUE)
xp <- seq(0.9*min(x), 1.1*max(x), length.out = 100)
lines(xp, pnorm(xp, mean(x), sd(x)))
```



Højde af studerende - Normal q-q plot

```
## q-q plot
qqnorm(x)
qqline(x)
```



Normal q-q plot

Metode 3.42 – Den formelle definition

The ordered observations $x_{(1)},\ldots,x_{(n)}$, called the sample quantiles, are plotted versus a set of expected normal quantiles $z_{p_1},\ldots,z_{p_n}$.

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

$$p_i = \frac{i-0.5}{n},\ i = 1,\ldots,n.$$

This is the default method in the `qqnorm` function in `R`, when $n > 10$, if $n \leq 10$ instead

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

is used.

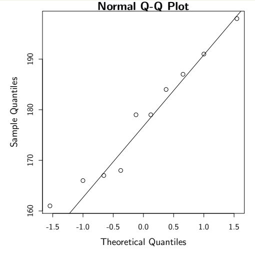
Normal q-q plot

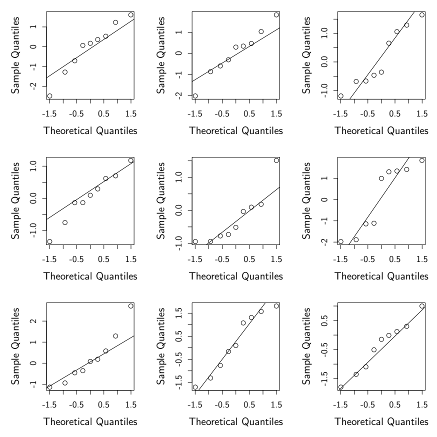
Vurder om de ligger på en ret linie:

Er observeret tydeligt forskellig simulerede normalfordelte stikprøver?

Simulerede:

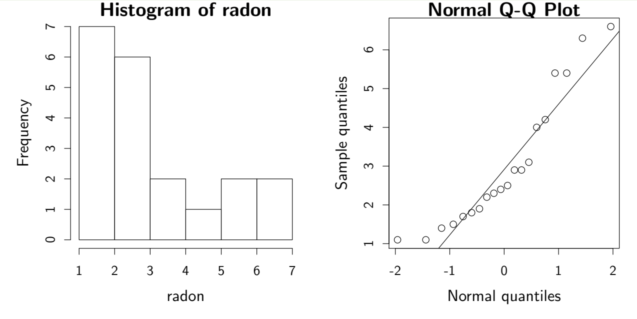
Observeret:





Eksempel - log-transformation af Radon data

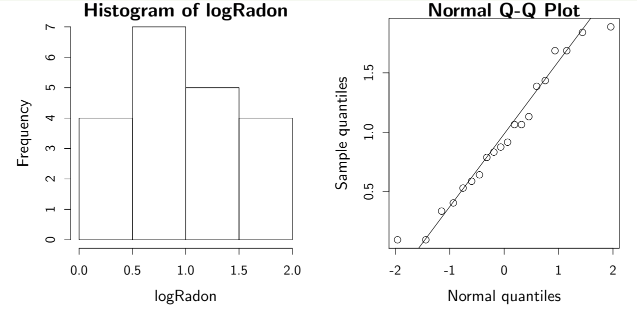
```r
## Reading in the data
radon<-c(2.4, 4.2, 1.8, 2.5, 5.4, 2.2, 4.0, 1.1, 1.5, 5.4, 6.3,
        1.9, 1.7, 1.1, 6.6, 3.1, 2.3, 1.4, 2.9, 2.9)
## A histogram and q-q plot
par(mfrow=c(1,2))
hist(radon)
qqnorm(radon,ylab = "Sample quantiles",xlab = "Normal quantiles")
qqline(radon)
```



Eksempel - Radon data - log-transformed are closer to a normal distribution

```r
## Transformer med naturlig logaritme
logRadon <- log(radon)

hist(logRadon)
qqnorm(logRadon, ylab="Sample quantiles", xlab="Normal quantiles")
qqline(logRadon)
```



- Midtvejsevaluering launches lige om lidt...skyd løs!
- Projekterne, går det fremad?