Inferens for populasjons-forventninga (7.1)

I kapittel 6 jobbet vi under en antagelse
om kjent populasjons-standardavvik σ

Nå i kapittel 7 skal vi lære t-prosedyrer:
prosedyrer for inferens for populasjons-
forventninga når populasjons-
standardavviket σ er ukjent
(ofte en mer realistisk antagelse)

Å lære om en populasjon fra et utvalg

- Anta $x_1,\ldots,x_n$ uavhengige fra $N(\mu,\sigma)$
- Observator $\bar{x}$ for å estimere ukjent forventningsverdi μ

σ kjent: Testobservator $z=(\bar{x}-\mu)/(\sigma/\sqrt{n})$

- $\sigma/\sqrt{n}$ kjent standardavvik for observator $\bar{x}$

σ ukjent: Testobservator $t=(\bar{x}-\mu)/(s/\sqrt{n})$

- $s/\sqrt{n}$ estimert standardavvik for observator $\bar{x}$
- Vi kaller SE = $s/\sqrt{n}$ *standardfeilen (standard error)* for $\bar{x}$

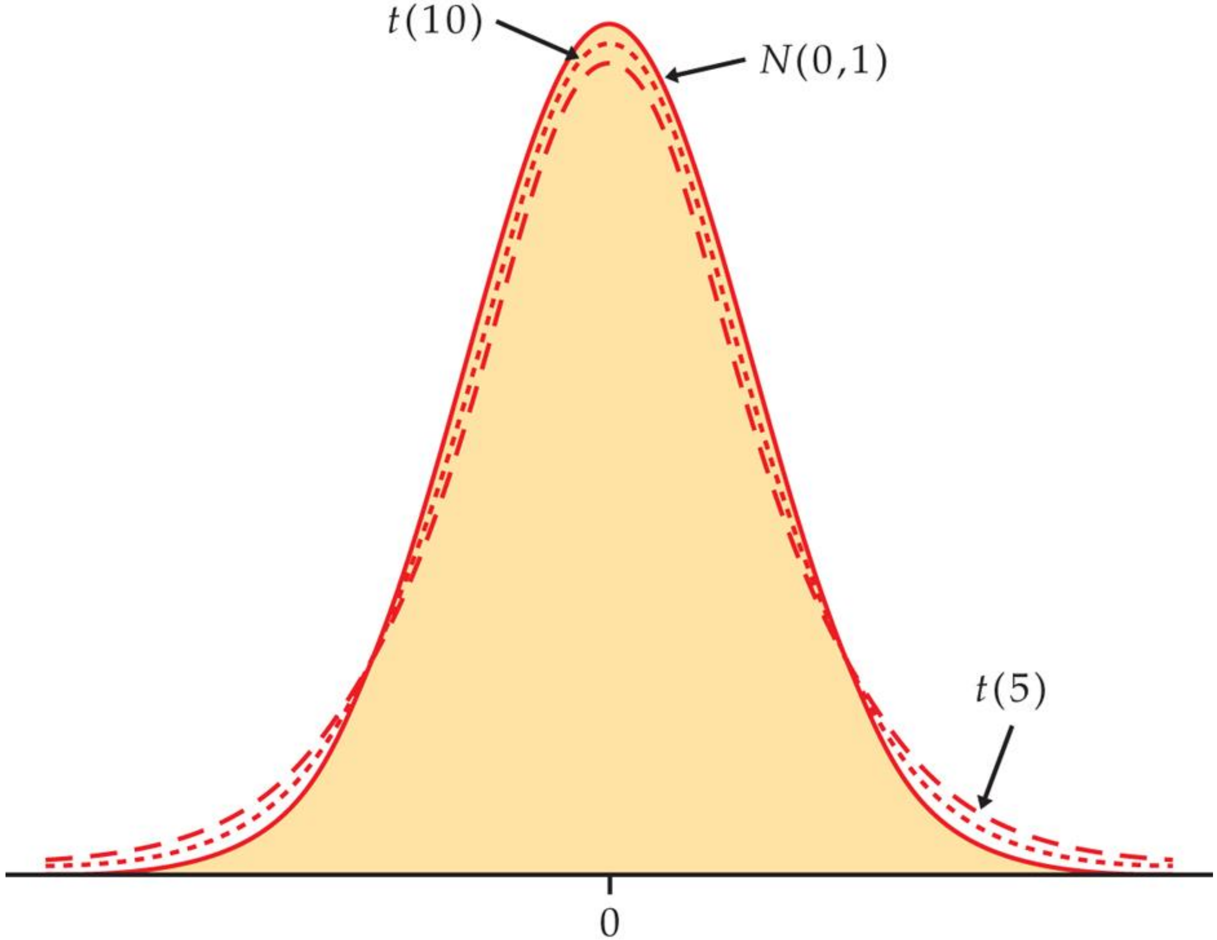
Når populasjonsstandardavviket er ukjent, bruker vi t-fordelinger

- $z=(\bar{x}-\mu)/(\sigma/\sqrt{n})$ er N(0,1)-fordelt -når σ har kjent verdi

- $t=(\bar{x}-\mu)/(s/\sqrt{n})$ er t-fordelt *med n-1 frihetsgrader*

T-fordelinger

- Formen ligner standard normalfordeling N(0,1), og fordelinga er symmetrisk om 0
- Større spredning enn N(0,1), lar oss ta hensyn til usikkerhet for σ
- Nærmere N(0,1) jo større antall frihetsgrader (større n)



t(10)
N(0,1)
t(5)
0

Konfidensintervall $[\bar{x} - m, \bar{x} + m]$ basert på ett enkelt tilfeldig utvalg

- σ *kjent*: $m = z^* \cdot \sigma/\sqrt{n}$
  – z^* er verdien slik at arealet mellom $-z^*$ og z^* i N(0,1) fordelingen er lik C

- σ *ukjent*: $m = t^* \cdot s/\sqrt{n}$
  – t^* er verdien slik at arealet mellom $-t^*$ og t^* i t(n-1) fordelingen er lik C
  – Eksakt hvis normalfordelte observasjoner (data)
  – Tilnærma riktig for stor *n* ellers

95% KI for forventa mengde C-vitaminer i mais (regneeksempel)

- Data: Det er målt mengde C-vitaminer (mg/100g) i 8 tilfeldig valgte mais
  - 26, 31, 23, 22, 11, 22, 14, 31
- Mellom-beregninger
  - $\bar{x} = 22.50$, $s = 7.19$, $n = 8$
  - $SE_{\bar{x}} = s/\sqrt{n} = 7.19/\sqrt{8} = 2.54$
- Med $n = 8$ og dermed $n-1 = 7$ frihetsgrader, blir $t^* = 2.365$ og intervallet $[\bar{x} - t^*s/\sqrt{n}, \bar{x} + t^*s/\sqrt{n}] =$ $[22.50-2.365 \cdot 2.54, 22.50+2.365 \cdot 2.54] = [16.5 , 28.5]$

- Dersom vi hadde ignorert usikkerheten i s og brukt gal faktor $z^* = 1.96$, hadde vi **feilaktig** konstruert det betydelig smalere intervallet $\bar{x} \pm z^*\sigma/\sqrt{n} = [17.5 , 27.5]$

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

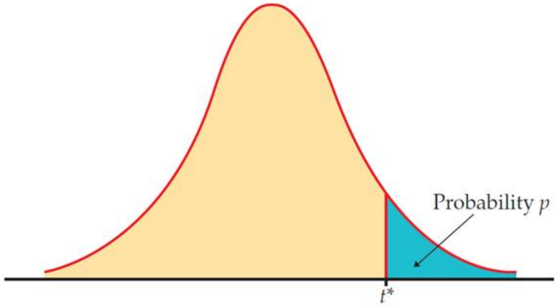


TABLE D

t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Ett-utvalgs t-test: hypotesetest om verdi av populasjons-forventningsverdi basert på et utvalg med ukjent populasjons-standardavvik

- Kap 6, σ kjent verdi: ett-utvalgs z-test
  Hypotesetester for μ basert på testobservator

  $$z = (\bar{x}-\mu_0)/(\sigma/\sqrt{n})$$

- Kap 7, σ ukjent verdi: ett-utvalgs t-test
  Hypotesetester for μ basert på testobservator

  $$t = (\bar{x}-\mu_0)/(s/\sqrt{n})$$

Hypotesetest om C-vitaminer i mais

(regneeksempel)

- Ønsker å teste H_0: $\mu = \mu_0 = 30$ mot H_a: $\mu = \mu_0 \neq 30$
- Mellom-beregninger
  - $\bar{x} = 22.50$, $s = 7.19$, $n = 8$
  - $SE_{\bar{x}} = s/\sqrt{n} = 7.19/\sqrt{8} = 2.54$
- Testobservator $t = (\bar{x}-\mu_0)/(s/\sqrt{n}) = (22.5-30)/2.54 = -2.95$
- Tosidig P-verdi = 0.021 (fra R; Tabell D angir 0.020<P-verdi<0.040)

- Hvis vi feilaktig hadde beregna p-verdien basert på normalfordelinga, ville vi fått p-verdi = 0.003
  Det er alt for lavt fordi usikkerheten i s blir ignorerert!

Husk: Er du i tvil, bruk tosidig hypotesetest

- Ensidig eller tosidig test må bestemmes ut fra problemstilling
- Det er ikke innafor å bestemme seg for å gjøre en ensidig test etter å ha sett på dataene

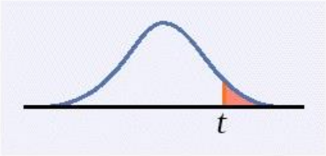
THE ONE-SAMPLE t TEST

Suppose that an SRS of size n is drawn from a population having unknown mean μ. To test the hypothesis $H_0: \mu = \mu_0$ based on an SRS of size n, compute the one-sample t statistic

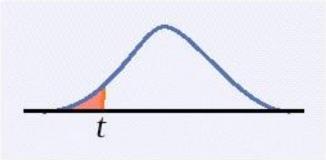
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

In terms of a random variable T having the $t(n-1)$ distribution, the P-value for a test of H_0 against

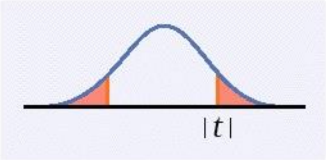
$H_a: \mu > \mu_0$ is $P(T \geq t)$



$H_a: \mu < \mu_0$ is $P(T \leq t)$



$H_a: \mu \neq \mu_0$ is $2P(T \geq |t|)$



These P-values are exact if the population distribution is normal and are approximately correct for large n in other cases.

Vi foretrekker ofte sammenlignende eksperimenter fremfor ett-utvalgs-analyser

- Beskyttende mot sammenblandede variable f.eks sammenligne ny medisin med placebo

- Studier med matchede par:
  - Individer matchet i par
  - Ønsker par som er mest mulig like, oppnås feks ved:
    - Tvillinger
    - To målinger fra samme individ
    - Venstre og høyre hånd

Triks for statistisk analyse av matchede par: Se på differansen innenfor hvert par

- **Regneeksempel: Gir forenklet måleprosedyre forskjellig svar sammenlignet med original prosedyre?**
- MeasureMind 3D Multisensor er en programvare for måling av kompliserte maskindeler. En bruker av systemet oppdager at ved å slå av en opsjon i prosedyren, reduseres måletiden med 10%. Spørsmålet er om endringen påvirker måle-resultatene.
  - Velger 51 ulike maskindeler. Måler dem med og uten opsjonen. Kaster mynt om hvilken metode som brukes først.
- Differansene er ett enkelt utvalg
- Kan bruke metoder for ett-utvalgs data

Målinger for de 20 første maskindelene. Målingene er i micrometer

TABLE 7.2 Parts Measurements Using Optical Software

Part	OptionOn	OptionOff	Diff	Part	OptionOn	OptionOff	Diff
1	118.63	119.01	0.38	11	119.03	118.66	−0.37
2	117.34	118.51	1.17	12	118.74	118.88	0.14
3	119.30	119.50	0.20	13	117.96	118.23	0.27
4	119.46	118.65	−0.81	14	118.40	118.96	0.56
5	118.12	118.06	−0.06	15	118.06	118.28	0.22
6	117.78	118.04	0.26	16	118.69	117.46	−1.23
7	119.29	119.25	−0.04	17	118.20	118.25	0.05
8	120.26	118.84	−1.42	18	119.54	120.26	0.72
9	118.42	117.78	−0.64	19	118.28	120.26	1.98
10	119.49	119.66	0.17	20	119.13	119.15	0.02

Histogram for differansen i 51 matchede par

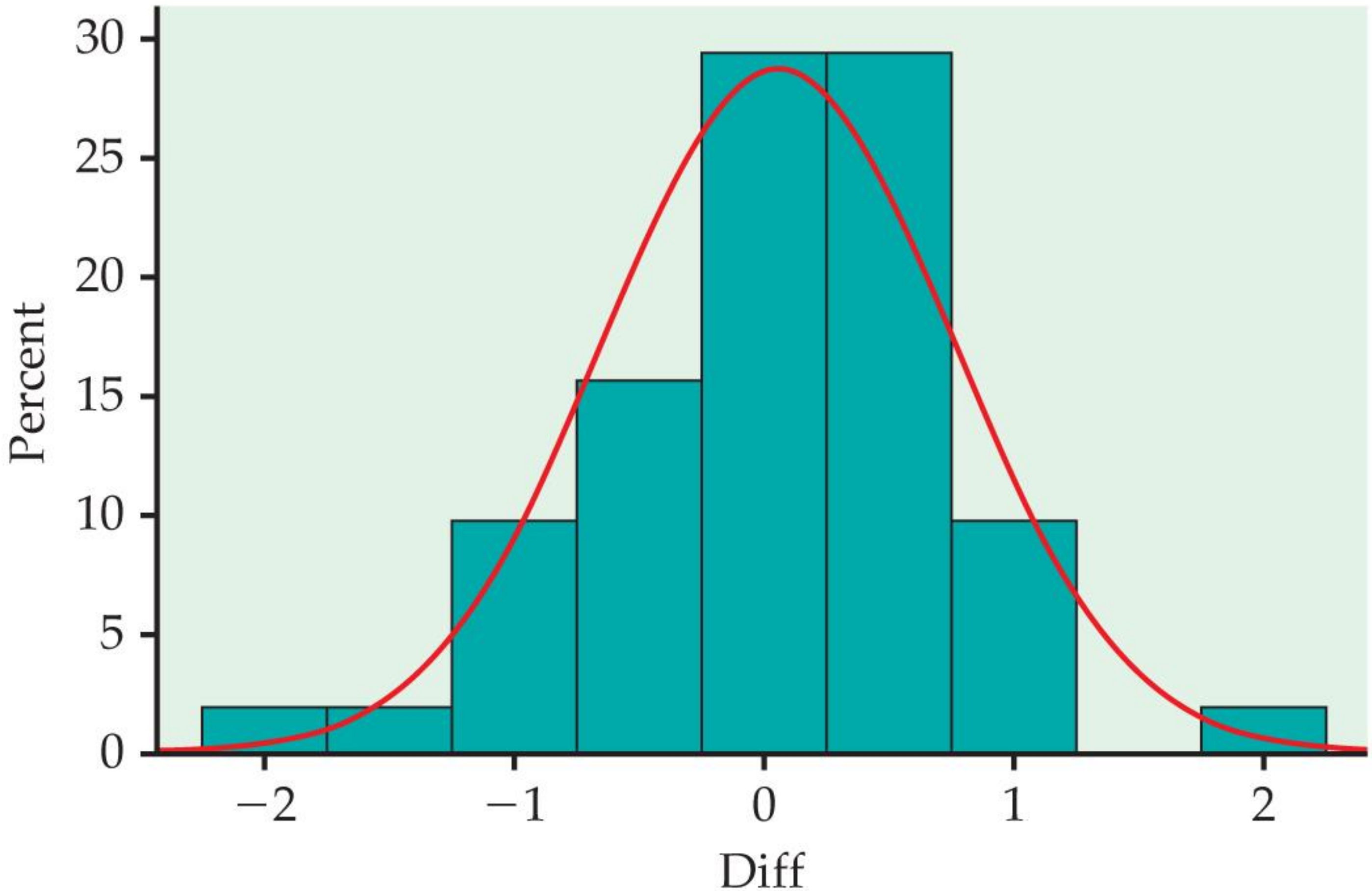


Figure 7.9

Fortsettelse av eksempel:

Analyse av matchede par

- X_i = differansen for maskindel nr. i
- Antar at X_i er $N(\mu,\sigma)$ og uavhengige

- H_0: $\mu = 0$ mot H_a: $\mu \neq 0$
  ($\mu = 0$ betyr ingen systematisk forskjell)
- $\bar{x} = 0.0504$, $s = 0.6943$, $n = 51$

- Testobs. $t = \bar{x} / (s/\sqrt{n}) = (0.0504-0)/(0.6943/\sqrt{51}) = 0.52$
- P-verdi (tosidig, df=50) = 0.6054 fra R, Tabell D: >0.25
- Forskjellen i målinger var ikke statistisk signifikant, og vi **konkluderer at det er OK å slå av opsjonen.**

Er t-tester robuste, altså insensitive til avvik fra antagelsen om normalfordelt populasjon?

- Nei, Ikke robust mot uteliggere
- Men, Robust mot ikke for store avvik fra normalfordeling
- Mer robust jo større *n*
  - *$n < 15$:* Trygt hvis data er nær normale
  - $n \geq 15$: Kan brukes hvis ingen uteliggere eller sterk skjevhet
  - $n \geq 40$: Kan brukes selv med sterk skjevhet

Kapittel 7.2: To-utvalgs t-test Sammenlikne forventningsverdi for to populasjoner med ukjent standardavvik

- F.eks: Effekt av økt kalsiuminntak på blodtrykk
  - En gruppe får økt kalsium i diett
  - Kontrollgruppe får placebo

- To-utvalgsproblem:
  - Sammenlikne respons fra to grupper
  - Hver gruppe tilfeldig utvalg fra to forskjellige populasjoner
  - Respons i de to gruppene uavhengige

Notasjon for 2 populasjoner og 2 utvalg

- Populasjonsparameterne:

Populasjon	Variabel	Forventning	Standardavvik
1	X_1	μ_1	σ_1
2	X_2	μ_2	σ_2

- Tall for utvalgene:

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

To uavhengige utvalg

- Vi er interessert i parameteren $\mu_1 - \mu_2$

- Naturlig og forventningsrett estimator

for $\mu_1 - \mu_2$ er $\bar{x}_1 - \bar{x}_2$

- Varians for $\bar{x}_1 - \bar{x}_2$ er lik $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

- Hvis populasjonsfordelingene er normale,

er $\bar{x}_1 - \bar{x}_2$ også normalfordelt

Observator for forventnings-differansen til uavhengige utvalg fra to populasjoner **med kjent standardavvik**

- Interesseparameter: $\mu_1 - \mu_2$
- Naturlig estimator: $\bar{x}_1 - \bar{x}_2$
- Varians for $\bar{x}_1 - \bar{x}_2$ er lik $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$
- Testobservatoren $Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ er N(0,1) fordelt

men kan bare brukes hvis vi har kjente verdier for standardavvikene σ_1, σ_2

Robert Warren/Getty Images

Eksempel 7.10 – med kjent standardavvik

- Klasse: 12 jenter (n_1) og 8 gutter (n_2).
  Høyde måles på 10 årsdagen. X_1: høyde jenter, X_2: høyde gutter
- Hva er sannsynlighet for at gjennomsnittlig høyde for jenter er større enn gjennomsnittlig høyde for gutter?
  Dvs. hva er $\mathbf{P(\bar{x}_1 > \bar{x}_2) = P(\bar{x}_1 - \bar{x}_2 > 0) = ?}$
- Forhåndskunnskap: X_1 er N(56.9,2.8), X_2 er N(56.0,3.5),
  da er differanse $\bar{x}_1 - \bar{x}_2$ også normalfordelt med
  - Forventning: $\mu_1 - \mu_2 = 56.9 - 56.0 = 0.9$
  - Varians: $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{2.8^2}{12} + \frac{3.5^2}{8} = 2.18$
  - Standardavvik: $\sqrt{2.18} = 1.48$

Får da

$$P(\bar{x}_1 - \bar{x}_2 > 0) = P\left(\frac{\bar{x}_1 - \bar{x}_2 - 0.9}{1.48} > \frac{0 - 0.9}{1.48}\right)$$
$$= P(Z > -0.61) = 0.7291$$

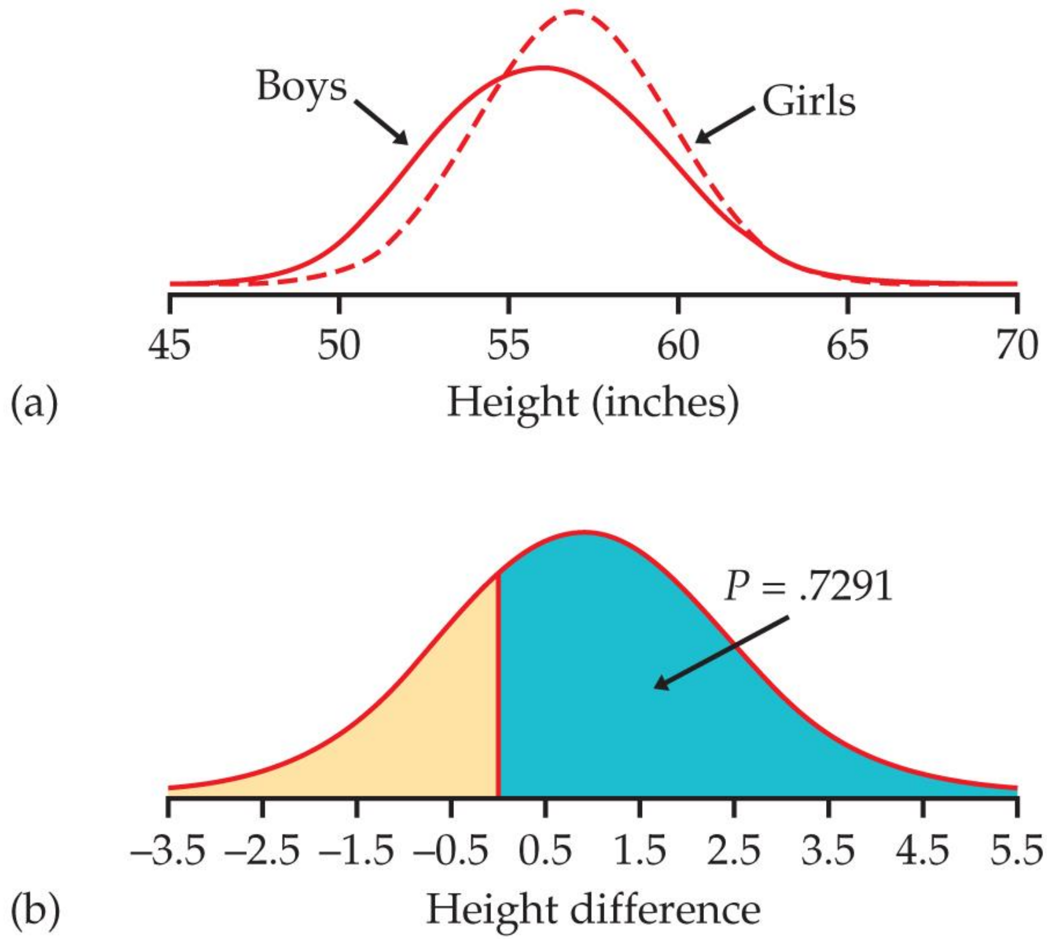


Figure 7.11
Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e,

To utvalg med ukjente populasjonsstandardavvik

- I praksis er ofte σ_1 og σ_2 ukjente
- Estimeres ved empiriske standardavvik s_1 og s_2.
  Vi får da testobservator

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

- Denne er egentlig 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: **Tilnærma** *t(k)*-fordelt for et passende tall *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)

Antall frihetsgrader beregnes av R og annen programvare med denne eller lignende formler:

$$\mathrm{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Dette gir ofte ikke et heltall.

NB: En enkel og grei tommelfingerregel (for eksempel på eksamen) er å velge k lik den minste av (n_1-1) og (n_2-1)

THE TWO-SAMPLE t CONFIDENCE INTERVAL

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. The **confidence interval for $\boldsymbol{\mu_1 - \mu_2}$** given by

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

has confidence level at least C no matter what the population standard deviations may be. The quantity

$$t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

is the **margin of error**. Here, t^* is the value for the $t(k)$ **density curve** with area C between $-t^*$ and t^*. The value of the degrees of freedom k is approximated by software, or we use the smaller of $n_1 - 1$ and $n_2 - 1$. Similarly, we can use either software or the conservative approach with **Table D** to approximate the value of t^*.

Evaluerer ny læringsmetode

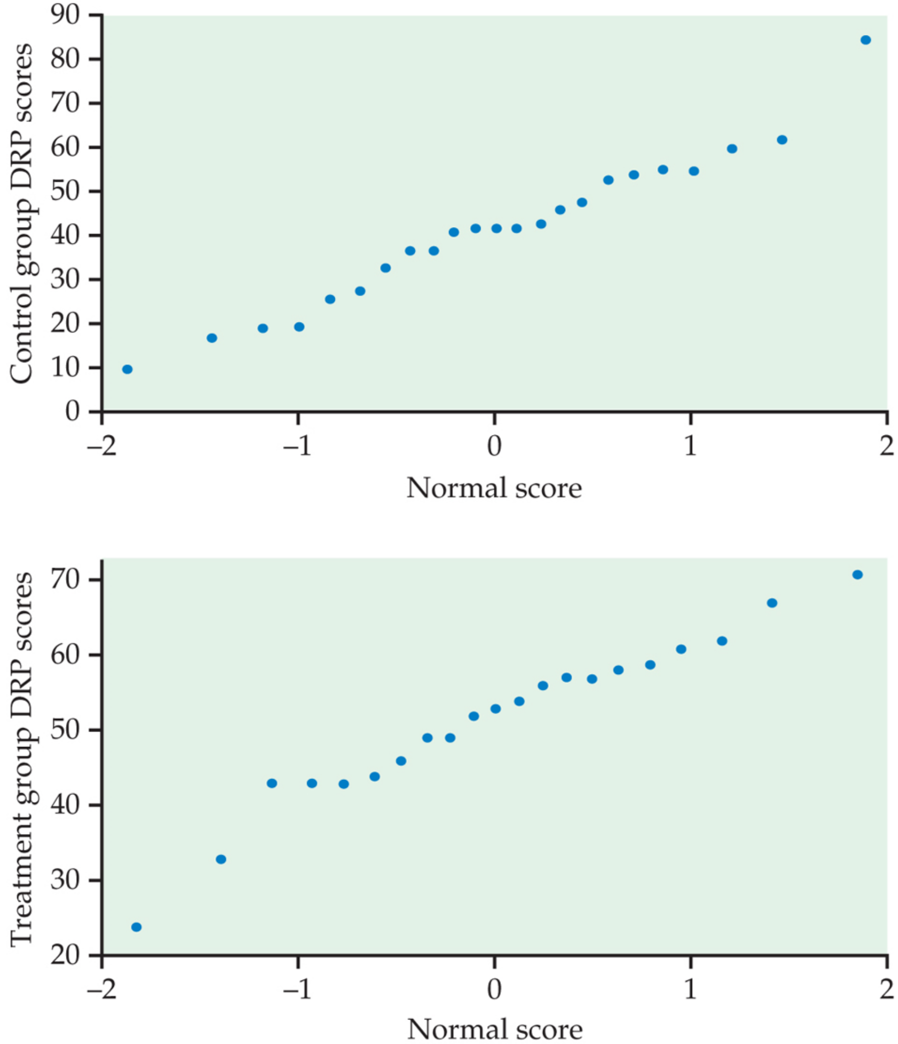

RICHARD HUTCHINGS/Science Source/Getty Images

- **Klasse 1 :** 21 elever prøver **ny** metode
- **Klasse 2 :** 23 elever følger dagens metode
- Klassene ble valgt slik at de var så like som mulig, slik at en eventuell effekt av forskjellen på metodene er minst mulig konfundert med andre forskjeller mellom klassene

TABLE 7.3 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	

Ny læringsmetode: (Normale) Kvantilplott



Ny læringsmetode - konfidensintervall

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

Gruppe	n	Gj.sn.	Std.avvik
1 (Behandling)	21	51.48	11.01
2 (Kontroll)	23	41.52	17.15

For å beskrive størrelsen av behandlingseffekten, lager vi et 95% konfidensintervall for $\mu_1 - \mu_2$.

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_1^2 / n_1 + s_2^2 / n_2} = \sqrt{11.01^2 / 21 + 17.15^2 / 23} = 4.31$$

Frihetsgrader min(n_1 -1, n_2 -1)=20 gir 97.5-persentil t*= 2.086 gir 95% KI:

$$\bar{x}_1 - \bar{x}_2 \pm t^* SE_{\bar{x}_1 - \bar{x}_2} = 51.48 - 41.52 \pm 2.086 \times 4.31 = [\mathbf{0.97}, \mathbf{18.95}]$$

Data-beregna ant. frihetsgrader=37.86 gir 97.5-persentil t*= 2.023, og 95% KI:

$$\bar{x}_1 - \bar{x}_2 \pm t^* SE_{\bar{x}_1 - \bar{x}_2} = 51.48 - 41.52 \pm 2.023 \times 4.31 = [1.24, 18.68]$$

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: $\mu_1 - \mu_2 = \Delta_0$, compute the **two-sample *t* statistic**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

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

Det vanligste er at $\Delta_0 = 0$

To-utvalgs *t*-tester

- **Større enn:** 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 = $P(T \geq t)$
- **Mindre enn:** 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 = $P(T \leq t)$
- **Tosida test:** 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 = $2P(T \geq |t|)$,

For alle alternativene, er T **tilnærma** t-fordelt med *k* frihetsgrader når begge populasjonene er normalfordelte

Ny læringsmetode

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

Gruppe	n	Gj.sn.	Std.avvik
1 (Behandling)	21	51.48	11.01
2 (Kontroll)	23	41.52	17.15

Håper å vise at leseferdigheter i behandlingsgruppa er bedre enn i kontrollgruppa: H_0: $\mu_1 = \mu_2$ mot H_a: $\mu_1 > \mu_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{(51.48 - 41.52) - 0}{\sqrt{11.01^2/21 + 17.15^2/23}} = 2.31$$

Frihetsgrader min(n_1-1, n_2-1) =20 gir p-verdi P(T>2.31) = 0.016 (eller mellom 0.01 og 0.02 fra Tabell D)

Data-beregnet ant. frihetsgrader=37.86 gir p-verdi P(T>2.31) = 0.013

To-utvalgs t-test er mer robust enn ett-utvalgs t-test

- Svært presise når utvalgene er omtrent like store, $n_1 \approx n_2$, og formene på fordelingene er ganske like hverandre
  - selv om fordelingene ikke er normale
- Større utvalg trengs når ulike former på fordelinger
- Velg så like utvalgsstørrelser som mulig!

Utfordringer ved små utvalg

- Ikke nok observasjoner til å sjekke fordeling
- Kun ekstreme uteliggere mulig å detektere
- Feilmarginer i konfidensintervaller blir store
- Hvis størrelsen på en effekt / forskjell er stor nok, er det ofte mulig å trekke konklusjoner likevel

Dersom vi kan anta like standardavvik $\sigma_1 = \sigma_2$

- Hvis vi kan anta $\sigma_1 = \sigma_2 = \sigma$, kan vi uttrykke variansen til $\bar{x}_1 - \bar{x}_2$ som

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

- Med kjent σ blir dermed z-observatoren

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

- Når σ er ukjent må den estimeres. Men begge de empiriske variansene s_1^2 og s_2^2 estimerer σ^2.

Vi beregner begge separat, og slår sammen estimatene er med formelen

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

t-test for to utvalg med like varianser

- Vi får da følgende t-observator

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

- Hvis dataene i begge utvalg er fra normalfordelinger (med lik varians) så er faktisk denne t-observatoren eksakt t-fordelt med $k = n_1 + n_2 - 2$ frihetsgrader.

- Vi kan "vinne" noen frihetsgrader på denne metoden – men metoden er bare gyldig under den strenge antagelsen $\sigma_1 = \sigma_2$

- Det finnes metoder for å teste om $\sigma_1 = \sigma_2$, men disse krever antagelse om normalfordeling.

t-metoder for to utvalg med like varianser og sammenslått (pooled) variansestimat

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

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

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

$$t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

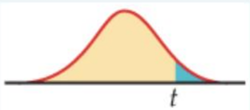
is the **margin of error**.

To test the hypothesis $H_0 : \mu_1 - \mu_2 = \Delta_0$, compute the **pooled two-sample *t* statistic**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

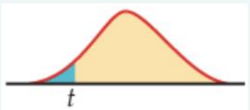
In terms of a random variable T having the $t(n_1 + n_2 - 2)$ distribution, the P-value for a test of H_0 against

H_a: $\mu_1 - \mu_2 > \Delta_0$ is $P(T \geq t)$



Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

H_a: $\mu_1 - \mu_2 < \Delta_0$ is $P(T \leq t)$



Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

H_a: $\mu_1 - \mu_2 \neq \Delta_0$ is $2P(T \geq |t|)$

Moore/McCabe/Craig, *Introduction to the Practice of Statistics*, 9e, © 2017 W. H. Freeman and Company

Gir økt inntak av kalsium **redusert** blodtrykk?

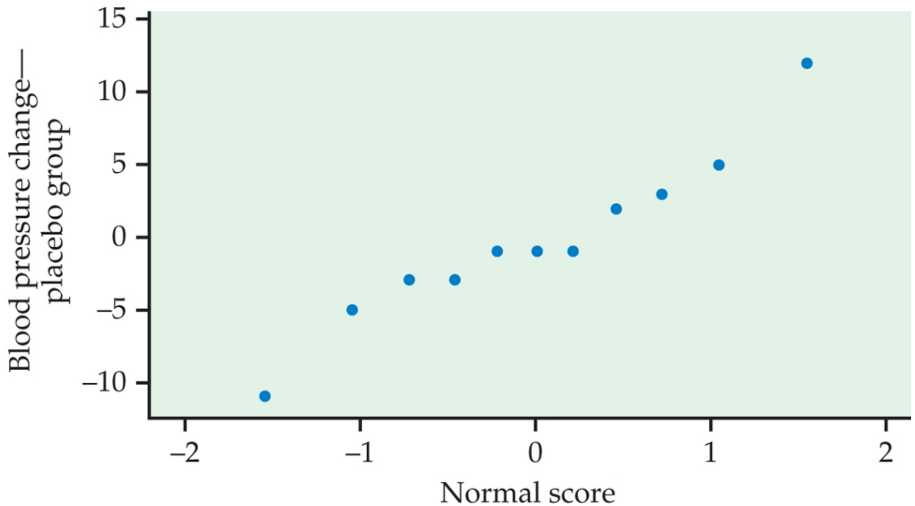
TABLE 7.4 Seated Systolic Blood Pressure (mm Hg)

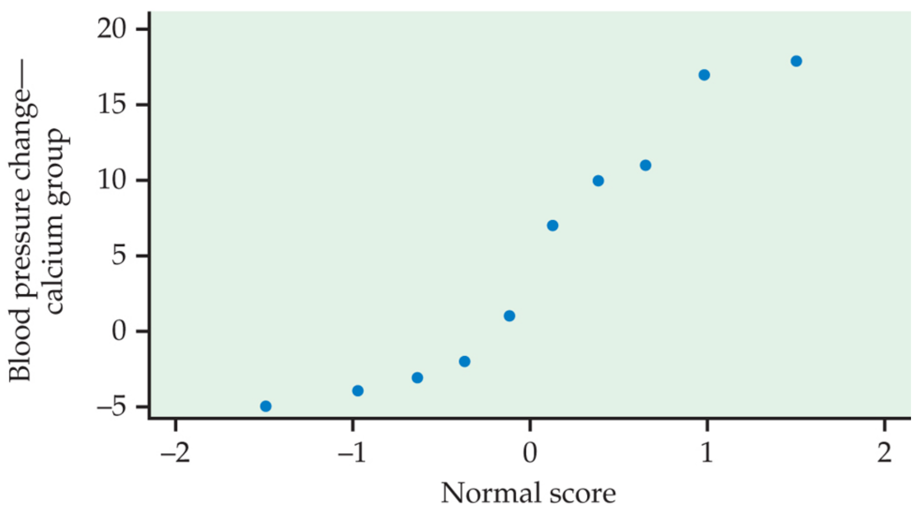
Calcium Group			Placebo Group		
Begin	End	Decrease	Begin	End	Decrease
107	100	7	123	124	−1
110	114	−4	109	97	12
123	105	18	112	113	−1
129	112	17	102	105	−3
112	115	−3	98	95	3
111	116	−5	114	119	−5
107	106	1	119	114	5
112	102	10	114	112	2
136	125	11	110	121	−11
102	104	−2	117	118	−1
			130	133	−3

- NB: Verdiene i kolonnen Decrease (nedgang) er positiv når det er en reduksjon, negativ ved oppgang

Eksempel: Kalsium og blodtrykk

- Kvantilplott: Ingen alvorlige avvik fra normalfordeling for begge grupper
- Må gjøre en signifikanstest for å besvare spørsmålet om økt inntak av kalsium reduserer 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 alternativ-hypotese**, da et tidligere forsøk med rotter hadde vist at kalsium ga en nedgang og det var ingen kjent grunn til å tro 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	8.473
2	Placebo	11	-0.273	5.901

- De empiriske standardavvikene utelukker ikke like populasjons-standardavvik, forskjell i estimatene kan lett være pga tilfeldigheter
- Vi velger her å **anta like standardavvik** i begge grupper

Eksempel: Kalsium og blodtrykk

- Vi vil anta like standardavvik i begge grupper
- Det sammenslåtte standardavviket blir

$$s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = 7.385$$

- Dette gir en t-observator

$$t = \frac{(\bar{x}_1-\bar{x}_2) - (\mu_1-\mu_2)}{s_{pooled}\sqrt{1/n_1 + 1/n_2}} = \frac{(5-(-0.273)) - 0}{7.385\sqrt{1/10 + 1/11}} = 1.634$$

- som under nullhypotesen H_0: $\mu_1 = \mu_2$ er t-fordelt med
  $k = n_1 + n_2 - 2 = 19$ frihetsgrader.
- P-verdi for testen (med H_a: $\mu_1 > \mu_2$) blir 5.9% og altså ikke signifikant.
- Tilsvarende blir 90% konfidensintervallet (med 95 persentil t^*=1.729)

$$\bar{x}_1 - \bar{x}_2 \pm t^* s_{pooled}\sqrt{1/n_1 + 1/n_2} = [-0.31\ ,\ 10.85]$$

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

Vi har snakket om konfidensintervall og t-test for to uavhengige utvalg med ukjent standardavvik

- For t-test har vi snakket om to ulike metoder, nemlig:
  - når de to populasjonene har **ulike** standardavvik
  - når vi antar at de to populasjonene har **like** standardavvik, og beregner *pooled standard deviation*
    - Antagelsen om like standardavvik må brukes med forsiktighet, spesielt hvis utvalgsstørrelsene er veldig forskjellige