

Anvendt medisinsk statistikk

Vår 2010

Diagnostiske tester

Eirik Skogvoll

1.amanuensis dr.med.

Enhet for Anvendt klinisk forskning (AKF)

Oversikt

- Malin Dögl: "Partus-testen", blir det fødsel innen 3 døgn?
- Grunnleggende om diagnostiske testers egenskaper
 - Vanlig 2*2 tabell, forutsetter "gullstandard" og binært utfall (ja/nei)
 - Sensitivitet og spesifisitet
- Beskrive testene sannsynlighetsteoretisk og statistisk
 - Betinget sannsynlighet
 - Binomisk fordeling
 - Presisjonsmål (konfidensintervall, asymptotiske og eksakte)
 - McNemars test for sammenlikning av to metoder mot samme "gullstandard"
- Anvendelse av tester i populasjoner med forskjellig sykdomsprevalens
 - Bayes regel: snur problemstillingen "opp-ned"
- ROC: testens egenskaper ved forskjellige "cut-off" verdier

Hierarki (Thornbury 1991)

- Level 1: Technical efficacy
(bildekvalitet, optimal blodprøve-timing etc.)
- Level 2: Diagnostic accuracy efficacy
(sensitivitet, spesifisitet, ROC)
- Level 3: Diagnostic thinking efficiency
(klinikerens endring av vurdering før/etter testresultat)
- Level 4: Therapeutic efficacy
(i hvilken grad test-resultatet faktisk påvirker behandlingen)
- Level 5: Patient outcome efficacy
(reduksjon i antall døde, bedret livskvalitet)
- Level 6: Societal efficacy
(kost-nytte i et samfunnmessig perspektiv)

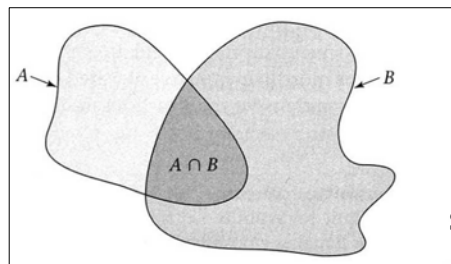
Betinget sannsynlighet

Definisjon (Rosner Def. 3.9, s. 55)

Sannsynlighet for hendelse B gitt hendelse A

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

Vi omdefinerer
utfallsrommet fra S til A



Rosner fig. 3.5,
s. 52

Tester: Grunnleggende egenskaper

Diagnostisk test – egenskaper

		Test resultat		
		Syk	Frisk	
Fasit	Syk	$1-\beta$	β	Falsk negativ
	Frisk	α	$1-\alpha$	
		Falsk positiv	Spesifisitet	

H_0 : pasienten er frisk vs. H_1 : pasienten er syk

α = False positive rate = Type I feil: forkaste H_0 når H_0 er sann

β = False negative rate = Type II feil: akseptere H_0 når H_0 er feil

Mål på testens "godhet"

- Sensitivitet (sensi- syk -itet) = true positive rate = $1 - \beta$
- Spesifisitet (spesi- frisk- itet) = true negative rate = $1 - \alpha$
- [Accuracy = Youden's index $1 - (\alpha + \beta)$]
- Receiver Operating Characteristic (ROC)
- [Likelihood ratio of a positive result (LR+)]
- [Vektet κ (kappa)]

Diagnostisk test – estimering fra et utvalg

		Test resultat		
		Syk	Frisk	
Fasit	Syk	s_1	s_0	n_1
	Frisk	r_1	r_0	n_0
		m_1	m_0	N

Diagram illustrating the relationship between the actual status (Fasit) and the test result (Test resultat). The table shows counts for True Positive (Sanne positive), False Positive (Falsk positive), True Negative (Sanne negative), and False Negative (Falske negative).

Binomisk fordeling

Binomisk fordeling

Binomisk forsøksrekke:

- n uavhengige forsøk
- to mulige utfall i hvert enkelt forsøk:
{suksess} eller {svikt = "failure"}
- $\Pr(\{\text{suksess}\}) = p$ (el. π), konstant fra forsøk til forsøk
- $\Pr(\{\text{svikt}\}) = q = (1-p)$ (el. $1-\pi$), også konstant
- X (tilfeldig variabel) = {antall suksesser ila. n forsøk}

Binomisk fordeling

Eksempel (Rosner expl. 4.23, s. 95)

Hva er sannsynligheten for at akkurat celle nr. 2 og celle nr. 5 blant totalt 5 celler i et blodutstryk er nøytrofil (©)?

$P(\{\text{nøytrofil}\}) = p = \pi = 0,6$

Fra multiplikasjonsloven ...

$$\begin{aligned} & \Pr(\{\bullet \text{ } \textcircled{C} \text{ } \bullet \text{ } \bullet \text{ } \textcircled{C}\}) \\ &= \Pr(\{\bullet\}) * \Pr(\{\textcircled{C}\}) * \Pr(\{\bullet\}) * \Pr(\{\bullet\}) * \Pr(\{\textcircled{C}\}) \\ &= q * p * q * q * p = (1-0,6) * 0,6 * (1-0,6) * (1-0,6) * 0,6 \\ &= (1-0,6)^3 * 0,6^2 = 0,023 \end{aligned}$$

Binomisk fordeling

Eksempel (Rosner expl. 4.24, s. 95)

Men hva med sannsynligheten for at 2 vilkårlige celler blant totalt 5 er nøytrofile?

Rosner tbl. 4.5,
s. 95,

$x = \text{nøytrofil}$
(tidl. kalt "©")
 $o = \text{annen celle}$

Possible orderings for 2 neutrophils out of 5 cells

xxooo	oxxoo	ooxox
xoxoo	oxoxo	oooxx
xooxo	oxoox	
xooox	ooxxo	

← Forrige ekspl.

Binomisk fordeling

Definisjon (Rosner Eq. 4.5, s. 93)

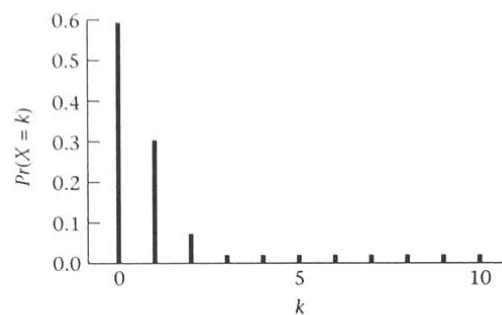
$$\Pr(X = k) = \binom{n}{k} p^k q^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \dots = \binom{n}{k} \pi^k (1-\pi)^{n-k}$$

Eksempel (Rosner expl. 4.25, s. 93)

Hva er sannsynligheten for å få 2 gutter blant 5 barn i en familie, gitt $\Pr(\{\text{gutt}\}) = 0,5$?

$$\Pr(X = 2) = \binom{5}{2} \cdot 0,5^2 0,5^3 = \frac{5!}{2! \cdot 3!} \cdot 0,5^5 = 10 \cdot 0,5^5 = 0,3125$$

Binomisk fordeling - eks. (Rosner Fig. 4.4, s. 102)



(a) $n = 10, p = .05$

Tabell 1 (Rosner, s. 817) gir eksakte verdier i binomisk fordeling med $n = 2, 3, \dots, 20$ og p (el. π) fra 0,05 til 0,5

Diagnostisk test – resultat fra undersøkelse

		Test resultat		
		Syk	Frisk	
Fasit	Syk	s_1	s_0	n_1
	Frisk	r_1	r_0	n_0
		m_1	m_0	N

Sanne positive (arrow to s_1)
 Falske negative (arrow to s_0)
 Falsk positive (arrow to r_1)
 Sanne negative (arrow to r_0)

Vi “ser lett” at:

Antallet som er test positive blant syke utgjør en binomisk forsøksrekke:

- n_1 uavhengige “forsøk” (personer)
- to mulige utfall i hvert enkelt forsøk: {positiv} eller {negativ}
- $\Pr(\{\text{positiv}|\text{syk}\}) = 1 - \alpha$, konstant sensitivitet fra person til person
- s_1 (tilfeldig variabel) = {antall test positive i.l.a. n forsøk}

Do. for test negative blant friske.

Estimering i binomisk fordeling

Punktestimat (Rosner kap. 6.8)

- Vi har observert X "suksesser" i n forsøk - hva er vårt beste estimat for p (el. π)?

$$\hat{p} = \hat{\pi} = \frac{X}{n} \quad (\text{Rosner eq. 6.16, s. 192})$$

Estimering i binomisk fordeling

Konfidensintervall (Rosner kap. 6.8)

- Konfidensintervall for p (el. π) er gitt som (p_1, p_2) og angir "plausible verdier for p , gitt våre observasjoner". Et mål på presisjonen av estimatet.
- Symmetrisk CI, basert på N-tilnærming (Rosner eq. 6.19):

$$(p_1, p_2) = \hat{p} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (\text{bare hvis } n\hat{p}\hat{q} \geq 5)$$

Eks. Observert 219 test pos. blant totalt 300 syke.

$$\hat{p} = \frac{219}{300} = 0.73 \quad \hat{q} = 0.27, \text{ dvs. } n\hat{p}\hat{q} = 59 \quad (\text{ok})$$

$$95\% \text{ CI} = (p_1, p_2) = \hat{p} \mp z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}, \text{ setter } \frac{\alpha}{2} = 0.025 \leftrightarrow z_{1-\frac{\alpha}{2}} = 1.96$$

$$\begin{aligned} \text{Dvs. } 95\% \text{ CI} &= 0.73 \mp 1.96 \cdot \sqrt{\frac{0.73 \cdot 0.27}{300}} = \\ &= 0.73 \mp 1.96 \cdot 0.0256 = 0.73 \mp 0.0502 = (0.68, 0.78) \end{aligned}$$

Estimering i binomisk fordeling

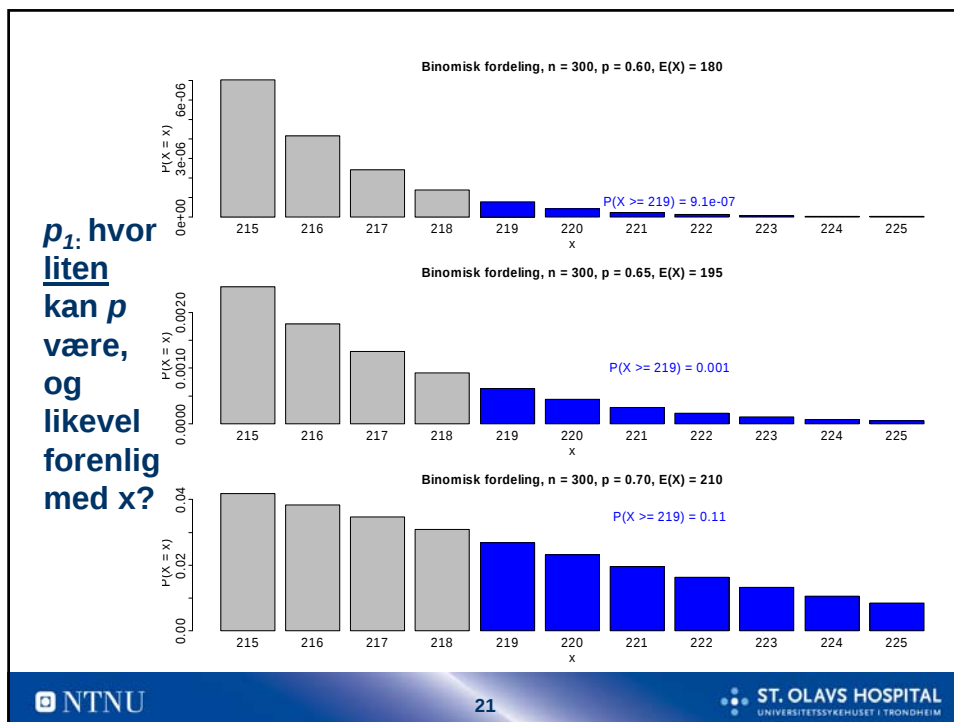
Eksakt konfidensintervall (Rosner kap. 6.8, eq. 6.20)

100 % $(1 - \alpha)$ konfidensintervall for p (el. π)

er gitt ved (p_1, p_2) der

$$\Pr(X \geq x \mid p = p_1) = \frac{\alpha}{2} = \sum_{k=x}^n \binom{n}{k} p_1^k (1 - p_1)^{n-k}$$

(p_1 angir nedre grensen for hvor *liten* p kan være, og likevel forenlig med den observerte x)



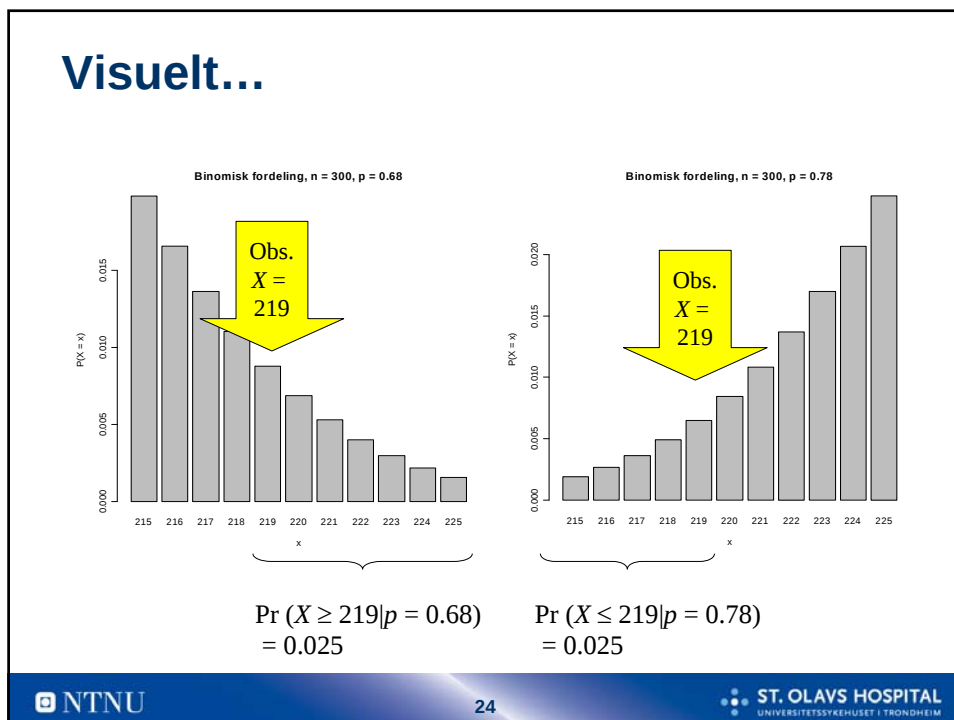
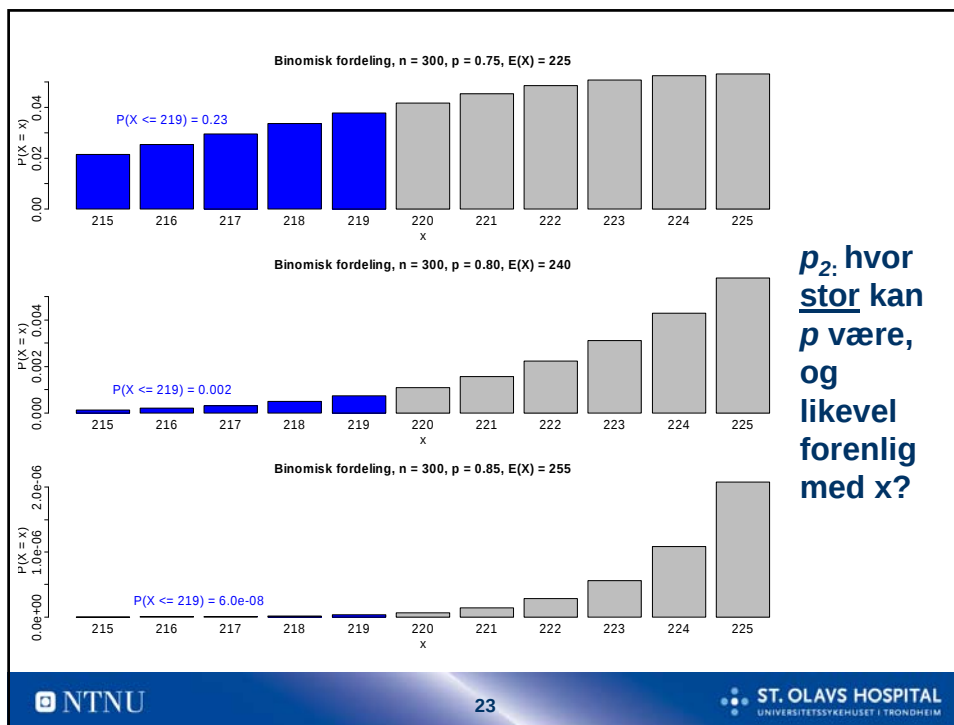
Estimering i binomisk fordeling

Eksakt konfidensintervall (Rosner kap. 6.8, eq. 6.20)

100 % $(1 - \alpha)$ konfidensintervall for p (el. π)
er gitt ved (p_1, p_2) der

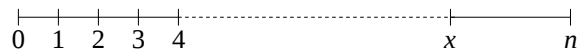
$$\Pr(X \leq x \mid p = p_2) = \frac{\alpha}{2} = \sum_{k=0}^x \binom{n}{k} p_2^k (1 - p_2)^{n-k}$$

(p_2 angir øvre grensen for hvor stor p kan være,
og likevel forenlig med den observerte x)



Eks. $n = 39$, obs. $x = 37$, $npq = 1.89$ Vi ønsker et tosidig eksakt 95 % CI for p , dvs $\alpha = 0.05$, gitt ved (p_1, p_2)

$$\Pr(X \geq x | p = p_1) = \frac{\alpha}{2} = \sum_{k=x}^n \binom{n}{k} p_1^k (1-p_1)^{n-k}$$



p_1 : hvor liten kan p tenkes å være, når vi har observert (en så stor) $x = 37$?



**La oss
prøve!**

$$P(X \geq 37 | p_1) = 0.025$$

$x =$	30	31	32	33	34	35	36	37	38	39
$p =$										
0.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.6	0.012	0.005	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.7	0.094	0.064	0.037	0.018	0.008	0.003	0.001	0.000	0.000	0.000
0.8	0.134	0.156	0.156	0.132	0.093	0.053	0.024	0.008	0.002	0.000
0.9	0.009	0.023	0.053	0.101	0.160	0.206	0.206	0.150	0.071	0.016

Mer nøyaktig...

$$P(X \geq 37 | p_1) = 0.025$$

x =	30	31	32	33	34	35	36	37	38	39
p =										
0.7	0.094	0.064	0.037	0.018	0.008	0.003	0.001	0.000	0.000	0.000
0.75	0.144	0.126	0.094	0.060	0.032	0.014	0.005	0.001	0.000	0.000
0.8	0.134	0.156	0.156	0.132	0.093	0.053	0.024	0.008	0.002	0.000
0.85	0.062	0.102	0.145	0.174	0.174	0.141	0.089	0.041	0.012	0.002
0.9	0.009	0.023	0.053	0.101	0.160	0.206	0.206	0.150	0.071	0.016
0.95	0.000	0.000	0.002	0.009	0.031	0.085	0.180	0.278	0.278	0.135

Fra StatXact: ESTIMATION OF BINOMIAL PARAMETER (PI)

Number of Trials = 39
Number of Successes = 37

Maximum Likelihood Estimate of PI = 0.9487

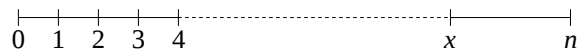
95.00% Confidence Interval for PI:

(Clopper-Pearson) = (0.8268 , 0.9937)

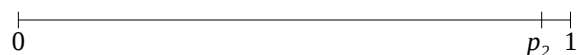
(Blyth-Still-Casella) = (0.8298 , 0.9908)

Eks. $n = 39$, obs. $x = 37$, $npq = 1.89$ Vi ønsker et tosidig eksakt 95 % CI for p , dvs $\alpha = 0.05$, gitt ved (p_1, p_2)

$$\Pr(X \leq x | p = p_2) = \frac{\alpha}{2} = \sum_{k=0}^x p_2^k (1 - p_2)^{n-k}$$



p_2 : hvor stor kan p tenkes å være, når vi har observert (en så liten) $x = 37$?



Krav:
 $P(X \leq 37 | p_2)$
 $= 0.025$

x=	30	31	32	33	34	35	36	37	38	39
P =										
0.98	0	0	0	0	0.001	0.006	0.035	0.140	0.362	0.455
0.985	0	0	0	0	0.000	0.002	0.018	0.095	0.329	0.555
0.99	0	0	0	0	0.000	0.001	0.006	0.051	0.266	0.676
0.995	0	0	0	0	0.000	0.000	0.001	0.015	0.161	0.822

Fra StatXact: ESTIMATION OF BINOMIAL PARAMETER (PI)

Number of Trials =39
 Number of Successes =37

Maximum Likelihood Estimate of PI = 0.9487

95.00% Confidence Interval for PI:

(Clopper-Pearson) = (0.8268 , 0.9937)

(Blyth-Still-Casella) = (0.8298 , 0.9908)

Sensitivitet & spesifisitet: utvalgsstørrelse

- Konfidensintervallet for p er bredest, lengst "på midten" nær $p = 0.5$
- Maksimal lengde av et symmetrisk $(1 - \alpha) \cdot 100\%$ konfidensintervall, basert på N-tilnærming:

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (\text{bare hvis } n\hat{p}\hat{q} \geq 5)$$

Konfidensintervall-lengde, N-tilnærming

$$l = 2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \text{og siden } \max(\sqrt{\hat{p}\hat{q}}) = 0.5$$

$$\text{vil max}(l) = \frac{2 \cdot z_{1-\frac{\alpha}{2}} \cdot 0.5}{\sqrt{n}}$$

$$\text{dvs. } l = \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \quad \text{slik at } n = \left(\frac{z_{1-\frac{\alpha}{2}}}{l} \right)^2$$

Eks. Ønsker max lengde $\approx 0,1$ (= 10 %) dvs. presisjonen er $p \pm 0.05$ («5 % i hver retning»).

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{l} \right)^2 = \left(\frac{1.96}{0.1} \right)^2 = 19.6^2 = 384$$

Konfidensintervall-lengde, eksakt

- N-tilnærming er veldig konservativ. Betydelig smalere CI "på kantene" (nær 0 eller 1).

Alternativer:

- Prøve seg fram med mulige verdier i f.eks. StatExact
- Spesielt tabellverk.
- Grafisk avlesning av Rosner fig. 7, s. 833

Sammenlikning av to metoder mot samme "gullstandard" (her: biopsi/ histologi)

- Estimert sensitivitet (MGG) = $37/39 = 0.948$ og sensitivitet (Pap) = $38/39 = 0.974$
Er de signifikant forskjellige?
- Problem: målt mot samme standard og således ikke uavhengige.... χ^2 test og Fisher's eksakt test er begge uegnet.
- Benytter i stedet McNemar's test (Rosner 10.4) og do. spesialprosedyre for beregning av konfidensintervall (f.eks. StatExact)

Pap og MGG blant 39 tilfeller av BCC

		MGG		
		+	-	
Pap	+	37	1	38
	-	0	1	1
		37	2	39

Konkordant (arrow from top-left box to cell 37)
 Diskordant (arrow from top-right box to cell 1)
 Diskordant (arrow from bottom-left box to cell 0)
 Konkordant (arrow from bottom-right box to cell 1)

MGG vs. Pap: test og konfidensintervall for differans

Fra StatExact:

CONFIDENCE INTERVAL FOR A DIFFERENCE OF TWO RELATED BINOMIAL PROPORTIONS
BASED ON THE STANDARDIZED STATISTIC AND INVERTING TWO 1-SIDED TESTS

Statistics based on the observed 2 by 2 table :

Observed proportion for population	MGG : piHat_1	=	0.9487
Observed proportion for population	Pap : piHat_2	=	0.9744
Observed difference of proportions	Pap-MGG	=	0.0256

Results:

Method	P-value		95.00% Conf. Interval for pi_2-pi_1	
	1-sided(Pr{T .LE. t})	2*1-sided		
Asymp	0.1587	0.3173	(	-0.0663, 0.1318)
Exact	0.2670	0.5340	(	-0.0678, 0.1348)

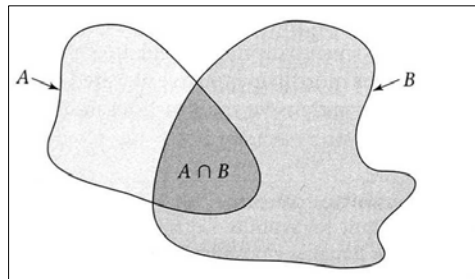
Bayes regel Bruk av tester i populasjoner med forskjellig prevalens

Total sannsynlighet

Rosner Eq. 3.6, s. 55

For to vilkårlige hendelser A og B er

$$\Pr(B) = \Pr(B|A) \cdot \Pr(A) + \Pr(B|\bar{A}) \cdot \Pr(\bar{A})$$



$\Pr(B)$ er nå gitt som en sum av to betingede sannsynligheter!

Total sannsynlighet

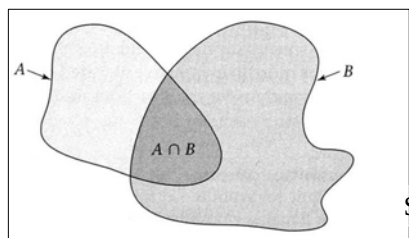
Bevis

$B = (B \cap A) \cup (B \cap \bar{A})$ A og $\bar{A}$ er gjensidig utelukkende

$$\Pr(B) = \Pr(B \cap A) + \Pr(B \cap \bar{A})$$

$$\text{Vi vet at } \Pr(B|A) = \frac{\Pr(B \cap A)}{\Pr(A)} \leftrightarrow (B \cap A) = \Pr(B|A) \cdot \Pr(A)$$

$$\text{Dvs. } \Pr(B) = \Pr(B|A) \cdot \Pr(A) + \Pr(B|\bar{A}) \cdot \Pr(\bar{A})$$



Litt sannsynlighetsregning igjen ...

$A = \{\text{positiv test}\}$

$B = \{\text{sykdom}\}$

$P(B) = \text{sykdomsprevalens}$

$P(A|B) = \text{sensitivitet}$

$P(A|\bar{B}) = \alpha = \text{"false positive rate"}$

$P(\bar{A}|B) = \beta = \text{"false negative rate"}$

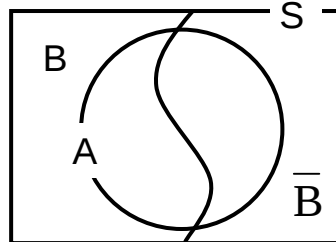
$P(\bar{A}|\bar{B}) = \text{spesifisitet}$

$P(\bar{A}|\bar{B}) + P(A|\bar{B}) = 1$ (hvorfor?)

$\Leftrightarrow P(A|\bar{B}) = 1 - P(\bar{A}|\bar{B}) = 1 - \text{spesifisitet}$

$P(B|A) = \text{PPV} = \text{PV}^+ = \text{"positiv prediktiv verdi"}$

$P(\bar{B}|\bar{A}) = \text{NPV} = \text{PV}^- = \text{"negativ prediktiv verdi"}$

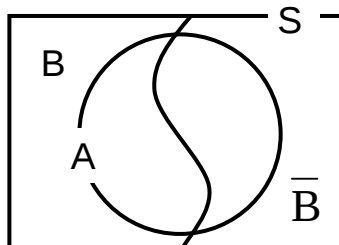


Bayes regel

Definisjon (Rosner Eq. 3.9) Bayes regel/ teorem

Kombinerer uttrykkene for betinget og total sannsynlighet:

$$\text{PPV} = \text{PV}^+ = P(B|A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{\Pr(A|B) \cdot \Pr(B)}{\Pr(A|B) \cdot \Pr(B) + \Pr(A|\bar{B}) \cdot \Pr(\bar{B})}$$



Et revolusjonerende
resultat ...

Vi har funnet en betinget
sannsynlighet ved hjelp av de
"motsatte" sannsynlighetene!

OBS! OBS!

Er testens egenskaper virkelig uavhengig av prevalens?

Ikke nødvendigvis.... Mulighet for bl.a. "spectrum bias"!

Testen kan være utviklet i en klinisk situasjon der sykdommen *presenterer seg annerledes* enn i andre situasjoner...

Situasjon og presentasjon kan m.a.o. henge sammen ...!

Bayes regel

Eksempel (Rosner expl. 3.26, s. 61)

Prevalens av hypertensjon = $\Pr(B) = 0,2$. Et apparat klassifiserer 84 % av hypertensive og 23 % av normotensive som hypertensive. Hva er PPV? NPV?

$\Pr(A|B) = 0,84$ (sensitivitet)

og $\Pr(A|\bar{B}) = 0,23$ ("false positive rate")

dvs. spesifisitet = $\Pr(\bar{A}|\bar{B}) = 1 - 0,23 = 0,77$

Fra Bayes' regel får vi

$$\begin{aligned} PV^+ &= \Pr(B | A) = \frac{\Pr(A | B) \cdot \Pr(B)}{\Pr(A | B) \cdot \Pr(B) + \Pr(A | \bar{B}) \cdot \Pr(\bar{B})} \\ &= \frac{\text{sens} \cdot \text{prevalens}}{\text{sens} \cdot \text{prevalens} + (1 - \text{spes}) \cdot (1 - \text{prevalens})} \\ &= \frac{0,84 \cdot 0,2}{0,84 \cdot 0,2 + 0,23 \cdot 0,8} = \frac{0,168}{0,352} = 0,48 \end{aligned}$$

og tilsvarende

$$\begin{aligned} PV^- &= \Pr(\bar{B} | \bar{A}) = \frac{\text{spes} \cdot (1 - \text{prevalens})}{\text{spes} \cdot (1 - \text{prevalens}) + (1 - \text{sens}) \cdot \text{prevalens}} \\ &= \frac{0,77 \cdot 0,8}{0,77 \cdot 0,8 + 0,16 \cdot 0,2} = \frac{0,616}{0,648} = 0,95 \end{aligned}$$

ROC (Receiver Operating Characteristic)

Testens egenskaper ved forskjellige "cut-off" verdier

ROC

Ratings of 109 CT images by a single radiologist

True disease status	CT rating by radiologist					Total
	Definitely normal (1)	Probably normal (2)	Questionable (3)	Probably abnormal (4)	Definitely abnormal (5)	
Normal	33	6	6	11	2	58
Abnormal	3	2	2	11	33	51
Total	36	8	8	22	35	109

Rosner
tbl. 3.3
og 3.4,
s. 65

Kriterium "1+": alle med rating 1 til 5 får diagnosen syk.
Finner samtlige syke, men identifiserer ingen friske.
Sensitivitet = 1, spesifisitet = 0, 'false positive rate' = 1.

ROC

Ratings of 109 CT images by a single radiologist

True disease status	CT rating by radiologist					Total
	Definitely normal (1)	Probably normal (2)	Questionable (3)	Probably abnormal (4)	Definitely abnormal (5)	
Normal	33	6	6	11	2	58
Abnormal	3	2	2	11	33	51
Total	36	8	8	22	35	109

Kriterium "2+": alle med rating 2 til 5 får diagnosen syk.
Finner 48/51 syke, og identifiserer 33/58 friske.
Sensitivitet = 0,94 Spesifisitet = 0,57 'False positive rate' = 0,43

ROC

Ratings of 109 CT images by a single radiologist

True disease status	CT rating by radiologist					Total
	Definitely normal (1)	Probably normal (2)	Questionable (3)	Probably abnormal (4)	Definitely abnormal (5)	
Normal	33	6	6	11	2	58
Abnormal	3	2	2	11	33	51
Total	36	8	8	22	35	109

Kriterium "3+": alle med rating 3 til 5 får diagnosen syk.
 Finner 46/51 syke, og identifiserer 39/58 friske.
 Sensitivitet = 0,90 Spesifisitet = 0,67 'False positive rate' = 0,33

ROC

Ratings of 109 CT images by a single radiologist

True disease status	CT rating by radiologist					Total
	Definitely normal (1)	Probably normal (2)	Questionable (3)	Probably abnormal (4)	Definitely abnormal (5)	
Normal	33	6	6	11	2	58
Abnormal	3	2	2	11	33	51
Total	36	8	8	22	35	109

Kriterium "5+": alle med rating 5 får diagnosen syk.
 Finner 33/51 syke, og identifiserer 56/58 friske.
 Sensitivitet = 0,65 Spesifisitet = 0,97 'False positive rate' = 0,03

ROC

Ratings of 109 CT images by a single radiologist

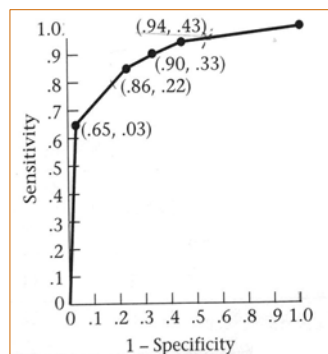
True disease status	CT rating by radiologist					Total
	Definitely normal (1)	Probably normal (2)	Questionable (3)	Probably abnormal (4)	Definitely abnormal (5)	
Normal	33	6	6	11	2	58
Abnormal	3	2	2	11	33	51
Total	36	8	8	22	35	109

Kriterium "6+": bare de med rating > 5 får diagnosen syk.
 Finner ingen syke, identifiserer alle som friske.
 Sensitivitet = 0 Spesifisitet = 1 'False positive rate' = 0

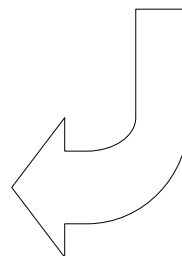
ROC

Resultatet
sammenfattes i
en tabell ...
(Rosner table
3.4, s. 65)

Test positive criteria	Sensitivity	Specificity	'False pos. rate'
1+	1.0	0	1
2+	.94	.57	0,43
3+	.90	.67	0,33
4+	.86	.78	0,22
5+	.65	.97	0,03
6+	0	1.0	0

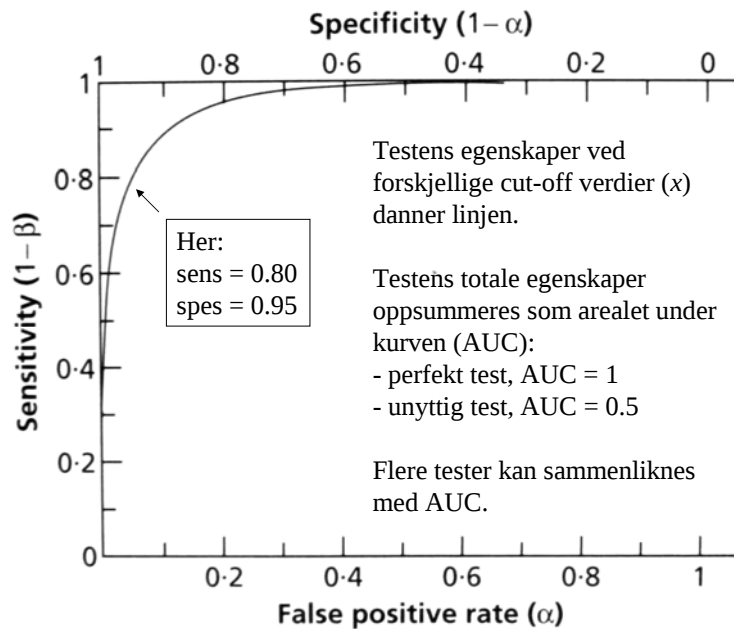
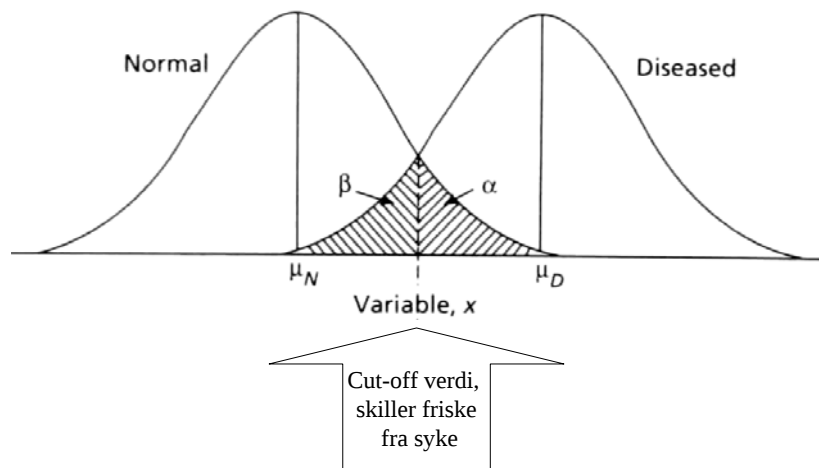


... og vises
som en ROC
kurve.
(Rosner fig.
3.7, s. 66)!



Kontinuerlig skala (x)

NB! Den valgte cut-off tillegger begge typer feil (α , β) like stor vekt, noe som ikke alltid er optimalt!



Litteratur

- BMJ-serien "Statistics notes", om diagnostiske tester: (1994; 308: 1552, 1994; 309: 102, 1994; 309: 188, 2004; 329: 168 – 169)
- Gluud & Gluud: Evidence based diagnostics. *BMJ* 2005;330;724-726
- Newcombe, RG. Improved confidence intervals for the difference between binomial proportions based on paired data. *Statistics in Medicine* 1998;17: 2635 - 2650
- Zhou, Obuchowski & McClish: *Statistical Methods in Diagnostic Medicine* (2002). Wiley
- B Rosner: Ch. 3: Probability, Ch.4: Discrete Probability Distributions. In *Fundamentals of Biostatistics* 6.ed (2006) Duxbury
- P Armitage, G Berry, JNS Matthews. Ch 19.9: Diagnostic tests. In *Statistical Methods in Medical Research* 4.ed (2002), Blackwell Science.
- Wulff & Gøtzsche: Ch.4. Diagnosis. In *Rational Diagnosis and Treatment* 3.ed (2000)