

Bài 5 KIỂM ĐỊNH MỘT PHÂN PHỐI VÀ BẢNG TƯƠNG LIÊN

1- Kiểm định một phân phối

Biến định tính X được chia thành p lớp, mỗi lớp có m_i cá thể (gọi là tần số thực tế).

Giả thiết H_0 cần kiểm định được thể hiện dưới dạng một dãy các tần suất f_i .

Lấy tổng $N = \sum m_i$ sau đó tính tần số lý thuyết $t_i = N \times f_i$

Tính các khoảng cách $(m_i - t_i)^2 / t_i$ sau đó lấy tổng được χ^2_{tn} .

Tính ngưỡng so sánh $\chi^2_{lt} = \chi^2_{inv}(\alpha, p-1)$.

Nếu $\chi^2_{tn} \leq \chi^2_{lt}$ thì chấp nhận H_0 , ngược lại thì bác bỏ H_0

Chấp nhận H_0 có nghĩa là chấp nhận giả thiết “dãy quan sát thực tế m_i phù hợp với dãy phân phối f_i ”

Đậu với 2 tính trội (Tỷ lệ 9:3:3:1)				Khoảng cách	
x	m_i	f_i	t_i	$(m_i - t_i)^2 / t_i$	
AB	59	9/16	$t_1 = N \times f_1$	64.69	0.5001
Ab	18	3/16	$t_2 = N \times f_2$	21.56	0.5886
bA	26	3/16	$t_3 = N \times f_3$	21.56	0.9132
ab	12	1/16	$t_4 = N \times f_4$	7.19	3.2223
N=	115			χ^2_{tn}	5.2242
				χ^2_{lt}	7.8147
$\chi^2_{tn} < \chi^2_{lt}$					
Kết luận : Chấp nhận H_0					
Tinh Chitest(day m, day t), nếu $< 0,05$ thì bác bỏ H_0					

2- Bảng tương liên

Cho bảng hai chiều với hai biến định tính X, Y biến X chia thành k lớp, Y chia thành l lớp. Trong bảng chứa các tần số thực tế m_{ij} .

Giả thiết H_0 : “Hai biến X (hàng) và Y (cột) không có quan hệ” hoặc giả thiết H_0 “Các đám đông ở hàng (X) có cùng tỷ lệ phân chia theo các cột (Y)”

Để chấp nhận hay bác bỏ H_0 cần phải:

Tính các tổng hàng TH_i , tổng cột TC_j , Tổng toàn bộ N.

Tính các số lý thuyết $t_{ij} = TH_i \times TC_j / N$

Tính các khoảng cách $(m_{ij} - t_{ij})^2 / t_{ij}$

$\chi^2_{tn} =$ Tổng tất cả các khoảng cách

$\chi^2_{lt} = \chi^2_{inv}(\alpha, df)$ với $df = (k-1)(l-1)$

Nếu $\chi^2_{tn} \leq \chi^2_{lt}$ thì chấp nhận H_0 , ngược lại thì bác bỏ H_0

Có thể tính Chitest (day m, day t) nếu kết quả $\geq \alpha$ chấp nhận H_0 , ngược lại thì bác bỏ H_0

2 ao và 5 loại cá						
Bảng số thực tế m _{ij}						
Ao	10-20cm	20-30	30-40	40-50	50-60	THi
Ao1	108	140	256	385	111	1000
Ao2	9	15	28	40	8	100
TCj	117	155	284	425	119	1100
Bảng số lý thuyết t _{ij}						N
106.3636 140.9091 258.1818 386.3636 108.1818						
10.63636 14.09091 25.81818 38.63636 10.81818						
Bảng khoảng cách						
0.0252 0.0059 0.0184 0.0048 0.0734						
0.2517 0.0587 0.1844 0.0481 0.7341						
Chi ² _{tn}		1.405		Chi ² _{lt}		9.4877
Chi ² _{tn} <		Chi ² _{lt}		Kết luận: Chấp nhận H_0		
Chitest(B4:F5,B8:F9)				0.8434 > 0,05		

Tai nạn và tuổi lái xe						
Bảng m _{ij}						
Tuổi lái xe						
Số tai nạn	21-30	31-40	41-50	51-60	61-70	TH _i
0	748	821	786	720	672	3747
1	74	60	51	66	50	301
2	31	25	22	16	15	109
trên 2	9	10	6	5	7	37
TC _j	862	916	865	807	744	4194

Bảng t _{ij}					
770.127	818.372	772.808	720.989	664.704	
61.865	65.741	62.080	57.918	53.396	
22.403	23.806	22.481	20.974	19.336	
7.605	8.081	7.631	7.119	6.564	

Bảng khoảng cách					
0.63576	0.00844	0.2252	0.00136	0.08009	
2.3803	0.50128	1.97767	1.12786	0.21602	
3.29908	0.05985	0.01029	1.17939	0.9724	
0.25602	0.45567	0.34865	0.63096	0.02901	

Chi ² _{tn}	14.3953	Chi ² _{lt}	21.0261
Kết luận: chấp nhận H ₀			
Chitest	0.27618 > 0,05		