

NTNU

KL MED8005
Medisinsk statistikk Del II, våren 2008
 - og -
St2303
Medisinsk statistikk, våren 2008

16 januar 2008:

- Praktisk om kursene
- Analyse av 2x2 tabeller (Avsnitt 10.1 – 10.3)

Stian Lydersen

1

NTNU

Praktisk om kursene

- Hjemmesider under
http://folk.ntnu.no/slyderse/medstat/medstatII_v08.html
http://folk.ntnu.no/slyderse/medstat/st2303/info_v08.html
- Forelesninger onsdager 1230-1415 / 1515
- Sted: se hjemmesiden
- Kursleder: Stian Lydersen
- Kurssekretær: Gerd Bromseth, tlf (735)50290, gerd.bromseth@ntnu.no
- Obligatoriske øvinger

2

NTNU

KL MED8005 Medisinsk statistikk del II
Innhold:

- Analyse av kategoriske data
 - 2x2 tabeller
 - Generelle tabeller (txc)
- Enkel lineær regresjon
- Multipl lineær regresjon
- Korrelasjon
- Variansanalyse (ANOVA)
- Kovariansanalyse (ANCOVA)

- Epidemiologiske metoder
- Logistisk regresjon
- Overlevelsesanalyse
 - Kaplan-Meier plot
 - Log rank test
 - Cox-regresjon

St2303
Medisinsk statistikk

3

NTNU

KL MED8006 Anvendt medisinsk statistikk

- Generaliserte lineære modeller (GLM)
- Repeterte målinger I
- Repeterte målinger II
- Analyse med manglende data
- Diagnostiske tester
- Metaanalyse. Bayesiansk statistikk.

Inngår i St2303 medisinsk statistikk

4

NTNU

Øvingene

- 3 øvinger, hver med 8 deloppgaver.
- Obligatoriske øvinger:
 - KL MED8005: Alle øvingene
 - St2303:
 - Øving 1 deloppgave 1-5.
 - Øving 3 deloppgave 3-8.
 - Evt egne oppgaver fra eget stoff.

5

NTNU

Øvingene (forts.)

- Grupper med inntil 8 personer. Hver gruppe får en veileder. Hver gruppe har en kontaktperson i gruppa.
- Gruppen arbeider gjennom hele øvingen før veiledning.
- 1 veiledning i 2 timer ca hver 3 uke (1 pr øving). Obligatorisk oppmøte! Veileder
 - Svarer på spørsmål, avklarer problemer.
 - Noterer oppmøte
- Hver gruppe leverer 1 øving til veileder, signert av alle.
- Veileder retter, godkjenner, returnerer til gruppekontakt.
- Hver deltaker sørger for å skaffe seg en kopi av godkjent øvingssett.

6

Evaluering KL MED8006

- Øvingene danner grunnlag for godkjent kurs
- Fremlegg av oppgaver (15 min)
 - Forevise komplett godkjent øvingssett
 - Noen få enkeltoppgaver trekkes ut og presenteres. Kandidaten blir stilt spørsmål knyttet til oppgaven.
 - Bestått / ikke bestått.

7

Evaluering St2303

- Øvingene obligatorisk. (Ikke midtsemesterprøve)
- Skriftlig eksamen.

8

Analyse av 2x2 tabeller (Avsnitt 10.1 – 10.3) Tre metoder beskrevet i Rosner:

- To-utvalgstest for binomiske andeler. Normalfordelings-tilnærming. Avsnitt 10.2.1.
- Pearsons kjikvadratetest. Avsnitt 10.2.2-10.2.4.
- Fisher's eksakte test. Avsnitt 10.3.

Ikke nevnt i Rosner: To metoder som generelt er bedre, og brukes i økende grad

- Eksakt betinget mid p
- Eksakt ubetinget

9

Tabell 10.1

Count		alder v første fødsel		Total
		>= 30 år	< 30 år	
status	kasus (brystkreft)	683	2537	3220
	kontroll (ikke brystkreft)	1498	8747	10245
Total		2181	11284	13465

10

Tabell 10.2

Count		hjerterinfarkt innen 3 år		Total
		ja	nei	
bruker	ja	13	4987	5000
p-pille	nei	7	9993	10000
Total		20	14980	15000

11

Tabell 10.9

Count		diett		Total
		mye salt	lite salt	
dødsårsak	ikke CVD	2	23	25
	CVD	5	30	35
Total		7	53	60

12

To-utvalgstest for binomiske andeler

Avsnitt 10.2.1

13

Eksempel 10.4

Beskytter lav alder ved første fødsel mot brystkreft?

D = kvinnen har (hatt) brystkreft

E = alder ved første fødsel var 30 år eller mer

$$P(D|E) \stackrel{?}{=} P(D|\bar{E})$$

Data fra en kassu – kontroll studie:

Av 3220 kvinner med brystkreft
hadde 683 (21.2%) alder første fødsel ≥ 30

Av 10245 kvinner uten brystkreft
hadde 1498 (14.6%) alder første fødsel ≥ 30

14

Altså: Vi ønsker å sammenlikne

$$P(D|E) \text{ og } P(D|\bar{E})$$

Men i en kassu – kontroll studie observeres

$$P(E|D) \text{ og } P(E|\bar{D}).$$

Imidlertid gjelder at

$$P(D|E) = P(D|\bar{E}) \Leftrightarrow P(E|D) = P(E|\bar{D})$$

15

Bevis-skisse (for spesielt interesserte):

For det første (lett å vise)

$$P(D|E) = P(D|\bar{E}) \Leftrightarrow \frac{P(D|E)/P(\bar{D}|E)}{P(D|\bar{E})/P(\bar{D}|\bar{E})} = 1$$

Siste kalles Odds Ratio (OR)

Dessuten gjelder alltid (uansett verdi på OR), vist av Cornfield (1956):

$$OR = \frac{P(D|E)/P(\bar{D}|E)}{P(D|\bar{E})/P(\bar{D}|\bar{E})} = \frac{P(E|D)/P(\bar{E}|D)}{P(E|\bar{D})/P(\bar{E}|\bar{D})}$$

Sykdoms odds ratio = Eksponerings odds ratio

16

Bevis-skisse (fortsettelse)

$$P(D|E) = P(D|\bar{E}) \Leftrightarrow \frac{P(D|E)/P(\bar{D}|E)}{P(D|\bar{E})/P(\bar{D}|\bar{E})} = 1$$

$$\Leftrightarrow \frac{P(E|D)/P(\bar{E}|D)}{P(E|\bar{D})/P(\bar{E}|\bar{D})} = 1 \Leftrightarrow P(E|D) = P(E|\bar{D})$$

q.e.d.

17

To grupper av størrelse n_1 og n_2 .

Observerer

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

$$H_0: p_1 = p_2 \text{ (eller } p_1 - p_2 = 0) \text{ mot } H_1: p_1 \neq p_2$$

Estimatorer for p_1 og p_2 :

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

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

18

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

$$\begin{aligned} \text{Var}(\hat{p}_1 - \hat{p}_2) &= \text{Var}(\hat{p}_1) + (-1)^2 \text{Var}(\hat{p}_2) \\ &= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \stackrel{\text{pga uavh.}}{\underset{\text{Under } H_0}{=}} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) p(1-p) \end{aligned}$$

Dermed fås

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

19

Likning 10.3 s 357 har en (omdiskutert) kontinuitetkorreksjon i telleren:

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2} \right)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \hat{p}(1-\hat{p})}}$$

20

Eksempel 10.5, brystkreft og alder ved første barn

$$z = \frac{0.212 - 0.146}{\sqrt{\left(\frac{1}{3220} + \frac{1}{10245} \right) 0.162(1-0.162)}} = 8.87$$

Forkast H_0 på nivå $\alpha=0.05$ hvis $|z| > z_{1-\alpha/2} = 1.960$

p-verdi = $2(1-\Phi(8.8)) < 0.001$

21

Pearsons kjikvadrattest.

Avsnitt 10.2.2-10.2.4.

22

Tabell 10.1

Count		alder v første fødsel		Total
		>= 30 år	< 30 år	
status	kasus (brystkreft)	683	2537	3220
	kontroll (ikke brystkreft)	1498	8747	10245
Total		2181	11284	13465

Er rader og kolonner uavhengige?
(Merk symmetri mellom rader og kolonner!)

Altså: Er D og E uavhengige:

D = "Kvinnen har (hatt) brystkreft"

E = "Alder ved første fødsel ≥ 30 år"

D og E er uavhengige hvis og bare hvis $\Pr(D \cap E) = \Pr(D)\Pr(E)$

23

Her er de observerte andelene

$$\hat{\Pr}(D) = \frac{3220}{13465} \quad \text{og} \quad \hat{\Pr}(E) = \frac{2181}{13465}$$

Hvis uavhengighet så er

$$\hat{\Pr}(D \cap E) = \hat{\Pr}(D) \times \hat{\Pr}(E) = \frac{3220}{13465} \times \frac{2181}{13465}$$

og forventet antall med $D \cap E$ lik

$$E_{11} = N \times \hat{\Pr}(D \cap E) = 13465 \times \frac{3220}{13465} \times \frac{2181}{13465} = \frac{3220 \times 2181}{13465} = 521.6$$

og det observerte antall er $O_{11}=683$

24

SPSS, Expected count:

status * alder v første fødsel Crosstabulation

		alder v første fødsel		Total
		>= 30 år	< 30 år	
status	kasus	Count 683	Count 2537	Count 3220
		Expected Count 521,6	Expected Count 2698,4	Expected Count 3220,0
	kontroll	Count 1498	Count 8747	Count 10245
		Expected Count 1659,4	Expected Count 8585,6	Expected Count 10245,0
Total		Count 2181	Count 11284	Count 13465
		Expected Count 2181,0	Expected Count 11284,0	Expected Count 13465,0

Person's kjikvadratobservator er definert som

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \frac{(O_{21} - E_{21})^2}{E_{21}} + \frac{(O_{22} - E_{22})^2}{E_{22}}$$

Under H_0 (rader og kolonner uavhengige) så er denne tilnærmet kjikvadratfordelt med 1 frihetsgrad dersom alle $E_{ij} > 5$.

Likning 10.5 er gitt med Yates' kontinuitetskorreksjon:

$$\chi^2_{cc} = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}$$

Eks 10.13 s 366

$$\begin{aligned} \chi^2 &= \frac{(|683 - 521.6| - 0.5)^2}{521.6} + \frac{(|2537 - 2698.4| - 0.5)^2}{2698.4} \\ &\quad + \frac{(|1498 - 1659.4| - 0.5)^2}{1659.4} + \frac{(|8747 - 8585.6| - 0.5)^2}{8585.6} \\ &= \frac{160.9^2}{521.6} + \frac{160.9^2}{2698.4} + \frac{160.9^2}{1659.4} + \frac{160.9^2}{8585.6} \\ &= 49.661 + 9.599 + 15.608 + 3.017 = 77.89 \approx \chi^2_1 \quad \text{under } H_0 \end{aligned}$$

Tabell 10.7 Generell kontingenstabell

a	b	a+b
c	d	c+d
a+c	b+d	n=a+b+c+d

Pearson's kjikvadrat:

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Med Yates kontinuitetskorreksjon:

$$\chi^2_{cc} = \frac{n \left(|ad - bc| - \frac{n}{2} \right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

z-test og Pearson's kjikvadrattest for 2x2 tabeller

- Det kan vises at z^2 fra test for sammenlikning av to binomiske proporsjoner er identisk lik χ^2 i Person's kjikvadrattest.
- Så de to testene er ekvivalente (gir alltid nøyaktig samme p-verdi).
- De kontinuitetskorrigererte versjonene er også ekvivalente.

31

Fisher's eksakte test

Avsnitt 10.3.

32

Example, two binomials:
Treatment of children with cardiac arrest.
(Perondi et al, NEJM, 2004)

Epinephrine treatment	survival at 24 hours		Total
	yes	no	
High dose	1	33	34
Standard dose	7	27	34
total	8	60	68

$$p_i = P(\text{Survival at 24 hours} \mid \text{Treatment no i})$$

$$H_0: p_1 = p_2 \quad \text{versus} \quad H_1: p_1 \neq p_2$$

33

Tabell 10.9

Count		diett		Total
		mye salt	lite salt	
dødsårsak	ikke CVD	2	23	25
	CVD	5	30	35
Total		7	53	60

$$p_i = P(\text{Diett m/ mye salt} \mid \text{Dødsårsak nr i})$$

$$H_0: p_1 = p_2 \quad \text{versus} \quad H_1: p_1 \neq p_2$$

34

Tabell 10.9 - dødsårsak * diett Crosstabulation

			diett		Total
			mye salt	lite salt	
døds årsak	ikke CVD	Count	2	23	25
	CVD	Expected Count	2,9	22,1	25,0
		Count	5	30	35
		Expected Count	4,1	30,9	35,0
Total		Count	7	53	60
		Expected Count	7,0	53,0	60,0

Minst en celle har forventet antall (expected count) < 5.
⇒ Pearson's kjikvadrattest bør ikke brukes.

35

Fisher's tea-drinker (Fisher, 1935)

Poured first	Guess poured first		Total
	Milk	Tea	
Milk	3	1	4
Tea	1	3	4
total	4	4	8

All marginal sums are fixed!

36

Fisher's eksakte test.

"For mathematical convenience, we shall assume that the margins (row sums and column sums) of this table are *fixed*"

Egentlig ser vi på betingede sannsynligheter (som gir en betinget p-verdi) gitt marginalsommene. Dette gir en test som garantert holder signifikansnivået, men kan være konservativ (dvs har lavere teststyrke enn nødvendig)

37

Fisher's eksakte test (forts)

Dersom dødsårsak og diett ikke er assosiert, er det helt tilfeldig hvem som hadde hatt saltholdig diett.

Tankeeksperiment: Trekk M_1 blant de N objektene (uten tilbakelegging), og la X være antall av de M_1 som var i gruppe 1 (her: ikke CVD).

38

	mye salt	lite salt	Total
Ikke CVD	X		N_1
CVD			N_2
Total	M_1	M_2	N

$$\Pr(X = a) = \frac{N_1! N_2! M_1! M_2!}{N! a! (N_1 - a)! (M_1 - a)! (M_2 - N_1 + a)!}$$

for alle heltall a som gir ikke-negative tall i alle cellene, dvs $a = \max(0, N_1 - M_2), \dots, \min(M_1, N_1)$

Kalles hypergeometrisk fordeling.

Rosner likn 10.8: Står $a = 0, \dots, \min(M_1, N_1)$ fordi Rosner først ordner rader og kolonner slik at $N_1 \geq N_2$ og $M_1 \geq M_2$.

39

Eksempel:

	mye salt	lite salt	Total
Ikke CVD	2	23	25
CVD	5	30	35
Total	7	53	60

$$\Pr(X = a) = \frac{25! 35! 7! 53!}{60! 2! 23! 5! 30!} = 0.252$$

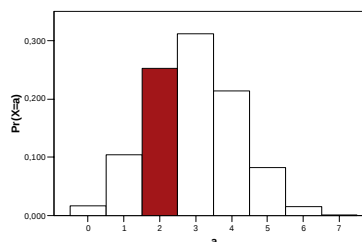
40

Tabell 10.12

Enumeration of all possible tables with fixed margins and their associated probabilities based on the hypergeometric distribution for Example 10.19

0	25	1	24	2	23	3	22
7	28	6	29	5	30	4	31
.017		.105		.252		.312	
4	21	5	20	6	19	7	18
3	32	2	33	1	34	0	35
.214		.082		.016		.001	

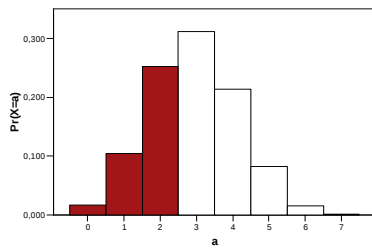
41



p-verdi = $\Pr(\text{Dette utfall eller noe mer ekstremt})$

Men hva er mer ekstremt?

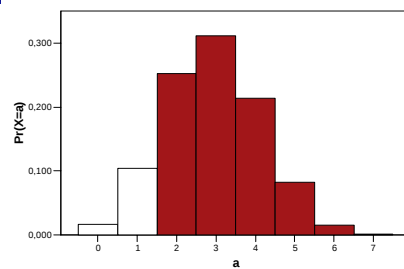
42



$H_0: p_1 = p_2$ versus $H_1: p_1 < p_2$

$$p_{\text{venstresidig}} = P(X \leq 2) = 0.375$$

43



$H_0: p_1 = p_2$ versus $H_1: p_1 > p_2$

$$p_{\text{høyresidig}} = P(X \geq 2) = 0.875$$

44

2-sidig p-verdi, som beskrevet i Rosner eqn 10.11 s 406:

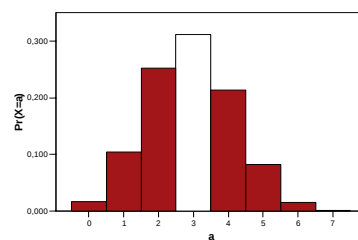
$$2\text{-sidig } p\text{-verdi} = 2 \times \min(p_{\text{venstresidig}}, p_{\text{høyresidig}}, 0.5) \\ = 2 \times \min(0.375, 0.875, 0.5) = 0.749$$

Dette er TST (Twice the smallest tail) prinsippet.

Merk at SPSS og Stata kun oppgir den ene ensidige p-verdien
 $p_{\text{ensidig}} = \min(p_{\text{venstresidig}}, p_{\text{høyresidig}})$

45

Tosidig p-verdi beregnet bl.a. i SAS, SPSS, MINITAB, Stata:



$$p\text{-verdi} = \sum_{\{i: \Pr(i) \leq \Pr(2)\}} \Pr(X = i) = \Pr(X \leq 2) + \Pr(X \geq 4) = 0.688$$

46

To alternativer for beregning av
 2-sidig p-verdi i Fisher's eksakte test:

$$2\text{-sidig } p\text{-verdi} = 2 \times p_{\text{ensidig}}$$

$$p\text{-verdi} = \sum_{\{i: \Pr(i) \leq \Pr(a)\}} \Pr(X = i)$$

Hvilken bør brukes?

47

Altman (1991), page 256:

"I feel that the second approach is more reasonable, but many statisticians recommend doubling the p-value obtained for one tail. ... The second approach will always give a value of P less than or equal to that obtained by the first method."

48

Agresti (2002), page 93:
 "Each approach has advantages and disadvantages ... To conduct a test of size 0.05 when one truly believes that the effect has a particular direction, it is safest to conduct the one-sided test at the 0.025 level to guard against criticism"

Kommentar: det siste tilsvarer $2 \times p_{\text{ensidig}} < 0.05$.

49

International Conference on Harmonization ICH E9 (1998):

"The approach of setting type I errors for one-sided tests at half the conventional type I error used in two-sided tests is preferable in regulatory settings. This promotes consistency with the two-sided confidence intervals that are generally appropriate for estimating the possible effect size of the difference between two treatments."

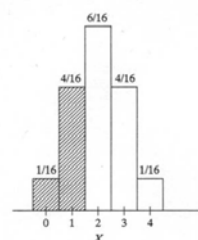
50

Om Yates' kontinuitetskorreksjon

51

Repetisjon om kontinuitetskorreksjon

If a random variable X has a binomial distribution with parameters $p = 1/2$ and $n = 4$ [$E(X) = np = 2$], then

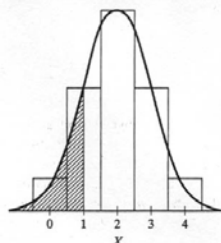


and, exactly,

$$G(1) = P(X \leq 1) = 1/16 + 4/16 = 5/16 = 0.3125.$$

52

An approximation using the normal distribution looks like:



$$\begin{aligned} E(X) &= np = 4(1/2) = 2 \\ \sigma_X^2 &= np(1-p) = 4(1/2)(1/2) = 1 \\ \sigma_X &= 1 \end{aligned}$$

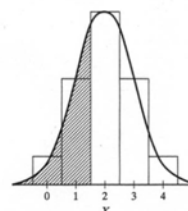
The approximate probability based on the normal distribution is

$$F(1) = P(X \leq 1) = P(Z \leq -1) = 0.159.$$

53

$$F(1 + 0.5) = P(X \leq 1 + 0.5) = P(X \leq 1.5) = P(Z \leq -0.5) = 0.309$$

(approximate + correction factor).



Altså:

Eksakt 0.3125

Normaltinnærmning uten korreksjon: 0.159

Normaltinnærmning med korreksjon: 0.309

54

Yates' korreksjon fremkommer ved å legge til eller trekke fra 0.5 fra hver celle slik at marginalsommene holdes konstant. Trekk fra (legg til) 0.5 fra a hvis $ad-bc > 0$ (< 0)

a - 0.5	b + 0.5	a+b
c + 0.5	d - 0.5	c+d
a+c	b+d	n=a+b+c+d

55

Da fås

$$\chi^2_{cc} = \frac{n[(a-0.5)(d-0.5) - (b+0.5)(c+0.5)]^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$= \frac{n[ad - bc - 0.5(a+b+c+d)]^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{n[ad - bc - 0.5n]^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$= \frac{n\left(|ad - bc| - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

56

Skal vi bruke Yates korreksjon?
(som er identisk med kontinuitetskorreksjonen i likning 10.3)

Selvin, S.: *Biostatistics. How it works*
Pearson, 2004. page 272-273:
“The use of this correction factor is equivocal and debate exists in the statistical literature on the appropriate balance between accuracy and power.”

57

Haviland, M. G. (1990). Yates's correction for continuity and the analysis of 2x2 contingency tables. *Statistics in Medicine*, **9**, 363-367.

Summary:

Despite recommendations to the contrary, medical researchers still routinely use the Yates-corrected chi-square statistic in analyses of 2 x 2 contingency tables. Research has shown that these 'corrected' statistics are overly conservative and that the conventional Pearson chi-square generally provides accurate control over type I error probabilities. This paper makes a straightforward argument against the use of Yates's correction for continuity and Fisher's exact probability test.

58

Haviland (1990) konkluderer:

Admittedly, a *single best strategy* for dealing with this problem across *all* experimental situations has not emerged. It is clear, however, that if a chi square test of association (independence or homogeneity) is considered appropriate, the conventional statistic is preferable to the one modified by Yates's method.

59

Agresti, A. *Categorical Data Analysis*. 2nd Ed, Wiley, 2002, page 103

“Since software now makes Fisher's exact test feasible even with large samples, this correction is no longer needed.”

Hirji, K. F. *Exact Analysis of Discrete Data*. Chapman & Hall / CRC, 2006, page 149

“An applied statistician today, in our view, may regard such corrections as historic curiosities, though Upton (1992) has a different opinion.”

60

NTNU

Tre metoder for analyse av 2x2 tabeller.

- To-utvalgstest for binomiske andeler:
 - Konfidensintervall for $p_1 - p_2$ kan også beregnes (Avsnitt 13.3)
- Pearson's kjikvadrattest.
 - Generaliserbar til rxc tabeller (Avsnitt 10.6)
- Fisher's eksakte test.
 - Garanterer at reelt signifikansnivå $\leq$ nominelt signifikansnivå α
 - Men har noe lavere styrke enn asymptotisk metode uten kontinuitetskorreksjon

61

NTNU

Hvilken metode bør brukes for testing i 2x2 tabeller?

- Store tabeller: (Cochran, 1954: Alle forventede antall >5)
 - Pearson's asymptotiske kjikvadrattest. (Identisk med z-test for 2 binomiske andeler.) Unngå Yates' kontinuitetskorreksjon.
- Ellers:
 - Fisher's eksakte test.
- Men: "Eksakt betinget mid p test" og "eksakt ubetinget" er bedre i små utvalg.
<http://folk.ntnu.no/slyderse/NSF%20conference%202007>

62

NTNU

Exact conditional mid p

better than

Asymptotic (e.g. Pearson χ^2)

Do not always preserve test size

Exact unconditional

better than

Exact conditional (e.g. Fisher exact)

Always preserve test size

63

NTNU

Unconditional tests in 2x2 tables

- Usually more powerful than conditional tests (Mehta and Hilton, TAS, 1993, Mehrotra et al, Biometrics, 2003).
- Free software: www.stat.ncsu.edu/exact/

64