■$global_settings, echo = FALSE, warning = FALSE, include = FALSE >>= library(knitr) library(car) library(ggplot2$

$- function(x)x < -as.character(match.call()[[2]])if(!require(x, character.only = TRUE))install.packages(pkgs = x, repos$

$opts_chunk$ set(fig.path='figure/', fig.align='center', fig.pos='H',fig.width=5, fig.height=4) $opts_knit$ set(root.dir

= getwd())

Statistical packages - report 2

Urszula Grochocińska, Marcin Mazurkiewicz

January 6, 2019

Contents

1 Introduction

This report concerns checking the hypothesis using theree tests: Student t-test, Welch t-test and Wilcoxon test. In our case the hypothesis hypothesis is two-sided (we are testing if the mean is equal).

1.1 Student t-test

In our case the test is used for testing hypothesis about mean with equal variances in the samples and with normal distribution. The test is based on t distribution and the significance level. The t distribution is specified by number of freedom degrees. With increasing number of freedom degrees this distribution converge to normal distribution.

     The t-test has asumption: independent observations, normal distribution of samples and equal variances for samples. The last one could be drop if the sample size is large enough. In the test we calculate the statistics t:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \tag{1}$$

where $\overline{x}_1$ and $\overline{x}_2$ are means, n_1 and n_2 are sample sizes and s^2 is calculated by formula:

$$s^2 = \frac{\sum_{i=1}^{n_1}(x_i - \overline{x}_1)^2 + \sum_{j=1}^{n_2}(x_j - \overline{x}_2)^2}{n_1 + n_2 - 2}. \tag{2}$$

Because of the fact that we have two-sided test we reject the null hypothesis in favor of the alternative if the t statistics is not in the interval $[t_{\alpha/2}, t_{1-\alpha/2}]$. The interval is based on t distribution with $n_1 + n_2 - 2$ number of freedom degrees.

1.2 Welch t-test

This test is an extension of Student t-test. Because of the fact that Student t-test has assumption about equal variances and normality there appear a need to construct a test to examinate a case with unequal variances.

Such test is for example Welch t-test. There is used a statistic t wchih is given by following formula:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \tag{3}$$

where

$$s_1^2 = \frac{\sum\limits_{i=1}^{n_1}(x_i - \bar{x}_1)^2}{n_1 - 1}, \; s_2^2 = \frac{\sum\limits_{i=1}^{n_2}(x_j - \bar{x}_2)^2}{n_2 - 1}. \tag{4}$$

The number of freedon degrees are given by:

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}. \tag{5}$$

The rest part of procedure (rejecting or not the null hypothesis) are like Student t-test.

1.3 Wilcoxon test

The last test taken into consideration is Wilcoxon test. It can be used as a substitution of Student t-test when the distribution of samples is not normal. This test is non-parametric and is used to compare two related samples.

The algorithm consist of few steps. Firstly we calculate the differences between pairs of observations $|x_i - y_i|$. Separately we calculate sign of this differences. Then we rank those differences and denote then R_i. Statistic W^+ is defined as sum of ranks R_i in which the difference $d_i < 0$. After all the statistics is compared with reference table. This comparison results in rejecting or not the hypothesis.

2 Task 1.

In the first task we consider two samples of 200 observations in each which have normal distribution with variance $\sigma = 2$. Firstly we make the function for calculation power function fo Student t-test and wilcoxon test.

After it we use this function and plot power functions for all three test.

```
>   powers_student <- vector(mode="numeric", length=0)
>   powers_welch <- vector(mode="numeric", length=0)
>   powers_wilcoxon <- vector(mode="numeric", length=0)
>   for (i in seq(-2,2,.1)){
+     powers_student = c(powers_student, t.power1(means = c(0, i), sds = c(2,2)))
+     powers_welch = c(powers_welch, t.power1(means = c(0, i), sds = c(2,2), var.equal = FALSE))
+     powers_wilcoxon = c(powers_wilcoxon, wilcoxon.power(means = c(0, i), sds = c(2,2)))
+   }
>   plot(seq(-2,2,.1), powers_student)
>   points(seq(-2,2,.1), powers_welch, col='red')
>   points(seq(-2,2,.1), powers_wilcoxon, col='blue')
>   legend("right", c('Student t-test', 'Welsh test', 'Wilcoxon test'))
```