NTNU
Det skapende universitet

Medisinsk statistikk, termin IC

av
Stian Lydersen, professor i medisinsk statistikk
Regionsenter for barn og unges psykiske helse (RBUP) Midt-Norge

Forelesning 15 desember 2011

www.ntnu.no

Innhold:

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

Hvorfor statistikk?

- For å kunne lese medisinsk litteratur inkl vitenskapelige artikler
- For å kunne utføre enkle statistiske analyser ifm hovedoppgaven

Litteratur

- Bowers, D: "Medical Statistics from Scratch". 2ed, Wiley 2008.
- Aalen, Odd. m.fl.: Statistiske metoder i medisin og helsefag. Gyldendal, 2006.
- Cambell, M, Machin, D, Walters, S: Medical statistics: A textbook for the health sciences. 4th ed, Wiley, 2007
- Gonick, L and Wollcott, S: "The Cartoon Guide to Statistics" Harper Collins, 1993

Editorial, NEJM, 1 January 2000:
Looking Back on the Millennium in Medicine.

The most important medical developments of the past millennium:

- Elucidation of Human Anatomy and Physiology
- Elucidation of the Chemistry of Life
- Application of Statistics to Medicine
- Development of Anesthesia
- Discovery of the Relation of Microbes to Disease
- Elucidation of Inheritance and Genetics
- Knowledge of the Immune System
- Development of Body Imaging
- Discovery of Antimicrobial Agents
- Development of Molecular Pharmacotherapy

"Any serious investigator in biological and medical sciences must have a grasp of the basic principles (of statistics). With modern computer facilities there is little need for familiarity with the techical detail of statistical calculations. However, a physician should understand when such calculations are valid, when they are not, and how they should be interpreted."

(Campbell and Machin, 2007)

"Statistikk": Forskjellige betydninger:

1. En samling tall –f.eks Statistisk årbok fra Statistisk sentralbyrå
2. Engelsk: "Statistic", norsk "observator" eller "testobservator": En funksjon av data som f.eks gjennomsnitt, maksimumsverdi eller Student's t observator.
3. (Matematisk) statistikk: En gren av matematikken med egen terminologi og metoder. Det vitenskapelige redskap for å trekke konklusjoner basert på data med elementer av usikkerhet.

Tre typer statistikk:

- Deskriptiv
  - Grafer
  - Oppsummerende tall
- Bekreftende
  - Hypotesetesting
  - Konfidensintervall
- Prediktiv

Deskriptiv statistikk:

Grafer og oppsummeringstall

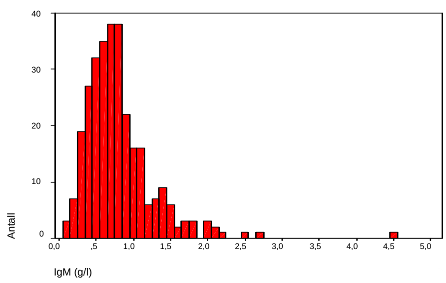
Typer data:

- Skalavariabel - f.eks høyde i cm
- Kategorisk variabel
  - Ordinal, f.eks "Føler du deg deprimert?" 1 = "Ikke i det hele tatt" , 2 = "Litt", 3 = "Endel", 4 = "Svært mye"
  - Nominal, f.eks Sivilstand: 1 = "ugift", 2 = "gift", 3 = "samboer" , 4 = "skilt", 5 = "enke(mann)"

Konsentrasjon av serum IgM (g/l) hos
298 friske barn, 6 mnd - 6 år gamle (Altman, 1991)

0,8	1,1	0,7	0,5	0,5	0,5	0,9	0,7	0,4	0,7	0,5	4,5	1,0	1,4	0,8
0,8	0,7	0,6	1,6	0,8	0,2	1,5	0,2	0,7	1,0	1,2	0,7	0,5	0,6	0,6
0,9	0,5	0,4	1,0	1,7	1,1	1,1	0,5	0,4	0,3	0,6	0,8	0,4	1,7	1,1
0,7	0,2	1,2	0,5	0,3	1,5	0,6	1,0	0,8	0,1	0,8	0,3	1,0	1,0	0,8
0,6	0,3	0,5	0,6	0,6	0,4	0,5	1,5	0,9	0,8	0,4	0,5	0,4	0,4	0,4
1,0	0,4	1,5	1,0	0,9	2,7	0,9	1,5	0,7	0,5	0,7	0,5	0,9	1,5	1,0
1,4	0,7	0,4	0,6	1,4	0,4	1,1	1,4	0,4	0,3	0,7	1,0	1,6	0,6	0,6
0,7	0,7	1,1	0,4	0,9	0,8	0,6	1,1	0,6	1,0	1,4	0,3	0,7	0,9	0,7
1,8	0,6	0,9	0,9	0,9	0,5	0,5	0,7	1,1	0,5	2,2	1,2	1,3	0,5	0,8
0,8	0,8	0,1	0,6	1,2	0,3	0,7	1,0	0,8	0,6	0,4	1,1	0,5	1,3	0,3
0,8	0,4	0,3	1,1	0,3	0,8	0,8	1,0	0,4	0,6	0,5	0,4	0,7	0,7	0,9
1,2	1,8	2,5	0,8	0,8	0,2	0,9	0,6	0,7	1,4	1,4	0,6	2,0	1,3	0,8
0,3	1,3	0,5	0,7	1,2	0,5	1,7	0,5	0,7	1,0	0,2	1,1	1,0	0,8	1,0
0,5	0,5	1,3	0,5	0,7	0,4	0,9	0,4	0,6	0,8	0,7	0,7	0,8	1,1	0,7
0,8	0,8	0,9	0,4	0,6	0,7	0,1	0,7	0,8	0,7	0,4	1,1	0,8	0,5	0,6
0,7	0,8	0,3	0,8	0,6	0,8	1,4	0,8	0,7	0,6	0,5	0,9	0,8	0,9	0,4
0,8	0,5	0,2	0,8	2,0	0,5	0,9	0,4	0,3	0,4	0,9	0,5	2,1	0,5	1,4
1,1	0,4	0,7	0,3	0,5	0,7	0,7	0,6	0,6	0,8	0,6	0,7	0,6	0,6	0,6
0,8	0,3	1,1	0,3	0,4	2,0	0,8	1,3	0,8	0,6	0,7	2,1	1,8	0,3	
0,3	0,2	0,9	1,3	0,6	0,7	0,9	0,6	0,6	1,1	0,3	0,7	0,6	0,9	

Histogram - IgM data:



13

Example: EORTC Quality of life questionnaire

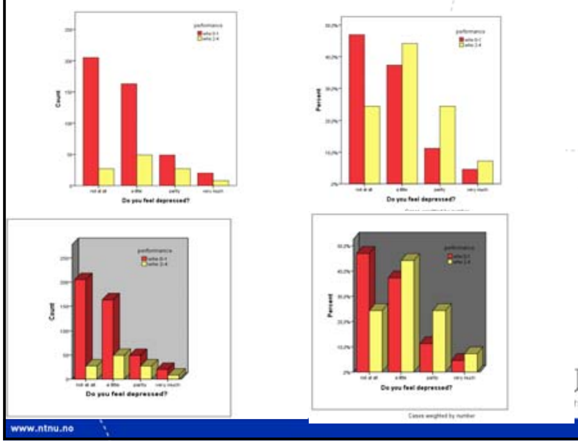
do you feel depressed? * performance Crosstabulation

Count

		performance		
		who 0-1	who 2-4	Total
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

NTNU Det skapende universitet

14



15

Noen nyttige grafer

- Én kategorisk variabel:
  - Bar chart (stolpediagram)
  - Pie chart (kakediagram)
- To kategoriske variable:
  - Clustered bar chart (klynget stolpediagram)

NTNU Det skapende universitet

16

Noen nyttige grafer (forts.)

- Én skalavariabel:
  - Histogram
  - Sammenlikne data med normalfordeling: Q-Q plot lettere å lese og tolke enn “normal curve overlay” i histogram
- To skalavariable:
  - Scatterplot

NTNU Det skapende universitet

17

Noen nyttige grafer (forts.)

- Én skalavariabel og én kategorisk variabel (sammenlikne skalavariabelen i to eller flere grupper):
  - Dot plot eller scatter plot (ved “få” observasjoner)
  - Box plot (ved “mange” observasjoner)

NTNU Det skapende universitet

18

Beskrivelse av fordelingen

- Skalavariable, evt også ordinale variable: sentrum og spredning:
  - Gjennomsnitt og standardavvik
  - Median og kvartiler
- Kategoriske data:
  - Frekvenstabell
  - Krysstabell

NTNU Det skapende universitet

19

Data: $x_1, x_2, \ldots, x_n$

Lettere å regne ut

Gjennomsnitt: $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$

(Empirisk) varians: $s^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \boxed{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\right]}$

(Empirisk) standardavvik: $s = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

20

Data sortert: $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$

Median:
$x_{((n+1)/2)}$ hvis n er oddetall
$(x_{(n/2)} + x_{(n/2+1)})/2$ hvis n er partall
Medianen deler tallmaterialet "på midten".
Like mange observasjoner under som over medianen.

Nedre kvartil, median, øvre kvartil:
Deler tallmaterialet i fire like store deler.

21

Eksempel: Antall dager i sykehus.

Behandling A:
26, 15, 37, 11, 13, 10, 17, 21, 131, 38
Sortert:
10, 11, 13, 15, 17, 21, 26, 37, 38, 131

Behandling B
141, 32, 115, 22, 26, 12, 203, 65, 40, 15, 32, 49, 243
Sortert:
12, 15, 22, 26, 32, 32, 40, 49, 65, 115, 141, 203, 243

Hva blir median og kvartiler for behandling A?

22

Statistics

Days_in_hospital

1	N	Valid		10
		Missing		0
	Mean			31,90
	Std. Deviation			36,238
	Percentiles	25		12,50
		50		19,00
		75		37,25
2	N	Valid		13
		Missing		0
	Mean			76,54
	Std. Deviation			75,835
	Percentiles	25		24,00
		50		40,00
		75		128,00

Hvilke(t) mål vil du bruke på sentrum og spredning i fordelingene?

23

Gjennomsnitt og standardavvik har gunstige matematiske egenskaper.

Eks:
Hvis gjennomsnitt og standardavvik for hvert av r utvalg er gitt, kan man beregne dem for det totale tallmaterialet:

Gjennomsnitt totalt: $\bar{x}_{total} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_r\bar{x}_r}{n_1 + n_2 + \ldots + n_r}$

Varians totalt: $s_{total}^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \ldots + (n_r-1)s_r^2}{n_1 + n_2 + \ldots + n_r - r}$

Standardavvik totalt: $s_{total} = \sqrt{s_{total}^2}$

24

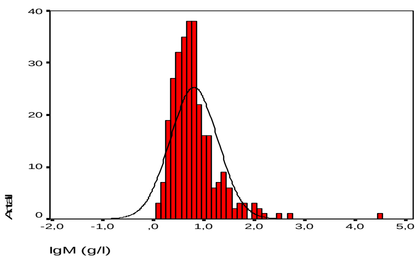
Normalfordelingen

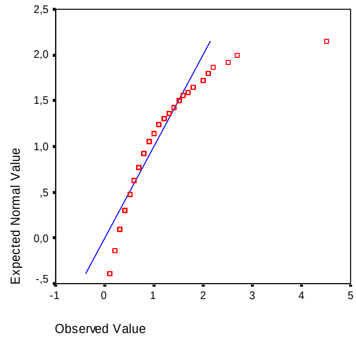
- I en del situasjoner er skalavariable (tilnærmet) normalfordelt, dvs symmetrisk og med en spesiell "klokkeformet" fasong.

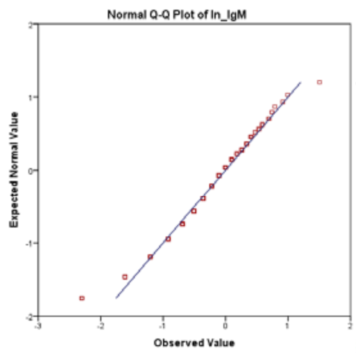
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Når data er normalfordelt:
  - Ca 68% ligger innen 1 standardavvik fra gjennomsnittet
  - Ca 95% ligger innen 2 standardavvik fra gjennomsnittet
- Visse metoder forutsetter at data er (tilnærmet) normalfordelt. F.eks Students t-test, vanlig regresjonsanalyse

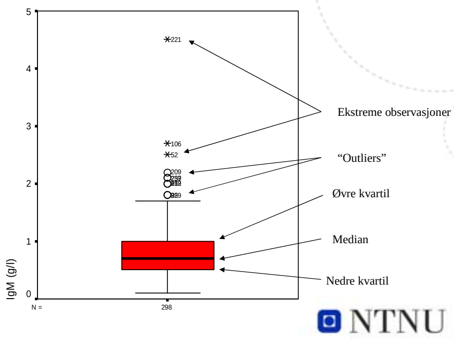
Histogram m/normalfordelingskurve IgM data



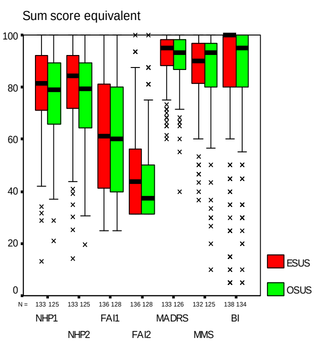




Box plot



Box plot - eksempel



Eksempel - IgM data:

- Gjennomsnitt: 0,803
- Standardavvik: 0,47

- Median (50% under): 0,7
- Nedre kvartil (25% under): 0,5
- Øvre kvartil (75% under): 1,0

31

Eksempel - EORTC data

	Performance status	
	who 0-1	who 2-4
Gjennomsnitt	1.73	2.14
Standardavvik	0.83	0.87
median	2	2

NTNU Det skapende universitet

www.ntnu.no

32

Valg av deskriptiv statisikk for sentrum og spredning i fordelingen

- Gjennomsnitt og standardavvik ELLER median og kvartiler ELLER begge deler?
  - Avhenger av målsettingen med analysen.
- Hvis data er symmetrisk fordelt (for eksempel normalfordelt):
  - Median = gjennomsnitt
- Hvis data ikke er normalfordelt:
  - Ganske vanlig å oppgi median og kvartiler. OK å oppgi gjennomsnitt og standardaavvik. Med standardavviket har ikke samme enkle tolkning som i normalfordelingen.
- OK å anta normalforelding i små datasett?
  - JA, hvis rimelig antakelse basert på annen kunnskap eller inspeksjon av data.

NTNU Det skapende universitet

www.ntnu.no

33

Hvordan sjekke om data avviker fra normalfordelingen?

- Hvis median avviker mye fra gjennomsnitt, så er data ikke symmetrisk fordelt (og dermed ikke normalfordelt). (Men ikke omvendt!)
- Statistisk test:
  - Kolmogorov-Smirnoff mye brukt men lite egnet.
  - Shapiro-Wilk noe bedre egnet.
- Histogram med normalfordelingskurve: Vanskelig å vurdere
- Q-Q plott: Velegnet!

NTNU Det skapende universitet

www.ntnu.no

34

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?

NTNU Det skapende universitet

www.ntnu.no

35

Enkel sannsynlighetsregning og diagnostiske tester

NTNU Det skapende universitet

www.ntnu.no

36

Sannsynlighet for gutt:

Antall levendefødsler	Antall gutter	Andel gutter
10	8	0,8
100	55	0,55
1000	525	0,525
10000	5139	0,5139
100000	51127	0,51127
3760358	1927054	0,51247
17989361	9219202	0,51248
34832051	17857857	0,51268

NTNU Det skapende universitet

www.ntnu.no

37

Sannsynlighet (Def 3.1)

- Et forsøk gjennomføres n ganger. Begivenheten A inntreffer n_A av gangene. Den relative hyppigheten n_A/n tenderer mot et tall når antall forsøk tenderer mot uendelig. Dette tallet, P(A), kalles *sannsynligheten* for A. Engelsk: *Probability*

www.ntnu.no

38

Sannsynlighetsmodell: Forsøk, utfallsrom, sannsynligheten til hvert enkeltutfall

a) Terningkast: P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6
b) Barnefødsel. P(jente) = 0.487, P(gutt) = 0.513
c) Behandling med penicillin. Realistisk for enkelte pasientgrupper: P(frisk) = 0.6377, P(forblir syk) = 0.3622, P(anafylaktisk sjokk) = 0.0001

www.ntnu.no

39

Aalen et al (2006), side 49: … Det er for eksempel mennesker som har kastet en terning svært mange ganger, og da har funnet ut at hvert av utfallene opptrer i omtrent 1/6 av tilfellene.

www.ntnu.no

40

3.4 Regneregler for sannsynlighet

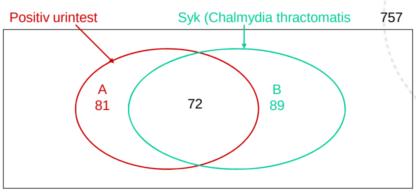
Regel 3.2 Komplementregelen
$P(A) + P(\bar{A}) = 1$

Regel 3.3 Addisjonsregelen:
For disjunkte A og B har vi
$P(A \cup B) = P(A) + P(B)$

Regel 3.4 Den generelle addisjonsregel:
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

www.ntnu.no

41



A = “Pasienten har positiv urintest” $P(A) = 81/757 = 10.7\%$

B= ”Pasienten er syk” $P(B) = 81/757 = 11.8\%$

$P(A \mid B) = 72/89 = 80.9\%$

$$P(A \mid B) = \frac{72/757}{89/757} = \frac{P(A \cap B)}{P(B)}$$

www.ntnu.no

42

Definisjon av betinget sannsynlighet for A gitt B:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Regel 3.5
Den generelle multiplikasjonsregelen

$$P(A \cap B) = P(A \mid B)P(B) \;\; = P(B \mid A)P(A)$$

www.ntnu.no

43

A og B er stokastisk uavhengige hvis
$P(A \mid B) = P(A)$

Regel 3.6 mm:
A og B er stokastisk uavhengige hvis og bare hvis
$P(A \cap B) = P(A)P(B)$

Regel 3.7:
Hvis $A_1, A_2, \ldots, A_n$ er stokastisk uavhengige så er
$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$

NTNU Det skapende universitet
www.ntnu.no

44

Eksempel:
Sannsynlighet for to gutter i to enkeltfødsler:

$P(G_1 \cap G_2) = P(G_1)P(G_2) = 0.513 \cdot 0.513 = 0.263$

NTNU Det skapende universitet
www.ntnu.no

45

Aalen et al, eksempel 3.4 s 56:
"… Vi antar da uavhengighet mellom hver fødsel med hensyn til barnets kjønn. Egentlig kan en ikke bare gå ut fra at det er avhengighet i dette tilfellet. Det bør undersøkes om en slik antakelse stemmer med virkeligheten. Det finnes flere undersøkelser om dette, og det viser seg at det ikke er full stokastisk uavhengighet med hensyn til barns kjønn i en familie. Enkelte familier har en tendens til å få jenter og andre en tendens til å få gutter. …"

NTNU Det skapende universitet
www.ntnu.no

46

Lippert, T, Skjærven, R, Salvesen, K. Å: Hvorfor får noen bare gutter eller bare jenter? Tidsskr Nor Lægeforen 2005; 125: 3414-7

Studie basert på kvinner som har født to, tre og fire barn i perioden 1967 – 2003 (Norsk fødselsregister): 540 699 kvinner og 1 382 974 fødsler.

Andel gutter 51.33%.

NTNU Det skapende universitet
www.ntnu.no

47

Lippert et al (2005):
"Det er ikke holdepunkter for at sannsynligheten for å få gutt eller jente avviker fra populasjonsgjennomsnittet hos noen spesielle foreldrepar. Den viktigste forklaringen på at det er flere rene gutte- og jentesøskenflokker enn statistisk fordeling forutsier, er at en del mødre med bare gutter eller bare jenter føder flere barn, i hva vi tror er et forsøk på å få et barn av motsatt kjønn."

NTNU Det skapende universitet
www.ntnu.no

48

Altså:
Barnets kjønn (ved enkeltfødsler) er uavhengig av kjønnsfordeling på eldre søsken, Norge 1967 – 2003.

NTNU Det skapende universitet
www.ntnu.no

49

Eksempel 3.8 Samme kjønn hos tvillinger

A = ”Tvillingparet har samme kjønn”

B = ”Tvillingparet er enegget”

$P(A \mid \bar{B}) = 1/2$

$P(A \mid B) = 1$

$P(B) = 1/3$

P(A)=?

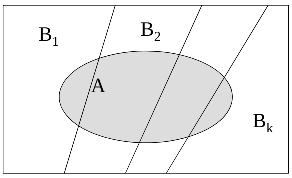
50

Regel 3.8 – Loven om total sannsynlighet

$$P(A) = P(A \mid B)P(B) + P(A \mid \bar{B})P(\bar{B})$$

51

Loven om total sannsynlighet – generelt:



$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

52

Regel 3.9 – Bayes’ lov

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \bar{B})P(\bar{B})}$$

53

Bayes’ lov og diagnostiske tester

Tradisjonell 2*2 tabell

		Sykdom +	Sykdom –	
Test	+	a [SP]	b [FP]	a + b
resultat	–	c [FN]	d [SN]	c + d
		a + c	b + d	a + b + c + d

A = {test positiv}, B = {syk}, SP = sann positiv, FP = falsk positiv, FN = falsk negativ, SN = sann negativ

54

$$Prevalens = P(B) = \frac{a+c}{a+b+c+d}$$

$$Sensitivitet = P(A|B) = \frac{a}{a+c}$$

$$Spesifisitet = P(\bar{A}|\bar{B}) = \frac{d}{b+d}$$

$$PPV = P(B|A) = \frac{a}{a+b}$$

$$NPV = P(\bar{B}|\bar{A}) = \frac{d}{c+d}$$

$$\left[Accuracy = \frac{a+d}{a+b+c+d} \right]$$

55

Performance of Chlamydia Rapid Test versus Polymerase chain reaction (PCR), at a genitourinary medicine clinic in London. (Nadala et al., BMJ, 2009).

Disease status (PCR result)	Test result Positive	Negative	Total
Diseased	72	17	89
Non-diseased	9	659	668
Total	81	676	757

Hva er testens sensitivitet og spesifisitet?
Hva er prevalens og PPV i dette paientutvalget?

NTNU
Det skapende universitet

56

Sensitivitet = 72/89 = 0.81
Spesifisitet = 659/668 = 0.987

Prevalens = 89/757 = 0.118
PPV = 72/81 = 0.89
NPV= 659/668 = 0.975

Hva blir PPV i en populasjon med prevalens 3%?

NTNU
Det skapende universitet

57

$$PPV = P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \bar{B})P(\bar{B})}$$
$$= \frac{sensitivitet \times prevalens}{sensitivitet \times prevalens + (1 - spesifisitet) \times (1 - prevalens)}$$

$$NPV = P(\bar{B} \mid \bar{A}) = \frac{P(\bar{A} \mid \bar{B})P(\bar{B})}{P(\bar{A} \mid \bar{B})P(\bar{B}) + P(\bar{A} \mid B)P(B)}$$
$$= \frac{spesifisitet \times (1 - prevalens)}{spesifisitet \times (1 - prevalens) + (1 - sensitivitet) \times prevalens}$$

NTNU
Det skapende universitet

58

Prevalens 3%:

$$PPV = \frac{0.81 \cdot 0.03}{0.81 \cdot 0.03 + (1 - 0.987)(1 - 0.03)} = 0.65$$

NTNU
Det skapende universitet

59

For example, consider a diagnostic test with sensitivity 0.95 and specificity 0.92, which in most settings would be regarded as a highly accurate test. In a population with prevalence 0.001, the above formulas give $PPV = 0.012$ and $NPV = 0.99995$. On the other hand, in a population with prevalence 0.1, the predictive values become $PPV = 0.57$ and $NPV = 0.994$. The positive predictive values are typically very low in a low prevalence population. As another example, consider a diagnostic test with sensitivity 0.95 and a very high specificity of 0.999. In a population with disease prevalence 0.001, we obtain $PPV = 0.49$ and $NPV = 0.99995$. Compared to the $PPV = 0.012$ obtained with a specificity of 0.92, the $PPV = 0.49$ obtained with a specificity of 0.999 may be acceptable in some situations. This illustrates that a diagnostic test to be used for screening purposes ought to have high specificity.

NTNU
Det skapende universitet