Непараметрические методы анализа качественных признаков для повторных наблюдений

1 Критерий McNemar

Рассмотрим данные такого рода: X_1 и X_2— наличие в ЭКГ эпизода ускоренного идиовентрикулярного ритма (УИР) до и после операции АКШ. У 57 больных этот эпизод не наблюдался ни до, ни после операции; у восьми больных он был только до операции, у двух больных появился после. Не было ни одного больного, у которого эпизод проявлял устойчивость. Проверяется гипотеза о том, что ситуации улучшения или ухудшения равновероятны. Имеется таблица сопряженности в виде матрицы 2×2

$X_1 \setminus X_2$	$-$	$+$	сумма
$-$	a	b	$a+b$
$+$	c	d	$c+d$
сумма	$a+c$	$b+d$	n

$X_1 \setminus X_2$	$-$	$+$	сумма
$-$	57	2	59
$+$	8	0	8
сумма	65	2	67

Нас интересует, насколько значимо различие между частотами b и c. Точная статистика критерия Мак Немара (McNemar's test) вычисляется как

$$\alpha_* = 2 \sum_{i=0}^{\min(b,c)} C_{b+c}^{i} \frac{1}{2^{b+c}} \, .$$

При малых значениях $\alpha_* < \alpha = 0.05$ гипотеза о равенстве $b = c$ отвергается и различие между ними нельзя объяснить случайностью.

В нашем случае $b = 2$, $c = 8$,

$$\alpha_* = 2(C_{10}^0 + C_{10}^1 + C_{10}^2)/2^{10} = 0.109375 > 0.05,$$

поэтому число больных, у которых исчез эпизод УИР после операции, незначимо больше тех, у кого этот эпизод в ЭКГ появился. Если важно, больше ли число улучшившихся, чем ухудшившихся, то разумнее использовать одностороннюю альтернативную гипотезу.

Преобразуем данные из таблицы сопряженности в исходные признаки.

```
> X<-data.frame(TransformData(matrix(c(57,8,2,0),ncol=2)))
> tab<-table(X);tab
```

```
   X2
X1  0  1
  0 57  2
  1  8  0
```

```
> nn<-c(tab[1,2],tab[2,1])
> n<-sum(nn)
> alpha_<-2*pbinom(min(nn),size=n,prob=0.5);alpha_
```

```
[1] 0.109375
```

```
> binom.test(2, n=10, p = 0.5,
+            alternative = c("two.sided", "less", "greater"),
+            conf.level = 0.95)
```

```
        Exact binomial test

data:  2 and 10
number of successes = 2, number of trials = 10,
p-value = 0.1094
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.02521073 0.55609546
sample estimates:
probability of success
                   0.2
```

Этот же алгоритм носит название критерий знаков. Пусть ξ_+, ξ_- случайное число соответственно положительных и отрицательных изменений $n = n_+ + n_-$, $m = min(n_+, n_-)$. Нулевая гипотеза состоит в том, что вероятность успеха ("плюс") p равна 0.5, $H_0 : p = 0.5$. При альтернативной гипотезе $H_1 : p > 0.5$ вычисляется значимость

$$\alpha_* = P(\xi_+ > n_+),$$

при альтернативной гипотезе $H_1 : p < 0.5$ вычисляется значимость

$$\alpha_* = P(\xi_+ < n_+),$$

при альтернативной гипотезе $H_1 : p \neq 0.5$ вычисляется значимость

$$\alpha_* = 2 \cdot P(min(\xi_+, \xi_-) < min(n_+, n_-)).$$

Иногда для критерия знаков используют статистику Фишера.

```
> r<-2;l<-10
> alpha_F<-pf(r/(l-r+1),2*(l-r+1),2*r)+1-pf((l-r)/(r+1),2*(r+1),2*(l-r))
> alpha_F
```

```
[1] 0.06542969
```

```
>
```

Помимо этой статистики используется выражение

$$\chi^2 = \frac{(b-(b+c)/2)^2}{(b+c)/2} + \frac{(c-(b+c)/2)^2}{(b+c)/2} = \frac{(b-c)^2}{b+c},$$

в котором добавляется поправка на непрерывность

$$\chi^2_* = \frac{(|b-c|-1)^2}{b+c}.$$

При справедливости нулевой гипотезы статистика χ^2 имеет распределение хи-квадрат с одной степенью свободы. Доверительный уровень вероятности равен $p = P\{\chi^2 > \chi^2_* = 3.6\} = 0.058$, следовательно, различие между c и b нельзя объяснить случайностью с уровнем значимости, большим 0.058. С поправкой на непрерывность

$$p = P\{\chi^2 > \chi^2_* = 2.5\} = 0.11.$$

```
> mcnemar.test(X[,1],X[,2],correct=FALSE)
```

```
	McNemar's Chi-squared test

data:  X[, 1] and X[, 2]
McNemar's chi-squared = 3.6, df = 1, p-value =
0.05778
```

```
> p.McN.chi.0<-1-pchisq((abs(nn[1]-nn[2]))^2/sum(nn),1);p.McN.chi.0
```

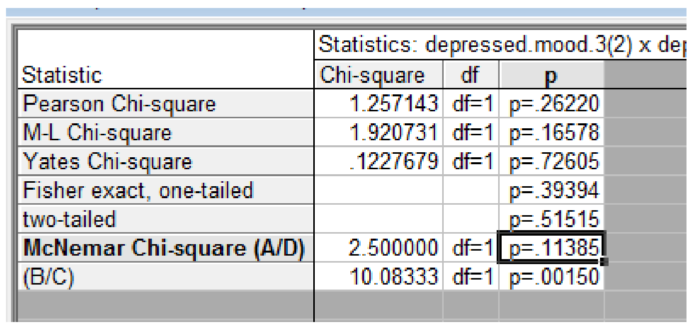
```
[1] 0.05777957
```

```
> mcnemar.test(X[,1],X[,2],correct=TRUE)
```

```
        McNemar's Chi-squared test with continuity
        correction

data:  X[, 1] and X[, 2]
McNemar's chi-squared = 2.5, df = 1, p-value = 0.1138
```

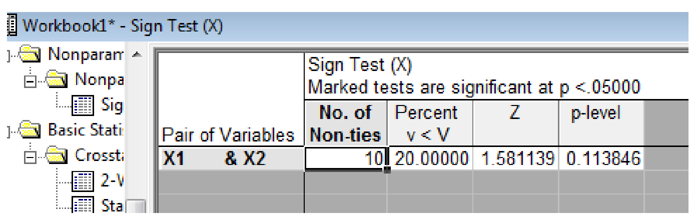
```
> p.McN.chi<-1-pchisq((abs(nn[1]-nn[2])-1)^2/sum(nn),1);p.McN.chi
```

```
[1] 0.1138463
```



Statistic	Statistics: depressed.mood.3(2) x dep		
	Chi-square	df	p
Pearson Chi-square	1.257143	df=1	p=.26220
M-L Chi-square	1.920731	df=1	p=.16578
Yates Chi-square	.1227679	df=1	p=.72605
Fisher exact, one-tailed			p=.39394
two-tailed			p=.51515
McNemar Chi-square (A/D)	2.500000	df=1	p=.11385
(B/C)	10.08333	df=1	p=.00150

Рис. 1: Критерий МакНемара.



Pair of Variables	Sign Test (X) Marked tests are significant at p <.05000			
	No. of Non-ties	Percent v < V	Z	p-level
X1 & X2	10	20.00000	1.581139	0.113846

Рис. 2: Критерий знаков.

1.1 Критерий Кохрена

Для категориальных данных, повторяющихся многократно, используется обобщение критерия Мак Немара в виде критерия Кохрена (Cochren's Q test). Пусть имеется s дихотомических признаков у n индивидов. Для определенности закодируем нулем ответы „нет“ и единицей ответы „да“, количество положительных ответов у i-го индивида обозначим через x_{i*}, в j-й момент через x_{*j}, $N = \sum\limits_{j=1}^{s} x_{*j} = \sum\limits_{i=1}^{n} x_{i*}$.

	X_1	X_2	$\ldots$	X_s	сумма
1	x_{11}	x_{12}	$\ldots$	x_{1s}	x_{1*}
2	x_{21}	x_{22}	$\ldots$	x_{2s}	x_{2*}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
n	x_{n1}	x_{n2}	$\ldots$	x_{ns}	x_{n*}
сумма	x_{*1}	x_{*2}	$\ldots$	x_{*s}	N

При отсутствии изменений в динамике наблюдений статистика

$$T = s(s-1)\frac{\sum\limits_{j=1}^{s}(x_{*j} - \frac{N}{s})^2}{\sum\limits_{i=1}^{n} x_{i*}(s - x_{i*})} \tag{1}$$

имеет распределение хи-квадрат с $s-1$ степенями свободы. Например, исследуется динамика появления эпизода УИР у $n = 52$ больных в $s = 4$ точках: до операции, через две недели после операции, через полгода и через год. Эпизод УИР появлялся у $N = 13$ больных однажды: $x_{j*} \in \{0, 1\}$, до операции у $x_{*1} = 6$ больных, после у $x_{*2} = 2$, затем $x_{*3} = 3$, $x_{*4} = 2$. Значение статистики (1) равно $\chi_* = 3.3$ со значимостью $p = P\{\chi^2 > \chi_*^2\} = 0.35$. Это свидетельствует о том, что снижение числа эпизодов УИР после операции можно объяснить случайностью.

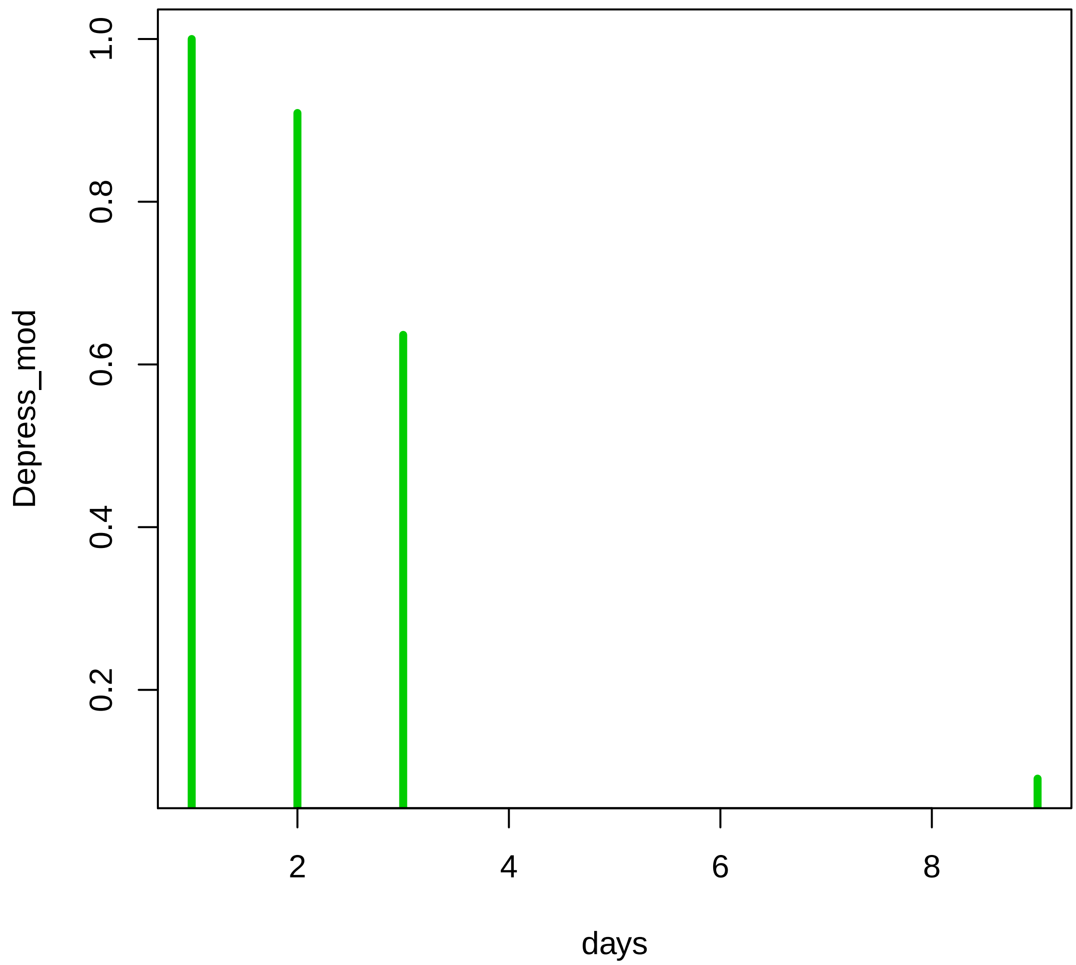
```
> dir<-"~/Documents/share/R/lessions/A_Fitting/days_AAC"
> XX<-data.frame(lapply(c(1,2,3,9),
+       function(x){
+         file<-paste(dir,"/day",x,".csv",sep="")
+         X<-data.frame(read.csv(file,sep=";", dec=",") )
+         colnames(X)<-paste(colnames(X),x,sep=".")
+         X
+         }))
```

```
> data<-XX[,paste(v[1],c(1,2,3,9),sep=".")]
> data<-apply(data,2,function(x)sapply(x,function(y){y[y>1]<-1;y}))
> data<-data.frame(na.omit(data))
> Data<-cbind(stack(data),subject=rep(seq(nrow(data)),ncol(data)))
> library("RVAideMemoire")
>  cch<-cochran.qtest(values~ind|subject, Data,
+             alpha = 0.05, p.method="bonferroni");cch
```

```
        Cochran's Q test

data:  values by ind, block = subject
```

```
Q = 42.8824, df = 3, p-value = 2.607e-09
alternative hypothesis: true difference in probabilities is not equal to 0
sample estimates:
proba in group depressed.mood.1
                     1.00000000
proba in group depressed.mood.2
                     0.90909091
proba in group depressed.mood.3
                     0.63636364
proba in group depressed.mood.9
                     0.09090909

        Pairwise comparisons using Wilcoxon sign test

                 depressed.mood.1 depressed.mood.2
depressed.mood.2        1.000e+00                -
depressed.mood.3        4.688e-02        4.219e-01
depressed.mood.9        1.144e-05        4.578e-05
                 depressed.mood.3
depressed.mood.2                -
depressed.mood.3                -
depressed.mood.9          0.00293

P value adjustment method: bonferroni
```

```
> plot(c(1,2,3,9),as.vector(cch$estimate),type="h",lwd=4,col=3,
+      xlab="days",ylab="Depress_mod")
```




```
> TT<-ncol(data)*(ncol(data)-1)*sum((colSums(data)-sum(sum(data))/ncol(data))^2)
> TT<-TT/sum(rowSums(data)*(ncol(data)-rowSums(data)))
> 1-pchisq(TT,ncol(data)-1)
```

```
[1] 2.606692e-09
```

```
> attach(Data)
> pairwise.wilcox.test(values, ind, p.adjust.method = "none",paired = TRUE)
```

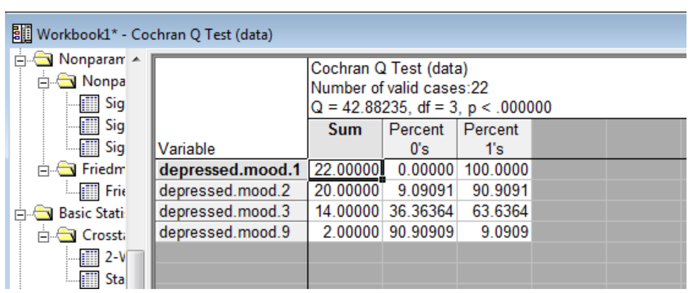
```
        Pairwise comparisons using Wilcoxon signed rank test
```

```
data:  values and ind

                   depressed.mood.1 depressed.mood.2
depressed.mood.2 0.34578          -
depressed.mood.3 0.00596          0.04108
depressed.mood.9 8.6e-06          2.5e-05
                   depressed.mood.3
depressed.mood.2 -
depressed.mood.3 -
depressed.mood.9 0.00063

P value adjustment method: none
```

```
> wilcox.test(data[,3],data[,4],paired=TRUE,exact=FALSE)$p.value
```

```
[1] 0.0006269252
```

```
> detach()
```



Cochran Q Test (data)
Number of valid cases:22
Q = 42.88235, df = 3, p < .000000

Variable	Sum	Percent 0's	Percent 1's
depressed.mood.1	22.00000	0.00000	100.0000
depressed.mood.2	20.00000	9.09091	90.9091
depressed.mood.3	14.00000	36.36364	63.6364
depressed.mood.9	2.00000	90.90909	9.0909

Рис. 3: Критерий Q-Кохрена.