Отчет по домашнему заданию

Фахртдинов Т. А.

19 декабря 2019 г.

Третья задача. Статистический анализ категориальных(независимых) признаков.

Имеется таблица сопряженности факторов "угнетенное депрессивное состояние при алкогольном абстинентном синдроме"и "слабость". Нужно проверить независимость признаков.

	Weakness 1	Weakness 2	Row totals
1	15	11	26
2	0	5	5
Totals	15	16	31

а) Критерий равенства частот:

```
a <- 15
b <- 11
c <- 0
d <- 5
n <- 31
Z <- (b*c - a*d)*sqrt(n) / sqrt((a + b)*(c + d)*(a + c)*(b + d))
Z

## [1] -2.364094
```

Если критерии независимы, то статистика Z будет иметь стандартное нормальное распределение.

```
p_val <- 2 * pnorm(-abs(Z))
p_val

## [1] 0.01807421
```

p-value $= 0.018 < 0.05$, отклоняем гипотезу о независимости.

б) Точный критерий Фишера:

```
chi_t_0 <- (a*d - b*c)^2 * (n) / ((a + b)*(c + d)*(a + c)*(d + b))
p_t <- 0
for (x in (-d+a):(a+c)) {
  b1 <- a + b - x
  c1 <- a + c - x
  d1 <- (c + d) - (a + c - x)
  chi_t <- n*(x * d1 - b1 * c1)^2 / ((x + b1)*(c1 + d1)*(x + c1)*(b1 + d1))
  chi_t
  chi_t - chi_t_0
  if (chi_t >= chi_t_0){
    temp <- factorial(a + b)*
            factorial(a + c)*
            factorial(c + d)*
            factorial(d + b)
    temp <- temp / (factorial(x)*
                    factorial(a + b - x)*
                    factorial(a + c - x)*
                    factorial(d - a + x)*
                    factorial(a + b + c  +d))
    p_t <- p_t + temp
  }
}

p_t

## [1] 0.04338154
```

Получили значение $0.04338 < 0.05$, отклоняем гипотезу о независимости.
Воспользуемся встроенной функцией fisher.test

```
fisher.test(tab)

##
##  Fisher's Exact Test for Count Data
##
## data:  tab
## p-value = 0.04338
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.9885589       Inf
## sample estimates:
## odds ratio
##        Inf
```

Полученное значение критерия совпадает со значением, которое мы получили ранее.

в) Критерий независимости Пирсона:

```
chi <- n * (a * d - b * c)^2 / ((a + b)*(c + d)*(a + c)*(d + b))
p_val_1 <- 1 - pchisq(chi, 1)
p_val_1

## [1] 0.01807421
```

Получаем p-value $= 0.01807421 < 0.05$, отклонем гипотезу о независимости. Значние критерия Пирсона должно совпадать с квадратом критерия равенства частот и действительно:

```
chi - Z^2

## [1] 0
```

Считаем коэффициенты неопределенности:

```
m <- matrix(c(15, 0, 11, 5), nrow = 2)
x <- addmargins(prop.table(m))
H_X <-  -sum(x[1:2, 3] * log2(x[1:2, 3]))
H_Y <- -sum(x[3, 1:2] * log2(x[3, 1:2]))
H_X_Y <- -sum(x[1, 1:2] * log2(x[1, 1:2])) - x[2, 2] * log2(x[2, 2])
I <- H_X + H_Y - H_X_Y
J_X <- I / H_Y * 100
J_Y <- I / H_X * 100
J <- 2 * I / (H_X + H_Y) * 100
```

Односторонние коэффициенты неопределенности:

```
c(J_X, J_Y)

## [1] 17.50476 27.44267
```

Симметричный коэффициент:

```
J

## [1] 21.37507
```