**Krotov Egor**

**BIG DATA SYSTEMS 181**

Assignment 2 on Descriptive Statistics

*Assignment 2.*

1. *Assignment on Con\_dence Intervals and Hypothesis Testing*

For the assignment i took the iROBOT company stock prices from Yahoo Finance. I selected the open prices of stocks from 1st January of 2010 to present day.

*> library(quantmod)*

*> dataIRBT <- getSymbols("IRBT", src = "yahoo", from ="2010-00-01", auto.assign = FALSE)*

*> data <- dataIRBT$IRBT.Close*

*(a)*

*> t.test(data, conf.level=0.97)*

*One Sample t-test*

*data: data*

*t = 86.598, df = 2202, p-value < 2.2e-16*

*alternative hypothesis: true mean is not equal to 0*

*97 percent confidence interval:*

*38.84455 40.84276*

*sample estimates:*

*mean of x*

*39.84365*

*(b)*

*> summary(data)*

*Index IRBT.Close*

*Min. :2010-01-04 Min. : 14.52*

*1st Qu.:2012-03-10 1st Qu.: 25.91*

*Median :2014-05-20 Median : 33.15*

*Mean :2014-05-18 Mean : 39.84*

*3rd Qu.:2016-07-26 3rd Qu.: 40.82*

*Max. :2018-10-02 Max. :117.71*

*> t.test(data, mu=33.15)*

*One Sample t-test*

*data: data*

*t = 14.548, df = 2202, p-value < 2.2e-16*

*alternative hypothesis: true mean is not equal to 33.15*

*95 percent confidence interval:*

*38.94138 40.74593*

*sample estimates:*

*mean of x*

*39.84365*

*(c)*

*> df=length(data)-1*

*> varIRBT=var(data)*

*> lower=varIRBT\*df/qchisq(0.05/2,df,lower.tail=FALSE)*

*> upper=varIRBT\*df/qchisq(1-0.05/2,df,lower.tail=FALSE)*

*> c(lower=lower,variance=varIRBT,upper=upper)*

*lower variance upper*

*439.9902 466.3561 495.1771*

*> c(lower=sqrt(lower),std.dev=sqrt(varIRBT),upper=sqrt(upper))*

*lower std.dev upper*

*20.97594 21.59528 22.25257*

*(f)* I took epil$trt from MASS

*> head(epil)*

*y trt base age V4 subject period lbase lage*

*1 5 placebo 11 31 0 1 1 -0.7563538 0.11420370*

*2 3 placebo 11 31 0 1 2 -0.7563538 0.11420370*

*3 3 placebo 11 31 0 1 3 -0.7563538 0.11420370*

*4 3 placebo 11 31 1 1 4 -0.7563538 0.11420370*

*5 3 placebo 11 30 0 2 1 -0.7563538 0.08141387*

*6 5 placebo 11 30 0 2 2 -0.7563538 0.08141387*

*> prop.test(112,236,conf.level=0.99)*

*1-sample proportions test with continuity correction*

*data: 112 out of 236, null probability 0.5*

*X-squared = 0.51271, df = 1, p-value = 0.474*

*alternative hypothesis: true p is not equal to 0.5*

*99 percent confidence interval:*

*0.3906510 0.5599267*

*sample estimates:*

*p*

*0.4745763*

*(e)*

*> prop.test(x=112,n=236,p=0.5)*

*1-sample proportions test with continuity correction*

*data: 112 out of 236, null probability 0.5*

*X-squared = 0.51271, df = 1, p-value = 0.474*

*alternative hypothesis: true p is not equal to 0.5*

*95 percent confidence interval:*

*0.4097148 0.5402790*

*sample estimates:*

*p*

*0.4745763*

*(f)* I took ObamaApproval from Using R

*> sum(ObamaApproval$approve[ObamaApproval$year==2013])*

*[1] 3360*

*> sum(ObamaApproval$disapprove[ObamaApproval$year==2013])*

*[1] 3277*

*> sum(ObamaApproval$approve[ObamaApproval$year==2010])*

*[1] 8200*

*> sum(ObamaApproval$disapprove[ObamaApproval$year==2010])*

*[1] 8416*

*> prop.test(x=c(3360,6637), n=c(8200,16616), conf.level=0.99)*

*2-sample test for equality of proportions with continuity correction*

*data: c(3360, 6637) out of c(8200, 16616)*

*X-squared = 2.3889, df = 1, p-value = 0.1222*

*alternative hypothesis: two.sided*

*99 percent confidence interval:*

*-0.006842106 0.027485741*

*sample estimates:*

*prop 1 prop 2*

*0.4097561 0.3994343*

*(g)*

*> prop.test(x=c(3360,6637), n=c(8200,16616), conf.level=0.95, alt="less")*

*2-sample test for equality of proportions with continuity correction*

*data: c(3360, 6637) out of c(8200, 16616)*

*X-squared = 2.3889, df = 1, p-value = 0.9389*

*alternative hypothesis: less*

*95 percent confidence interval:*

*-1.00000000 0.02131514*

*sample estimates:*

*prop 1 prop 2*

*0.4097561 0.3994343*

*Approval (first) vs Disapproval (second) гипотеза о равенстве подтвердилась*