

定性数据统计分析作业 (7)

钟瑜 222018314210044

2020 年 12 月 18 日

4. 某种疾病有两种治疗方法, 一是按病因服标准剂量的药, 二是按症状服药。考虑到治疗效果与性别没有关系, 但可能与年龄有关, 所以分 20 到 35 岁, 以及 36 到 50 岁两个年龄组。治疗结果如下:

年龄	治疗效果	治疗方法	
		按病因服标准剂量的药	按症状服药
20~35	迅速痊愈	22	19
	缓慢痊愈	16	17
	未痊愈	2	4
36~50	迅速痊愈	19	15
	缓慢痊愈	61	72
	未痊愈	20	13

试求各种独立性检验问题的解。

解. 计算独立性检验的 p 值:

```
1 > #高维列联表独立性检验
2 > mut_independence=function(x){
3 +   r=dim(x)[1];r      #行
4 +   c=dim(x)[2];c      #列
5 +   t=dim(x)[3];t      #层
6 +   n=sum(x)
7 +   temp11=0
8 +   temp21=0
9 +   temp22=0
10 +  temp23=0
11 +  temp31=0
12 +  temp32=0
13 +  temp33=0
14 +  for (i in 1:r) {
15 +    for (j in 1:c){
16 +      for (k in 1:t){
17 +        ni__=sum(x[i,,])
```

```

18 +         n_j_=sum(x[,j,])
19 +         n__k=sum(x[, ,k])
20 +         n_jk=sum(x[,j,k])
21 +         ni_k=sum(x[i, ,k])
22 +         nij_=sum(x[i,j,])
23 +         temp11=temp11+x[i,j,k]*log(ni__*n_j_*n__k/(x[i,j,k]*n^2))
24 +
25 +         temp21=temp21+x[i,j,k]*log(ni__*n_jk/(x[i,j,k]*n))
26 +         temp22=temp22+x[i,j,k]*log(n_j_*ni_k/(x[i,j,k]*n))
27 +         temp23=temp23+x[i,j,k]*log(n__k*nij_/(x[i,j,k]*n))
28 +
29 +         temp31=temp31+x[i,j,k]*log(ni_k*nij_/(x[i,j,k]*ni__))
30 +         temp32=temp32+x[i,j,k]*log(nij_*n_jk/(x[i,j,k]*n_j_))
31 +         temp33=temp33+x[i,j,k]*log(ni_k*n_jk/(x[i,j,k]*n__k))
32 +
33 +     }
34 + }
35 + }
36 + A_B_C=-2*temp11;A_B_C
37 + A_BC=-2*temp21;A_BC
38 + B_AC=-2*temp22;B_AC
39 + C_AB=-2*temp23;C_AB
40 + AB_AC=-2*temp31;AB_AC
41 + BA_BC=-2*temp32;BA_BC
42 + CA_CB=-2*temp33;CA_CB
43 +
44 + cat('\n(A,B,C) p值: ',1-pchisq(A_B_C, r*c*t-r-c-t+2))
45 + cat('\n(A,BC) p值: ',1-pchisq(A_BC, (r-1)*(c*t-1)))
46 + cat('\n(B,AC) p值: ',1-pchisq(B_AC, (c-1)*(r*t-1)))
47 + cat('\n(C,AB) p值: ',1-pchisq(C_AB, (t-1)*(r*c-1)))
48 + cat('\n(AB,AC) p值: ',1-pchisq(AB_AC, r*(c-1)*(t-1)))
49 + cat('\n(BA,AC) p值: ',1-pchisq(BA_BC, c*(r-1)*(t-1)))
50 + cat('\n(CA,CB) p值: ',1-pchisq(CA_CB, t*(r-1)*(c-1)))
51 + }
52 > dim(x)=c(3,2,2);x
53 , , 1
54
55 [,1] [,2]
56 [1,] 22 19
57 [2,] 16 17
58 [3,] 2 4

```

```

59
60 , , 2
61
62 [,1] [,2]
63 [1,] 19 15
64 [2,] 61 72
65 [3,] 20 13
66
67 > mut_independence(x)
68
69 (A,B,C) p 值: 5.943482e-06
70 (A,BC) p 值: 2.246753e-06
71 (B,AC) p 值: 0.5773336
72 (C,AB) p 值: 2.083189e-06
73 (AB,AC) p 值: 0.6496818
74 (BA,AC) p 值: 6.493699e-07
75 (CA,CB) p 值: 0.4325714

```

(A,B,C) p 值: 5.943482e-06
 (A,BC) p 值: 2.246753e-06
 (B,AC) p 值: 0.5773336
 (C,AB) p 值: 2.083189e-06
 (AB,AC) p 值: 0.6496818
 (BA,AC) p 值: 6.493699e-07
 (CA,CB) p 值: 0.4325714

可以看出, A, B 和 C 之间仅有相关关系。它们相互之间不独立, 其中任意一个和另外两个也不独立, 其中任意一个给定后另外两个条件独立。(B,AC) p 值: 0.57733>0.1 不独立, B 与 AC 相关。

6. 为了解男性和女性对两种类型的饮料的偏好有没有差异, 分别在年青人和老年人中作调查。调查数据如下:

		偏好饮料 A	偏好饮料 B
年青人	男性	37	26
	女性	11	23
老年人	男性	30	43
	女性	31	11

试分析这批数据, 关于男性和女性对这两种类型的饮料的偏好有没有差异的问题, 你有什么看法? 为什么?

解.

```

1  congruence_test=function(x,alternative="twoside")
2  #适用于列联表的相合性检验问题
3  #x为列联表矩阵; alternative对应于备择假设'twoside'相合,'greater'正相合,or
   'less'负相合
4  {
5      n=sum(x)
6      G=0;H=0
7      r=nrow(x)
8      c=ncol(x)
9      r1=r-1;c1=c-1
10     for (i in 1:r1){
11         for (j in 1:c1){
12             G=G+x[i,j]*sum(x[(i+1):r,(j+1):c])
13         }
14     }
15     for (i in 1:r1){
16         for (j in 2:c){
17             H=H+x[i,j]*sum(x[(i+1):r,1:(j-1)])
18         }
19     }
20     z=G-H
21     TA=sum(rowSums(x)*(rowSums(x)-1)/2)
22     TB=sum(colSums(x)*(colSums(x)-1)/2)
23     #TAB=G+H+TA+TB-n*(n-1)/2
24     Cn2=n*(n-1)/2
25     #计算各系数的值
26     Kendall_TAO=z/sqrt((Cn2-TA)*(Cn2-TB))
27     Gamma=(G-H)/(G+H)
28     d_BA=(G-H)/(Cn2-TA)
29     d_AB=(G-H)/(Cn2-TB)
30
31     #近似公式,表示sigma的平方
32     sigma_2=(n^3-sum(rowSums(x)^3))*(n^3-sum(colSums(x)^3))/(9*n^3)
33
34     #构建U统计量
35     U=z/sqrt(sigma_2)
36     if(alternative=="twoside")
37     {p_value=1-pchisq(U^2, 1)}
38     else
39     {
40         if(alternative=="greater")

```

```

41         {p_value=pnorm(-U)}
42         else if(alternative=="less")
43         {p_value=pnorm(U)}
44         else{cat("please input:\n alternative= 'twoside','greater
           ',or'less'")}
45     }
46     cat('【各种相关系数】\n')
47     cat('Kendall_TAO=',Kendall_TAO,'\n')
48     cat('Gamma=',Gamma,'\n')
49     cat('d_BA=',d_BA,'\n')
50     cat('d_AB=',d_AB,'\n\n')
51     cat('【相合性检验】\n')
52     cat('U检验统计量的值',U,'\n')
53     cat('p_value=',p_value)
54 }
55
56 > x=c(37,11,26,23,30,31,43,11)
57 > dim(x)=c(2,2,2);x
58 , , 1
59
60 [,1] [,2]
61 [1,]  37  26
62 [2,]  11  23
63
64 , , 2
65
66 [,1] [,2]
67 [1,]  30  43
68 [2,]  31  11
69
70 > mut_independence(x)
71
72 (A,B,C) p值:  0.001036864
73 (A,BC)  p值:  0.0004155325
74 (B,AC)  p值:  0.0003747595
75 (C,AB)  p值:  0.0005111978
76 (AB,AC) p值:  0.0001482043
77 (BA,AC) p值:  0.0001652176
78 (CA,CB) p值:  0.0001191498
79
80 > #各个独立性检验的p值都很小，故这三个属性相互之间仅有相关关系

```

```

81 > #下对年龄进行分层，讨论相合性
82 > A1=x[,1]
83 > A2=x[,2]
84
85 > congruence_test(A1,alternative="greater")
86 【各种相关系数】
87 Kendall_TAO= 0.2517214
88 Gamma= 0.4969217
89 d_BA= 0.2637722
90 d_AB= 0.2402211
91
92 【相合性检验】
93 U检验统计量的值 2.479168
94 p_value= 0.006584467
95
96 > #拒绝原假设，认为是正相合，即年轻人中，男性偏好饮料A，女性偏好饮料B
97 > congruence_test(A2,alternative="less")
98 【各种相关系数】
99 Kendall_TAO= -0.3156117
100 Gamma= -0.6031269
101 d_BA= -0.3271363
102 d_AB= -0.304493
103
104 【相合性检验】
105 U检验统计量的值 -3.384558
106 p_value= 0.0003564646
107 > #拒绝原假设，认为是负相合，及老年人中，男性偏好饮料B，女性偏好饮料A

```