SAINT PETERSBURG STATE UNIVERSITY

Faculty of Applied Mathematics and Control Processes

Mathematical Game Theory and Statistical Decisions Department

Applied Statistics in R

Laboratory work № 3

Professor: Parilina Elena M.

Student: Orlov Ivan M., 19.M09-пу

Saint Petersburg

2019

1. The distributions of the weights of coins in different packages were checked for sameness using a non-parametric Wilcoxon test:

```
euroweight.dat <- read.delim("~/00_mag/ASR/2/Datasets/euroweight.dat.txt", header = FALSE)
library("car")

euro = subset(euroweight.dat, select = c(-1))
names(euro) = c("Weight", "Batch")

batches = list()

for (i in unique(euro$Batch)) {
  batches[i] = subset(euro[euro$Batch == i, ])
}

test_dat = function(dat, test_func) {
  vec = c()
  for (d1 in dat)
    for (d2 in dat) {
      vec = append(vec, test_func(d1, d2)$p.value > 0.05)
    }
  return(vec)
}

batches_distr = test_dat(batches, wilcox.test)
print(matrix(batches_distr, nrow = length(batches)))
```

The results are:

```
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]
[1,]  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE
[2,]  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE
[3,] FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE
[4,] FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
[5,] FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
[6,]  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE
[7,]  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE
[8,]  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE
```

Out of 64 comparisons 38 returned false and (excluding 8 “batch i to batch i” tests) only 18 returned true, which means that most of the batches do not have the same distribution.

2. The hypotheses that the correlation between “sepal length” and “sepal width” and “petal length” and “petal width” inside each class of iris and is significant.

```
iris <- read.csv("~/00_mag/ASR/2/Datasets/iris.txt", header=FALSE)
library("car")
names(iris) = c("Sepal_len", "Sepal_wid", "Petal_len", "Petal_wid", "Class")

iris_by_class = split(iris, iris$Class)

research_cor = function(name_x, name_y, data) {
  print(paste(name_x, name_y, sep = "-"))
  print(cor.test(data[[name_x]], data[[name_y]], method = "kendall"))
  print(cor.test(data[[name_x]], data[[name_y]], method = "spearman"))
}

to_check = names(iris_by_class[[1]][-5])

for (ir in iris_by_class) {
  print(unique(ir$Class))
  research_cor(to_check[1], to_check[2], ir)
  research_cor(to_check[3], to_check[4], ir)
}
```

Results (p-values) are the following, with highlighed false H_0.:

Var	Test	Iris-setosa	Iris-versicolor	Iris-virginica
Petal_len - Petal_wid	Kendell	5.437e-09	0.0001142	0.003222
	Spearman	7.162e-11	0.0001184	0.002011
Sepal_len - Sepal_wid	Kendell	0.1198	1.286e-09	0.009399
	Spearman	0.1124	1.229e-11	0.009591

3. The hypotheses that the correlation between “nicotine” and “weight” and “nicotine” and “carbon monoxide” is significant.

```
cigarettes.dat <- read.delim("~/00_mag/ASR/3/Datasets/cigarettes.dat.txt", header=FALSE)

cigarettes = subset(cigarettes.dat, select = c(-1,-2))

names(cigarettes) = c("Nicotine", "Weight", "CO")

research_cor = function(name_x, name_y, x, y) {
  print(paste(name_x, name_y, sep = "-"))
  print(cor.test(x, y, method = "kendall"))
  print(cor.test(x, y, method = "spearman"))
}

for (var in names(cigarettes)[-1])
  research_cor("Nicotine", var, cigarettes$Nicotine, cigarettes[[var]])
```

Results are the following, with highlighed false H_0.:

Var	Test	P-value
Nicotine-Weight	Kendell	0.1289
	Spearman	0.166
Nicotine-CO	Kendell	7.725e-07
	Spearman	6.068e-09

Correlation between nicotine in cigarette and weight of it is significant while correlation between nicotine and carbon monoxide in a cigarette is insignificant.

4. Suppose in a coin tossing, the chance to get a head or tail is 50 %. In a real case, we have 100 coin tossings, and get 48 heads, is our original hypothesis true?

```
binom.test(x = 48, n = 100, p = 0.5, alternative = "less")
---
data:  48 and 100
number of successes = 48, number of trials = 100, p-value = 0.3822
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.5667681
sample estimates:
probability of success
                  0.48
```

No proofs of H_0 being false were found with alternative hypothesis “the coin was fixed in a way that makes it less likely to land a head”.

5. Did a fair coin produce 8 heads in 10 flips? By “fair” we mean the coin with equal probabilities of appearance of both sides.

```
binom.test(x = 8, n = 10, p = 0.5, alternative = "greater")
---
data:  8 and 10
number of successes = 8, number of trials = 10, p-value = 0.05469
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.4930987 1.0000000
sample estimates:
probability of success
                   0.8
```

No proofs of H_0 being false were found with alternative hypothesis “the coin was fixed in a way that makes it more likely to land a head”, even though it was really close.