F21, Chi-två test forts..

Christian Tallberg

Statistiska institutionen

Stockholms universitet

Exempel 2 (Körner sid 230):

Vad är sannolikhetsfördelningen för antalet pojkar i en slumpmässigt vald trebarnsfamilj? Vi har ett slumpmässigt stickprov med 400 trebarnsfamiljer där antalet pojkar fördelas på följande sätt:

Antal pojkar	0	1	2	3
Frekvens	56	135	142	67

Vi definierar den stokastiska variabeln X = "Antalet pojkar i en trebarnsfamilj". Det vi vill testa är följande hypoteser:

$$H_0 : X \text{ är } Bin(n = 3; p = 0.5)$$
$$H_1 : X \text{ är inte } Bin(n = 3; p = 0.5)$$

för signifikansnivån $\alpha = 0.05$.

Antal pojkar	Obs antal pojkar	Sannolikh under H_0	Förv antal pojkar (under H_0)
0	56	0.125	400·0.125 = 50
1	135	0.375	400·0.375 = 150
2	142	0.375	400·0.375 = 150
3	67	0.125	400·0.125 = 50
Summa	400	1	400

Testvariabeln är

$$\chi^2 = \sum \frac{(O-E)^2}{E} \sim approx\ \chi^2(a-1) - \text{ fördelad}$$
då H_0 är sann.

H_0 förkastas om $\chi^2_{obs} > 7.81$.

Det observerade testvärdet är

$$\begin{aligned}\chi^2_{obs} &= \frac{(56-50)^2}{50} + \frac{(135-150)^2}{150} + \frac{(142-150)^2}{150} \\ &\quad + \frac{(67-50)^2}{50} \\ &= 8.43.\end{aligned}$$

Då $8.43 > 7.81$ förkastas nollhypotesen. Det kan anses statistiskt påvisat att X inte är binomialfördelad med parametrarna $n = 3$ och $p = 0.5$.

Exempel 2 Körner (sid 230):

Föregående exempel med distinktionen att vi nu inte har någon uppfattning om parametern p (se kap 9.2). p skattas då från stickprovet med

$$\hat{p} = \frac{0 \cdot 56 + 1 \cdot 135 + 2 \cdot 142 + 3 \cdot 67}{3 \cdot 400} = \frac{620}{1200} = 0.52$$

Det vi vill testa är följande hypoteser:

$$H_0 \;:\; X \text{ är } Bin\,(n = 3; p = 0.52)$$
$$H_1 \;:\; X \text{ är inte } Bin\,(n = 3; p = 0.52)$$

för signifikansnivån $\alpha = 0.05$.

Antal pojkar	Obs antal pojkar	Sannolikh under H_0	Förv antal pojkar (under H_0)
0	56	0.111	$400 \cdot 0.111 = 44.4$
1	135	0.359	$400 \cdot 0.359 = 143.6$
2	142	0.389	$400 \cdot 0.389 = 155.6$
3	67	0.141	$400 \cdot 0.141 = 56.4$
Summa	400	1	400

Testvariabeln är

$$\chi^2 = \sum \frac{(O-E)^2}{E} \sim approx\ \chi^2\,(a-1-1) - \text{ fördelad}$$
då H_0 är sann.

H_0 förkastas om $\chi^2_{obs} > 5.99$.

Det observerade testvärdet är

$$\begin{aligned}\chi^2_{obs} &= \frac{(56-44.4)^2}{50} + \frac{(135-143.6)^2}{150} + \frac{(142-155.6)^2}{150} \\ &\quad + \frac{(67-56.4)^2}{50} \\ &= 6.73.\end{aligned}$$

Då $6.73 > 5.99$ förkastas nollhypotesen. Det kan anses statistiskt påvisat att X inte är binomialfördelad med parametrarna $n = 3$ och $p = 0.52$.

Analys av korstabeller (oberoendetest)

I en korstabell redovisas två variabler (eller fler) samtidigt. Variablerna kan vara kvantitativa eller kvalitativa. Man vill undersöka om (utfallet av) den ena variabeln är oberoende av (utfallet av) den andra variabeln.

Uppgift 906 (Körner sid 242):

Man vill undersöka om attityden till arbetsförhållandena skiljer sig mellan två företag. Vi har följande data från ett slumpmässigt stickprov:

Attityd	Företag A	Företag B
Positiv	62 (39.3)	43 (38.4)
Neutral	58 (58.9)	65 (57.6)
Negativ	11 (32.7)	20 (32.0)
Summa	131	128

Vi vill testa följande hypoteser:

H_0 : Attityd är *oberoende* av företag.

H_1 : Det finns ett samband mellan attityd och företag.

för (t.ex) $\alpha = 0.01$. Räkna om till absoluta tal.

Attityd	Företag A	Företag B	Summa
Positiv	20.4 (27.6)	48.6 (41.4)	69.0
Neutral	78.0 (62.9)	79.2 (94.3)	157.2
Negativ	21.6 (29.5)	52.2 (44.3)	73.8
Summa	120	180	300

För att beräkna det förväntade antalet i varje grupp under nollhypotesen kan man resonera på följande sätt. Den totala andelen oavsett företag som har uppfattningen "positiv" är

$$\frac{69}{300}.$$

Om attityden till arbetsförhållandena är oberoende av företag blir det förväntade antalet som är positiva i företag A

$$120 \cdot \frac{69}{300},$$

och det förväntade antalet som är positiva i företag B

$$180 \cdot \frac{69}{300},$$

osv.

Alternativt kan man resonera på följande sätt. Den totala andelen i företag A oavsett attityd är

$$\frac{120}{300}.$$

Om attityden till arbetsförhållandena är oberoende av företag blir det förväntade antalet som tillhör företag A i gruppen "positiva"

$$69 \cdot \frac{120}{300}.$$

Den totala andelen i företag B oavsett attityd är

$$\frac{180}{300}.$$

Om attityden till arbetsförhållandena är oberoende av företag blir det förväntade antalet som tillhör företag B i gruppen "positiva"

$$69 \cdot \frac{180}{300},$$

osv.

Generellt ser man nu att vi får en räkneformel där det förväntade antalet i varje cell blir produkten av marginalerna delat med totala antalet observationer.

Testvariabeln är

$$\chi^2 = \sum \frac{(O-E)^2}{E} \sim approx\ \chi^2 - \text{ fördelad}$$
$$\text{med } (r-1)(k-1) \text{ f.g. då } H_0 \text{ är sann.}$$

H_0 förkastas om $\chi^2_{obs} > 9.21$.

Det observerade testvärdet är

$$\begin{aligned}\chi^2_{obs} &= \frac{(20.4-27.6)^2}{27.6} + \frac{(48.6-41.4)^2}{41.4} \\ &+ \frac{(78-62.9)^2}{62.9} + \frac{(79.2-94.3)^2}{94.3} \\ &+ \frac{(21.6-29.5)^2}{29.5} + \frac{(52.2-44.3)^2}{44.3} \\ &= 12.73\end{aligned}$$

Då $12.73 > 9.21$ förkastas nollhypotesen. Det kan anses statistiskt påvisat att det finns ett beroende mellan attityd och företag.