Отчет по домашнему заданию

Фахртдинов Т. А.

25 декабря 2019 г.

Четвертая задача. Статистический анализ категориальных (зависимых) признаков. 3 вариант.

```
SH1 <- c(8.1, 6.2, 6.11, 5.23, 5.2, 8.12, 6.21, 6.21, 5, 6, 8.16, 4.96, 5.36, 5.92, 8.32, 8.08, 6.52, 5.2, 7, 5, 6.3, 6.1, 5.95, 5.2, 8.5, 8.2, 7.3, 6.45, 5.2, 6.2, 4.95, 6.45, 4.95, 6.55, 6.5, 5.9, 6.2, 6.5, 7.7, 7.2, 7, 7.4, 7.8, 7.2, 7)

SH2 <- c(8.9, 8.16, 8.1, 7.44, 7.46, 8.91, 8.1,8.41, 7.44, 8.43, 8.92, 6.48, 7.2, 7.4, 8.9, 8.9, 8.1, 6.88, 8.6, 7.2, 8.4, 8.4, 8.4, 7.4, 8.91, 8.9, 8.8, 8.1, 7.4, 8.4, 7.3, 7.9, 7.36, 8.1, 8, 8.4, 8.2, 8.3, 8, 8.1, 8.2, 8.6, 8.5, 8.6, 8.5)
```

Строим таблицу:

```
tab <- matrix(c(0, 0, 0, 0), nrow = 2)
tab[1, 1] <- length(SH1[SH1 <= median(SH1)])
tab[2, 1] <- length(SH1[SH1 > median(SH1)])
tab[1, 2] <- length(SH2[SH2 <= median(SH1)])
tab[2, 2] <- length(SH2[SH2 > median(SH1)])
```

Получаем:

```
## Sum

## 23 0 23

## 22 45 67

## Sum 45 45 90
```

Нас интересует насколько значимо различие между частотами значения в ячейке [1,2] и [2,1]. H_0 : tab[1,2] = tab[2,1] (Ситуация улучшения и ухудшения после лечения равновероятны) Посчитаем точную статистику критерия Мак Немара:

```
a <- tab[1, 2]
b <- tab[2, 1]
alfa <- 0
for (i in 0:min(a, b))
{
    alfa <- alfa + choose(a + b, i) * (1 / 2^(a + b))
}
alfa <- 2 * alfa
alfa
## [1] 4.768372e-07</pre>
```

alfa < 0.05, отклоняем гипотезу H_0 , о том, что ситуация улучшения и ухудшения после лечения равновероятны.

Проверим результат используя функцию из пакета exact2x2 mcnemar.exact:

```
## Loading required package: exactci
## Loading required package: ssanv
```

```
mcnemar.exact(tab)

##

## Exact McNemar test (with central confidence intervals)

##

## data: tab

## b = 0, c = 22, p-value = 4.768e-07

## alternative hypothesis: true odds ratio is not equal to 1

## 95 percent confidence interval:

## 0.0000000 0.1825538

## sample estimates:

## odds ratio

## 0
```

Значение критерия совпадает с посчитанным нами.