Kapittel 7: Inferens for forventninger-ukjent standardavvik

- 7.1: Inferens for forventningen i en populasjon
- 7.2: Inferens for å sammenligne to forventninger

7.1 Inferens for forventningen i en populasjon

- t-fordelingen
- Ett-utvalgs t-konfidensintervall
- Ett-utvalgs t-test
- Matchede par t-prosedyrer
- Robusthet

Ødelegger lagring søtheten av Cola?

En cola-produsent ønsker å teste hvor mye søtheten til en ny cola-drikk påvirkes av lagring. Søthets-tapet som følge av lagring ble bedømt av 10 profesjonelle smakstestere (ved å sammenligne søtheten før og etter lagring):

Smakstester	Søthets-tap
1	2.0
2	0.4
3	0.7
4	2.0
5	–0.4
6	2.2
7	–1.3
8	1.2
9	1.1
10	2.3

Vi ønsker altså å teste om lagring medfører tap i søthet, altså:

H_0: μ = 0 versus H_a: μ > 0

der μ er forventet tap av søthet

Dette ser kjent ut. Men siden dette er en ny cola-oppskrift har vi ikke noe populasjonsdata fra tidligere, og vi kjenner ikke populasjonsparameteren σ.

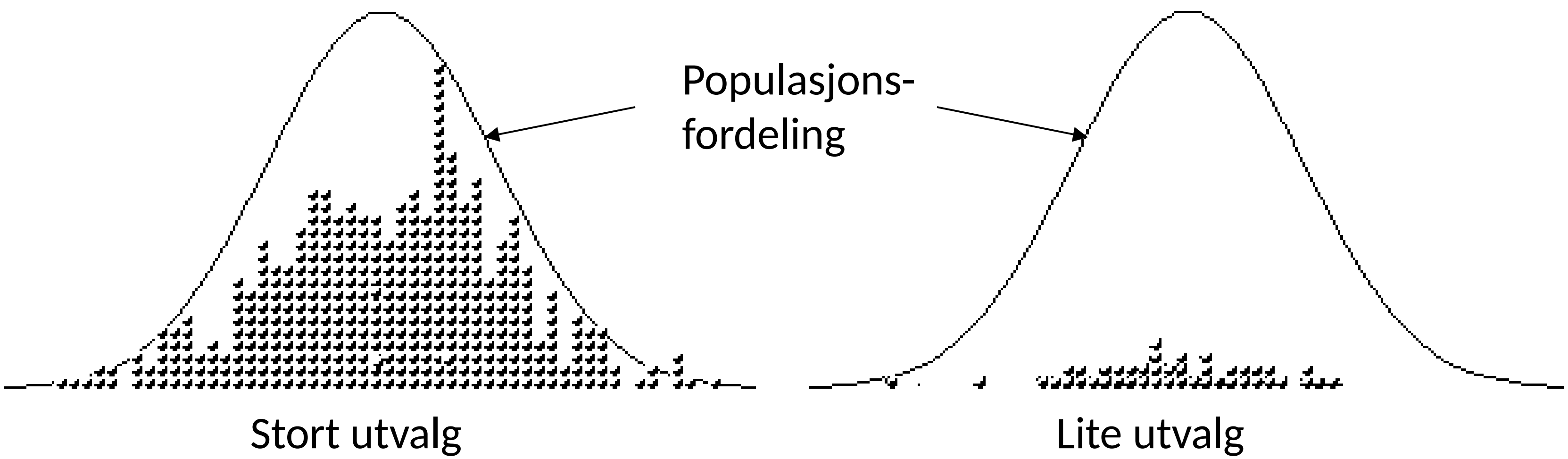
Dette er en veldig vanlig situasjon for relle data.

Når σ er ukjent

Empirisk standardavvik s gir oss et estimat for populasjonens standardavvik σ.

- Når utvalgsstørrelsen er stor, er det sannsynlig at utvalget representerer populasjonen godt. Da er s et godt estimat for σ.

- Men hvis utvalgsstørrelsen er liten, er s et dårlig estimat for σ.



Husk empirisk standardavvik

$$s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

der n-1 kaltes antall frihetsgrader (degrees of freedom, df)

En populasjon

- Anta $x_1,\ldots,x_n$ uavhengige fra $N(\mu,\sigma)$
- Statistikk: $\bar{x}$
- σ kjent: $z=(\bar{x}-\mu)/(\sigma/\sqrt{n})$
- $\sigma/\sqrt{n}$: standardavvik for statistikk $\bar{x}$
- $s/\sqrt{n}$: estimert standardavvik for statistikk $\bar{x}$
- Kalles *standardfeil*

Standard Error, SE = $s/\sqrt{n}$

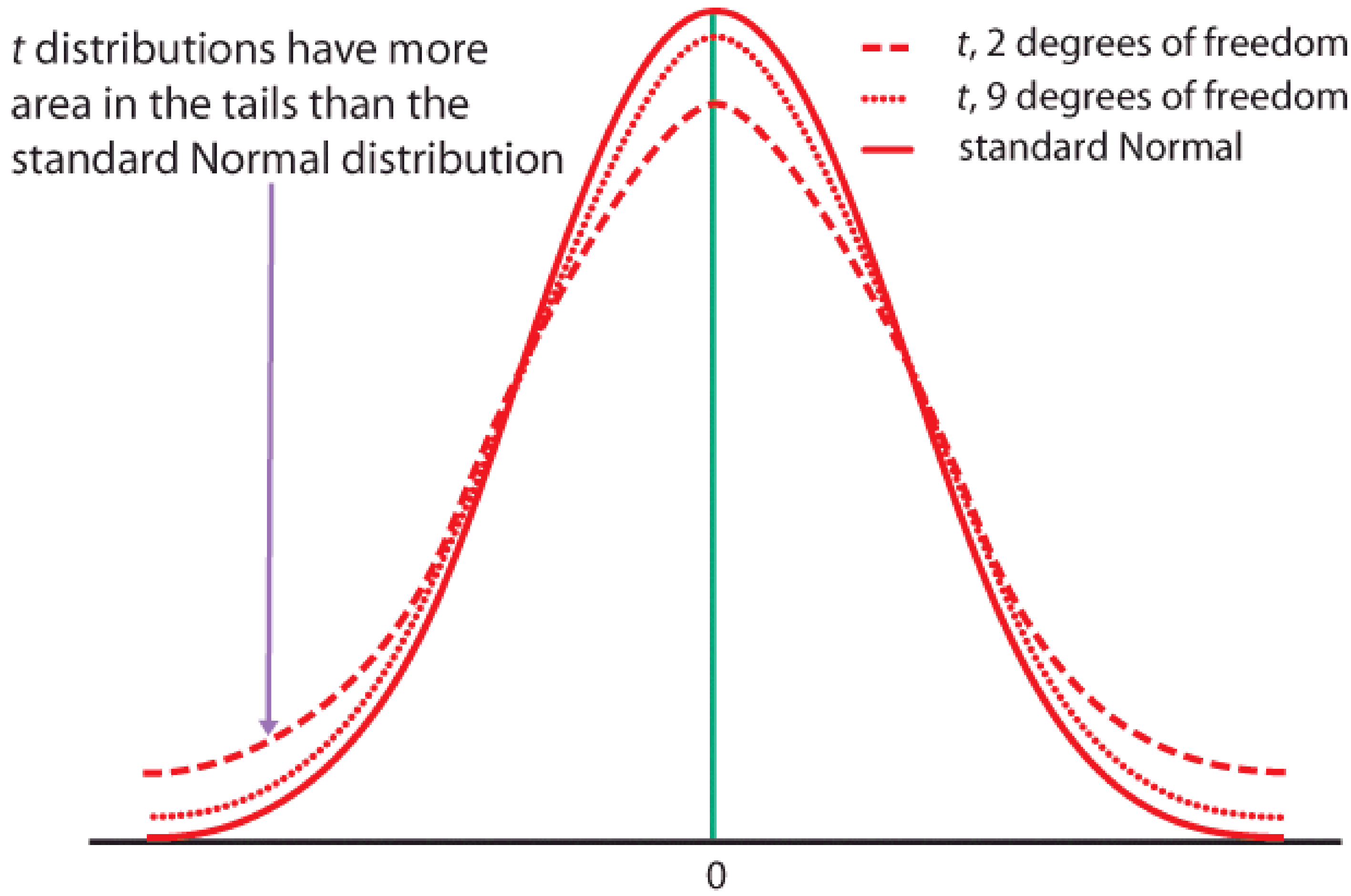
t-fordeling

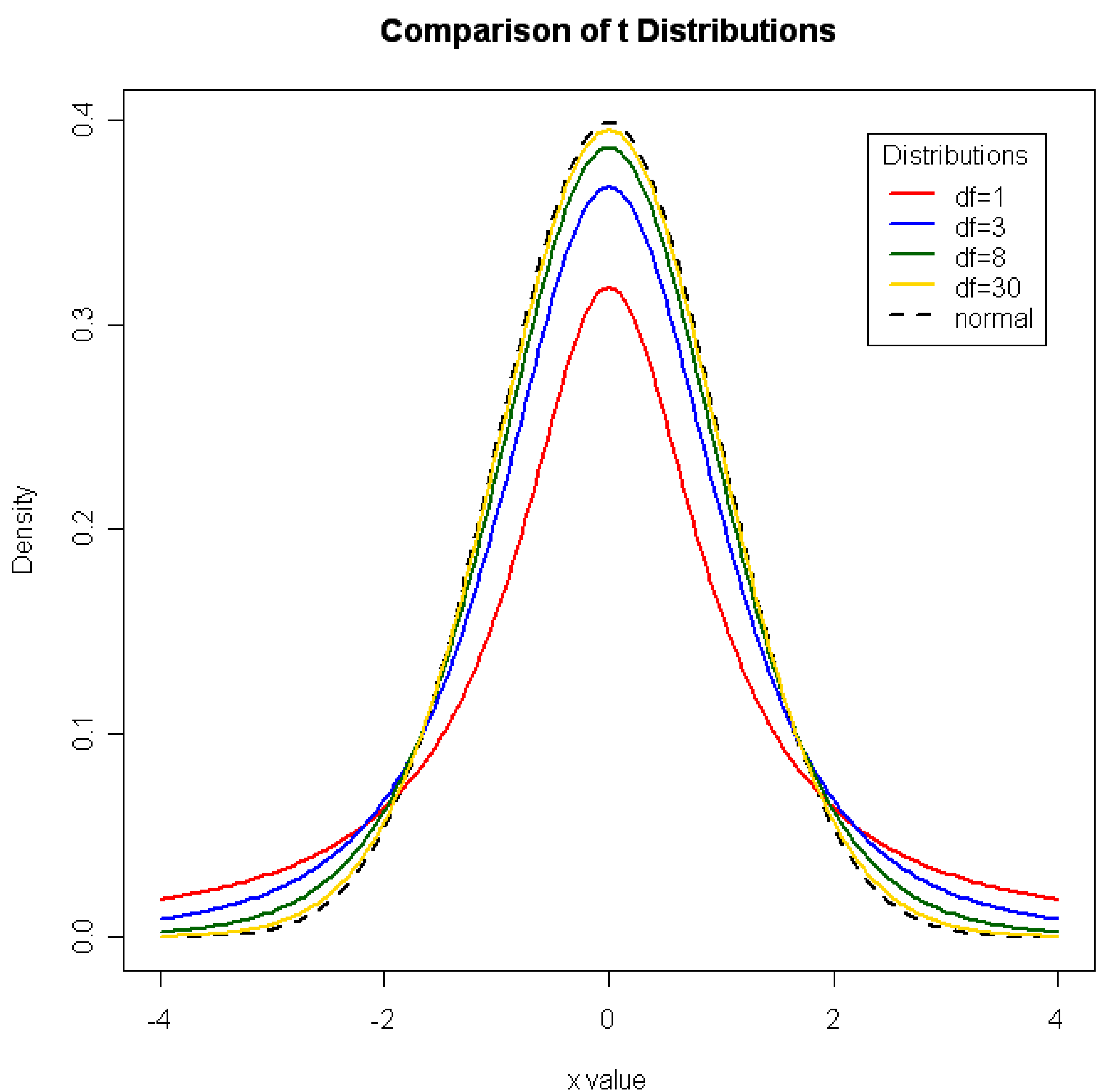
- $z=(\bar{x}-\mu)/(\sigma/\sqrt{n})$ er N(0,1)-fordelt
- $t=(\bar{x}-\mu)/(s/\sqrt{n})$ er *t-fordelt* med *n-1 frihetsgrader*

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

Kalles *ett-utvalgs t-statistikk*
(eller ett-utvalgs t-observator)

- Form som normalfordeling
- Ekstra spredning/usikkerhet pga. ukjent σ
- Nærmer seg N(0,1) når *n* vokser


t distributions have more area in the tails than the standard Normal distribution
t, 2 degrees of freedom
t, 9 degrees of freedom
standard Normal
0

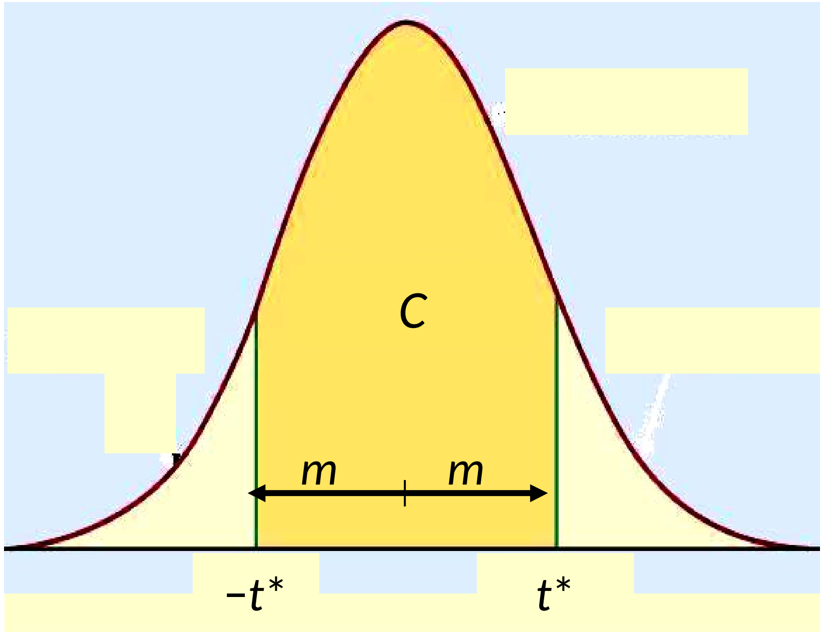


Comparison of t Distributions
Distributions
df=1
df=3
df=8
df=30
normal
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x value

Ett-utvalgs t-konfidensintervall

- σ *kjent:* $[\bar{x}-z^*\sigma/\sqrt{n}, \bar{x}+z^*\sigma/\sqrt{n}]$
- z^* er verdien slik at arealet mellom $-z^*$ og z^* i $N(0,1)$ fordelingen er C

- σ *ukjent:* $[\bar{x}-t^*s/\sqrt{n}, \bar{x}+t^*s/\sqrt{n}]$
- t^* er verdien slik at arealet mellom $-t^*$ og t^* i $t(n-1)$ fordelingen er C



- Feilmarginen er $m = t^*s/\sqrt{n}$
- Eksakt hvis normalfordelte data
- Tilnærmet riktig ellers

Når σ er ukjent, bruker vi *t*-fordeling med "*n*−1" frihetsgrader (degrees of freedom df).

Tabell D viser *z*-verdier og *t*-verdier knyttet til typiske P-verdier/ konfidensnivåer

Når σ er kjent, bruker vi normalfordeling og den standardiserte z-verdien.

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

	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	[illegible]	[illegible]	[illegible]	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	[illegible]	[illegible]	[illegible]	[illegible]	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	[illegible]	[illegible]	[illegible]	[illegible]	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	[illegible]	[illegible]	[illegible]	[illegible]	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	[illegible]	[illegible]	[illegible]	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	[illegible]	[illegible]	[illegible]	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	[illegible]	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*											

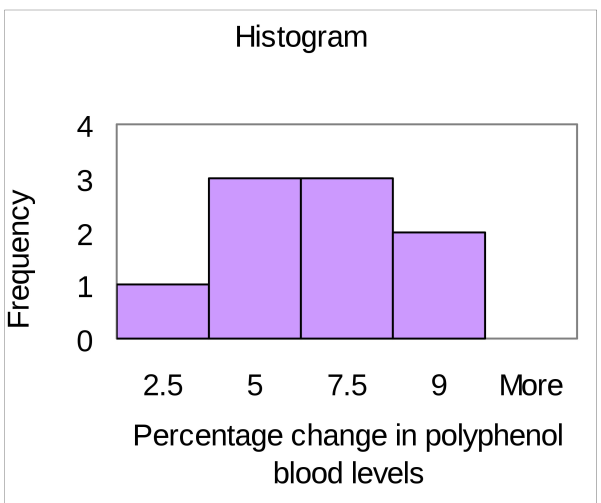
Moderate mengder rødvin

Å drikke rødvin i moderate mengder kan beskytte mot hjerteinfarkt. Polyphenolene i rødvinen virker på kolesterolet, noe som kanskje kan motvirke hjerteinfarkt.

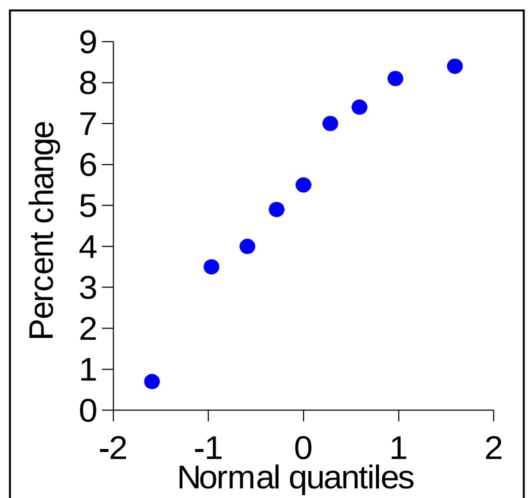
For å undersøke om moderat inntak av rødvin øker forventet nivå av polyphenoler i blodet skulle en gruppe på 9 tilfeldig valgte friske menn drikke en moderat mengde rødvin hver dag i 2 uker. Nivået av polyphenoler i blodet ble målt både før og etter disse 2 ukene, og den prosentvise endringen var som følger:

Først: Er dataene tilnærmet normalfordelte?

0.7 3.5 4 4.9 5.5 7 7.4 8.1 8.4



0	7
1	
2	
3	5
4	09
5	5
6	
7	04
8	14



Det er en lav verdi, men generelt ser det ut til at man kan anta at dataene kan antas å være tilnærmet normalfordelte.

Hva er 95% konfidensintervallet for forventet prosentvis økning av nivå av polyphenoler i blodet?

Utvalgs-gjennomsnitt = 5.5; s = 2.517; df = $n - 1 = 8$

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

Utvalgsfordelingen er en t-fordeling med $n - 1$ frihetsgrader.

For df = 8 og C = 95%: $t^* = 2.306$.

Feilmarginen m er: $m = t^* s/\sqrt{n} = 2.306 * 2.517/\sqrt{9} \approx 1.93$.

Med 95% sikkerhet kan vi si at forventet prosentvis økning av nivå av polyphenoler i blodet for friske menn som drikker en moderat mengde rødvin hver dag er mellom 3.6% og 7.4%. *Viktig: Konfidensintervallet viser hvor stor endringen av nivået av ployphenoler I blodet er, men ikke om det har en innvirkning på menns helse!*

Ett-utvalgs t-test

Fremgangsmåten for å teste en hypotese er som tidligere:

1. Formuler null- og alternativ-hypoteser (H_0 versus H_a)
2. Velg signifikansnivå α
3. Beregn t-observator og antall frihetsgrader (beregnet ved å anta at parameterverdien gitt av H_0 er sann)
4. Finn ønsket sannsynlighet fra *Tabell D* (sannsynligheten for at et utfall er like ekstremt eller mer ekstremt enn faktisk utfall)
5. Oppgi P-verdi og formuler en konklusjon

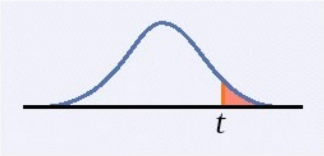
THE ONE-SAMPLE *t* TEST

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

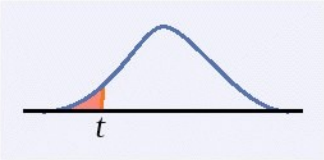
$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

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

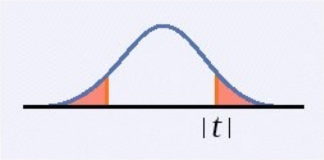
$H_a: \mu > \mu_0$ is $P(T \geq t)$



$H_a: \mu < \mu_0$ 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.

Definition, pg 454
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

Søthet av cola (*fortsettelse*)

Er det bevis for at lagring medfører søthets-tap for den nye cola-oppskriften på et 0.05 signifikansnivå ($\alpha = 5\%$)?

H_0: $\mu = 0$ versus H_a: $\mu > 0$ (ensidig test)

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

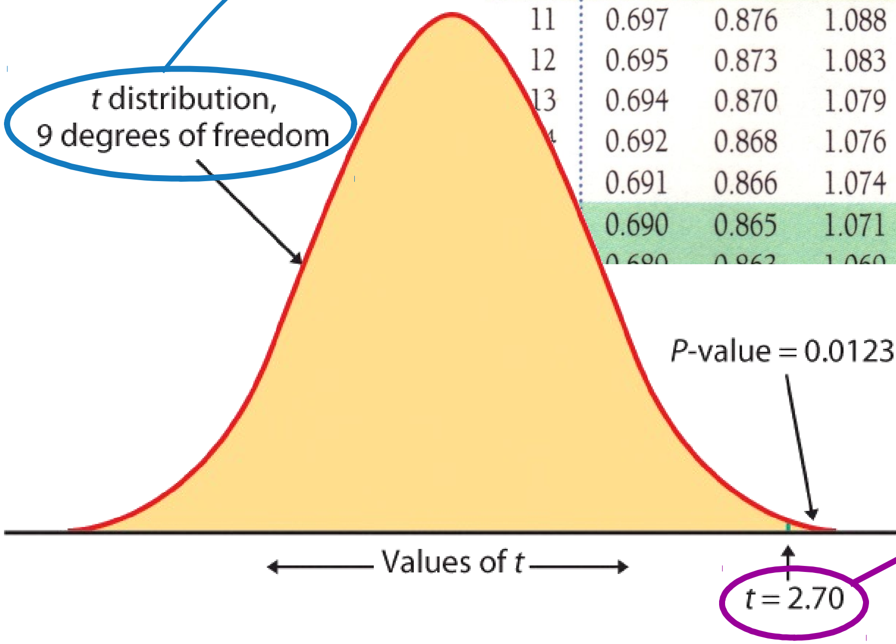
Antall frihetsgrader: df=10-1=9

Taster	Sweetness loss
1	2.0
2	0.4
3	0.7
4	2.0
5	-0.4
6	2.2
7	-1.3
8	1.2
9	1.1
10	2.3

Average	**1.02**
Standard deviation	**1.196**
Degrees of freedom	***n* − 1 = 9**

	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
4	0.692	0.868	1.076	1.345	1.761	2.145	2.26[illegible]	2.624	2.977	3.326	3.787	4.140
	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015

For df = 9 ser vi bare på denne linjen i tabellen



Den beregnede verdi av t er 2.7. Vi finner de to nærmeste t-verdiene:

$2.398 < t = 2.7 < 2.821$

så

$0.02 >$ øvre hale p > 0.01

For en en-sidig H_a, er dette P-verdien (mellom 0.01 og 0.02);

for en to-sidig H_a, hadde P-verdien vært det dobbelte (mellom 0.02 og 0.04).

Søthet av cola (*fortsettelse*)

Er det bevis for at lagring medfører søthets-tap for den nye cola-oppskriften på 0.05 signifikansnivå ($\alpha = 5\%$)?

H_0: $\mu = 0$ versus H_a: $\mu > 0$ (ensidig test)

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

– $2.398 < t = 2.70 < 2.821$, altså er $0.02 > p > 0.01$.
  $p < \alpha$, altså er resultatet signifikant.

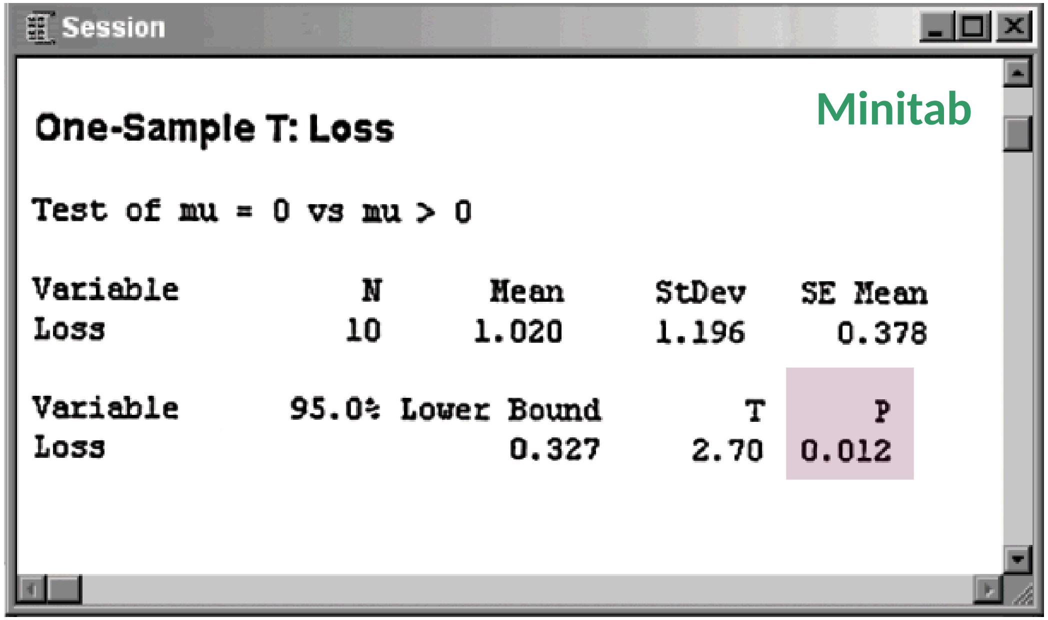
Taster	Sweetness loss
1	2.0
2	0.4
3	0.7
4	2.0
5	-0.4
6	2.2
7	-1.3
8	1.2
9	1.1
10	2.3
Average	**1.02**
Standard deviation	**1.196**
Degrees of freedom	***n* − 1 = 9**

	Upper tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
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

t-testen har en signifikant p-verdi. Vi forkaster H_0.

Det er et signifikant forventet tap av søthet som følge av lagring.

Søthet av cola (*fortsettelse*)



One-Sample T: Loss

Minitab

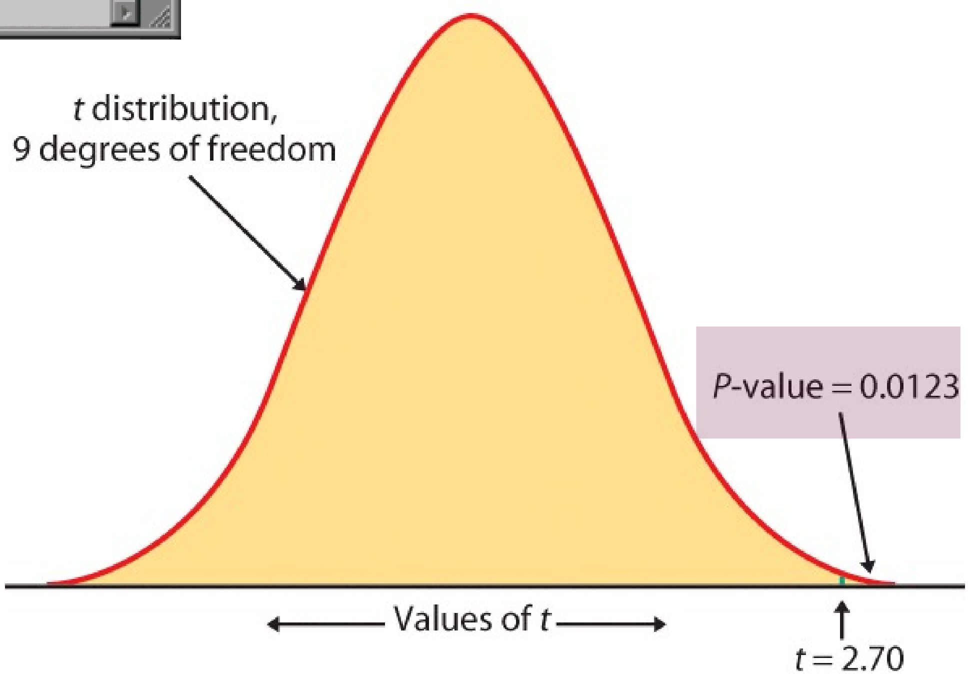
Test of mu = 0 vs mu > 0

Variable	N	Mean	StDev	SE Mean
Loss	10	1.020	1.196	0.378

Variable	95.0% Lower Bound	T	P
Loss	0.327	2.70	0.012

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

$$df = n - 1 = 9$$



- Parrede (matchede) t-prosedyrer

Noen ganger vil vi sammenligne behandlinger på de samme individene. Dette gir oss observasjoner som ikke er uavhengige – de er parret eller matchede to og to:

  - Eks. Før og etter behandling (blodtrykk før og etter behandling med betablokker, søtsmak før og etter lagring)
  - Eks. Tvillingstudier, begrenser effekten av genetiske forskjeller ved å se på en variabel i sett av tvillinger
  - Eks. Ved å bruke folk som matcher hverandre i alder, kjønn, utdanning i sosiale studier, kan man kansellere ut effekten av slike underliggende lurevariable

I slike situasjoner kan vi bruke par-data til å teste forskjell i forventning mellom de to fordelingene. Vi studerer variabelen $X_{\text{diff}} = (X_1 - X_2)$, og tester

$$\boxed{H_0\colon \mu_{\text{diff}} = 0\ ;\ \ H_a\colon \mu_{\text{diff}} > 0 \text{ (eller } <0\text{, eller } \neq 0\text{)}}$$

Dette er det samme som å teste i en ett-utvalgs-situasjon . Data for individ nr *i* er da $x_{\text{diff}} = x_{1i} - x_{2i}$

Søthet av cola *(om igjen)*

Søthets-tapet pga lagring ble bedømt av 10 profesjonelle smakstestere (som sammenlignet søtheten **før og etter**):

Smakstester	Søthets-tap
1	2.0
2	0.4
3	0.7
4	2.0
5	–0.4
6	2.2
7	–1.3
8	1.2
9	1.1
10	2.3

Vi ønsker å teste om lagring medfører tap I søthet, atltså:

H_0: $\mu = 0$ versus H_a: $\mu > 0$

Selv om teksten ikke sa det eksplisitt, er dette et pre-/post-test design og variabelen er differensen i søthet av cola før (X_1) minus etter lagring (X_2).

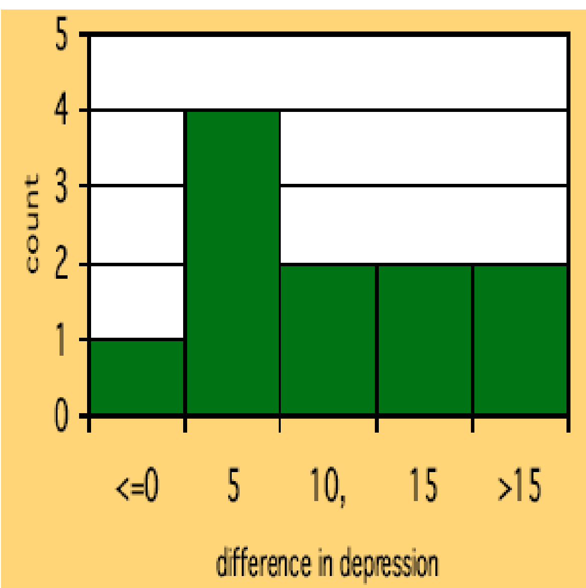
En parret signifikanstest er altså akkurat det samme som en ett-utvalgs signifikanstest.

Gir koffeinavholdenhet økt depresjon?

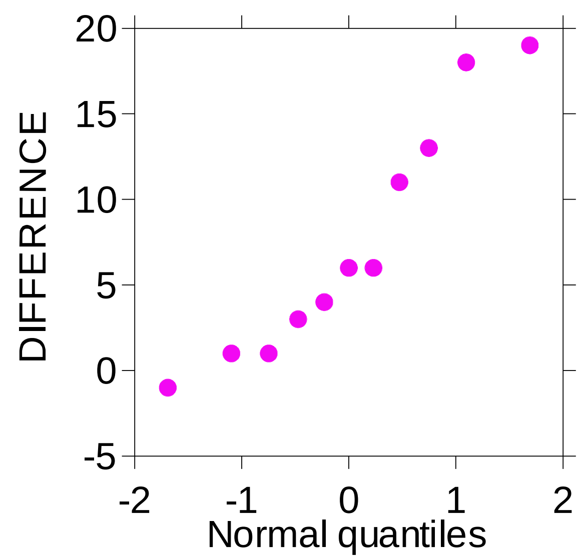
Subject	Depression with Caffeine	Depression with Placebo	Placebo - Cafeine
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	-1

11 individer som er diagnostisert som koffeinavhengige blir forhindret fra å innta koffein-holdig mat og drikke og får i stedet daglige piller. Noen ganger inneholder disse pillene koffein og noen ganger inneholder de en placebo. Forskjellen i depresjonsnivå når individene får koffein kontra når de får placebo ble evaluert.

- Det er 2 datapunkter for hvert individ, men vi ser bare på differansen i depresjonsnivå ($x = x_{2i} - x_{1i}$)
- Utvalgsfordelingen ser ut til å være passende for en *t*-test.



11 “difference” data points.



Gir koffeinavholdenhet økt depresjon?

For hvert individ i utvalget har vi beregnet forskjellen i depresjonsnivå (placebo minus koffein).

Det var 11 "differanse"-verdier, altså er df = $n - 1 = 10$.

Vi beregner at $\bar{x} = 7.36; s = 6.92$

H_0: $\mu_{\text{diff}} = 0$; H_0: $\mu_{\text{diff}} > 0$

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{7.36}{6.92/\sqrt{11}} = 3.53$$

Subject	Depression with Caffeine	Depression with Placebo	Placebo - Cafeine
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	-1

For df = 10: $3.169 < t = 3.53 < 3.581$ ➔ $0.005 > p > 0.0025$

Koffeinavholdenhet gir en signifikant økning av depresjon.

•Robusthet

t-prosedyrene er eksakt riktige når populasjonen er eksakt normalfordelt. I praksis vil vi ikke alltid ha eksakt normalfordeling, men t-prosedyrene er **robuste** i forhold til mindre avvik fra normalitet – resultatene blir ikke så gale selv om normalitetsantakelsen ikke holder. Viktige faktorer er:

- **Tilfeldig utvalg.** Utvalget **må** være et SRS fra populasjonen.
- **Uteliggere og skjevhet.** Påvirker gjennomsnittet og derfor også t-prosedyrene. MEN, betydningen av dette avtar med økende antall observasjoner på grunn av sentralgrenseteoremet (CLT).

Spesielt:

- Når $n < 15$, må data være tilnærmet normalfordelte og uten uteliggere
- Når $15 > n > 40$, er det ok med noe skjevhet, men ikke uteliggere
- Når $n > 40$, er t-observatoren ok selv med sterk skjevhet i underliggende fordeling

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^*.

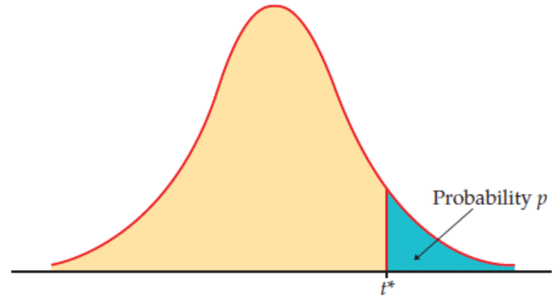


TABLE D

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											

7.2 Sammenligning av to forventinger

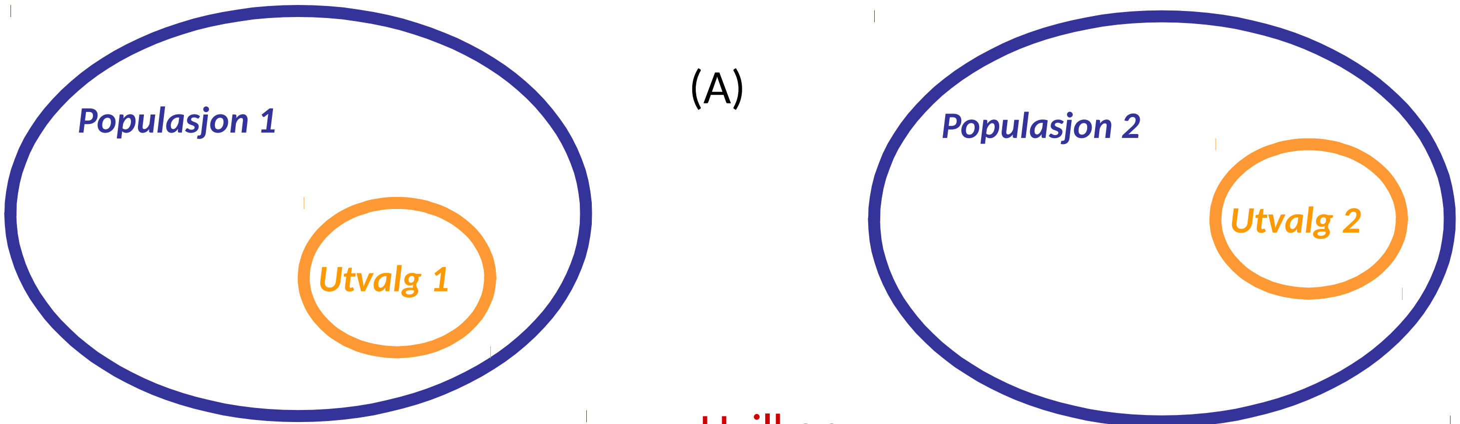
- To-utvalgs z-statistikk
- To-utvalgs t-prosedyrer
- To-utvalgs t-tester
- To-utvalgs t-konfidensintervall
- Robusthet
- To-utvalgs t-prosedyrer når variansene er like

Sammenlikning av to forventninger

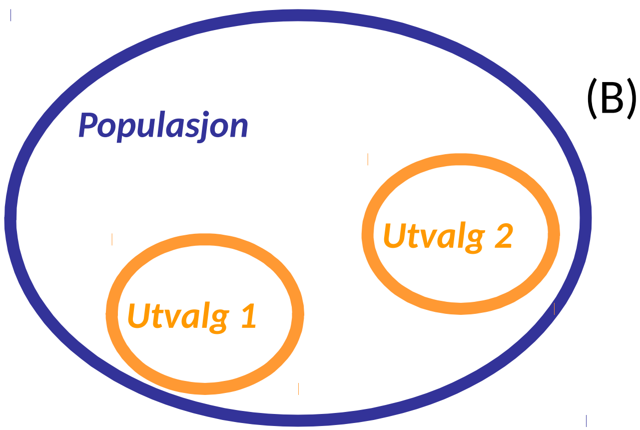
- Effekt av økt kalsium på blodtrykk
  - En gruppe får økt kalsium i diett
  - Kontrollgruppe får placebo

- 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



Hvilken situasjon har vi?



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

Statistikk for to-utvalg med kjente standardavvik

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 $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

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

Men veldig uvanlig at man kjenner standardavvikene!

To-utvalg – ukjente varianser

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

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

- Ikke *t*-fordelt! En *t*-fordeling erstatter en N(0,1)-fordeling bare når et enkelt standardavvik σ i en *z*-statistikk erstattes med ett 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)

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 $n_2 - 1$.

Eks. ny læringsmetode for lesing

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

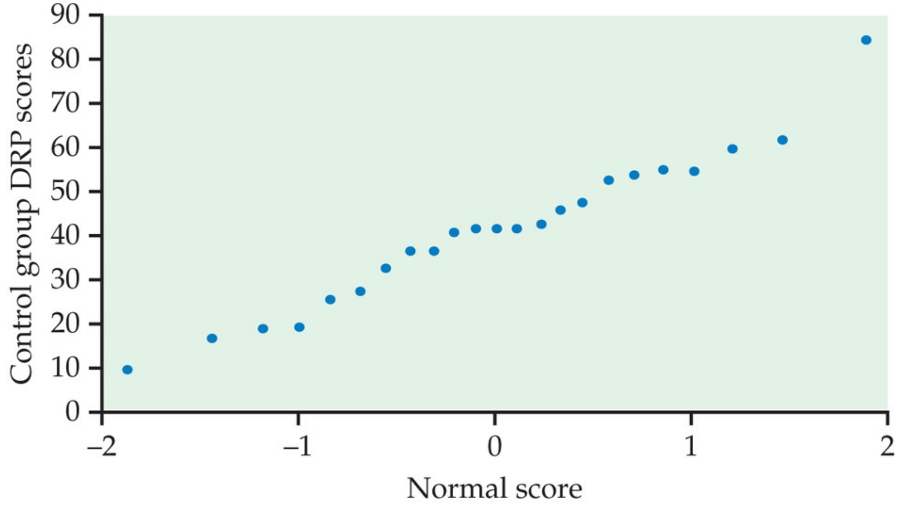
TABLE 7.4

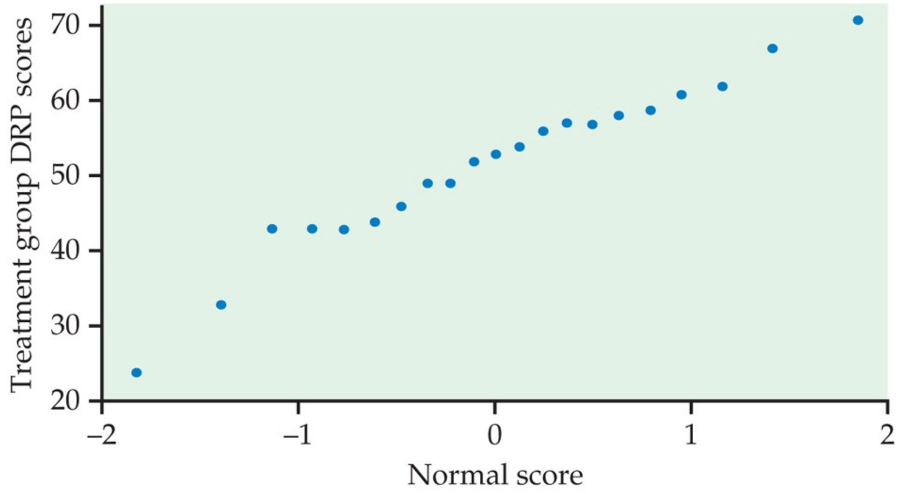
DRP scores for third-graders

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
© 2005 W. H. Freeman and Company

Ny læringsmetode: (Normale) Kvantilplott





Ny læringsmetode for lesing

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

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

- Frihetsgrader: Minste av n_1-1=20 og n_2-1=22

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

- 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]

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 $n_2 - 1$.

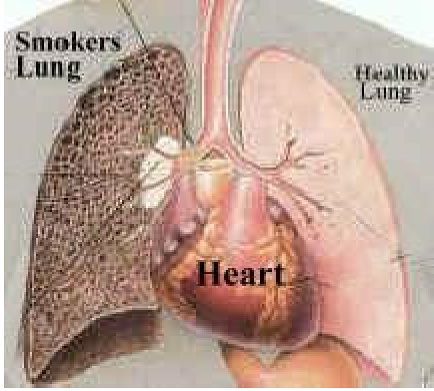
Definition, pg 489
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

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?



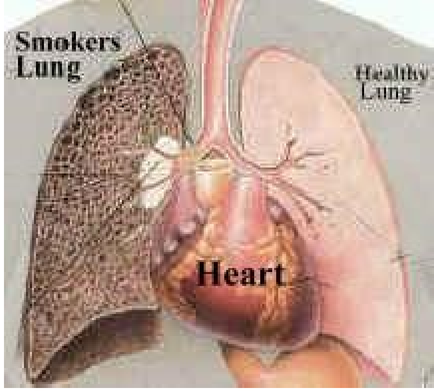
Forsert vital kapasitet (FVC) er volumet (i milliliter) av luft som en person kan puste ut etter maksimal innpust.

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

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

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

I tabell D, for antall frihetsgrader=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.

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

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-statistikk 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-statistikken med en t-statistikk ved å bytte ut σ med et *samlet* estimat på σ, da er t-statistikken eksakt t-fordelt
- (Erstatter da bare ett enkelt standardavvik i en z-statistikk med ett estimert standardavvik)

Statistikk 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 σ²
- 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-statistikken

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

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 $H_0\colon \mu_1 = \mu_2$, 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

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

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

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

Definition, pg 499
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

Eksempel: Kalsium og blodtrykk

- Gir økt inntak av kalsium redusert blodtrykk?

TABLE 7.5

Seated systolic blood pressure

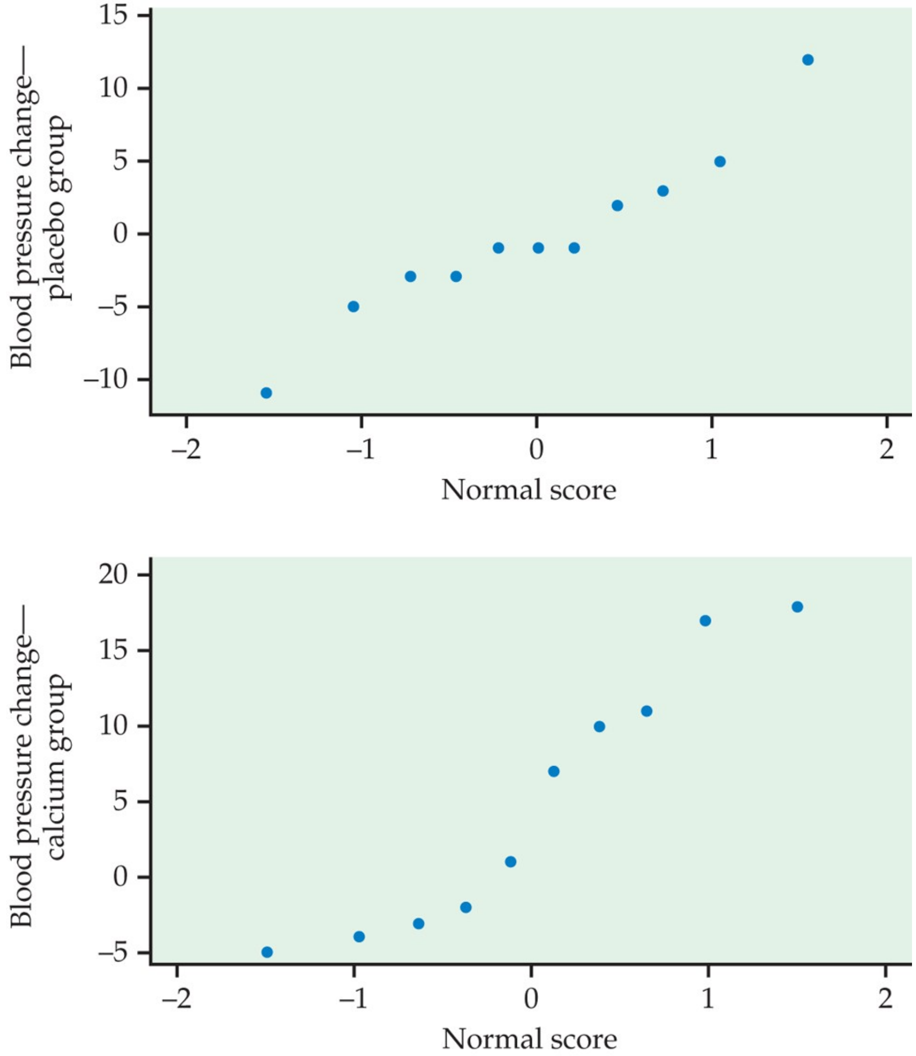
Calcium group			Placebo group		
Begin	End	Decrease	Begin	End	Decrease
107	100	7	123	124	−1
110	114	−4	109	97	12
123	105	18	112	113	−1
129	112	17	102	105	−3
112	115	−3	98	95	3
111	116	−5	114	119	−5
107	106	1	119	114	5
112	102	10	114	112	2
136	125	11	110	121	−11
102	104	−2	117	118	−1
			130	133	−3

Table 7-5
Introduction to the Practice of Statistics, Fifth Edition
© 2005 W. H. Freeman and Company

- 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-statistikken 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