

Практика №2 Статистический анализ качественных признаков по таблицам сопряженности

Долаева А.Р., г. 20.М04-мм

06/04/2021

Вариант (с) Депрессия и гиперимия (depressed.mood.1, hyperemia.1), слабость и бессонница (weakness.1, insomnia.1)

Функции вычисления энтропии и неопределенности

```
Entropy<-function(x){
  p<-x/sum(x);
  pp<-p[p!=0]
  -sum(p*log(p,2))
}

Uncertain<-function(tab){
  Hxy<-Entropy(as.vector(tab));Hxy
  Hx<-Entropy(rowSums(tab));Hx
  Hy<-Entropy(colSums(tab));Hy
  I<-Hx+Hy-Hxy;I
  c(I/Hx*100, I/Hy*100, 2*I/(Hx+Hy)*100)
}
```

1. Построить две таблицы сопряженности по двум парам качественных признаков

1.1 (depressed.mood.1, hyperemia.1)

```
data <- read.table(file = "data.csv", header = TRUE, sep = ",")
X1<-data$depressed.mood.1;X1
```

```
## [1] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 2 1 1 1 2
```

```
Y1<-data$hyperemia.1;Y1
```

```
## [1] 2 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 0 2 1 1 1 1 2 2 2 1 2 1 1 1 2
```

```
Y1<-ifelse(Y1<2,1,2)
```

```
tab1<-table(depressed=X1, hyperemia=Y1);tab1
```

```
##           hyperemia
## depressed  1  2
##           1 22  6
##           2  2  4
```

1.2 (weakness.1, insomnia.1)

```
X2<-data$weakness.1;X2
```

```
## [1] 1 1 2 2 2 2 2 1 1 1 1 2 2 2 2 1 2 0 1 0 2 1 2 1 1 2 0 2 1 2 1 1 1 2
```

```
X2<-ifelse(X2<2,1,2)
Y2<-data$insomia.1;Y2
```

```
## [1] 1 2 1 0 1 0 2 2 0 2 1 0 2 2 1 2 1 2 1 2 0 2 2 0 1 0 1 2 2 2 2 1 2
```

```
Y2<-ifelse(Y2<2,1,2)
tab2<-table(weakness=X2, insomia=Y2);tab2
```

```
##          insomia
## weakness  1  2
##          1  8 10
##          2  9  7
```

2. Вычислить условные вероятности и указать на значимость их различия по критерию хи-квадрат и по точному критерию Фишера

2.1 (depressed.mood.1, hyperemia.1)

```
tab1[,2]/rowSums(tab1)
```

```
##          1          2
## 0.2142857 0.6666667
```

```
fisher.test(tab1)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  tab1
## p-value = 0.04775
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.7686783 92.8517921
## sample estimates:
## odds ratio
##  6.807227
```

```
chisq.test(tab1)
```

```
## Warning in chisq.test(tab1): Chi-squared approximation may be incorrect
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tab1
## X-squared = 2.9354, df = 1, p-value = 0.08666
```

При слабой депрессии сильная гиперемия наблюдается в 21% случаев, а при ее усилении в более чем 66% случаев. Критерии дали разные результаты: по критерию Фишера $pvalue = 0,047 < 0,05$ - а значит, различие в частотах нельзя объяснить случайностью; по критерию хи-квадрат $pvalue = 0,086 > 0,05$ - результаты не значимы.

2.2 (weakness.1, insomia.1)

```
tab2[,2]/rowSums(tab2)
```

```
##          1          2
## 0.5555556 0.4375000
```

```
fisher.test(tab2)$p.value
```

```
## [1] 0.7318573
```

```
chisq.test(tab2)
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data: tab2  
## X-squared = 0.11806, df = 1, p-value = 0.7312
```

При низкой слабости сильная бессонница наблюдается в более чем 54% случаев, а при ее увеличении в более чем 43% случаев. Различие в частотах можно быть случайностью: по критерию Фишера $pvalue = 0,732 > 0,05$; по критерию хи-квадрат $pvalue = 0,731 > 0,05$

3. Вычислить коэффициенты неопределенности

3.1 (depressed.mood.1, hyperemia.1)

```
Uncertain(tab1)
```

```
## [1] 14.07345 10.82576 12.23780
```

Для признака депрессия неопределенность меньше, чем для признака гипермия. Коэффициент неопределенности в случае зависимой переменной депрессия 14% больше, чем в случае зависимой переменной гипермия 10%. Для признака депрессия неопределенность меньше, чем для признака гипермия. Большая доля информации о депрессия определяется тем, есть гипермия или нет, 14%. Если гипермия несильная, то высокая депрессия встречается у 1 из 12 человек, при сильной гипермии шансы 2 из 5.

3.2 (weakness.1, insomnia.1)

```
Uncertain(tab2)
```

```
## [1] 1.006751 1.004236 1.005492
```

Для признаков слабость и бессонница наблюдается высокая неопределенность, что говорит о сильной зависимости значений 9/17 и 7/17. Либо различия в показаниях случайны, либо требуется большая выборка.

4. Применить критерии Мак-Немара и Кохрена для проверки значимых изменений во времени

```
library(nonpar)  
library(plyr)  
Name<-"depressed.mood"  
A<-sapply(colnames(data),function(x)substr(x,1,nchar(Name)))  
data.<-apply(data[,names(A[A==Name])],2,function(x)ifelse(x<2,1,2))  
tab<-table(data.[,1],data.[,2])  
tab
```

```
##  
##      1  2  
##  1 27  1  
##  2  3  3
```

```
mcnemar.test(data.[,1],data.[,2])
```

```
##  
## McNemar's Chi-squared test with continuity correction  
##  
## data: data.[, 1] and data.[, 2]  
## McNemar's chi-squared = 0.25, df = 1, p-value = 0.6171
```

Для дисперсии проверяем гипотезу о том, что количество ухудшившихся (1) равно количеству улучшившихся (3) на второй день. По критерию Мак-Немара: Значимость $p-value = 0.6171 > 0.05$, результаты не значимы.

```
data.<-na.omit(data.[,seq(3)])-1
colnames(data..)

## [1] "depressed.mood.1" "depressed.mood.2" "depressed.mood.3"
cochrans.q(data., alpha = 0.05)

##
## Cochran's Q Test
##
## H0: There is no difference in the effectiveness of treatments.
## HA: There is a difference in the effectiveness of treatments.
##
## Q = 6.33333333333333
##
## Degrees of Freedom = 2
##
## Significance Level = 0.05
## The p-value is 0.0421438435092764
## There is enough evidence to conclude that the effectiveness of at least two treatments differ.
##
```

По критерию Кохрена $p\text{-value} = 0.042 < 0.05$, в течение трех дней у больных существенное уменьшилась депрессия.

```
Name<-"hyperemia"
A<-sapply(colnames(data),function(x)substr(x,1,nchar(Name)))
data.<-apply(data[,names(A[A==Name])],2,function(x)ifelse(x<1,0,1))
tab<-table(data.[,1],data.[,2])
tab
```

```
##
##      0  1
##  0  0  1
##  1 10 23
```

```
mcnemar.test(data.[,1],data.[,2])
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data: data.[, 1] and data.[, 2]
## McNemar's chi-squared = 5.8182, df = 1, p-value = 0.01586
```

Для гипермии проверяем гипотезу о том, что количество ухудшившихся (1) равно количеству улучшившихся (10) на второй день. По критерию Мак-Немара: Значимость $p\text{-value} = 0.01 < 0.05$, поэтому можно говорить о значимом улучшении на второй день по признаку гипермии.

```
data.<-na.omit(data.[,seq(3)])-1
colnames(data..)

## [1] "hyperemia.1" "hyperemia.2" "hyperemia.3"
cochrans.q(data., alpha = 0.05)
```

```
##
## Cochran's Q Test
##
## H0: There is no difference in the effectiveness of treatments.
```

```
## HA: There is a difference in the effectiveness of treatments.
##
## Q = -5.77272727272727
##
## Degrees of Freedom = 2
##
## Significance Level = 0.05
## The p-value is 1
##
##
```

По критерию Кохрена $p\text{-value} = 1 > 0.05$, улучшение по признаку гипермии на третий день не значимо, можно объяснить случайностью.

```
Name<-"weakness"
A<-sapply(colnames(data),function(x)substr(x,1,nchar(Name)))
data.<-apply(data[,names(A[A==Name])],2,function(x)ifelse(x<2,1,2))
tab<-table(data.[,1],data.[,2])
tab
```

```
##
##      1  2
##  1 16  2
##  2 11  5
```

```
mcnemar.test(data.[,1],data.[,2])
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data: data.[, 1] and data.[, 2]
## McNemar's chi-squared = 4.9231, df = 1, p-value = 0.0265
```

Для слабости проверяем гипотезу о том, что количество ухудшившихся (2) равно количеству улучшившихся (11) на второй день. По критерию Мак-Немара: Значимость $p\text{-value} = 0.02 < 0.05$, поэтому можно говорить о значимом улучшении на второй день по признаку слабости.

```
data.<-na.omit(data.[,seq(3)])-1
colnames(data..)
```

```
## [1] "weakness.1" "weakness.2" "weakness.3"
```

```
cochrans.q(data.., alpha = 0.05)
```

```
##
## Cochran's Q Test
##
## H0: There is no difference in the effectiveness of treatments.
## HA: There is a difference in the effectiveness of treatments.
##
## Q = 11.4545454545455
##
## Degrees of Freedom = 2
##
## Significance Level = 0.05
## The p-value is 0.00325594498568582
## There is enough evidence to conclude that the effectiveness of at least two treatments differ.
##
```

По критерию Кохрена $p\text{-value} = 0.003 < 0.05$, улучшение по признаку слабости на третий день существенно.

```
Name<-"insomia"
A<-sapply(colnames(data),function(x)substr(x,1,nchar(Name)))
data.<-apply(data[,names(A[A==Name])],2,function(x)ifelse(x<1,0,1))
tab<-table(data.[,1],data.[,2])
tab
```

```
##
##      0  1
##    0  6  1
##    1 12 15
```

```
mcnemar.test(data.[,1],data.[,2])
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  data.[, 1] and data.[, 2]
## McNemar's chi-squared = 7.6923, df = 1, p-value = 0.005546
```

Для бессоницы проверяем гипотезу о том, что количество ухудшившихся (1) равно количеству улучшившихся (12) на второй день. По критерию Мак-Немара: Значимость $p\text{-value} = 0.005 < 0.05$, поэтому можно говорить о существенном улучшении.

```
mcnemar.test(data.[,1],data.[,2])
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  data.[, 1] and data.[, 2]
## McNemar's chi-squared = 7.6923, df = 1, p-value = 0.005546
```

```
data.<-na.omit(data.[,seq(3)])-1
colnames(data..)
```

```
## [1] "insomia.1" "insomia.2" "insomia.3"
```

```
cochrans.q(data.., alpha = 0.05)
```

```
##
##  Cochran's Q Test
##
## H0: There is no difference in the effectiveness of treatments.
## HA: There is a difference in the effectiveness of treatments.
##
## Q = -2
##
## Degrees of Freedom = 2
##
## Significance Level = 0.05
## The p-value is 1
##
##
```

По критерию Кохрена $p\text{-value} = 1 > 0.05$, улучшение по признаку бессоница на третий день не значимо.