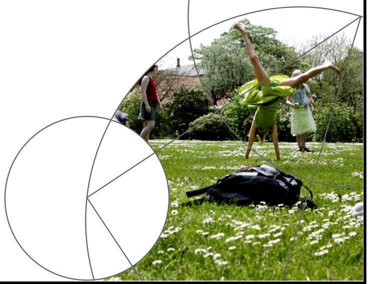
Faculty of Life Sciences

# *t*-test

## Hypoteser, teststørrelser og *p*-værdier

**Claus Ekstrøm**
E-mail: ekstrom@life.ku.dk





# Program

- Resumé og hængepartier fra sidst.
- Eksempel: effekt af foder på hormonkoncentration
  - repetition af statistisk model og konfidensinterval
  - test af hypotese
  - sammenhæng mellem konfidensintervaller og hypotesetest
- Hypotesetest: begreber og resume
- Lineær regression (stearinsyre og fordøjelighed): test af hypotese(r)
- Midtvejsevaluering ca. 9.40.





# Hormonkoncentration: data

Forsøg om effekt af fodertype på koncentration af hormon:

- Ni køer har fået foderet i en periode
- Hormonkoncentrationen målt før og efter
- Spørgsmål: har foderet en effekt på hormonkoncentrationen?

| Cow | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Initial ($\mu$g/ml) | 207 | 196 | 217 | 210 | 202 | 201 | 214 | 223 | 190 |
| Final ($\mu$g/ml) | 216 | 199 | 256 | 234 | 203 | 214 | 225 | 255 | 182 |
| Difference, $y$ | 9 | 3 | 39 | 24 | 1 | 13 | 11 | 32 | -8 |

Statistisk analyse:

- Statistisk model, parametre, estimation, konfidensinterval
- Test af hypotese





# Hormonkonc.: statistisk model og konfidensinterval

Lad os se på differencerne, $y_1, \ldots, y_9$.

- Statistisk model?
- Parametre? Estimater? Standard error?
- Konfidensinterval? Fortolkning?
- Konklusion: har foderet en effekt på hormonkoncentrationen?

# Hypotese

Hvis foderet ikke har nogen effekt, så er der ikke systematisk forskel på "før og efter" — dette svarer til at $\mu = 0$.

Vil derfor teste hypotesen (nulhypotesen)

$$H_0 : \mu = 0$$

Hypotesen er en ekstra restriktion på den statistiske model.

- Under modellen: $y_i \sim N(\mu, \sigma^2)$, uafhængige
- Hvis $H_0$ er sand: $y_i \sim N(0, \sigma^2)$, uafhængige





# Ideen i et hypotesetest

Hypotese $H_0 : \mu = 0$

Vi har estimatet — "bedste gæt" — $\hat{\mu} = \bar{y}$.

- Hvis $\hat{\mu} = \bar{y}$ ligger langt fra nul, tyder det på at $H_0$ er falsk.
- Hvis $\hat{\mu} = \bar{y}$ ligger tæt nul, tyder det ikke på at $H_0$ er falsk.

Men hvad er "langt fra" og hvad er "tæt på"?

- Værdien $\hat{\mu} = 13.78$ alene er ikke nok! Hvis vi målte i $\mu g/l$ i stedet ville vi have fået 0.01378 i stedet. Det lyder lille, men er jo helt den samme forskel.
- Skal tage højde for variationen i data!
- Skyldes forskellen i stikprøven en reel effekt eller skyldes den blot tilfældigheder? Hvad ville der ske hvis vi gentog eksperimentet?





# Ideen i et hypotesetest

Måler "langt fra"/"tæt på" følgende måde:

Hvis $H_0$ virkelig er sand — dvs. $\mu$ er nul — hvor sandsynligt er det så at få et $\hat{\mu}$ der ligger lige så langt eller længere fra nul end de 13.78 som vi faktisk fik?

- Hvis det er meget usandsynligt at få noget der passer dårligere med $H_0$ end det vi fik, så tyder det på at hypotesen er falsk.
- Hvis det er ret sandsynligt at få noget der passer dårligere end det vi faktisk fik, så tyder det ikke på at hypotesen er falsk.

Dette er grundtanken i hypotesetestet! Lad os være mere præcise...





# *t*-teststørrelsen

Statistisk model: $y_i \sim N(\mu, \sigma^2)$.

Husk at $\hat{\mu} = \bar{y}$ er normalfordelt med middelværdi $\mu$ og spredning $\sigma/\sqrt{n}$.

Hvis hypotesen $H_0 : \mu = 0$ er sand:

- $\hat{\mu} = \bar{y}$ er normalfordelt med middelværdi 0 og spredning $\sigma/\sqrt{n}$.
- Standardisér og erstat $\sigma$ med $s$:

$$T = \frac{\bar{y} - 0}{s/\sqrt{n}} = \frac{\bar{y} - 0}{\mathrm{SE}(\bar{y})} \sim t_{n-1}$$

Vi fik $\bar{y} = 13.78$ og $s = 15.25$. Så er $\mathrm{SE}(\bar{y}) = 15.25/\sqrt{9} = 5.08$ og

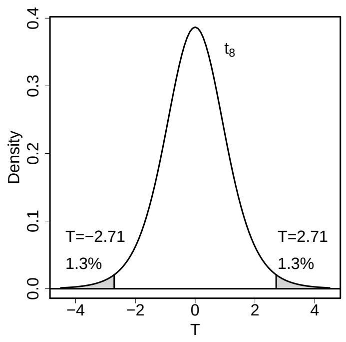
$$T_{\text{obs}} = \frac{13.78 - 0}{5.08} = 2.71$$

Kan nu bruge *t*-fordelingen til at sige om dette er langt fra eller tæt på nul!

## $p$-værdi og konklusion på test

$p$-værdien er sandsynligheden for at få en værdi af $T$ der ligger lige så langt eller længere væk fra nul end det vi fik:

$$p = P\big(|T| \geq |T_{\text{obs}}|\big) = P\big(|T| \geq 2.71\big) = 2 \cdot P\big(T \geq 2.71\big) = 0.026,$$




```
> pt(2.71,df=8)
[1] 0.986671
```

Hvis $H_0$ er sand er det altså ikke særligt sandsynligt at få en så stor værdi af $T$ som vi fik $\rightarrow$ $H_0$ afvises.

Skiller normalt ved 5%.

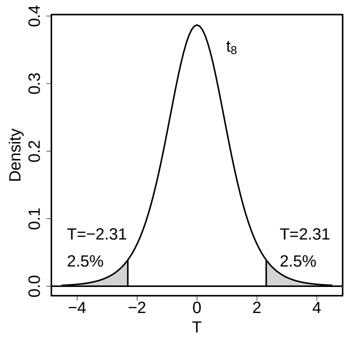
- $p < 0.05$: $H_0$ afvises
- $p \geq 0.05$: $H_0$ kan ikke afvises





## Signifikansniveau og kritiske værdier

Hvis vi bruger de 5% som signifikansniveau, så bliver det afgørende om $T_{\text{obs}}$ er større eller mindre end 97.5%-fraktilen i $t_{n-1}$-fordelingen.




```
> qt(0.975, df=8)
[1] 2.306004
```





## Hormonkoncentration: konklusion

Vi afviser hypotesen om at der ikke er en effekt.

Vi har med en vis sikkerhed påvist en effekt ($p = 0.026$).

Stigningen i hormonkoncentrationen estimeret til 13.78 med 95% konfidensinterval $(2.06, 25.49)$.





## Konfidensinterval og hypotesetest

I eksemplet gav konfidensintervallet og hypotesetestet samme konklusion:

- Nul ligger ikke i 95%-konfidensintervallet
- Vi afviser $H_0$ med en $p$-værdi mindre end 5%.

Sådan er det altid: nul ligger udenfor 95%-konfidensinterval hvis og kun hvis hypotesen $H_0 : \mu = 0$ kan afvises på 5% signifikansniveau.

- Nul ligger udenfor 95%-CI hvis og kun hvis $|\hat{\mu} - 0| > t_{0.975,n-1}\mathrm{SE}(\hat{\mu})$
- $H_0$ afvises hvis og kun hvis $|T| = \frac{|\hat{\mu}-0|}{\mathrm{SE}(\hat{\mu})} > t_{0.975,n-1}$.

# Hypotesetest: begreber og resumé

Hypotese

- Hypotese: simplificering af den statistiske model, restriktioner på parametrene. Her $H_0: \mu = 0$.
- Alternativ hypotese. Som regel blot det modsatte, her $H_A: \mu \neq 0$.

Teststørrelse og $p$-værdi

- Teststørrelse: Funktion af data der måler hvor godt data er i overensstemmelse med hypotesen.

  Her $T = \frac{\hat{\mu}-0}{\mathrm{SE}(\hat{\mu})}$ med både små og store værdier kritiske.
- $p$-værdi: sandsynligheden for — hvis $H_0$ er sand — at få en værdi af teststørrelsen der passer mindst lige så dårligt med hypotesen som den observerede værdi.

  Her:
  $p = P\big(|T| \geq |T_{\mathrm{obs}}|\big) = P\big(|T| \geq |T_{\mathrm{obs}}|\big) = 2 \cdot P\big(T \geq |T_{\mathrm{obs}}|\big).$





# Hypotesetest: generelt og resumé

Konklusion

- Afvise/ikke-afvise: vi afviser hypotesen hvis $p$-værdien er lille, typisk hvis $p < 0.05$. Hvis $p \geq 0.05$ kan vi ikke afvise hypotesen.
- Signifikansniveau som regel 5% — men vær ikke religiøs!
- Husk at kvantificere en evt. effekt: estimat og konfidensinterval.
- Samme konklusion fra test og konfidensinterval.

Type I og type II fejl:

| | Afvise | Ikke afvise |
|---|---|---|
| $H_0$ sand | type I | OK |
| $H_0$ falsk | OK | type II |

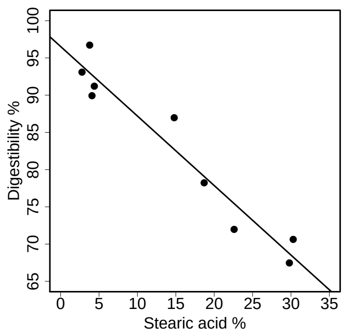
Hvis vi bruger signifikansniveau 5%, så laver vi type I fejl med 5% sandsynlighed!





# Lineær regression: stearinsyre og fordøjelighed

Statistisk model: $y_i = \alpha + \beta \cdot x_i + e_i$ hvor $e_1, \ldots, e_9 \sim N(0, \sigma^2)$




Vil teste hypotesen at der ikke er sammenhæng mellem stearinsyre-indhold og fordøjelighed.

- Hvordan "ser den rette linie ud" under hypotesen?
- Hvad er hypotesen, udtrykt ved parametrene i modellen?





# Lineær regresison: test for ingen sammenhæng

Hvad er:

- hypotesen, den alternative hypotese?
- teststørrelsen, $p$-værdien?
- konklusionen?

```
> model1 <- lm(ford~ssyre)
> summary(model1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 96.53336    1.67518   57.63 1.24e-10 ***
ssyre       -0.93374    0.09262  -10.08 2.03e-05 ***

Residual standard error: 2.97 on 7 degrees of freedom
```

## Lineær regression: test af en anden hypotese

En (opdigtet) fysiologisk teori siger at den forventede fordøjelighed er 78% for et stearinsyreindhold på 20%

NB: lidt anderledes end i noterne.

Undersøg om data i modstrid med denne teori!

- Forventet fordøjelighed ved 20% stearinsyre? Estimat?
- Hvad er hypotesen?
- Teststørrelse? $p$-værdi?
- Konklusion?





## Resumé: $t$-test

- Hypotese, $H_0 : \theta = \theta_0$ hvor $\theta$ er en parameter eller en kombination af parametre, og $\theta_0$ er en fast værdi.
  - Fx. $\mu = 0$ eller $\beta = 0$ eller $\alpha + \beta \cdot 20 = 78$.
- Alternativ hypotese, $H_A : \theta \neq \theta_0$
- Teststørrelse,

$$T = \frac{\hat{\theta} - \theta_0}{\mathrm{SE}(\hat{\theta})} \sim t_{n-p}$$

  hvor $p$ er antal middelværdiparametre i modellen
- $p$-værdi:
  $p = P\big(|T| \geq |T_{\text{obs}}|\big) = P\big(|T| \geq |T_{\text{obs}}|\big) = 2 \cdot P\big(T \geq |T_{\text{obs}}\big)$
- 95%-konfidensinterval indeholder præcis de værdier $\mu_0$ for hvilke hypotesen $H_0 : \theta \neq \theta_0$ ikke vil blive afvist på 5% signifikansniveau.
- Husk at kvantificere resultaterne: $\hat{\theta}$ og 95%-konfidensinterval.





## Dagens hovedpunkter

- Hypoteser: restriktion af parametre i en model
  - Nulhypotese, alternativ hypotese
- Hvordan tester man en hypotese?
- Sammenhæng mellem hypotese-test og konfidensinterval
- Fortolkning af $p$-værdien.





## Ordliste

| Engelsk | Dansk |
|---|---|
| alternative hypothesis | alternativ hypotese |
| critical values | kritiske værdier |
| hypothesis test | hypotesetest |
| (null) hypothesis | (nul)hypotese |
| quantile | fraktil |
| $p$-value | $p$-værdi |
| signifance level | siknifikansniveau |
| test statistic | teststørrelse |
| reject | afvise |