Kapitel 8
Chi-i-anden (χ^2) prøven

Peter Tibert Stoltze
stat@peterstoltze.dk

Elementær statistik
F2011

Indledning

- Forskelle mellem stikprøver undersøges med
  - z-test eller t-test for data målt på interval- eller ratioskala
  - Mann-Whitney eller Wilcoxon for data målt på ordinalskala
- For data målt på nominalskala kan vi benytte χ^2-testet til at undersøge om
  - én stikprøve er fordelt svarende til et bestemt mønster (goodness-of-fit test)
  - to eller flere stikprøver er forskelligt fordelt på kategorierne (test for uafhængighed)
- Baseres på normeret forskel mellem observeret antal O og forventet antal E i de forskellige kategorier

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Goodness-of-fit med én stikprøve — eksempel

- Vi starter med et eksempel: På audiologopædi er der 110 studerende, der er fordelt med 100 kvinder og 10 mænd
- De i alt 5000 indskrevne ved fakultetet er fordelt med 3500 kvinder og 1500 mænd
- Skulle kønsfordelingen hos aud.log. være som på fakultetet som helhed ville vi forvente

$$E_K = 110 \cdot \frac{3500}{5000} = 77 \quad \text{og} \quad E_M = 110 \cdot \frac{1500}{5000} = 33$$

- Teststørrelsen beregnes nu som

$$\chi^2 = \frac{(100-77)^2}{77} + \frac{(10-33)^2}{33} = 22,90$$

- Jo større forskel mellem observeret og forventet antal, jo større bidrag til teststørrelsen

Eksempel *(fortsat)*

- Teststørrelsen skal sammenlignes med en χ^2-fordeling med $2-1=1$ frihedsgrader, da der er to grupper
- Kritiske værdier for χ^2-fordeling med $df = 1$

p	5%	1%	0,1%
χ^2	3,84	6,63	10,83

- Konklusionen på testet er, at stikprøven er signifikant forskellig fra modellen ($p < 0,1\%$), dvs. aud.log. ligner ikke resten af fakultetet

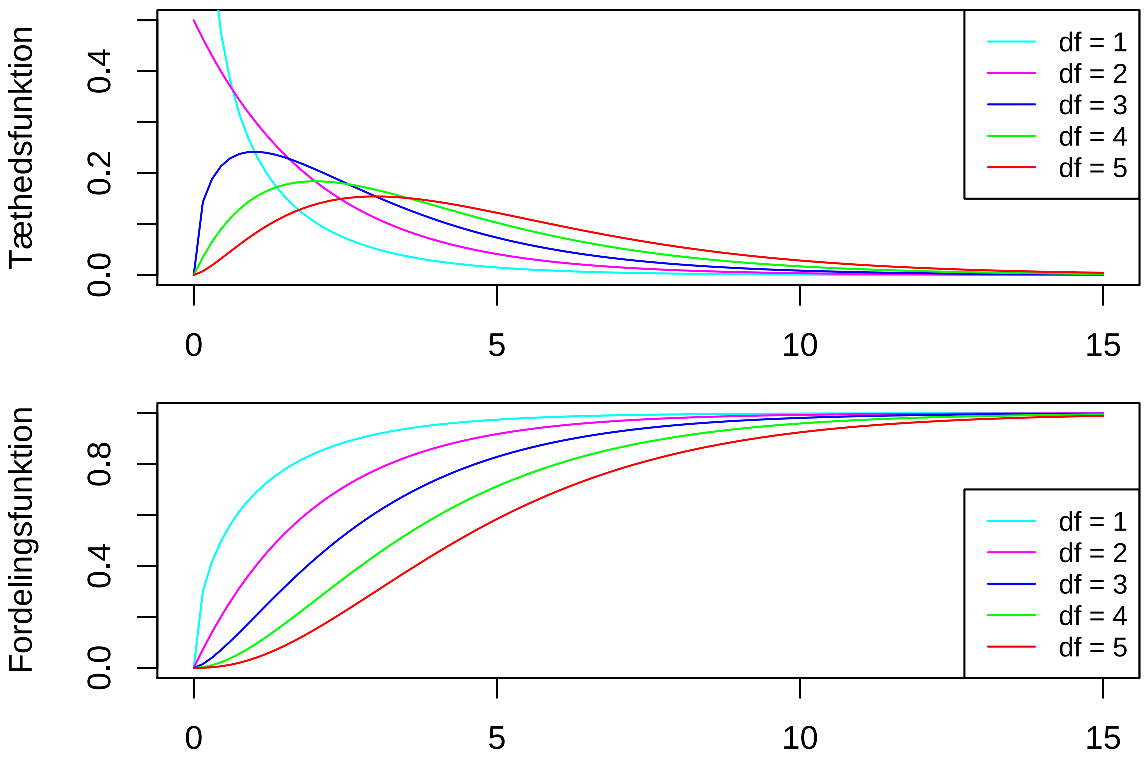
Lidt om χ^2-fordelingen

- Lad $Z_1, \ldots, Z_r$ være uafhængige og standardnormalfordelte stokastiske variable
- Fordelingen af den afledte stokastiske variabel

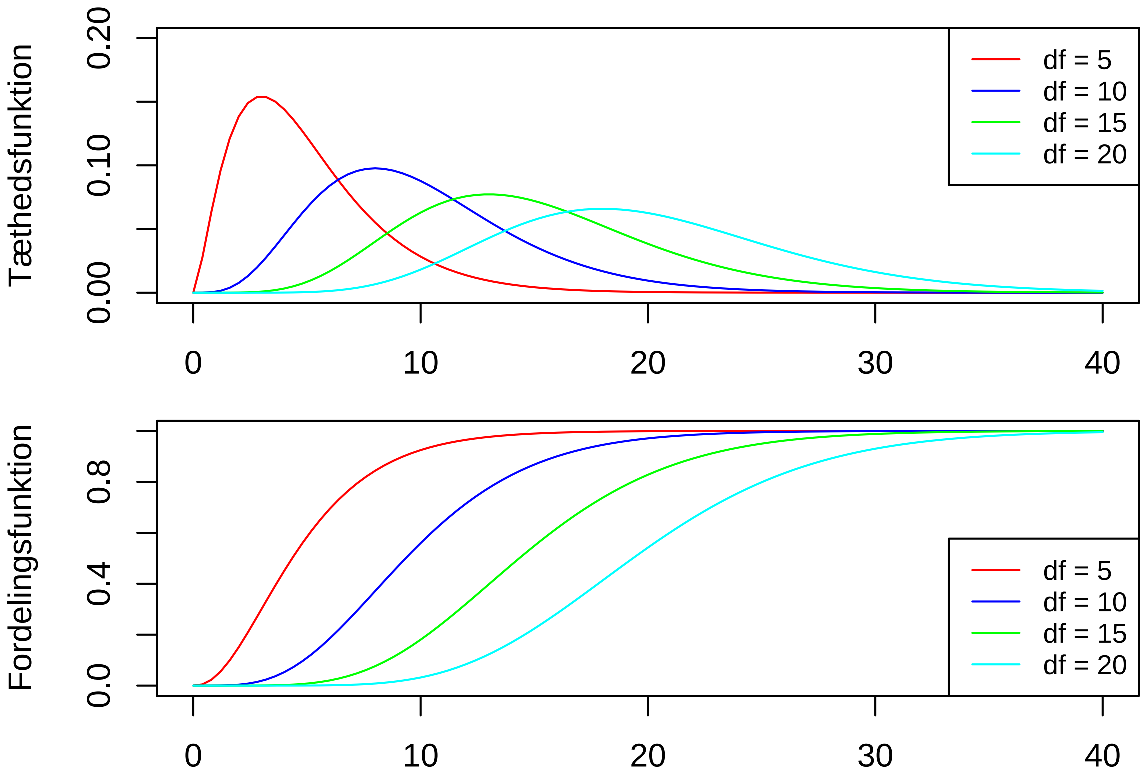
$$Y = Z_1^2 + \ldots + Z_r^2$$

  kaldes en χ^2-fordeling med r frihedsgrader
- Vi benytter ofte den korte skrivemåde $Y \sim \chi^2(r)$

Lidt om χ^2-fordelingen



Lidt om χ^2-fordelingen



Én stikprøve med mere end to grupper

- Samme princip ved mere end to grupper, idet antal frihedsgrader er antal grupper minus én
- Ingen celler må have $E < 1$, og højst $1/5$ af cellerne må have $E < 5$ — kan dette ikke overholdes må man lave passende sammenlægninger
- *Goodness-of-fit* kan også benyttes til sammenligning af stikprøve med teoretisk fordeling (kapitel 9)

Test for uafhængighed

- Vi kan sammenligne om n stikprøver er fordelt forskelligt over m kategorier ved at benytte et χ^2-test for uafhængighed
- Nulhypotesen er, at der ikke er forskel, og alternativhypotesen er, at der er forskel
- Bemærk dog: Er data korrelerede (fx. før- og efterscore) benyttes i stedet McNemars test (senere)

Test for uafhængighed *(fortsat)*

- Samme princip ved beregning af teststørrelsen, nemlig summation over $(O - E)^2/E$
- Data ordnes i en $n \times m$ tabel (matrix)
- Forventet antal beregnes for en celle som rækkesum gange søjlesum divideret med totalsum
- Antal frihedsgrader er $(n-1)(m-1)$

Test for uafhængighed *(fortsat)*

O	Gode	Middel	Dårlige	$\sum$
Piger	60	15	5	80
Drenge	30	15	15	60
$\sum$	90	30	20	140

E	Gode	Middel	Dårlige	$\sum$
Piger	51,4	17,1	11,4	79,9
Drenge	38,6	12,9	8,6	60,1
$\sum$	90,0	30,0	20,0	140,0

χ^2	Gode	Middel	Dårlige	$\sum$
Piger	1,43	0,27	3,62	5,31
Drenge	1,90	0,36	4,82	7,08
$\sum$	3,33	0,63	8,44	12,40

Test for uafhængighed *(fortsat)*

- Teststørrelsen skal sammenlignes med en χ^2-fordeling med $(2-1)(3-1) = 2$ frihedsgrader
- Kritiske værdier er

p	5%	1%	0,1%
$\chi^2(2)$	5,99	9,21	13,82

- Vi kan se at $0,01 > p > 0,001$ (faktisk er $p = 0,002$)
- Vi kan altså afvise, at drenge og piger fordeler sig ens på de tre læsekategorier
- Sandsynligheden for, at de faktisk er ens, og at den observerede forskel skyldes en tilfældighed, er $1/500$

Enkeltsidet alternativhypotese

- Retningsbestemt alternativ kun muligt for $df = 1$, dvs. ved et *goodness-of-fit* test for 2 kategorier eller uafhængighedstest for en 2×2 tabel
- Ønsker vi meget at få en retningsbestemt hypotese, så kan vi lægge beslægtede kategorier sammen
- p ved retningsbestemt alternativ får vi ved at dele p fra tabellen med 2...

Korrektion ved små frekvenser

- Hvis $E < 5$ kunne vi slå celler sammen, men det forudsætter at $df > 1$
- Hvis vi har $df = 1$ kan vi lave en Yates korrektion (på engelsk: Yates correction for continuity):

$$O > E : O \to O - \frac{1}{2}$$

$$O < E : O \to O + \frac{1}{2}$$

- Et eksempel kan være 8 kast med en mønt, der resulterer i 7 krone og én plat:

Data	Krone	Plat	χ^2	p
Forventet	4	4		
Observeret	7	1	4,500	3,4 %
Med korrektion	6,5	1,5	3,125	7,7 %

McNemars test for ændringer

- Benyttes når data er korrelerede, typisk i form af fordelingen før/efter på samme to kategorier
- Kan essentielt ordnes i en 2×2 tabel:

	Efter		
Før	Kategori 1	Kategori 2	$\sum$
Kategori 1	x_{11}	x_{12}	$x_{1\cdot}$
Kategori 2	x_{21}	x_{22}	$x_{2\cdot}$
$\sum$	$x_{\cdot 1}$	$x_{\cdot 2}$	$x_{\cdot\cdot}$

- Ændringerne (x_{12} og x_{21}) er de interessante, og det er dem, der benyttes til beregning af χ^2 teststørrelsen

McNemar's test *(fortsat)*

- Vi beregner teststørrelsen som

$$\chi^2_{McNemar} = \frac{(|x_{12} - x_{21}| - 1)^2}{x_{12} + x_{21}}$$

- Nulhypotesen for testen er, at antal ændringer fra kategori 1 til kategori 2 (x_{12}) er det samme som fra kategori 2 til 1 (x_{21}), hvilket netto svarer til *ingen ændring*
- Alternativhypoten er enten $x_{21} > x_{12}$ (*forbedring*) eller også $x_{21} < x_{12}$ (*forværring*)

McNemar's test *(fortsat)*

- Valg af alternativhypotese afhænger af data, der afgør hvilken af hypoteserne der er relevant

- Teststørrelsen sammenlignes med en χ^2-fordeling med én frihedsgrad (2 kategorier minus 1) idet vi benytter énhalet signifikanssandsynlighed ("dividerer p med 2") på grund af den retningsbestemte alternativhypotese

McNemars signifikanstest for ændringer — eksempel

Antal	Efter		
Før	Rigtig	Forkert	$\sum$
Rigtig	9	4	13
Forkert	12	5	17
$\sum$	21	9	30

- Vi beregner χ^2-teststørrelsen på baggrund af cellerne med ændringer (rigtig før$\rightarrow$ forkert efter og omvendt)

$$\chi^2 = \frac{(|x_{12} - x_{21}| - 1)^2}{x_{12} + x_{21}} = \frac{(|4 - 12| - 1)^2}{4 + 12} = \frac{49}{16} = 3,063$$

- Med $df = 1$ finder vi $p = 0,040$ énhalet og kan konkludere, at den observerede forbedring er signifikant

χ^2 med Excel

- `chidist(x;df)` beregnes som $P(X > x)$, hvor X følger en χ^2 fordeling med df frihedsgrader
- Den giver os altså vores p for en given værdi af teststørrelsen samt antallet af frihedsgrader
- Nørdet: Fordelingsfunktionen for χ^2-fordelingen får vi altså ved `1-chidist(x;df)`
- `chiinv(p;df)` er den omvendte funktion af `chidist`, så den kan vi bruge til at beregne kritiske værdier med
- `chitest(actual;expected)` kan vi bruge til at få beregnet p med det samme
- Supernørdet: Vil vi gerne kende teststørrelsen kan vi bruge `chiinv( chitest(actual;expected);df)`