Homework 7

Pascal Pilz, k12111234

2023-11-15

Exercise 16

Assumption

```
n <- 21
k <- 18
p <- k/n

cat(sprintf("n*p*(1-p) = %.2f", n*p*(1-p)))
```

```
## n*p*(1-p) = 2.57
```

Since $n \cdot p \cdot (1 - p) < 9$ we cannot use the asymptotic test. Therefore, we use the exact binomial test.

Exact Binomial Test

```
e <- 0.6 # Expected cure rate

p1 <- 0
for (i in 0:k) {
  p1 <- p1 + choose(n, i) * e^i * (1-e)^(n-i)
}
cat(sprintf("p1: %.3f", p1), "\n")
```

```
## p1: 0.998
```

```
p2 <- 0
for (i in k:n) {
  p2 <- p2 + choose(n, i) * e^i * (1-e)^(n-i)
}
cat(sprintf("p2: %.3f", p2), "\n")
```

```
## p2: 0.011
```

```
p_val <- 2*min(p1, p2)
a <- 0.05 # Type 1 error
cat(sprintf("The null hypothesis %s be rejected. p.value %.3f",
            if (p_val < a) {"can"} else {"can not"}, p_val))
```

```
## The null hypothesis can be rejected. p.value 0.022
```

Exercise 17

Since our data is ordinal we can perform a Mann-Whitney-U test to check if the two groups differ significantly.

```
A <- c(5,7,6,4,6,0)
B <- c(3,10,4,5,6,0)

library(coin)
```

```
## Loading required package: survival
```

```
data<-c(A,B)
group<-as.factor(c(rep("A",length(A)),rep("B",length(B))))
wilcox_test(data~group,distribution = "exact")
```

```
##
##  Exact Wilcoxon-Mann-Whitney Test
##
## data:  data by group (A, B)
## Z = 0.40531, p-value = 0.7359
## alternative hypothesis: true mu is not equal to 0
```

And so we can see that the null hypothesis cannot be reject, the data are likely from the same distribution.

Exercise 18

Since we have $n = 300 > 40$ and $e_{ij} \geq 5$ we use a Chi-square test.

Expected data

```r
A <- c(70,7,13,10)
B <- c(150,20,20,10)

A_p <- A / sum(A) * 100
A_e <- c(A+B) * sum(A)/(sum(A)+sum(B))
A_e_p <- A_e / sum(A_e) * 100

B_p <- B / sum(B) * 100
B_e <- c(A+B) * sum(B)/(sum(A)+sum(B))
B_e_p <- B_e / sum(B_e) * 100

cat(" Expected A:          ", sprintf("%6.2f", A_e), "\n",
    "Observed A:          ", sprintf("%6.2f", A), "\n",
    "Expected A percent: ", sprintf("%5.1f%%", B_e_p), "\n",
    "Observed A percent: ", sprintf("%5.f%%", A_p), "\n",
    "\n",
    "Expected B:          ", sprintf("%6.2f", B_e), "\n",
    "Observed B:          ", sprintf("%6.2f", B), "\n",
    "Expected B percent: ", sprintf("%5.1f%%", B_e_p), "\n",
    "Observed B percent: ", sprintf("%5.f%%", B_p), "\n")
```

```
##  Expected A:           73.33   9.00  11.00   6.67
##  Observed A:           70.00   7.00  13.00  10.00
##  Expected A percent:   73.3%   9.0%  11.0%   6.7%
##  Observed A percent:     70%     7%    13%    10%
##
##  Expected B:          146.67  18.00  22.00  13.33
##  Observed B:          150.00  20.00  20.00  10.00
##  Expected B percent:   73.3%   9.0%  11.0%   6.7%
##  Observed B percent:     75%    10%    10%     5%
```

Measure of divergence and test

```r
# Type 1 error
a <- 0.05

test_statistic <- sum( ( c(A, B) - c(A_e, B_e) )^2 / c(A_e, B_e) )
critical_value <- qchisq(1-a, 1)

cat(sprintf("The null hypothesis %s be rejected.",
            if (test_statistic > critical_value) {"can"} else {"can not"}))
```

```
## The null hypothesis can be rejected.
```