![NTNU – Det skapende universitet]

Medisinsk statistikk, termin IC

av
Stian Lydersen, professor i medisinsk statistikk
Regionalt kunnskapssenter for barn og unge
- Psykisk helse og barnevern (RKBU Midt-Norge)

Forelesning 7 januar 2013
Oppdatert 7 januar 2013

Læringsmål statistikk (3 og 7 januar 2013)

8.1.5 redegjøre for følgende begreper innenfor beskrivende statistikk: gjennomsnitt (mean), median, percentiler, standardavvik (SD), standardfeil (SEM), frekvenstabell og krysstabell, og tolke hva disse forklarer om enkle eksempeldatasett

8.1.6 redegjøre for hva som fremstilles i graftypene histogram, stolpediagram, Box-plott og spredningsplott.

8.1.7 redegjøre for begrepene konfidensintervall, nullhypotese, p-verdi, teststyrke, type I og type II-feil.

8.1.8 redegjøre for normalfordeling og binomisk fordeling, og velge egnet metode mellom uparet og paret T-test, uparet og paret ikke-paramterisk test, kjikvadrat-test, og tilhørende konfidensintervaller.

Innhold:

- Deskriptiv statistikk
- Enkel sannsynlighetsregning
- Randomiserte kontrollerte studier: Randomisering
- Populasjon og tilfeldige utvalg
- Statistisk inferens: Hypotesetesting og konfidensintervaller

Eksempel - postoperativ kvalme

Behandling * Kvalmeklasse Krysstabell

			Kvalme		
			lite eller ingen	betydelig	Total
Behandling	Nei	Antall	18	12	30
		%	60,0%	40,0%	100,0%
	Ja	Antall	24	5	29
		%	82,8%	17,2%	100,0%
Total		Antall	42	17	59
		%	71,2%	28,8%	100,0%

Risikoreduksjon i dette utvalget: 82,8% - 60,0% = 22,8%

Hva kan vi si om effekt av behandling i en populasjon av aktuelle pasienter?

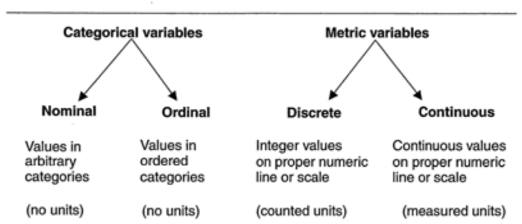


Figure 1.1 Types of variable (Bowers, 2008)

4.1 Sannsynlighetsfordeling (for tellevariabler)

- Stokastisk forsøk: Vet ikke utfallet på forhånd. Men vet mengden mulige utfall på forhånd.
- Stokastisk variabel (tilfeldig variabel) Tallstørrelse knyttet til utfallet. Vet ikke verdien på den før forsøket er utført.
- Sannsynlighetsfordeling: Sannsynlighetene for de mulige verdiene.

Forventningen til X
Forventningsverdien
(expectation, expected value, mean)

$$E(X) = \sum_{\text{alle } x_i} x_i P(X = x_i)$$

"tyngdepunktet"

NTNU
Det skapende universitet
www.ntnu.no

Variansen til X

$$Var(X) = \sum_{\text{alle } x_i} (x_i - E(X))^2 P(X = x_i)$$

$$\boxed{= \sum_{\text{alle } x_i} x_i^2 P(X = x_i) - (E(X))^2}$$

Lettest ved "håndregning"

Standardavviket til X
(Standard deviation)

$$SD(X) = \sqrt{Var(X)}$$

NTNU
Det skapende universitet
www.ntnu.no

De store talls lov:
Når antall observasjoner *n* vokser (mot uendelig), vil:

- $n_A/n \rightarrow P(A)$

$\bar{x} \rightarrow E(X)$

$s^2 \rightarrow Var(X)$

NTNU
Det skapende universitet
www.ntnu.no

Binomisk forsøksrekke. Definisjon:

- De enkelte forsøk er uavhengige av hverandre
- I hvert forsøk registreres hvorvidt hendelsen A inntreffer eller ikke
- Sannsynligheten for A, p=P(A), er den samme i hvert forsøk.

NTNU
Det skapende universitet
www.ntnu.no

Binomisk forsøksrekke - eksempler

- Barnefødsler: Kjønn på etterfølgende enkeltfødsler ved et sykehus
- Terningkast. P(sekser)=1/6.
- Behandling av en bestemt sykdom: Pasienten blir frisk.

NTNU
Det skapende universitet
www.ntnu.no

Binomisk fordeling:
X suksesser blant n forsøk, P(suksess)=p i hvert forsøk:

$$X \sim bin(n, p)$$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E(X) = np$$

$$Var(X) = np(1-p)$$

NTNU
Det skapende universitet
www.ntnu.no

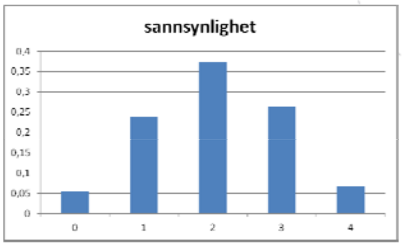
13

Eksempel: X=antall gutter blant 4 enkeltfødsler: p=0.513

$$E(X) = 4 \cdot 0.513 = 2.052$$

$$P(X=2) = \binom{4}{2} 0.513^2 (1-0.513)^{4-2} = 0.374$$

14



Antall gutter blant 4 uavhengige fødsler

15

Randomiserte kontrollerte studier

16

Randomiserte kontrollerte studier (Randomized controlled trials – RCT)

- Sammenlikne to (eller flere) behandlinger A og B
- Trekke tilfeldig (randomisere) om pasienten skal ha behandling A eller B. Datamaskinbasert. Forseglede konvolutter ikke anbefalt.
- Dobbelt blindt forsøk: Verken pasient eller behandler vet hvilken behandling pasienten får.

17

Randomisering

- Bør være tilfeldig og uforutsigbar

Alternativer
- Enkel randomisering
- Blokkrandomisering
- Stratifisert randomisering
- Minimering

18

Enkel randomisering

- Trekk med samme sannsynlighet (vanligvis ½) hver gang
- Eksempel, 30 pasienter: BBABBABBBBBBABAAABABBABBAABBBA
- Fordeler: Enkelt og fullstendig uforutsigbart
- Ulempe(r)?

Blokkrandomisering

- Randomiser halvparten av pasientene i hver blokk til behandling A
- Eksempel med blokkstørrelse b=4, 30 pasienter
- BAAB|BABA|BABA|BBAA|ABAB|AABB|BAAB|BA
- Fordel: Omtrent like mange i hver behandlingsgruppe
- Ulempe(r)?

Stratifisert randomisering

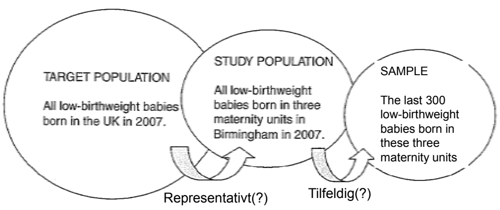
- Strata kan f.eks være klinikk (i en multisenterstudie), kjønn, alder over 75 år.
- Blokkrandomisering innen hvert stratum
- Eksempel: 30 pasienter. 2 sentre og 2 kjønn gir 4 strata
  - Senter 1, menn: BAAB|BABA
  - Senter 1, kvinner: AA
  - Senter 2, menn: BBAA|ABAB|AABB|BAAB|BA
  - Senter 2, kvinner: AA
- Statifiser i få strata (om noen), bare meget viktige prognostiske faktorer.
- Tommelfingerregel: Unngå flere strata enn n/(4b)

Minimering ”Minimization”

- Mulig å balansere flere faktorer enn ved stratifiserting
- Bland and Altman: ”Treatment allocation by minimisation”, BMJ 2005.
- Brukes i økende grad i små studier

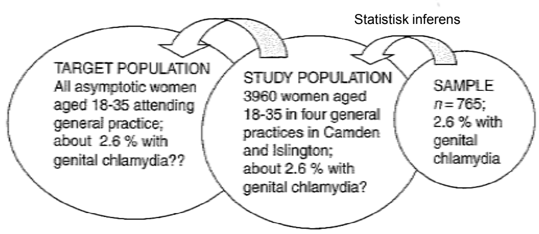
Populasjon og tilfeldige utvalg

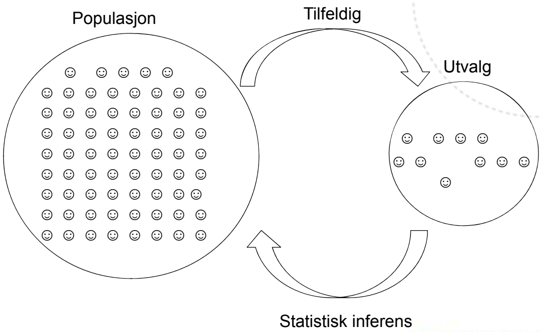
Statistisk modell og utvalg



I noen studier er målpopulasjonen og studiepopulasjonen den samme (ideelt)

Figur fra Bowers(2008)





Populasjon	Utvalg
Tilfeldig (stokastisk) variabel X	Observasjoner $x_1, \ldots, x_n$
Forventningsverdi $E(X) = \mu$	Gjennomsnitt $\bar{x}$
Varians σ^2	(Utvalgs)varians, empirisk varians s^2
Standardavvik σ	(Utvalgs)standardavvik, empirisk standardavvik s
Sannsynlighetsfordeling: F.eks Normalfordeling: $X \sim N(\mu, \sigma^2)$ Binomisk fordeling $X \sim bin(n, p)$	

Statistisk inferens (bekreftende statistikk)

- Trekke slutninger om (en eller flere parameter(e) i) en populasjon basert på analyse av et tilfeldig utvalg:
  - Estimat
  - Konfidensintervall
  - Hypotesetesting / P-verdi

Eksempel - postoperativ kvalme

Behandling * Kvalmeklasse Krysstabell

			Kvalme: lite eller ingen	Kvalme: betydelig	Total
Behandling	Nei	Antall	18	12	30
		%	60,0%	40,0%	100,0%
	Ja	Antall	24	5	29
		%	82,8%	17,2%	100,0%
Total		Antall	42	17	59
		%	71,2%	28,8%	100,0%

Differanse i suksessannsynlighet i dette utvalget:
82,8% - 60,0% = 22,8%

Hva kan vi si om effekt av behandling i en populasjon av aktuelle pasienter?

Estimert sannsynlighet for suksess:

kontrollgruppen: $\hat{p}_1 = 18/30 = 60\%$

behandlingsgruppen: $\hat{p}_2 = 24/29 = 83\%$

Hvis vi behandler 100 pasienter vil vi forvente hhv

100x60%=60 og 100x83%=83 suksesser.

Forventet differanse: 83-60=23

NNT - Number needed to treat
(=NNTB – Number needed to benefit)

The number of patients that a physician would have to treat with a new treatment in order to avoid one event that would otherwise have occurred with a standard treatment. … (Simon Day: Dictionary for clinical trials, 2nd edition, Wiley, 2007)

NNT beregnes som 1 delt på

differansen mellom sannsynlighetene i de to gruppene:

$$\widehat{NNT} = \frac{1}{\hat{p}_2 - \hat{p}_1} = \frac{1}{0.83 - 0.60} = 4.4$$

Hvis vi behandler 4.4 pasienter vil vi forvente hhv

4.4x60% = 2.6 suksesser

4.4x83% = 3.6 suksesser

Forventet differanse 3.6 – 2.6 = 1

Hypotesetesting

– Sett opp nullhypotese og alternativ hypotese. Eksempel:
  - H_0: Sannsynligheten for suksess er lik i gruppene
  - H_1: Sannsynligheten for suksess er forskjellig

Hypotesetesting:

Nullhypotese:
H_0: $p_1 = p_2$

Alternativ hypotese
H_1: $p_1 \neq p_2$ (tosidig)

eller

H_1: $p_1 < p_2$ (ensidig)

Ensidige alternativ hypoteser brukes nesten aldri i medisinsk forskning.

Sannheten

Beslutning	H_0	H_1
Aksepter H_0	OK	P(Type II feil $\|H_1$) $=\beta$
Forkast H_0 (påstå H_1)	P(Type I feil $\|H_0$) $=\alpha$	P(OK $\|H_1$) $=1-\beta$ =testens styrke(funksjon)

P(Type I feil) = P(Forkaste $H_0|H_0$) $=\alpha$
kalles testens signifikansnivå

P(Type II feil) = P(Akseptere $H_0|H_1$) $=\beta$

P(Forkaste $H_0|H_1$) $=1-\beta$
kalles testens styrke (power)

Varierende notasjon:
Noen lærebøker bruker β for styrke og $(1-\beta)$ for P(Type II feil)

Hypotesetesting og p-verdi

- P-verdien (signifikanssannsynlighet, sig.) er sannsynligheten for å få de observerte verdier eller noe mer ekstremt, gitt at H_0 er sann.
- P-verdien er <u>ikke</u> sannsynligheten for at H_0 er sann!
- Forkast H_0 hvis p-verdi $\leq \alpha$
  – Dette garanterer P(Type I feil) $\leq \alpha$

Kryssende interesser:

- Ønsker lav α og lav β.
- MEN: Desto lavere α, desto lavere teststyrke (høy β)
- I praksis: Sett α til et "lavt" tall, vanligvis 0.05 eller 0.01.
- H_0 og H_1 er ikke likeverdige. Hvis vi er "i tvil", aksepteres H_0.
- I rettsvesenet:
  - H_0: Tiltalte er uskyldig
  - H_1: Tiltalte er skyldig

"We find the defendant not guilty but not all that innocent, either."

Konfidensintervall: Et mål på usikkerhet i estimatet

Et $(1-\alpha)$ konfidensintervall (θ_l, θ_h) for en parameter θ (for eksempel $\theta = p_1 - p_2$) har egenskapen

$$P(\theta_l < \theta < \theta_h) = 1-\alpha$$

$(1-\alpha)$ kalles konfidenskoeffisienten.
Vanligvis er $(1-\alpha) = 0.95$

Hva betyr et $(1-\alpha)$ konfidensintervall?

Hvis det beregnes 95% konfidensintervall for mange forsøk, vil i det lange løp 95% av intervallene dekke den sanne verdien

Det er IKKE 95% sannsynlighet for at konfidensintervallet dekker den sanne verdien

Sammenheng mellom konfidensintervall og hypotesetest:
Hvis $(1-\alpha)$ konfidensintervallet for θ inneholder θ_0,
vil vi ikke forkaste $H_0: \theta = \theta_0$ på signifikansnivå α

(Generelt: Konfidensintervallet består av de verdier θ_0 som ikke ville blitt forkastet ved hypotesetesting på nivå α)

Fra "Vancouver-retningslinjene" :

Statistics
"… When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as the use of P values, which fails to convey important information about effect size. …"

ICMJE – International Committe of Medical Journal Editors
http://www.icmje.org/#prepare, januar 2013

43

Kaasbøll J, Lydersen S, Indredavik M: (Pain, 2012)
"Psychological symptoms in children of parents with chronic pain – the HUNT study"

Results adjusted for age

Parents with chronic pain	Number of children	Risk for conduct problems: Odds ratio (OR) estimate	Conf. int.	P-value
None	801	1 (ref.)		
Only mother	289	1.30	1.02 to 1.67	0.036
Only father	216	0.99	0.74 to 1.31	0.93
Both parents	117	1.36	0.96 to 1.93	0.087

NTNU
Det skapende universitet

www.ntnu.no

44

Eksempel - postoperativ kvalme

Behandling * Kvalmeklasse Krysstabell

			Kvalme: lite eller ingen	Kvalme: betydelig	Total
Behandling	Nei	Antall	18	12	30
		%	60,0%	40,0%	100,0%
	Ja	Antall	24	5	29
		%	82,8%	17,2%	100,0%
Total		Antall	42	17	59
		%	71,2%	28,8%	100,0%

Pearson's kjikvadrattest, tosidig alternativ:
p-verdi = 0.054

NTNU
Det skapende universitet

www.ntnu.no

45

To grupper av størrelse n_1 og n_2.

Observerer

$X_1 \sim \text{bin}(n_1, p_1)$ og $X_2 \sim \text{bin}(n_2, p_2)$

H_0: $p_1=p_2$ (eller $p_1-p_2=0$) mot H_1: $p_1 \neq p_2$.

Estimatorer for p_1 og p_2:

$$\hat{p}_1 = \frac{X_1}{n_1} \quad \text{og} \quad \hat{p}_2 = \frac{X_2}{n_2}$$

Forkaster H_0 hvis $\hat{p}_1 - \hat{p}_2$ avviker "mye" fra 0.

NTNU
Det skapende universitet

www.ntnu.no

46

Under H_0 er $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{Var(\hat{p}_1 - \hat{p}_2)}}$ tilnærmet standard normalfordelt.

$$Var(\hat{p}_1 - \hat{p}_2) \overset{\text{pga uavh.}}{=} Var(\hat{p}_1) + (-1)^2 Var(\hat{p}_2)$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \overset{\text{Under } H_0}{=} \left(\frac{1}{n_1} + \frac{1}{n_2}\right) p(1-p)$$

Dermed fås

$$z \approx \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}(1-\hat{p})}} \quad \text{hvor} \quad \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

NTNU
Det skapende universitet

www.ntnu.no

47

Generelt:

$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{Var(\hat{p}_1 - \hat{p}_2)}}$ er tilnærmet standard normalfordelt.

$$Var(\hat{p}_1 - \hat{p}_2) \overset{\text{pga uavh.}}{=} Var(\hat{p}_1) + (-1)^2 Var(\hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Dermed fås

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{1}{n_1} p_1(1-p_1) + \frac{1}{n_2} p_2(1-p_2)}} \approx \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{1}{n_1} \hat{p}_1(1-\hat{p}_1) + \frac{1}{n_2} \hat{p}_2(1-\hat{p}_2)}}$$

NTNU
Det skapende universitet

www.ntnu.no

48

Så

$$\Pr(-z_{1-\alpha/2} \le z \le z_{1-\alpha/2}) \approx 1-\alpha$$

$$\Pr\left(-z_{1-\alpha/2} \le \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{1}{n_1} \hat{p}_1(1-\hat{p}_1) + \frac{1}{n_2} \hat{p}_2(1-\hat{p}_2)}} \le z_{1-\alpha/2}\right) \approx 1-\alpha$$

Løser den mhp $p_1 - p_2$ og får et tilnærmet
1-α konfidensintervall for $p_1 - p_2$

NTNU
Det skapende universitet

www.ntnu.no

Tilnærmet 1-α konfidensintervall for $p_1 - p_2$ (Wald intervallet)

$$\hat{p}_1 - \hat{p}_2 \mp z_{1-\alpha/2}\sqrt{\frac{1}{n_1}\hat{p}_1(1-\hat{p}_1) + \frac{1}{n_2}\hat{p}_2(1-\hat{p}_2)}$$

Tilnærmingen er OK bare hvis n_1 og n_2 er "store"

Agresti & Caffo (2000) konfidensintervall for p_1-p_2:

Beregn estimert risikodifferanse som før:

$$\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

Legg til 1 i hver celle i 2x2 tabellen før du beregner "vanlig" asymptotisk konfidensintervall:

$$\tilde{p}_1 = \frac{\tilde{X}_1}{\tilde{n}_1} = \frac{X_1 + 1}{n_1 + 2}, \quad \tilde{p}_2 = \frac{\tilde{X}_2}{\tilde{n}_2} = \frac{X_2 + 1}{n_2 + 2}$$

Bedre tilnærmet konfidensintervall:

$$\tilde{p}_1 - \tilde{p}_2 \mp z_{1-\alpha/2}\sqrt{\frac{1}{\tilde{n}_1}\tilde{p}_1(1-\tilde{p}_1) + \frac{1}{\tilde{n}_2}\tilde{p}_2(1-\tilde{p}_2)}$$

Agresti & Caffo (2000) intervallet:

- Lett å beregne
- Legg til 4 observasjoner (1 suksess og 1 fiasko i hver av gruppene) og beregn Wald intervallet som om dette var observasjonene
- Gode egenskaper (dekningsgrad)
- Anbefalt i nyere innføringsbøker i statistikk

Eksempel: Postoperativ kvalme

$$\hat{p}_1 - \hat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2} = \frac{24}{29} - \frac{18}{30} = 0.8276 - 0.6000 = 0.2276$$

$$\tilde{p}_1 = \frac{\tilde{x}_1}{\tilde{n}_1} = \frac{x_1 + 1}{n_1 + 2} = \frac{24+1}{29+2} = 0.8064$$

$$\tilde{p}_2 = \frac{\tilde{x}_2}{\tilde{n}_2} = \frac{x_2 + 1}{n_2 + 2} = \frac{18+1}{30+2} = 0.5938$$

95% Agresti-Caffo konfidensintervall:

$$\tilde{p}_1 - \tilde{p}_2 \mp z_{1-0.05/2}\sqrt{\frac{1}{\tilde{n}_1}\tilde{p}_1(1-\tilde{p}_1) + \frac{1}{\tilde{n}_2}\tilde{p}_2(1-\tilde{p}_2)}$$

$$= 0.2276 \mp 1.96\sqrt{\frac{1}{31}0.8064(1-0.8064) + \frac{1}{32}0.6(1-0.6)} = (-0.007, 0.432)$$

(95% Wald konfidensintervall: (0.005, 0.45))

Er konfidensintervallet konsistent med hypotesetesten (p=0.054)?

Eksempel: Antall dager i sykehus.

Behandling A:
26, 15, 37, 11, 13, 10, 17, 21, 131, 38
$\bar{x}_A = 31.90$, $s_A^2 = 36.24^2$, median = 19

Behandling B
141, 32, 115, 22, 26, 12, 203, 65, 40, 15, 32, 49, 243
$\bar{x}_B = 76.54$, $s_A^2 = 75.86^2$, median = 40

Hva kan vi si om forskjell mellom A og B i populasjonen?

Student's t-test og konfidensintervall for to uavhengige utvalg.

- n_1 observasjoner, antas uavh. $N(\mu_1, \sigma_1^2)$
- n_2 observasjoner, antas uavh. $N(\mu_2, \sigma_2^2)$
- H_0: $\mu_1 = \mu_2$ mot H_1: $\mu_1 \neq \mu_2$
- Ekvivalent: H_0: $\mu_1 - \mu_2 = 0$ mot H_1: $\mu_1 - \mu_2 \neq 0$
- Antar foreløpig lik varians, $\sigma_1^2 = \sigma_2^2 = \sigma^2$
- Ikke brukbar ved observasjoner som avviker mye mer fra gjennomsnittet enn forventet I normalfodelingen.

Estimator for $\mu_1 - \mu_2$: $\quad \overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

Altså: $\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

Hvis $\sigma_1 = \sigma_2 = \sigma$ så er $\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$

Men σ^2 er ukjent og estimeres ved "pooled estimate of the variance":

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n_1}(X_{i1} - \overline{X}_1)^2 + \sum_{i=1}^{n_2}(X_{i2} - \overline{X}_2)^2\right]$$
$$= \frac{n_1 - 1}{n_1 + n_2 - 2}S_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2}S_2^2$$

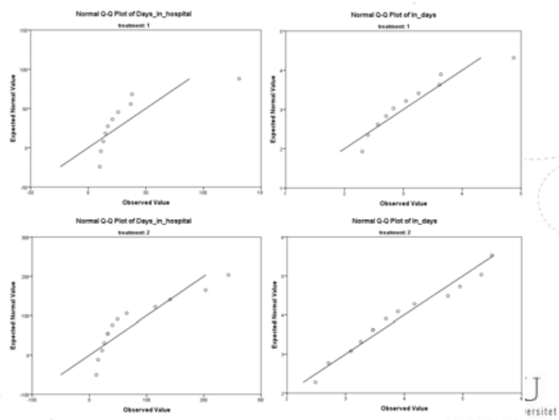
Vi bruker at

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

Antall dager forutsatt (tilnærmet) normalfordelt

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						95% Confidence Interval of the Difference	
		F	Sig.	Sig. (2-tailed)	Mean Difference	Lower	Upper
Days_in_hospital	Equal variances assumed	6,452	,019	,102	-44,638	-98,907	9,631
	Equal variances not assumed			,079	-44,638	-94,948	5,671



Antall dager er ikke normalfordelt.

Logaritmen til antall dager er tilnærmet normalfordelt

Logaritmen (ln) til antall dager

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
						95% Confidence Interval of the Difference	
		F	Sig.	Sig. (2-tailed)	Mean Difference	Lower	Upper
ln_days	Equal variances assumed	1,080	,311	,050	-,77890	-1,55894	,00114
	Equal variances not assumed			,044	-,77890	-1,53530	-,02250

Tolkning:

$$\ln(\widehat{MedianA}) - \ln(\widehat{MedianB}) = -0.7789$$

$$\ln\left(\frac{\widehat{MedianA}}{\widehat{MedianB}}\right) = -0.7789$$

$$\frac{\widehat{MedianA}}{\widehat{MedianB}} = e^{-0.7789} = 0.459$$

95% konfidensintervall: $(e^{-1.5353}, e^{-0,0225}) = (0.215, 0.978)$

Ikke-parametriske metoder:
Forutsetter ingen parametrisk fordeling:
Basert på rangordningen av observasjonene, "glemmer" originaldata

Behandling A sortert:
10, 11, 13, 15, 17, 21, 26, 37, 38, 131
Rang:
1, 2, 4, 5.5, 7, 8, 10.5, 14, 15, 20
Gjennomsnittsrang: 8.7

Behandling B sortert:
12, 15, 22, 26, 32, 32, 40, 49, 65, 115, 141, 203, 243
Rang:
3, 5.5, 9, 10.5, 12.5, 12.5, 16, 17, 18, 19, 21, 22, 23
Gjennomsnittsrang: 14.54

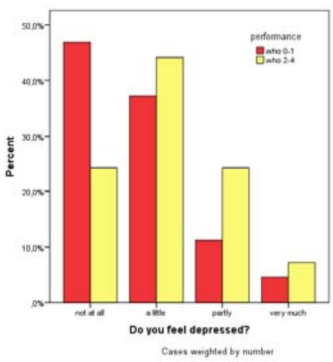
Wilcoxon-Mann-Whitney's test for to uavhengige utvalg: P=0.042

Example: EORTC Quality of life questionnaire

do you feel depressed? * performance Crosstabulation

Count

		performance		Total
		who 0-1	who 2-4	
do you feel depressed?	1: not at all	205	27	232
	2: a little	163	49	212
	3: partly	49	27	76
	4: very much	20	8	28
Total		437	111	548



Eksempel - EORTC data

Er du deprimert?	Performance status	
	who 0-1	who 2-4
Gjennomsnitt	1.73	2.14
Standardavvik	0.83	0.87
median	2	2

Observert differanse: 2.14 – 1.73 = 0.41
Er det forskjell på forventet depresjons-skåre mellom de to gruppene?

Student's T-test: 95% KI (0.23, 0.59), p-verdi < 0.001

Wilcoxon-Mann-Whitney (Ikke-parametrisk test): p-verdi < 0.001

Metodene vi har sett på, forutsetter uavhengige observasjoner i gruppene. Hva med matchede par?

Matchede par. Eksempel fra Box, Hunter & Hunter: Statistics for Experimenters 2nd ed. (2005)

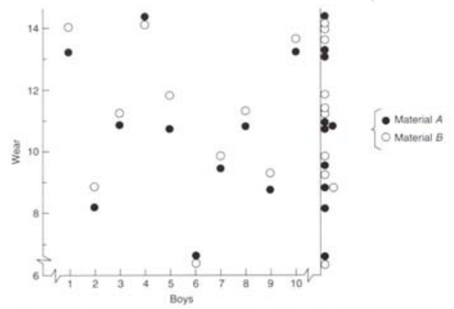


Figure 3.7. Data on two different materials A and B, used for making soles of boys' shoes.

Matchede par - eksempler fra medisinsk forskning:

Overkrysningsstudier
Kontralateral design

Matchede par:

- Skalavariable:
  - Regn ut differansen for hvert individ. Bruk en ett-utvalgsmetode for å teste om forventet differanse er 0:
    - Student's ettutvalgs t-test
    - Ekvivalent: "paired samples t-test" – programvaren regner ut differansene.

    eller
    - Wilcoxon's signed rank ikkeparamertrisk test
    - Ekvivalent: "Related samples …"
- Dikotome variable (to mulige utfall):
  - McNemar's test

Student't t-test eller ikke-parametrisk metode?

- Hvis data er normalfordelt:
  - Ikke-parametriske metoder har tilnærmet (dvs nesten) like høy teststyrke som t-testen i middels store og store datasett
  - Ikke-paramteriske metoder er vesentlig svakere enn t-testen i små datasett.
- Hvis data ikke er normalfordelt:
  - T-testen OK hvis ikke ekstreme observasjoner
  - Bruke t-testen på transformerte data?
  - Bruk en ikke-parametrisk test
- Student's t-test: Kan også beregne tilhørende konfidensintervall.

Oppsummering - valg av metode:

Type variabel	To uavhengige utvalg	Matchede par
Normalfordelt (evt etter transformasjon), eller data uten ekstreme observasjoner	Student's toutvalgs t-test: Anta lik varians – eller ikke anta lik varians	Student's ettutvalgstest på differansene Ekvivalent: Paret t-test
Vilkårlig fordelt skalavariabel eller ordinal variabel	(Wilcoxon-) Mann-Whitney's test	Wilcoxon's signed rank test på differansene
Dikotom (to mulige utfall)	Pearson's kjikvadrattest hvis forventet antall i alle celler >5. Agresti-Caffo konfidensintervall	McNemar's test