



Johan Koskinen, Statistiska institutionen, Stockholms universitet

Finansiell statistik, vt-05

F19 logistisk regression och icke-parametriska test

Logistisk regression

För exemplet: Y_i 1 om haft uppgång och 0 om nedgång

förklarande variabel x_i

blir alltså skattnigen

$$\hat{p}_{x_i} = \frac{e^{0,564+0,759x_i}}{1+e^{0,564+0,759x_i}}$$

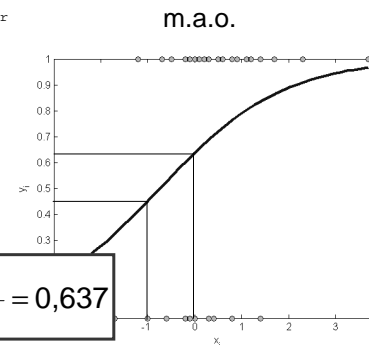
Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	0.563839	0.359835	1.57	0.117			
diffvinst	0.759181	0.412498	1.84	0.066	2.14	0.95	4.8

om man ökar värdet på den oberoende variabeln från -1 till 0 ökar den skattade sannolikheten för uppgång

$$\hat{p}_{-1} = \frac{e^{0,564-0,759}}{1+e^{0,564-0,759}} = 0,451$$

$$\hat{p}_0 = \frac{e^{0,564}}{1+e^{0,564}} = 0,637$$



Logistisk regression

Oddset för sannolikheter för 1 relativt sannolikheter för 0 ges av

$$\frac{\hat{p}_{x_{i1}, \dots, x_{ik}}}{1 - \hat{p}_{x_{i1}, \dots, x_{ik}}} = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 - e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}} = e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}$$

Oddskvoten mellan oddsen för $x_{i1}, x_{i2}, \dots, x_{ik}$ och $x_{i1}^*, x_{i2}^*, \dots, x_{ik}^*$

$$\frac{\frac{\hat{p}_{x_{i1}, \dots, x_{ik}}}{1 - \hat{p}_{x_{i1}, \dots, x_{ik}}}}{\frac{\hat{p}_{x_{i1}^*, \dots, x_{ik}^*}}{1 - \hat{p}_{x_{i1}^*, \dots, x_{ik}^*}}} = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{e^{\beta_0 + \beta_1 x_{i1}^* + \dots + \beta_k x_{ik}^*}}$$

Y_i 1 om barnet dog inom 1 mån och 0 annars

x_{i1} 1 om prenatal behandling på klinik A och 0 annars

x_{i2} 1 om mor på prenatal behandling mindre än 1 mån och 0 annars
barn dog

Klinik	mor närvarande	< 1 mån	> 1 mån	
A	< 1 mån	3	176	1,68
	> 1 mån	4	293	1,35
B	< 1 mån	17	197	7,94
	> 1 mån	2	23	8,00

$$\frac{\hat{p}_{x_{i1}, x_{i2}=0}}{1 - \hat{p}_{x_{i1}, x_{i2}=0}} = e^{\beta_2}$$

$$\frac{\hat{p}_{x_{i1}, x_{i2}=1}}{1 - \hat{p}_{x_{i1}, x_{i2}=1}}$$

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-2.54849	0.560623	-4.55	0.000			
klinik a	-1.69911	0.530662	-3.20	0.001	0.18	0.06	0.52
moder	0.110376	0.561015	0.20	0.844	1.12	0.37	3.35

obs! ej signifikant

χ^2 - test

Bl.a. har vi sett att för stickprov med normalfördelade variabler

$$\frac{(n-1)s^2}{\sigma^2} \in \chi^2(n-1)$$

generellt för många typer av data är summan av termer

$$\frac{(\text{observerade värden} - \text{förväntade värden})^2}{\text{förväntade värden}}$$

approximativt

$$\chi^2(v)$$



Johan Koskinen, Department of Statistics

2005-05-12

5

χ^2 - test

Lite abstrakt: enligt nollhypotes ges

förväntade värden

om de observerade värdena skiljer sig "mycket" från de förväntade

förkastar vi nollhypotesen

lämpligt mått på hur de observerade värdena skiljer sig från de förv.

$$\sum \frac{(\text{observerade värden} - \text{förväntade värden})^2}{\text{förväntade värden}}$$



Johan Koskinen, Department of Statistics

2005-05-12

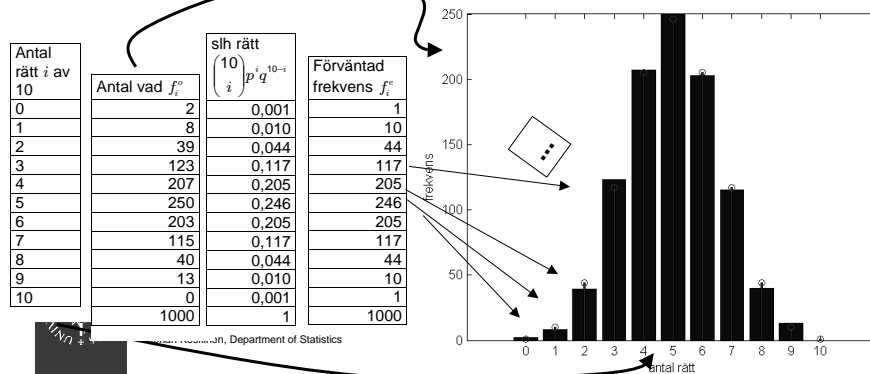
6

χ^2 - test, exempel

1000 gånger har det satsats på vinnaren i 10 matcher

H_0 : antal rätt beskrivs bra av en Binomial($p = 1/2$)

H_1 : antal rätt beskrivs inte bra av en Binomial($p = 1/2$)



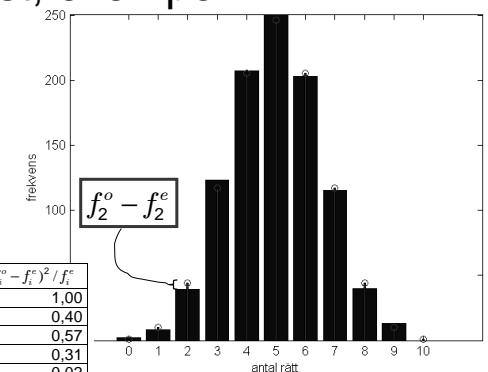
χ^2 - test, exempel

bra anpassning?

beräkna teststatistikan

$$\sum_{i=0}^{10} \frac{(f_i^o - f_i^e)^2}{f_i^e} = 4,68$$

Antal rätt i av 10	Antal vad f_i^o	slh rätt $\binom{10}{i} p^i q^{10-i}$	Förväntad frekvens f_i^e	$(f_i^o - f_i^e)^2 / f_i^e$
0	2	0,001	1	1,00
1	8	0,010	10	0,40
2	39	0,044	44	0,57
3	123	0,117	117	0,31
4	207	0,205	205	0,02
5	250	0,246	246	0,07
6	203	0,205	205	0,02
7	115	0,117	117	0,03
8	40	0,044	44	0,36
9	13	0,010	10	0,90
10	0	0,001	1	1,00
	1000	1	1000	4,68



Johan Koskinen, Department of Statistics

2005-05-12

8

χ^2 - test, exempel

Hur stort är det observerade värdet på tesstatistikan jämfört med det man borde få under H_0

Teststatistikan skall följa

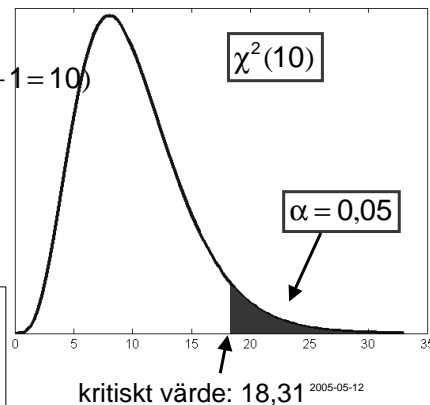
$$\sum_{i=0}^{10} \frac{(f_i^o - f_i^e)^2}{f_i^e} \in \chi^2(\# \text{ termer} - 1 = 10)$$

eftersom observerade värdet

$$\sum_{i=0}^{10} \frac{(f_i^o - f_i^e)^2}{f_i^e} = 4,68 < 18,31$$

förkastar vi inte

H_0 : antal rätt beskrivs bra av en Binomial($p = 1/2$)



9

χ^2 - test, för oberoende

För 2x2 -tabell "A mot B" testar man

H_0 : egenskaperna A & B är oberoende av varandra

H_1 : egenskaperna A & B är inte oberoende av varandra

		B		summa	
		ja	nej		
A	ja	a	b	a+b	(a+b)/s
	nej	c	d	c+d	(c+d)/s
summa		a+c	b+d	a+b+c+d	
relfrekv		(a+c)/s	(b+d)/s	s=a+b+c+d	1



Johan Koskinen, Department of Statistics

2005-05-12

10

χ^2 - test, för oberoende

Om A & B är oberoende händelser är $P(A \cap B) = P(A)P(B)$

alltså, använder vi marginalerna som skattningar får vi

		B		
		ja	nej	
A	ja	$f_1^e = s \frac{c_1 r_1}{s^2}$	$f_2^e = s \frac{c_2 r_1}{s^2}$	r_1
	nej	$f_3^e = s \frac{c_1 r_2}{s^2}$	$f_4^e = s \frac{c_2 r_2}{s^2}$	r_2
		c_1	c_2	s

frihetsgraderna är nu

(antal kolumner-1)

×

(antal rader-1)

11

χ^2 - test, för oberoende

Hur stort är det observerade värdet på tesstatistikan jämfört med det man borde få under H_0 : oberoende mellan rader och kolumner

Teststatistikan skall följa

$$\sum_{i=1}^n \frac{(f_i^o - f_i^e)^2}{f_i^e} \in \chi^2((r-1)(k-1))$$

Om stor skillnad mellan observerade och förväntade cellfrekvenser förkastar vi H_0



Johan Koskinen, Department of Statistics

2005-05-12

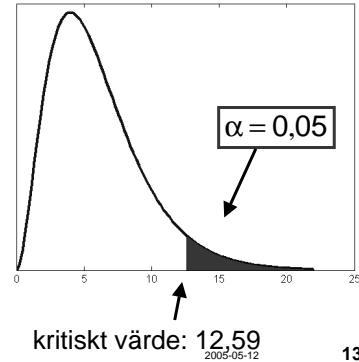
12

χ^2 - test, för oberoende

Chi-Square Test: A; B; C; F

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

	A	B	C	F	Total
Naturvet	12	36	34	8	90
	9.00	27.00	45.00	9.00	
	1.000	3.000	2.689	0.111	
Humaniora	10	24	46	10	90
	9.00	27.00	45.00	9.00	
	0.111	0.333	0.022	0.111	
Ekonomi	8	30	70	12	120
	12.00	36.00	60.00	12.00	
	1.333	1.000	1.667	0.000	
Total	30	90	150	30	300
Chi-Sq = 11.378; DF = 6; P-Value = 0.077					



Johan Koskinen, Department of Statistics