Kap17: Kontingenstabeller

30. maj 2002

En kontingens tabel:
$$\begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1J} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{iJ} \\ \vdots & & \vdots & & \vdots \\ x_{I1} & \cdots & x_{Ij} & \cdots & x_{IJ} \end{bmatrix} \begin{matrix} x_{1\cdot} \\ \\ x_{i\cdot} \\ \\ x_{I\cdot} \end{matrix}$$
$$x_{\cdot 1} \quad x_{\cdot j} \quad x_{\cdot J} \quad \sum = x_{\cdot\cdot}(=n)$$

i = rækkenr, j = søjlenr. (husk "i" kommer før "j" i alfabetet, som "r" kommer før "s"!!!)

Kontingenstabel er: - Krydstabellering af en stikprøve m. *to el. flere kategoriserede variable* (opstilling i stratificerede grupper)

Eksempelvis: - Personer opstillet efter køn og alder

NB! Hvis Ho testes og ikke forkastes men der er stærk SYSTEMATIK i residualernes fortegn SKAL Ho FORKASTES

	Begreb	(A) Uafhængighedstabel		(C) Sammenligning af binomialfordel.
1	Kendetegn	Kont.tabel, hvor *stikprøvestørrelsen* (n = x..) er eksogent/naturligt givet	Kont.tabel, hvor *række- eller søjlesummeringer* (=stratumstørrelser: $x_{i\cdot}$ eller $x_{\cdot j}$) eksogent/naturligt givet	Tabel hvor J=2 (2 søjler/rækker), så x'erne er X_{i1} og X_{i2}
2	Eksempel	Efterstratificeret simpelt tilfældigt udvalgt stikprøve, f.eks. 200 pers., der inddeles efter køn (mænd vs. kvinder i søjler) og efter alder (i rækker).	Udvalg af 100 mænd og 100 kvinder (søjler), der inddeles efter alder (i rækkerne)	
3	Simultan fordeling af de stok. var.	$X_{11},\dots,X_{IJ} \sim Mult(n, p_{11},\dots,p_{IJ})$ Forudsætning: X_{ij} er stokastisk uafhængige	$X_{i1},\dots,X_{iJ} \sim Mult(n_i, p_{i1},\dots,p_{iJ}),\ i=1,\dots,I$ Forudsætning: X_{ij} er stokastisk uafhængige	$X_{ij} \sim Bin(n, p_i)$ for i = 1,...,I og j = 1,2 Forudsætning: X_{ij} er stokastisk uafhængige
4	Den marginale ssh.	$\sum_{i=1}^{I} p_{ij} = p_{\cdot j}$ hhv. $\sum_{j=1}^{J} p_{ij} = p_{i\cdot}$	$\sum_{i=1}^{I} p_{ij} = p_{\cdot j}$ hhv. $\sum_{j=1}^{J} p_{ij} = p_{i\cdot}$	
5	Estimat for 4	$\hat{p}_{i\cdot} = \frac{x_{i\cdot}}{n}$ og $\hat{p}_{\cdot j} = \frac{x_{\cdot j}}{n}$	$\hat{p}_{i\cdot} = \frac{x_{i\cdot}}{n}$ og $\hat{p}_{\cdot j} = \frac{x_{\cdot j}}{n}$	
6	**Forventet antal**	$n \cdot \hat{p}_{\cdot i}\hat{p}_{j\cdot} = \frac{x_{\cdot i}x_{j\cdot}}{b}$	$n \cdot \hat{p}_{\cdot i}\hat{p}_{j\cdot} = \frac{x_{\cdot i}x_{j\cdot}}{b}$	$n_i \cdot \hat{p}$
7	Hypotese	*Uafhængighedshypotesen:* $H_0: p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \forall$ i, j vs. H_1:H_0 falsk	*Homogenitetshypotesen:* $H_0: p_{1j} = \dots = p_{Ij} = p_j$ for j=1,...,J vs. H_1:H_0 falsk	*Homogenitetshypotesen:* $H_0: p_{i1} = \dots = p_{I1} = p$ idet p_{i2} = følger af $p_{i2}=1-p_{i1}$
8	Forudsætning for test	**Forventede værdier >3 (gerne>5)** og hypotesen forkastes ofte på 5 pct. niveau hvis, der er for få celler jf. s 578.	**Forventede værdier >3 (gerne>5)**	Forventede værdier >3 (gerne>5)
9	Betydning af hypostesen i ord	Stikprøvens fordeling af (eks.) køn er uafhængig af aldersfordelingen	Om fordelingen af køn på hhv. mænd og kvinder er ens for alle strata (Da de I strata er homogene mht. fordelingen alder)	
10	(teoretisk) Teststørrelse	$Q = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(X_{ij} - \frac{X_{i\cdot}X_{\cdot j}}{n}\right)^2}{\frac{X_{i\cdot}X_{\cdot j}}{n}}$	$Q = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(X_{ij} - n_i \cdot \frac{X_{\cdot j}}{n}\right)^2}{n_i \cdot \frac{X_{\cdot j}}{n}}$	$Q = \sum_{i=1}^{I} \frac{(X_i - n_i\hat{p})^2}{n_i\hat{p}(1-\hat{p})}$, idet $\hat{p} = \frac{X_1 + \dots + X_I}{n}$ *og* $s_i^2 = n_i\hat{p}(1-\hat{p})$
11	Observeret værdi af **teststørrelsen** *Bemærk:* $n_i = x_{i\cdot}$ så teststr. er ens for (A) og (B)	$q = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(x_{ij} - \frac{x_{i\cdot}x_{\cdot j}}{n}\right)^2}{\frac{x_{i\cdot}x_{\cdot j}}{n}} \sim \chi^2_{1-\alpha}[(I-1)(J-1)]$	$q = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(x_{ij} - n_i \cdot \frac{x_{\cdot j}}{n}\right)^2}{n_i \cdot \frac{x_{\cdot j}}{n}} \sim \chi^2_{1-\alpha}[(I-1)(J-1)]$	$q = \sum_{i=1}^{I} \frac{(x_i - n_i\hat{p})^2}{n_i\hat{p}(1-\hat{p})} \sim \chi^2_{1-\alpha}$(I-1), p estimeres ved $\hat{p} = \frac{x_1 + \dots + x_I}{n}$ *og* $s_i^2 = n_i\hat{p}(1-\hat{p})$

Kommentarer og spørgsmål til: statistik@hotmail.com

Udarbejdet af Philip Jagd