

Basal statistik  
Esben Budtz-Jørgensen  
8. november 2011

# Forsøgsplanlægning Stikprøvestørrelse

1–51

# Planlægning af et studie

- Videnskabelig hypotese
- Endpoints
- Instrumentelle/eksponerings variable
- Variationskilder
- Design af studie
  - Eksperimentelt eller observationelt
  - Tværsnits- eller longitudinel undersøgelse
  - Prospektivt eller retrospektivt

## Variationskilder

- Observatørvariation
- Instrumentvariation
- Målefejl
- Dag-til-dag variation indenfor individer
- Biologisk variation mellem individer
- Tilfældig variation

## Stikprøvestørrelse

“Hvor mange observationer skal vi bruge?”

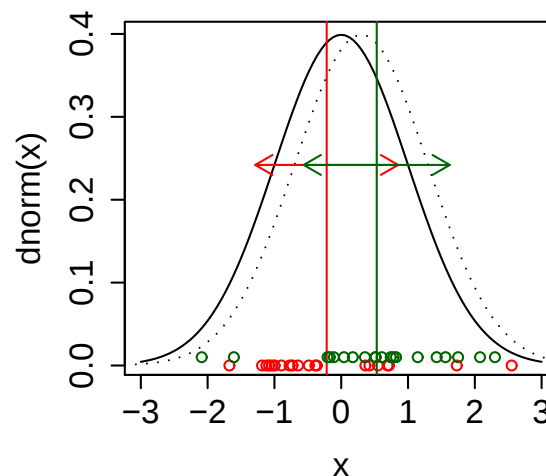
Det kommer an på

- Forsøgsplanen
- Hvor store effekter man leder efter
- Hvor sikker man vil være på at finde dem

## Repetition: Sammenligning af to grupper

To uafhængige stikprøver:  $x_{1,1}, \dots, x_{1,n_1}$  og  $x_{2,1}, \dots, x_{2,n_2}$

Model:  $x_{1,1}, \dots, x_{1,n_1} \sim N(\mu_1, \sigma^2)$  og  $x_{2,1}, \dots, x_{2,n_2} \sim N(\mu_2, \sigma^2)$

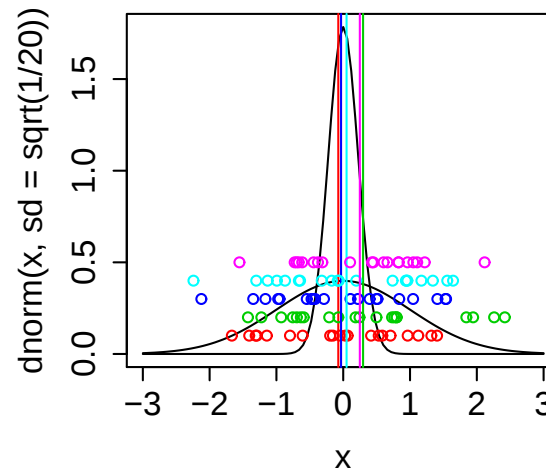


En normalfordeling i hver gruppe, men middelværdien kan være forskellig

# Estimation

$\mu_1$  og  $\mu_2$  angiver populationsmiddelværdier. Disse parametre er ukendte, men på baggrund af data kan vi give et skøn (estimat) over deres værdi:

$$\hat{\mu}_1 = \bar{x}_1 = (x_{1,1} + \cdots + x_{1,n_1})/n_1, \quad \hat{\mu}_2 = \bar{x}_2 = (x_{2,1} + \cdots + x_{2,n_2})/n_2$$



## $t$ -testet

Teststørrelsen

$$t = \frac{\bar{x}_2 - \bar{x}_1}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

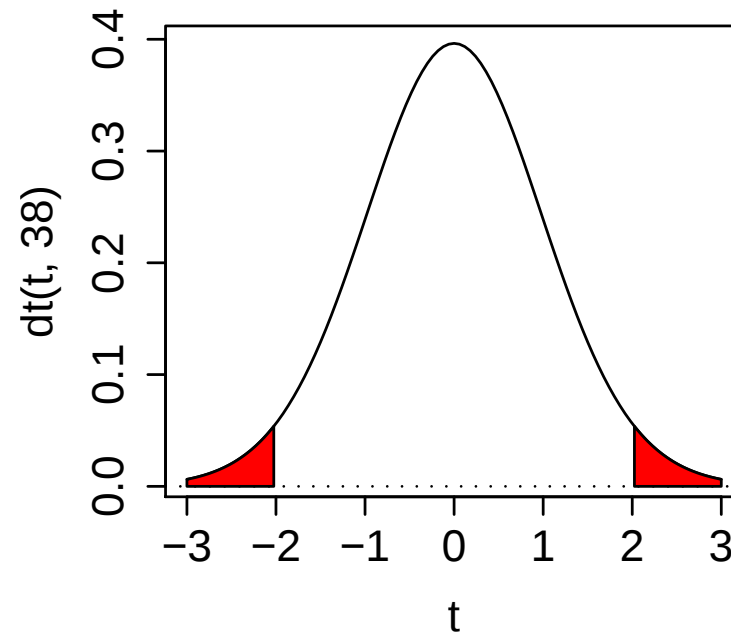
hvor  $s^2$  er et poolet varians estimat:

$$s^2 = \hat{\sigma}^2 = \frac{n_1-1}{n_1+n_2-2} s_1^2 + \frac{n_2-1}{n_1+n_2-2} s_2^2$$

Hvis hypotesen er sand følger teststørrelsen en  $t$ -fordeling med  $n_1 + n_2 - 2$  frihedsgrader.

$H_0$  forkastes når  $|t|$  er større end ca. 1.96

## $t$ -testet: p-værdien



*Hvis* der ingen gruppeforskel er, så er der lille sandsynlighed for at få en  $t$ -værdi som er mindre end -1.96 eller større end 1.96.

## Type I og Type II fejl

Et hypotesetest kan give forkert udslag på to måder:

Type I fejl: Forkaste nulhypotesen selv om den er sand

Type II fejl: Acceptere nulhypotesen selv om den er forkert

Sandsynlighed for fejl:  $\alpha$  resp.  $\beta$

$\alpha$ : ssh for type I fejl, Signifikansniveau (0.05, fx.)

$\beta$ : ssh for type II fejl

$1 - \beta$ : Styrke - sandsynlighed for at finde forskellen

Bemærk. Styrken afhænger af hvor stor den faktiske forskel er. Hvis forskellen er stor, er der en stor chance for at vi forkaster hypotesen.

## Simulationsstudie

I SAS simuleres observationer fra model uden forskel, dvs

Model:  $x_{1,1}, \dots, x_{1,25} \sim N(0, 1)$  og  $x_{2,1}, \dots, x_{2,25} \sim N(0, 1)$

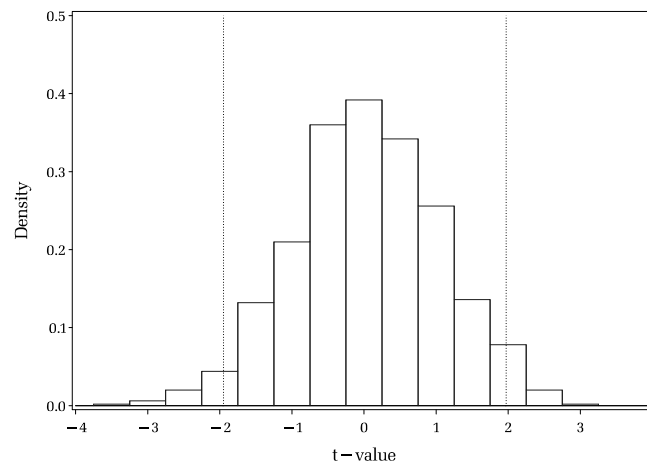
Der laves 1000 datasæt. I hvert beregnes  $t$  og gemmes.

Bemærk, at middelværdien i begge grupper er 0, så her er hypotesen sand.

# Simulationsstudie

Model:  $x_{1,1}, \dots, x_{1,25} \sim N(0, 1)$  og  $x_{2,1}, \dots, x_{2,25} \sim N(0, 1)$

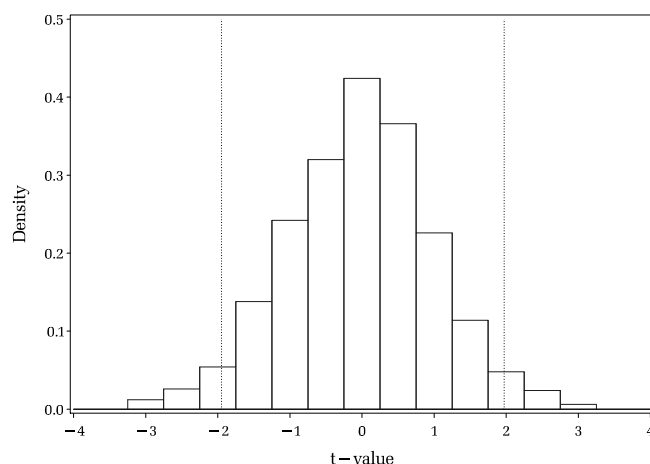
Fordeling af  $t$ :



Sandsynlighed for forkast:  $0.031 + 0.024 = 0.055$

## Hvad sker der hvis observationsantallet fordobles?

Model:  $x_{1,1}, \dots, x_{1,50} \sim N(0, 1)$  og  $x_{2,1}, \dots, x_{2,50} \sim N(0, 1)$



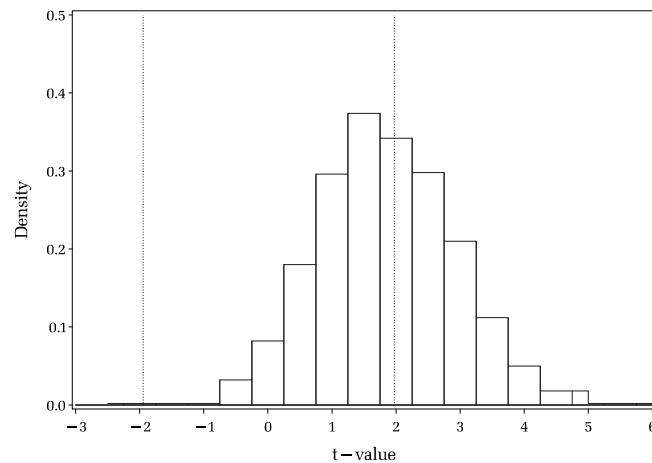
Sandsynlighed for forkast:  $0.023 + 0.030 = 0.0530$

**Fordelingen er (stort set) uændret:**

Risiko for type 1 fejl afhænger ikke af observationsantallet. Den er konstant 5% (signifikansniveau).

## Nu introduceres en niveauforskel

Model:  $x_{1,1}, \dots, x_{1,25} \sim N(0, 1)$  og  $x_{2,1}, \dots, x_{2,25} \sim N(0.5, 1)$



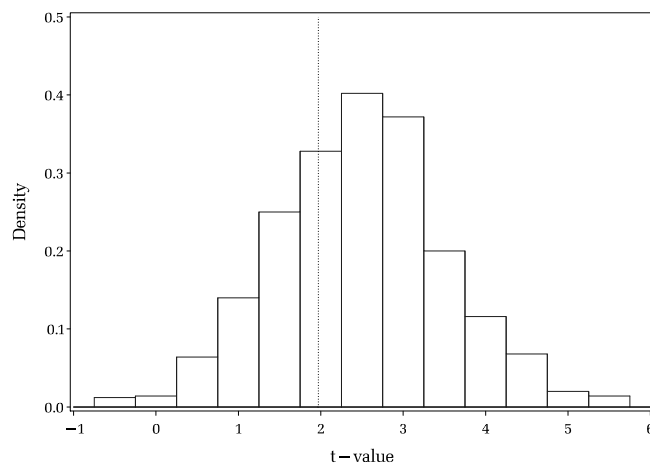
Bemærk: fordelingen er flyttet til højre

Sandsynlighed for forkast:  $0 + 0.439 = 0.439$

dvs styrken er ca. 44%

$n$  fordobles

Model:  $x_{1,1}, \dots, x_{1,50} \sim N(0, 1)$  og  $x_{2,1}, \dots, x_{2,50} \sim N(0.5, 1)$

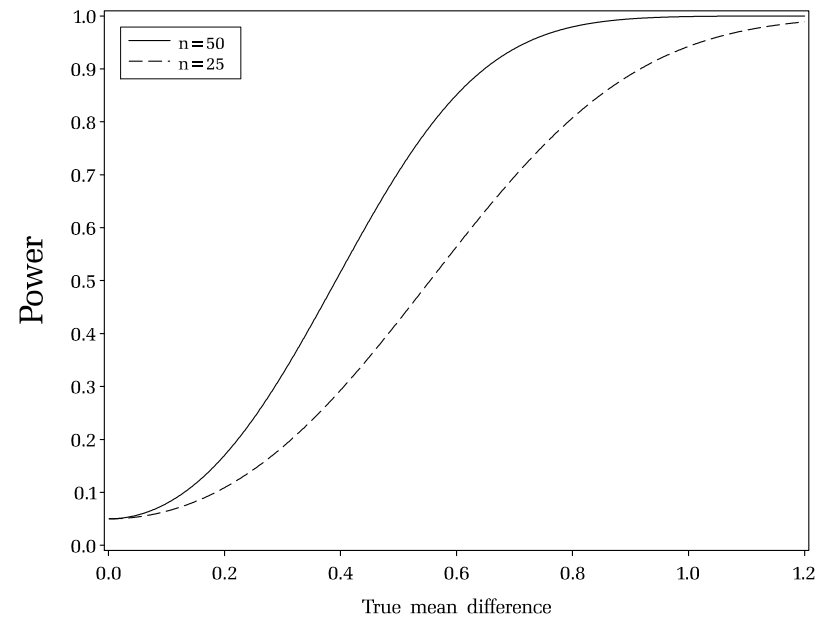


Bemærk: fordelingen er flyttet endnu mere til højre

Sandsynlighed for forkast:  $0 + 0.693 = 0.693$

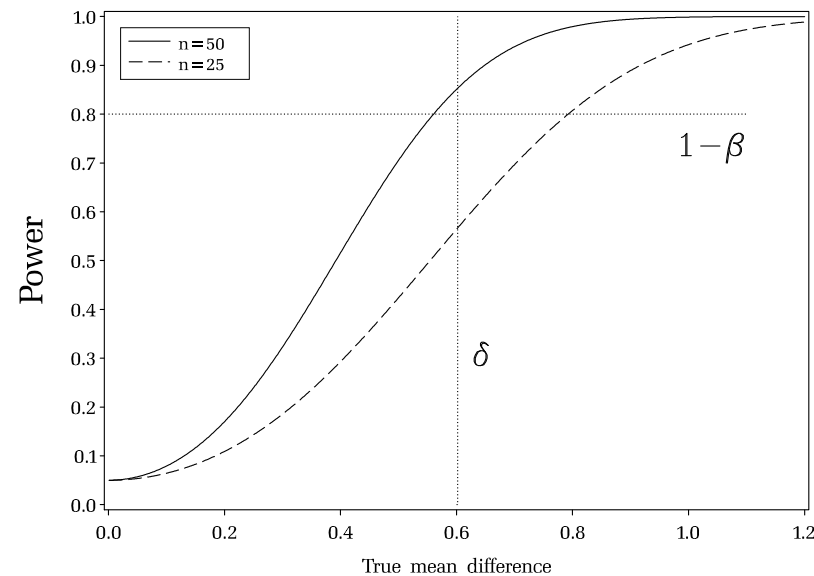
Styrken afhænger af observationsantallet

## Styrken som funktion af observationsantal og effekt



Styrken vokser som funktion af den *sande forskel mellem grupperne* og *antallet af observationer*

## Hvor mange observationer skal man bruge?



Styrken ( $1 - \beta$ ) og effekten ( $\delta$ ) fastsættes. Herefter kan man beregne hvor mange observationer der er nødvendige. Det svarer til at finde den kurve der går gennem skæringspunktet på tegningen - dvs et sted mellem 25 og 50 (i hver gruppe).

## Matematisk formel for $n$

(NB: Disse antager at spredningen er kendt. Pas på med meget små  $n$ )

Generel formel for vilkårlige  $\alpha$  og  $1 - \beta$ :

$$\begin{aligned} n &= 2 \times (z_{1-\alpha/2} + z_{1-\beta})^2 \times (\sigma/\delta)^2 \\ &= 2 \times f(\alpha, 1 - \beta) \times (\sigma/\delta)^2 \quad \text{næste side} \end{aligned}$$

- jo *større* gruppe-forskellen er jo *mindre* er det nødvendige antal observationer
- jo *større* variation indenfor grupper jo *større* er det nødvendige antal observationer

## En nyttig tabel

| Power: $1 - \beta$ |       |       |       |      |
|--------------------|-------|-------|-------|------|
| $\alpha$           | 0.95  | 0.9   | 0.8   | 0.5  |
| 0.1                | 10.82 | 8.56  | 6.18  | 2.71 |
| 0.05               | 12.99 | 10.51 | 7.85  | 3.84 |
| 0.02               | 15.77 | 13.02 | 10.04 | 5.41 |
| 0.01               | 17.81 | 14.88 | 11.68 | 6.63 |

$$f(\alpha, 1 - \beta) = (z_{1-\alpha/2} + z_{1-\beta})^2$$

(Fraktiler i normalfordelingen  $z_p = \Phi^{-1}(p)$ ,  $z_{0.975} = 1.96$ , osv.)

## Eksempel

Antag vil finde en forskel på 5 ( $\delta = 5$ )

Vi være 90% procent sikre på at finde forskellen ( $1 - \beta = 0.90$ )

når vi tester på et 5%'s niveau ( $\alpha = 0.05$ )

Spredning inden for grupper er 10 ( $\sigma = 10$ )

Indsæt i formel

$$n = 2 \times f(0.05, 0.9) \times (\sigma/\delta)^2 = 2 \times 10.51 \times (10/5)^2 = 84.1$$

Det vil sige 85 i hver gruppe.

## I SAS Analyst

Statistics → Sample Size → Two-Sample t-test

I vinduet specificeres herefter:

**Calculate:** N per group (alternativt bestemmes styrken for fast N)

**Group 1 mean:** 0, **Group 2 mean:** 5

(eller 95 og 100, kun forskellen betyder noget)

**Standard deviation:** 10, **Alpha:** 0.05

**Power: From:** 0.8 **To:** 0.95 **By:** 0.05

**Tails:** 2-sided

# Output

Two-Sample t-Test  
Group 1 Mean = 0      Group 2 Mean = 5  
Standard Deviation = 10      Alpha = 0.05      2-Sided Test

| Power | N per Group |
|-------|-------------|
| 0.800 | 64          |
| 0.850 | 73          |
| 0.900 | 86          |
| 0.950 | 105         |

Bemærk at det nødvendige antal observationer ( $n$ ) vokser som funktion af den ønskede styrke.

For Power= 90% ses  $n=86$ . Før fik vi 85? Forskellen skyldes den at 'vores' formel ser bort fra estimationsusikkerheden for spredningen  $\sigma$ . Herved bliver  $n$  lidt for lille.

# Opsummering

Skal specificere

- Klinisk relevant forskel  $\delta$
- Standardafvigelsen indenfor grupper ( $\sigma$ )
- Signifikansniveau  $\alpha$  (ofte 5%)
- Styrke  $1 - \beta$  (ofte 80% eller 90%)

Hvordan fastsættes standardafvigelsen?

- Lignende undersøgelser
- Pilotforsøg
- Start forsøget og se på de første observationer
- Specificer *standardiseret forskel*  $\delta/\sigma$

## Eksempel

Påvirker mælkeindtaget børns højde?

5-års børn i to grupper:

- Ekstra mælk i et år
- Kontrolgruppe

Børn på 5 år vokser ca. 6 cm/år med en SD på 2 cm.

Vil kunne opdage en forskel på 0.5 cm.

Signifikansniveau 1%, styrke 90%.

$$n = 2 \times (2\text{cm}/0.5\text{cm})^2 \times f(0.01, 0.9) = 2 \times 4^2 \times 14.88 = 476 \text{ pr. gruppe}$$

## Parret $t$ -test

Niveauet i to grupper sammenlignes ved at teste om differenserne indenfor par ( $D_i = X_{1i} - X_{2i}$ ) har middelværdi 0.

$n$  kan beregnes ved

$$n = f(\alpha, 1 - \beta) \times (\sigma/\delta)^2$$

men bemærk at

- $n$  angiver antallet af *par*
- $\sigma$  angiver standard deviationen for differenser

## $n$ ved sammenligning af hyppigheder

$$n = 2 \times f(\alpha, 1 - \beta) \times \frac{\bar{p}(1 - \bar{p})}{(p_2 - p_1)^2}$$

$p_1$  hyppighed i gruppe 1

$p_2$  hyppighed i gruppe 2

$\bar{p} = (p_1 + p_2)/2$

$n$  antal i hver gruppe

Baseret på normalfordelingsapproximation, og  $p_1$  ikke alt for langt fra  $p_2$ . Pas især på hvis  $n$  beregnes til at være så lille at der er forventede antal under 5.

**Beregning i SAS se side 26**

## Eksempel

To måder til afvænning af rygere

- Nyt tyggegummi
- Sædv. kontrolgruppe

Normalt holder 15% op efter 6 måneder. Håber på at komme op på 30%.

Ønsker at kunne finde forskellen med en styrke på 80% ved signifikansniveau 5%

Dvs.  $p_1 = 0.30$ ,  $p_2 = 0.15$ ,  $\bar{p} = 0.225$ ,  $\alpha = 0.05$ ,  $1 - \beta = 0.8$

$$n = \frac{2 \times 0.225 \times 0.775}{0.15^2} \times 7.85 = 121.6$$

dvs. 122 obs. i hver gruppe.

## Sammenligning af hyppigheder: power i SAS

SAS Analyst har menuer til power-beregninger

- Mere præcise formler ... men kun for kontinuerte data
- For sammenligning af hyppigheder kan man “snyde” ved at benytte t-test proceduren for normalfordelte data. Her indsættes  $p_1$  og  $p_2$  som “means” og som spredning bruges  $\sqrt{\bar{p}(1 - \bar{p})}$ .
- Begrundelse: sammenligning af hyppigheder er jo sammenligning af middelværdier af 0/1-variable. Middelværdien i grupperne er hhv  $p_1$  og  $p_2$ . Så mangler vi bare variationen indenfor grupper. For 0/1-variable variable er variansen  $p(1 - p)$ . Som  $p$  bruger vi gennemsnittet  $\bar{p} = (p_1 + p_2)/2$

## Simpel regression

Model:  $Y_i = a + bx_i + \epsilon_i$ ,  $\epsilon_i \sim N(0, \sigma^2)$ ,  $i = 1, \dots, n$

$b$ : er regressionskoefficienten

Stikprøveberegning for testet  $H_0 : b = 0$

$$n = f(\alpha, 1 - \beta) \times \left( \frac{\sigma}{b\sigma_x} \right)^2,$$

hvor  $\sigma_x$  er standardafvigelsen for kovariaten ( $x$ ). Dvs jo *større* spredning der er i kovariaten jo *mindre* er det nødvendige antal observationer.

# Nomogram

Bogen, s. 456

- Tegn linje mellem “Standardized difference” og “Power”, aflæs N der hvor den midterste linje skæres for det valgte signifikansniveau
- Kontinuert skema, kan nemt bede om fx. en styrke på 0.85.
- Nemt at “vende om” og finde fx. differens for valgt N og power

## MEN

- Svært at aflæse nøjagtigt
- Meget let at løbe sur i hvornår der skal ganges med  $\sqrt{2}$  og om det nu er antal personer eller antal par osv.
- Foretrækker formler eller computer programmer

## Strukturerede forsøg

- Minimere effekt af variationskilder
- Variation indenfor/mellem personer
- “Sin egen kontrol”
- Store styrkegevinster
- Men pas på: Kræver tilpassede analysemetoder
- f.eks parret t-test, flersidede variansanalyser
- “Standardmetoder” kan være *vildt* forkerte

## Parrede data

**Table 9.3** Mean daily dietary intake over 10 pre-menstrual and 10 post-menstrual days (Manocha *et al.*, 1986)

| Subject | Dietary intake (kJ) |                | Difference |
|---------|---------------------|----------------|------------|
|         | Pre-menstrual       | Post-menstrual |            |
| 1       | 5260                | 3910           | 1350       |
| 2       | 5470                | 4220           | 1250       |
| 3       | 5640                | 3885           | 1755       |
| 4       | 6180                | 5160           | 1020       |
| 5       | 6390                | 5645           | 745        |
| 6       | 6515                | 4680           | 1835       |
| 7       | 6805                | 5265           | 1540       |
| 8       | 7515                | 5975           | 1540       |
| 9       | 7515                | 6790           | 725        |
| 10      | 8230                | 6900           | 1330       |
| 11      | 8770                | 7335           | 1435       |
| Mean    | 6753.6              | 5433.2         | 1320.5     |
| SD      | 1142.1              | 1216.8         | 366.7      |

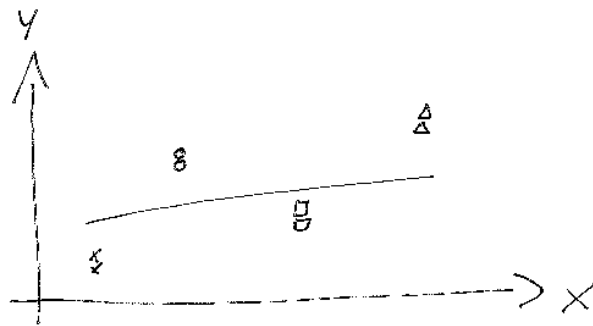
Two sample:  $t = \frac{1320.5}{\sqrt{1142^2 + 1216^2} / \sqrt{11}} = 2.62,$

Paired:  $t = \frac{1320.5}{366.7 / \sqrt{11}} = 11.99$

Note, the Two-sample test will give a p-value which is too high

## Dobbelt bestemmelser

Sammenhang mellem  $X$  og  $Y$ ?



$Y$  er målt 2 gange i hver person

Dobbelt bestemmelser er ikke uafhængige!

~~Ans:~~ Ignoreres dette bliver  $p$ -værdien for lille

## Crossover design

- Ny behandling
- Sammenligne med placebo/standard behandling
- Give personen begge behandlinger
- Sammenligninger *indenfor* individer
- Men der kan være en tidseffekt (spontan bedring)
- Bytte rundt på rækkefølgen
- — tilfældigt/systematisk

## Eksempel

Patienter med “Raynaud’s fænomen”. Test af Nicardipine (Ca-kanal blokker).

- Nicardipine i 2 uger. Tælle antal episoder.
- 1 uges washout
- Placebo i 2 uger. Tælle antal episoder.
- — eller omvendt
- 10 personer i hver gruppe.

## Nicardipine først

|      | Per. 1<br>N | Per. 2<br>P | (2) – (1) | (1) + (2) | (N) – (P) |
|------|-------------|-------------|-----------|-----------|-----------|
| Gr.I | 16          | 12          | -4        | 28        | 4         |
|      | 26          | 19          | -7        | 45        | 7         |
|      | 8           | 20          | 12        | 28        | -12       |
|      | 37          | 44          | 7         | 81        | -7        |
|      | 9           | 25          | 16        | 34        | -16       |
|      | 41          | 36          | -5        | 77        | 5         |
|      | 52          | 36          | -16       | 88        | 16        |
|      | 10          | 11          | 1         | 21        | -1        |
|      | 11          | 20          | 9         | 31        | -9        |
|      | 30          | 27          | -3        | 57        | 3         |
| Mean | 24.0        | 25.0        | 1.0       | 49.0      | -1.0      |
| SD   | 15.61       | 10.84       | 9.87      | 25.0      | 9.87      |

| Placebo først |             |             |           |           |           |
|---------------|-------------|-------------|-----------|-----------|-----------|
|               | Per. 1<br>P | Per. 2<br>N | (2) – (1) | (1) + (2) | (N) – (P) |
| Gr.II         | 18          | 12          | -6        | 30        | -6        |
|               | 12          | 4           | -8        | 16        | -8        |
|               | 46          | 37          | -9        | 83        | -9        |
|               | 51          | 58          | 7         | 109       | 7         |
|               | 28          | 2           | -26       | 30        | -26       |
|               | 29          | 18          | -11       | 47        | -11       |
|               | 51          | 44          | -7        | 95        | -7        |
|               | 46          | 14          | -32       | 60        | -32       |
|               | 18          | 30          | 12        | 48        | 12        |
|               | 44          | 4           | -40       | 48        | -40       |
| Mean          | 34.3        | 22.3        | -12.0     | 56.6      | -12.0     |
| SD            | 14.99       | 19.14       | 16.34     | 30.24     | 16.34     |

## Statistisk analyse? Hvorfor ikke parret $t$ -test?

Et parret  $t$ -test er baseret på behandlingsdifferensen: Nicardipine - Placebo. Hvis der er en periodeeffekt vil disse differenser give et forkert billede af effekten af behandling.

Antag f.eks at responsen falder i periode 2. I gruppe I undervurderes behandlingseffekten fordi Placebo gives i periode 2 hvor der er et lavere antal anfald. I gruppe II vil behandlingseffekten tilgængæld blive overvurderet.

Det samlede estimat for behandlingseffekten fås ved at tage gennemsnit over alle personer. Hvis der ikke er lige mange personer i hver gruppe er denne analyse biased. I vores eksempel er der faktisk lige mange i hver gruppe, men det parrede  $t$ -test er in-efficient fordi modellen ikke tager højde for *periode*.

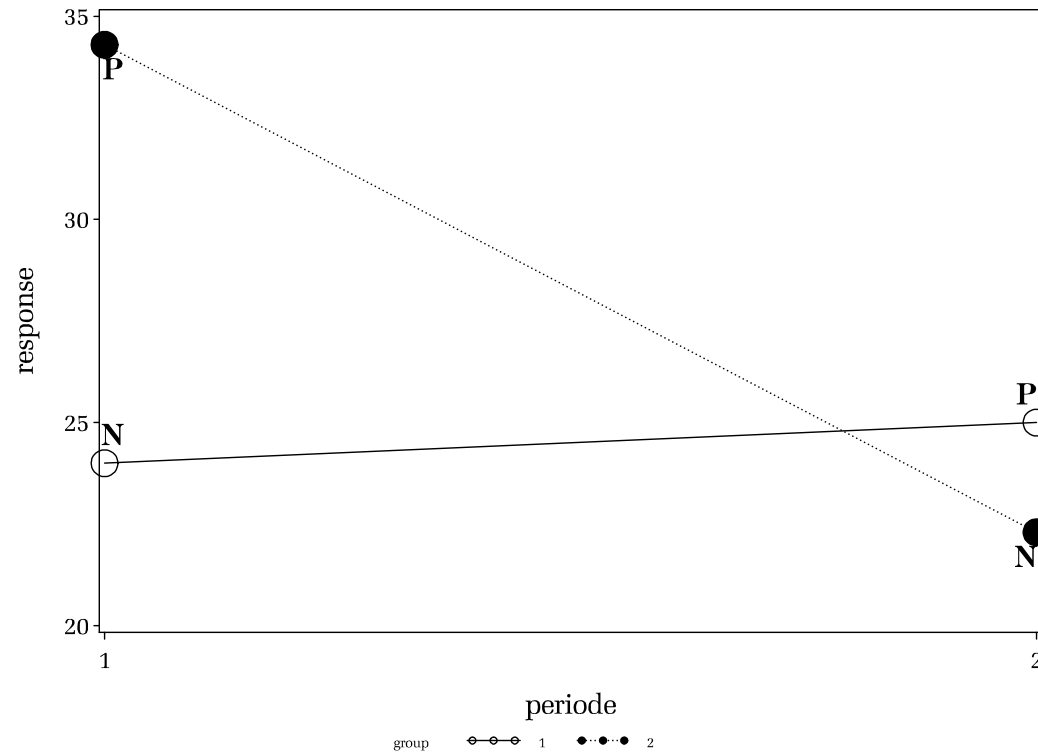
## En anden ide

Vi vil afgøre om der er en behandlingseffekt og vi er bekymrede for om der også kunne være en periodeeffekt.

Smart design: de to grupper har fået behandlingen i forskellige perioder.

hvis der kun er en periode-effekt, må periodeudviklingen i de to grupper være ens. Omvendt, hvis der er en behandlingseffekt, så bliver periodeudviklingerne forskellige fordi den ene gruppe starter med behandling mens den anden slutter med behandling.

Vi kan undersøge behandlingseffekten ved at se om periode-udviklingen er den samme i de to grupper



I gruppe 2 ser der ud til at være en stræk behandlingseffekt.  
eller får patienterne det bare bedre som tiden går?

## Underliggende additiv model

$$E(Y_{i,j}) = \alpha_i + \beta beh_{i,j} + \delta per_{i,j}, \text{ } i: \text{ person, } j=\text{periode}$$

hvor  $beh_{i,j}$  er en kovariat der er 1 hvis behandling=nicardipine, 0 ellers og  $per_{i,j}$  er en kovariat der er 1 hvis periode=2, 0 ellers

$\alpha_i$  : person's  $i$  responsniveau med placebo i periode 1.

$\beta$ : behandlingseffekt (nicardipine-placebo),

$\delta$ : periodeeffekt (periode2-periode1)

*Hvordan kan vi bestemme  $\beta$  udfra data?*

3-sidet variansanalyse (husk at inddrage person), men vi kan klare os med  $t$ -test.

## Test af behandlingseffekten baseres på periodedifferenserne

$PD_g$  : måling 2 minus måling 1 i gruppe  $g = 1, 2$  (*periode-differens*)

$$E(PD_1) = (\alpha_i + 0 + \delta) - (\alpha_i + \beta + 0) \quad (1)$$

$$= \delta - \beta \quad (2)$$

$$E(PD_2) = (\alpha_i + \beta + \delta) - (\alpha_i + 0 + 0) \quad (3)$$

$$= \delta + \beta \quad (4)$$

De to differens-niveauer er kun ens hvis behandlingseffekten er nul ( $\beta = 0$ ).

Ide: vi kan teste nulhypotesen “ingen behandlingseffekt” ved et to-stikprøve  $t$ -test på periode-forskellene.

Et forkast betyder at der er en signifikant behandlingseffekt.

## Eksempel

Antag at behandlingen nedsætter responsen med 6 ( $\beta = -6$ ) og periode 2 nedsætter responsen med 3 ( $\delta = -3$ ).

$PD_g$  : måling 2 minus måling 1 i gruppe  $j = 1, 2$

$$E(PD_1) = (\alpha_i + 0 - 3) - (\alpha_i - 6 + 0) \quad (5)$$

$$= -3 + 6 \quad (6)$$

$$E(PD_2) = (\alpha_i - 6 - 3) - (\alpha_i + 0 + 0) \quad (7)$$

$$= -3 - 6 \quad (8)$$

hvis vi trækker de to niveauer fra hinanden forsvinder periodeeffekt

$E(D_2) - E(D_1) = (-6 - 6) = -12 = 2 \times \beta$  og derfor:

$$\beta = [E(D_2) - E(D_1)]/2$$

dvs behandlingseffekten kan estimeres ved gruppe-forskellen i periodedifferensniveau delt med 2

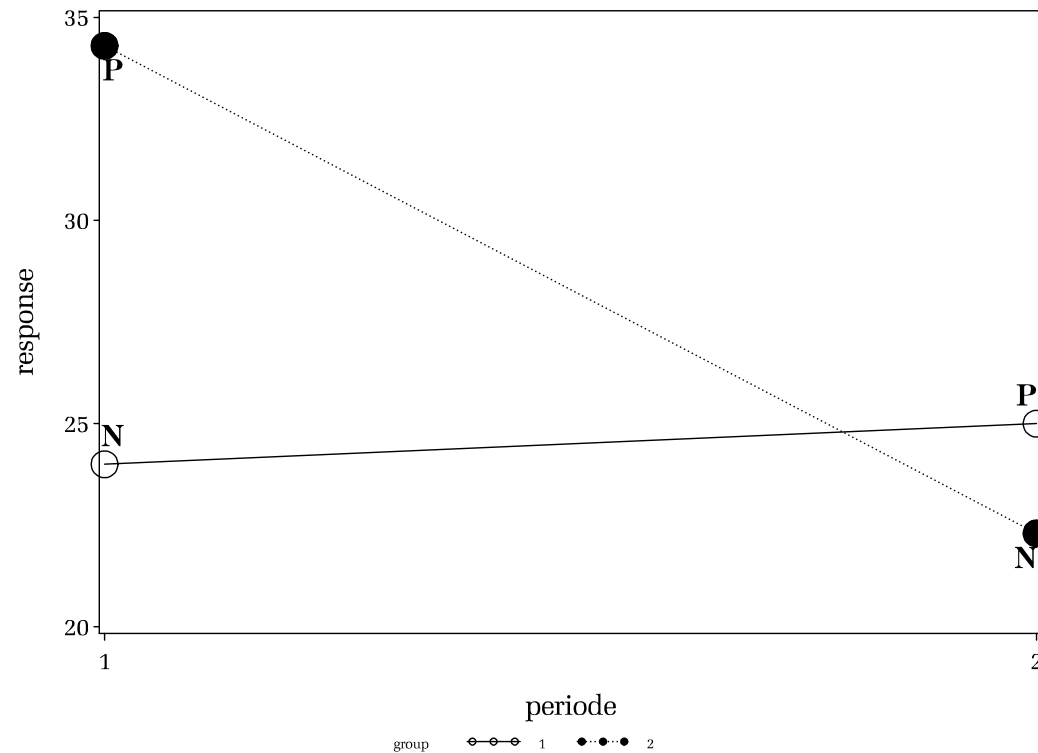
## Sig det lige en gang til!

Vi vil afgøre om der er en behandlingseffekt og vi er bekymrede for om der også kunne være en periodeeffekt.

Smart design: de to grupper har fået behandlingen i forskellige perioder.

To ting ændres når vi bevæger os fra periode 1 til periode 2: tiden går og behandlingen ændres.

Hvis behandlingen ikke har nogen effekt vil personerne udvikle sig ens i de to grupper. Vi kan derfor teste om der er en behandlingseffekt ved at testes om personerne udvikler sig ens i de to grupper. Dvs  $t$ -test på periode-differenser. Dette test er robust overfor en periode-effekt, der er det naive parrede  $t$ -test ikke.



I gruppe 2 ses en forberedring i responsen. Hvis dette alene skulle skyldes en periodeeffekt ville vi forvente at se den samme udvikling i gruppe 1. Men her er niveauet blevet dårligere over tid. Det at patienterne udvikler sig forskelligt i de to grupper er en indikation af at behandlingen har en effekt.

## Sammenligning af grupperne

|      | Per. 1 | Per. 2 | (2) – (1) | (1) + (2) | (N) – (P) |
|------|--------|--------|-----------|-----------|-----------|
|      | N      | P      |           |           |           |
| Mean | 24.0   | 25.0   | 1.0       | 49.0      | -1.0      |
| SD   | 15.61  | 10.84  | 9.87      | 25.0      | 9.87      |
|      | P      | N      |           |           |           |
| Mean | 34.3   | 22.3   | -12.0     | 56.6      | -12.0     |
| SD   | 14.99  | 19.14  | 16.34     | 30.24     | 16.34     |

## Behandlingseffekt i eksemplet

Periodedifferenser:  $\bar{P}D_1 = 1.0$  og  $\bar{P}D_2 = -12.0$

I gruppe 1 har man fået flere anfald fra periode 1 til 2. I gruppe 2 er antallet gået ned. Det tyder på en behandlingseffekt. I gruppe 2 sluttede man med aktiv behandling.

Estimeret effekt  $\hat{\beta} = [\bar{P}D_2 - \bar{P}D_1]/2 = (-12.0 - 1.0)/2 = -6.5$

Er effekten signifikant? ( $t$ -test på periodedifferenser)

Fælles SD:  $s = \sqrt{\frac{(10-1) \times 9.87^2 + (10-1) \times 16.34^2}{10+10-2}} = 13.50$

$$\text{se}(\bar{P}D_1 - \bar{P}D_2) = 13.50 \times \sqrt{\frac{1}{10} + \frac{1}{10}} = 6.037$$

$$t = \frac{13.0}{6.037} = 2.154 \quad \text{DF} = 18 \quad \text{P} = 0.045$$

## Uden korrektion for periodeeffekt: parret $t$ -test

Hvis vi ignorerer en mulig periodeeffekt kan behandlingseffekten analyseres i et parret  $t$ -test

Gennemsnit af N-P =  $-6.5$ ,  $t = 2.03$ ,  $p = 0.056$

I dette tilfælde bliver behandlingseffekten det samme uanset om der korrigeres for *periode* eller ej. Det er fordi fordelingen af (*behandling* er den samme i de to perioder (*behandling* og *periode* er uafhængige)). Er det ikke opfyldt vil det parrede  $t$ -test være biased.

Bemærk dog at behandlingseffekten ikke er signifikant i den ukorrigerede analyse. Endnu et eksempel på at man kan vinde styrke ved at inddrage kovariater.

## Periodeeffekt

$$E(Y_{i,j}) = \alpha_i + \beta beh_{i,j} + \delta per_{i,j},$$

$BD_g$ : Nicardipine - Placebo i gruppe  $g = 1, 2$

$$E(BD_1) = E(-PD_1) = (\alpha_i + \beta + 0) - (\alpha_i + 0 + \delta) \quad (9)$$

$$= \beta - \delta \quad (10)$$

$$E(BD_2) = E(PD_2) = (\alpha_i + \beta + \delta) - (\alpha_i + 0 + 0) \quad (11)$$

$$= \beta + \delta \quad (12)$$

De to differens-niveauer er kun ens hvis periode-effekten er nul ( $\delta = 0$ ).

Vi kan teste nulhypotesen “ingen periodeeffekt” ved et to-stikprøve  $t$ -test på behandlings-forskellene.

Periodeeffekt:  $\delta = [E(BD_2) - E(BD_1)]/2$

## Periodeeffekt, beregning

Behandlingsdifferenser (N-P):  $\bar{B}D_1 = -1.0$  og  $\bar{B}P_2 = -12.0$

I gruppe 1 fik man lidt færre anfald med behandling end placebo. I gruppe 2 fik man en del færre anfald med behandling. Det tyder på en periodeeffekt. I gruppe 2 sluttede man med behandling, dvs antallet af anfald falder i periode 2.

Estimeret effekt  $\hat{\delta} = [E(BD_2) - E(BD_1)]/2 = (-12.0 - (-1.0))/2 = -5.5$

Er effekten signifikant? ( $t$ -test på behandlingsdifferenser)

Fælles SD:  $s = \sqrt{\frac{(10-1) \times 9.87^2 + (10-1) \times 16.34^2}{10+10-2}} = 13.50$

$$\text{se}(\bar{B}D_1 - \bar{B}D_2) = 13.50 \times \sqrt{\frac{1}{10} + \frac{1}{10}} = 6.037$$

$$t = \frac{11.0}{6.037} = 1.822 \quad \text{DF} = 18 \quad \text{P} = 0.085$$

## Carry-over effekten

En svaghed i dette design er at en gavnlig effekt af behandlingen i periode 1 måske kan opretholdes og gøre placebo-værdien 'for god'. Dette kaldes carry-over effekten.

Om denne effekt er tilstede kan testes ved for hver patient at begrænse summen af de to målinger og derefter sammenligne de to grupper.

## Model for sum

Under modellen gælder:

$$E(Y_1 + Y_2) = 2\alpha_i + \beta + \delta$$

Carry-over effekten ( $\mu$ ) påvirker niveauet i gruppe I

$$\text{Gr I: } E[Y_1 + Y_2] = 2\alpha_i + \beta + \delta + \mu$$

$$\text{Gr II: } E[Y_1 + Y_2] = 2\alpha_i + \beta + \delta$$

Hvis  $\mu = 0$  er der ikke carry-over. Testet udføres som et uparret  $t$ -test, men det kræver at personniveauerne  $\alpha_i$  antages normalfordelt.

## Test for carryover-effekt

$t$ -test for (1)+(2):

Fælles SD:

$$\sqrt{\frac{(10-1) \times 25.0^2 + (10-1) \times 30.24^2}{10+10-2}} = 27.74$$

Forskel på gennemsnit

$$\bar{S}_1 - \bar{S}_2 = 49.0 - 56.6 = -7.6$$

$$\text{se}(\bar{S}_1 - \bar{S}_2) = 27.74 \times \sqrt{\frac{1}{10} + \frac{1}{10}} = 12.41$$

$$t = \frac{7.6}{12.41} = 0.613 \quad \text{DF} = 18 \quad \text{P} = 0.54$$

Dvs. carry-over effekten er ikke statistisk signifikant, men bemærk at testet er svagt fordi variationen mellem personer indgår i summerne.