7.2 Sammenligning av to forventinger

- To-utvalgs z-observator
- To-utvalgs t-prosedyrer
- To-utvalgs t-tester
- To-utvalgs t-konfidensintervall
- Robusthet
- To-utvalgs t-prosedyrerår variansene er like

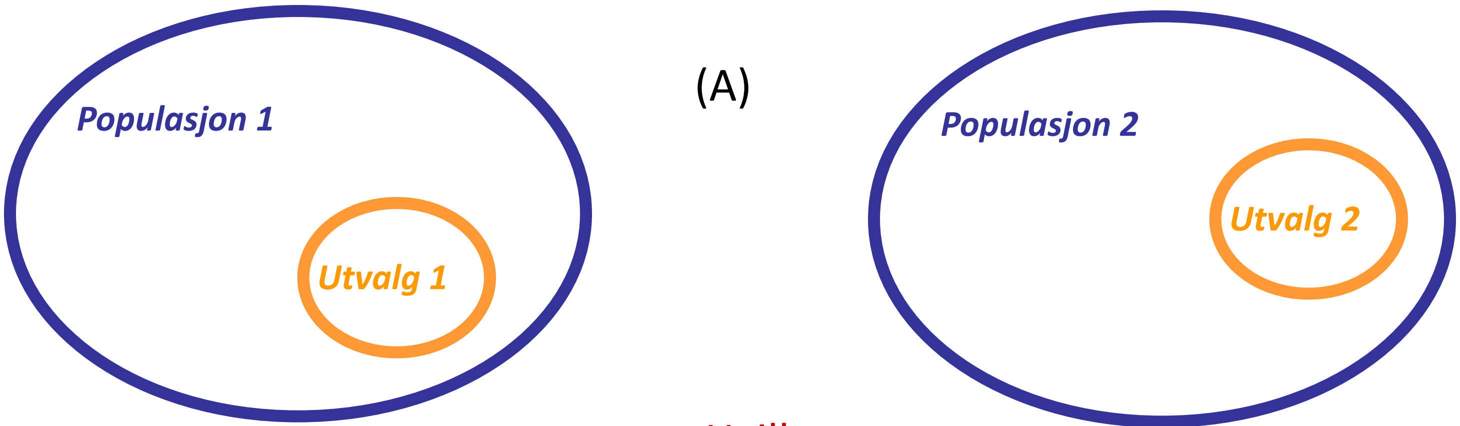
Sammenlikning av to forventninger

- Effekt av økt kalsium på blodtrykk
  - En gruppe får økt kalsium i diett
  - Kontrollgruppe får placebo

- Bank: To kampanjer for å øke bruk av kredittkort
  - Et tilfeldig utvalg eksponert for kampanje 1
  - Et annet tilfeldig utvalg eksponert for kamp. 2

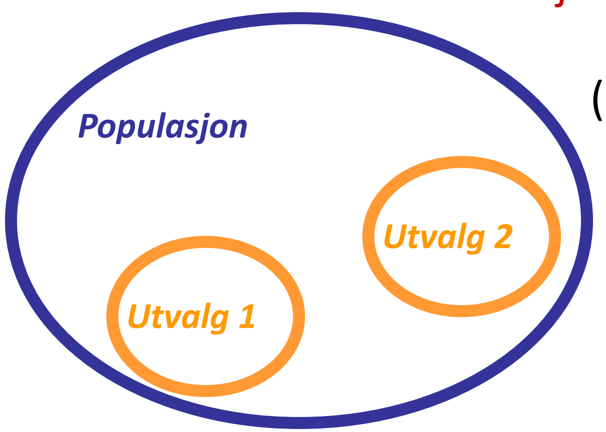
•Sammenligning av to utvalg

(A)



Hvilken situasjon har vi?

(B)



Vi sammenligner ofte to behandlinger på to uavhengige utvalg.

Er forskjellene vi observerer et resultat av variasjon i utvalgene (B), eller er det en ekte forskjell, dvs. at utvalgene kommer fra to forskjellige populasjoner (A)?

Notasjon

Populasjon	Variabel	Forventning	Standard avvik
1	x_1	μ_1	σ_1
2	x_2	μ_2	σ_2

Populasjon	Utvalgsstørrelse	Gjennomsnitt	Standard avvik
1	n_1	$\bar{x}_1$	s_1
2	n_2	$\bar{x}_2$	s_2

Observator for to-utvalg

Av interesse: $\mu_1 - \mu_2$

Naturlig estimat: $\bar{x}_1 - \bar{x}_2$

$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Varians $\bar{x}_2 - \bar{x}_2$ er $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

$$\boxed{z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \text{ er } N(0,1)}$$

To-utvalg – ukjente varianser

- I praksis: σ_1 og σ_2 ukjente
- Estimeres ved s_1 og s_2, får da

$$\text{Observator } t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Ikke *t*-fordelt! En *t*-fordeling erstatter en N(0,1)-fordeling bare når et enkelt standardavvik σ i en *z*-observator erstattes med et empirisk standardavvik s
- Men: Tilnærmet *t(k)*-fordelt med passende antall frihetsgrader *k*
- To måter å estimere *k* på: En metode som beregnes fra data (med dataprogram), eller den minste verdien av (n_1-1) og (n_2-1)

THE TWO-SAMPLE t SIGNIFICANCE TEST

Suppose that an SRS of size n_1 is drawn from a normal population with unknown mean μ_1 and that an independent SRS of size n_2 is drawn from another normal population with unknown mean μ_2. To test the hypothesis $H_0\colon \mu_1 = \mu_2$, compute the **two-sample t statistic**

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

and use P-values or critical values for the $t(k)$ distribution, where the degrees of freedom k are either approximated by software or are the smaller of $n_1 - 1$ and $n_2 - 1$.

Definition, pg 489

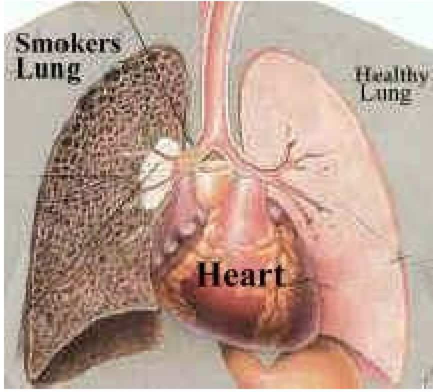
Introduction to the Practice of Statistics, Fifth Edition

To-utvalgs *t*-tester

- For å teste H_0: $\mu_1=\mu_2$ mot H_a:$\mu_1>\mu_2$(H_a:$\mu_1-\mu_2>0$)
  - P-verdi = $P(T \geq t)$
- For å teste H_0: $\mu_1=\mu_2$ mot H_a:$\mu_1<\mu_2$(H_a:$\mu_1-\mu_2<0$)
  - P-verdi = $P(T \leq t)$
- For å teste H_0: $\mu_1=\mu_2$ mot H_a:$\mu_1\neq\mu_2$(H_a:$\mu_1-\mu_2\neq 0$)
  - P-verdi = $2P(T \geq |t|)$,

der T er t-fordelt med *k* frihetsgrader

Ødelegger røyking hos foreldrene lungene til barna deres?



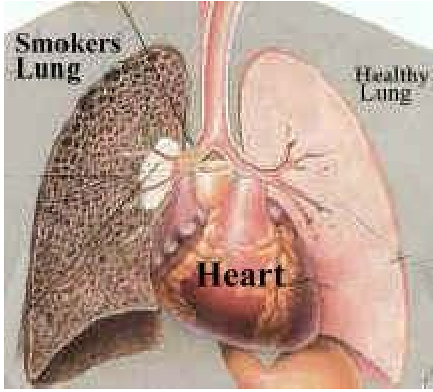
Forsert vital kapasitet (FVC) er volumet (i milliliter) av luft som en person kan puste ut etter maksimal innpust.

FVC ble målt for et utvalg av barn som ikke hadde vært utsatt for røyking hos foreldrene og et utvalg av barn som hadde vært utsatt for røyking hos foreldrene.

Foreldrene røyker	FVC $\bar{x}$	s	n
Ja	75.5	9.3	30
Nei	88.2	15.1	30

Vi ønsker å finne ut om røyking hos foreldrene senker barns lungekapasitet målt av FVC-testen.

Er forventet FVC lavere i populasjonen av barn som har vært utsatt for røyking hos foreldrene?



H_0: $\mu_{røyk} = \mu_{ikke\text{-}røyk}$ <=> $(\mu_{røyk} - \mu_{ikk\text{-}røyk}) = 0$

H_a: $\mu_{røyk} < \mu_{ikke\text{-}røyk}$ <=> $(\mu_{røyk} - \mu_{ikke\text{-}røyk}) < 0$ (en-sidig)

Forskjellen I utvalgs-gjennomsnittene følger tilnærmet en t-fordeling:

$$t\left(0,\ \sqrt{\frac{s_{smoke}^2}{n_{smoke}} + \frac{s_{no}^2}{n_{no}}}\right),\ df\ \ 29$$

Vi beregner t-observatoren:

Røyking hos foreldrene	FVC $\overline{x}$	s	n
Ja	75.5	9.3	30
Nei	88.2	15.1	30

$$t = \frac{\overline{x}_{røyk} - \overline{x}_{ikke-røyk}}{\sqrt{\frac{s_{røyk}^2}{n_{røyk}} + \frac{s_{ikke-røyk}^2}{n_{ikke-røyk}}}} = \frac{75.5 - 88.2}{\sqrt{\frac{9.3^2}{30} + \frac{15.1^2}{30}}}$$

$$t = \frac{-12.7}{\sqrt{2.9 + 7.6}} \approx -3.9$$

I tabell D, for df 29, finner vi:

$|t| > 3.659$ => $p < 0.0005$ (en-sidig)

Det er en veldig signifikant forskjell, og vi forkaster H_0 (p< 0.0005).

Lungekapasiteten blir signifikant svekket hos barn av røykende foreldre.

Eks. ny læringsmetode for lesing

(s.432)

- 21 elever prøver ny metode
- 23 elever følger dagens metode

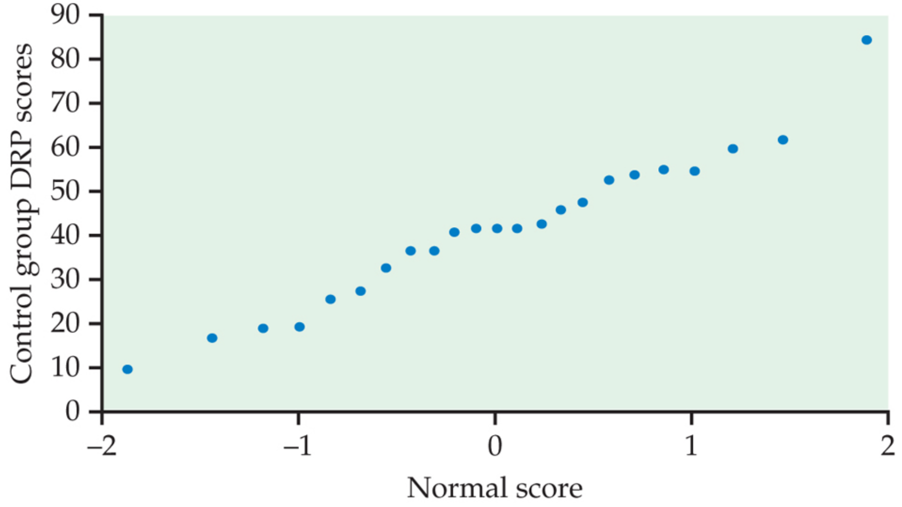
TABLE 7.4

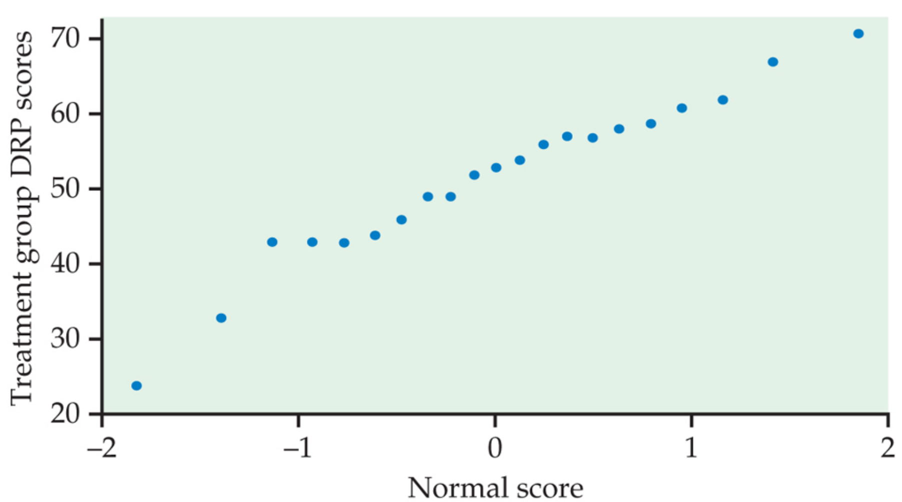
DRP scores for third-graders

Treatment group				Control group			
24	61	59	46	42	33	46	37
43	44	52	43	43	41	10	42
58	67	62	57	55	19	17	55
71	49	54		26	54	60	28
43	53	57		62	20	53	48
49	56	33		37	85	42	

Table 7-4
Introduction to the Practice of Statistics, Fifth Edition

Ny læringsmetode: (Normale) Kvantilplott





Ny læringsmetode for lesing

- 21 elever prøver ny metode
- 23 elever følger dagens metode
- H_0: $\mu_1 = \mu_2$ mot H_a: $\mu_1 > \mu_2$

Gruppe	n	$\bar{x}$	s
Behandling	21	51.48	11.01
Kontroll	23	41.52	17.15

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{51.48 - 41.52}{\sqrt{\frac{11.01^2}{21} + \frac{17.15^2}{23}}} = 2.31$$

Frihetsgrader:
Minste av $n_1 - 1 = 20, n_2 - 1 = 22$
$P(t > 2.31)$ er mellom 0.01 og 0.02
fra Tabell D

To-utvalgs *t*-konfidensintervall

Når vi har to tilfeldige, uavhengige utvalg av størrelser n_1 og n_2 fra to normalfordelte populasjoner med ukjente forventninger μ_1 og μ_2, har følgende konfidensintervall for $(\mu_1 - \mu_2)$

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

minst konfidensnivå C. Her er t^* verdien slik at arealet mellom -t* og t* i t(k)-fordelingen er C. Antall frihetsgrader k beregnes av et dataprogram eller vi bruker den minste verdien av $n_1 - 1$ and $n_2 - 1$.

Ny læringsmetode

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 9.96 \pm t^* \sqrt{\frac{11.01^2}{21} + \frac{17.15^2}{23}} = 9.96 \pm 4.31 t^*$$

- Konservativt valg: t(20)-fordeling: t*=2.086

  95% KI: [0.97,18.95]

- Minitab: t(37)-fordeling: t*=2.026,

  95% KI: [1.23,18.69]

Egenskaper to-utvalgs t-test

- Mer robust enn ett-utvalgs t-test
- Svært presise når $n_1 \approx n_2$ og formene på fordelingene er ganske like
- Større utvalg trengs når ulike former på fordelinger
- Velg så like utvalgsstørrelser som mulig!

Lik varians: Pooled t-metoder to-utvalg

- Hittil har vi benyttet oss av at to-utvalgs t-observator er tilnærmet t-fordelt når vi bruker de estimerte standardavvikene s_1 og s_2
- Dersom de to underliggende normale populasjons-fordelingene har *samme* standardavvik σ (dvs $\sigma_1 = \sigma_2 = \sigma$), og vi erstatter den N(0,1)-fordelte z-observatoren med en t-observator ved å bytte ut σ med et *samlet* estimat på σ, da er t-observatoren eksakt t-fordelt
- (Erstatter da bare ett enkelt standardavvik i en z-observator med ett estimert standardavvik)

Observator for to-utvalg

Av interesse: $\mu_1 - \mu_2$

Naturlig estimat: $\bar{x}_1 - \bar{x}_2$

$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Varians $\bar{x}_2 - \bar{x}_2$ er $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \text{ er } N(0,1)$$

- Dersom $\sigma_1 = \sigma_2 = \sigma$ og σ er ukjent, erstatter vi bare ett enkelt standardavvik σ i en z-observator med ett estimert standardavvik, da er t-observatoren eksakt t-fordelt

t-metoder for samlede («pooled») to-utvalg

- Begge de empiriske variansene s_1^2 og s_2^2 estimerer σ^2
- Den beste måten å utnytte begge disse til å estimere σ, er å ta et gjennomsnitt vektet med de respektive frihetsgradene n_1 og n_2
- Det samlede estimatet av σ er da

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

- t-observatoren

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

er da eksakt t-fordelt med frihetsgrader $n_1 + n_2 - 2$

t-metoder for samlede («pooled») to-utvalg

THE POOLED TWO-SAMPLE t PROCEDURES

Suppose that an SRS of size n_1 is drawn from a normal population with unknown mean μ_1 and that an independent SRS of size n_2 is drawn from another normal population with unknown mean μ_2. Suppose also that the two populations have the same standard deviation. A level C confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Here t^* is the value for the $t(n_1 + n_2 - 2)$ density curve with area C between $-t^*$ and t^*.

To test the hypothesis $H_0: \mu_1 = \mu_2$, compute the pooled two-sample t statistic

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

In terms of a random variable T having the $t(n_1 + n_2 - 2)$ distribution, the P-value for a test of H_0 against

$$H_a: \mu_1 > \mu_2 \quad \text{is} \quad P(T \geq t)$$

$$H_a: \mu_1 < \mu_2 \quad \text{is} \quad P(T \leq t)$$

$$H_a: \mu_1 \neq \mu_2 \quad \text{is} \quad 2P(T \geq |t|)$$

Definition, pg 499
Introduction to the Practice of Statistics, Fifth Edition

Eksempel: Kalsium og blodtrykk

- Gir økt inntak av kalsium redusert blodtrykk?

TABLE 7.5

Seated systolic blood pressure

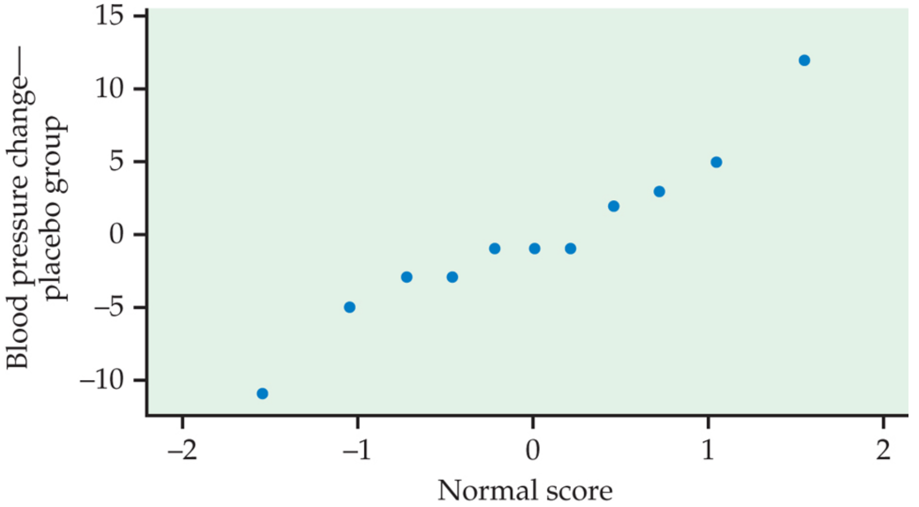
Calcium group			Placebo group		
Begin	End	Decrease	Begin	End	Decrease
107	100	7	123	124	−1
110	114	−4	109	97	12
123	105	18	112	113	−1
129	112	17	102	105	−3
112	115	−3	98	95	3
111	116	−5	114	119	−5
107	106	1	119	114	5
112	102	10	114	112	2
136	125	11	110	121	−11
102	104	−2	117	118	−1
			130	133	−3

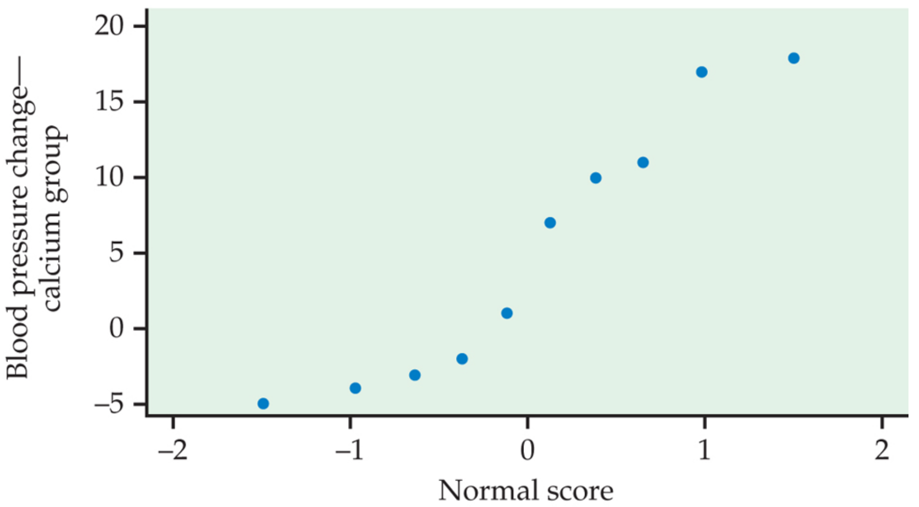
Table 7-5
Introduction to the Practice of Statistics, Fifth Edition

- NB: Nedgang (decrease) er positiv når det er en reduksjon, negative ved oppgang

Eksempel: Kalsium og blodtrykk

- Kvantilplott: Ser ut til å passe ganske bra med normalfordeling for begge grupper
- Kalsium-gruppen har en litt kort venstre hale
- Må gjøre en signifikanstest for å besvare spørsmålet: gir økt inntak av kalsium redusert blodtrykk?





Eksempel: Kalsium og blodtrykk

- Gruppe 1 med forventning μ_1 fikk kalsium-tilskudd, gruppe 2 med forventning μ_2 fikk placebo
- Naturlig med ensidig null-hypotese, for forsøk med rotter hadde vist at kalsium ga en nedgang, ingen grunn til å ro at det skulle øke blodtrykket. Hypoteser: H_0: $\mu_1 = \mu_2$, H_a: $\mu_1 > \mu_2$
- Numeriske beskrivelser av data

Gruppe	Behandling	n	$\bar{x}$	s
1	Kalsium	10	5.000	8.743
2	Placebo	11	-0.273	5.901

- De empiriske standardavvikene utelukker ikke like populasjons-standardavvik, forskjell i estimatene kan lett være pga. tilfeldigheter
- Vi vil anta like populasjons-standardavvik

Eksempel: Kalsium og blodtrykk

- Den samlede empiriske variansen er

$$s_p^2 = \frac{(n_1-1)\,s_1^2+(n_2-1)\,s_2^2}{n_1+n_2-2} = \frac{(10-1)*8.743^2+(11-1)*5.901^2}{10+11-2} = 54.536$$

som gir $s_p = 7.385$. Den samlede to-utvalgs t-observatoren er

$$t = \frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{5.000-(-0.273)}{7.385\sqrt{\frac{1}{10}+\frac{1}{11}}} = 1.634$$

- *P*-verdi=P(T≥1.634) der T er t(19)-fordelt (10+11-2=19)
- Tabell D sier oss at *P*-verdien er mellom 0.05 og 0.10. Analyse i Minitab gir *P*-verdi=0.059
- Konklusjon: Forsøket fant bevis for at kalsium muligens reduserer blodtrykket, men resultatet er ikke på 5%- (eller 1%-)nivå

Antagelsen om like populasjons-standardavvik

- Vanskelig å verifisere
- Derfor kan t-metodene for samlede («pooled») to-utvalg noen ganger være risikable å bruke
  - Ganske robuste mot både ikke-normalitet og ulike populasjons-standardavvik når utvalgsstørrelsene n_1 og n_2 er ganske like
  - Når utvalgsstørrelsene n_1 og n_2 er veldig forskjellige, er metodene sensitive for ulike populasjons-standardavvik, og man skal være forsiktige med å bruke dem med mindre utvalgsstørrelsene er store