

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Faglig kontakt under eksamen
 Stian Lydersen tlf 73867270 / 92632393

Eksamen ST2303 Medisinsk statistikk
 Tirsdag 6 desember 2005 kl 0900-1300

Tillatte hjelpemidler: Godkjent kalkulator. Tabeller og formler i statistikk, Institutt for matematiske fag. 1 håndskrevet A4-ark (stemplet ark fra Institutt for matematiske fag)

Eksamenssettet består av 4 sider norsk versjon, 4 sider engelsk versjon, og 1 side vedlegg (Tabell over kjikvadratfordelingen, fra Rosner)

Oppgave 1

a)

Overaktivitet av urinblæremuskelen (detrusor overactivity) kan skyldes blæremuskel hyper-refleksi (DH, detrusor hyperreflexia) eller urinblære ustabilitet (BI, bladder instability). Petersen & al (Br J Urol, 1997) undersøkte om en isvann-test kan brukes som diagnostisk test for å identifisere DH hos slike pasienter. Resultatene var:

		DH(+)	BI(-)	Total
Isvann- Test	+	39	3	42
	-	21	17	38
Total		60	20	80

Sensitivitet for testen er altså $39/60=65\%$. Beregn testens spesifisitet. Hva menes med positiv prediktiv verdi og negativ prediktiv verdi? Beregn disse for det gitte pasientutvalget.

b)

En annet diagnostisk test for samme tilstand er blitt undersøkt av en annen forskningsgruppe, med følgende resultat (ikke de samme pasientene som i a):

		DH(+)	BI(-)	Total
Test B	+	47	5	48
	-	13	15	32
Total		60	20	80

Beregn denne testens sensitivitet og spesifisitet. Er det grunnlag for å hevde at test B har bedre sensitivitet enn isvann-testen? Gjennomfør en hypotesetest, beregn p-verdien, og begrunn valg av metode for hypotesetesten.

c)

Tenk deg nå at undersøkelsen var blitt gjort med 60 pasienter med DH(+), der alle pasientene blir testet med begge testene, og med følgende resultat:

		Isvann-test		Total
		+	-	
Test B	+	38	9	47
	-	1	12	13
Total		39	21	60

Beregn p-verdien for en sammenlikning av sensitivitetene til test B og isvann-testen i denne situasjonen. Sammenlikne med svaret i b) og kommenter. Hva er fordelene ved denne studiedesignen i forhold til designen i a) og b)?

Oppgave 2

En gruppe pasienter med spiseforstyrrelsen over-spising (“binge-eating”) ble randomisert til medikamentet fluvoxamin eller placebo i et ni ukers dobbelt-blindt forsøk. På slutten av forsøket ble hver pasient klassifisert i en av 4 kategorier: Ingen respons, moderat respons, mye respons, eller bedring (remisjon). Tabellen viser en kryssklassifisering av resultatene:

treatment * response Crosstabulation

Count		response				Total
		no response	moderate response	marked response	remission	
treatment	placebo	22	7	3	11	43
	fluvoxamine	15	7	3	15	40
Total		37	14	6	26	83

For å teste om behandlingsgruppe og respons er assosiert, foreslås tre forskjellige analysealternativer av forskjellige forskere:

- i) Bruk Pearson’s kji-kvadrattest for assosiasjon. Testobservator: 1.834.
- ii) Bruk en kji-kvadrattest for trend. Testobservator: 1.734.
- iii) Bruk en logistisk regresjonsmodell med responskategori som kovariat og behandlingsgruppe som avhengig variabel. Estimert og estimert standardavvik for B er 0.225 og 0.170, når responsene er nummerert fra 0 (ingen respons) til 3 (remisjon), og behandling er kodet som fluvoxamine (1) og placebo (0).

a)
For hver av de foreslåtte metodene i) til iii), diskuter hvor passende den er i denne situasjonen. Er asymptotiske beregninger akseptable her – hvorfor eller hvorfor ikke?

b)
For hver av de foreslåtte metodene, hvilken fordeling har testobservatoren under H_0 ? Beregn den asymptotiske p-verdien for hver av dem så nøyaktig som mulig.

Oppgave 3.

a)
Hva er forskjellen mellom sensurerte data (som er vanlig ved levetidsdata) og manglende data ("missing data")?

b)
En mulig antakelse for fordelingen av manglende data er MAR ("missing at random"). Nevn to alternative antakelser. Hva går hver av disse tre antakelsene ut på? Hvordan kan man, eller kan man ikke, for et gitt datasett med manglende verdier, finne ut hvilken antakelse som er realistisk?

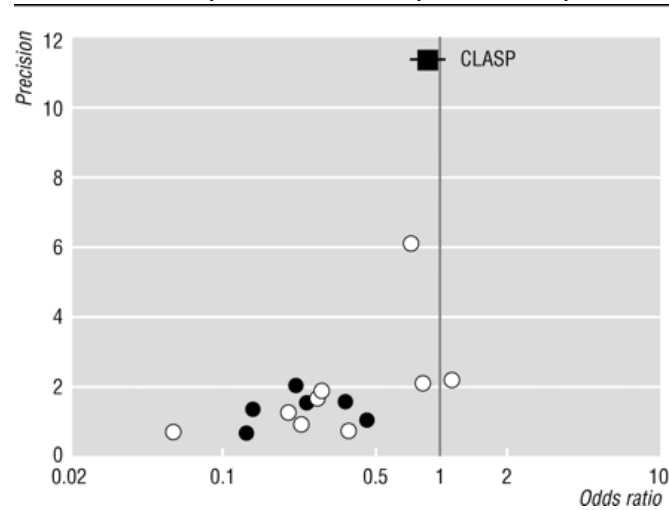
c)
En mye brukt fremgangsmåte ved manglende data er såkalt "complete case" analyse, også kalt "case deletion" eller "listwise deletion". Hva går denne ut på? Hva er ulempe(n)e med denne fremgangsmåten for hver av de tre antakelsene ovenfor, med tanke på forventningskjevhet ("bias") og/eller analysens styrke?

Oppgave 4

a)
Hvilke kilder til variasjon mellom enkeltstudier tas i betraktning i meta-analyse? Hva er forskjellen mellom en "fixed effect model" og en "random effect model"?

b)
Nedenstående traktplott "Funnel plot" stammer fra en meta-analyse av effekt av lavdose aspirin for å forebygge preeklampsi (svangerskapsforgiftning). Hvordan tolker du plottet – og hvilke konklusjoner trekker du?

**Funnel plot of trials of low dose aspirin
in the prevention of pre-eclampsia**



- Imperiale and Stollenwerk's 1991 meta-analysis
- Trials published in subsequent years up to 1993
- ! 1994 CLASP (collaborative low-dose aspirin study in pregnancy)

Oppgave 5

a)

Anta en Cox-modell med kovariater $\underline{x} = (x_1, \dots, x_p)$.

Utled et uttrykk for

$\log(-\log(S(t, \underline{x}_2))) - \log(-\log(S(t, \underline{x}_1)))$ der $\underline{x}_1 = (x_{11}, \dots, x_{1p})$ og $\underline{x}_2 = (x_{21}, \dots, x_{2p})$.

Hvordan kan dette resultatet brukes til å sjekke proporsjonal hazard antakelsen i praksis? Nevn en annen metode for å sjekke proporsjonal hazard antakelsen.

b)

Du vil sammenlikne overlevelse i to grupper, og den eneste forklaringsvariablen er gruppetilhørighet. Under hvilke(n) forutsetning(er) vil Log-rang testen og testing for β i en Cox-modell gi samme resultat?

The Norwegian University of Science and Technology
Department of Mathematical Sciences

Academic contact during the exam:
 Stian Lydersen, tel 73867270 / 92632393

Exam ST2303 Medical statistics
 Tuesday 6 December 2005, 0900-1300

Permitted aids: Approved calculator, Tabeller og formler i statistikk, Institutt for matematiske fag. 1 hand written A4 sheet (stamped sheet from Institutt for matematiske fag)

The exam consists of 4 pages Norwegian version, 4 pages English version, and 1 page appendix (Table of the chi square distribution, from Rosner)

Problem 1

a)

Overactivity of the urine bladder muscle (detrusor overactivity) can be caused by bladder muscle hyperreflexia (DH, detrusor hyperreflexia) or bladder instability (BI). Petersen & al (Br J Urol, 1997) investigated whether an ice water test can be used as diagnostic test to identify DH among such patients. The results were:

		DH(+)	BI(-)	Total
Ice water	+	39	3	42
Test	-	21	17	38
Total		60	20	80

The sensitivity for the test is thus $39/60=65\%$. Compute the specificity of the test. What is meant by positive predictive value and negative predictive value? Compute these for the given patient sample.

b)

Another diagnostic test for the same condition has been investigated by another research group, with the following results (not the same patients as in a):

		DH(+)	BI(-)	Total
Test B	+	47	5	48
	-	13	15	32
Total		60	20	80

Compute the sensitivity and specificity of this test. Can we conclude that test B has better sensitivity than the ice water test? Perform a hypothesis test, compute the p-value, and explain your choice of method for the hypothesis test.

c)

Assume that the investigation had been done on 60 patients with DH(+), where both tests were done on all the patients, with the following results:

		Ice water test		Total
		+	-	
Test B	+	38	9	47
	-	1	12	13
Total		39	21	60

Compute the p-value for comparing the sensitivities of Test B and the Ice water test in this situation. Compare with the answer in b) and comment. What are the advantages of this study design compared to the design in a) and b)?

Problem 2

A group of patients with a binge-eating disorder were randomly assigned to take either the experimental drug fluvoxamine or a placebo in a nine week long double-blind clinical trial. At the end of the trial the condition of each patient was classified into one of four categories: no response, moderate response, marked response, or remission. The following table shows a cross-classification of the data.

treatment * response Crosstabulation

Count		response				Total
		no response	moderate response	marked response	remission	
treatment	placebo	22	7	3	11	43
	fluvoxamine	15	7	3	15	40
Total		37	14	6	26	83

In order to test if there is an association between treatment group and condition, three different alternatives for analysis are suggested by different researchers:

- i) Use a Pearson's chi square test for association. Test statistic: 1.834.
- ii) Use a Chi Square test for trend. Test statistic: 1.734.
- iii) Use a logistic regression model with response category as covariate and treatment as dependent variable. The estimate and estimated standard deviation for B is 0.225 and 0.170, when responses are numbered from 0 (no response) to 3 (remission), and treatment is coded as fluvoxamine (1) and placebo (0).

a)

For each of the proposed analysis methods i) to iii), discuss its appropriateness in this situation. Are asymptotic calculations acceptable here – why or why not?

b)

For each of the three proposed methods, what is the distribution of the test statistic under H_0 ? Compute the asymptotic p-value corresponding to each of them as accurately as possible.

Problem 3.

a)

What is the difference between censored data (which is common for life time data) and missing data?

b)

A possible assumption for the distribution of missing data is MAR ("missing at random"). State two alternative assumptions. Explain what these 3 different assumptions mean. How can you, or can you not, for a given data set with missing data, find out which assumption is realistic?

c)

A much used method with missing data is "complete case" analysis, also called "case deletion" or "listwise deletion". How is it performed? What are its disadvantages under each of the 3 assumptions above, with respect to bias and/or the power of the analysis?

Problem 4

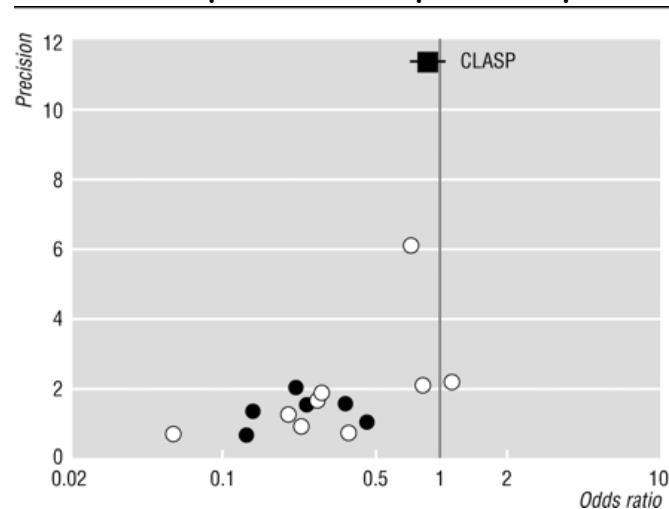
a)

Which sources of variation are considered in a meta analysis? What is the difference between a "fixed effect model" and a "random effect model"?

b)

The Funnel plot underneath is from a meta analysis of the effect of low dose aspirin on preventing pre-eclampsia. How do you interpret the plot and which conclusions do you draw?

Funnel plot of trials of low dose aspirin in the prevention of pre-eclampsia



- Imperiale and Stollenwerk's 1991 meta-analysis
- Trials published in subsequent years up to 1993
- ! 1994 CLASP (collaborative low-dose aspirin study in pregnancy)

Problem 5

a)

Assume a Cox-model with covariates $\underline{x} = (x_1, \dots, x_p)$.

Derive an expression for

$\log(-\log(S(t, \underline{x}_2))) - \log(-\log(S(t, \underline{x}_1)))$ where $\underline{x}_1 = (x_{11}, \dots, x_{1p})$ og $\underline{x}_2 = (x_{21}, \dots, x_{2p})$.

How can this method be used to check the proportional hazard assumption in practice? Name another method for checking the proportional hazard assumption.

b)

You are going to compare survival in two groups, and the only explanatory variable is group. Under which assumption(s) will the Logrank test and testing for β in a Cox-model give the same result?

Appendix: Table for the chi square distribution, from Rosner