Applied statistics Homework 2

Yapi Donatien Achou

March 21, 2019

Contents

1 Problem 2.1

We will learn about the following new R functions, as well as new functions

```
read.csv()
t.test()
shapiro.test()
read.table()
```

1.1 Part1: t-test

When the test statistic follows the normal distribution (gaussian distribution), a t-test is commonly applied to test the mean of a normally distributed population. This is achieved in R by using the command t.test(). For that you need a sample X, the mean μ of the population from which the sample X is drown, the mean μ_x of the sample, and the significance level α. Assuming a significance level $\alpha = 5\%$, the test is given by

```
t.test(X-μ_X, alternative=two.side, conf.level=0.95)
```

More option are

```
t.test(x,y=NULL,
alternative=c(two.sided,less, greater ),
mu=0, paired=FALSE, var.equal=FALSE,
conf.level=0.95,
formula, data,subset,na.action...
)
```

If the pvalue is greater than the significance level α, we can conclude that the null hypothesis is plausible

1.2 Part 2: Hypotheses test formulation

We would like to test the hypotheses that the average yield of barley is greater than 150

$$\begin{aligned} H_0 &: \mu = 150 \\ H_1 &: \mu > 150 \end{aligned} \tag{1}$$

```
Barley <- read.csv("Barley.csv")

mean <- mean(Barley$barley)
alpha <- 0.1
level <- 1-alpha
test <- t.test(Barley$barley, mu=150, alternative="greater",
           conf.level=level)
>>
>>
        One Sample t-test

data:  Barley$barley
t = 1.3607, df = 399, p-value = 0.08719
alternative hypothesis: true mean is greater than 150
90 percent confidence interval:
 150.1199      Inf
sample estimates:
mean of x
 152.1175
```

The one sided 90% confidence interval suggests that the mean barley yield is likely to be greater than 150.1199. The pvalue of 0.08719 is less than the significance level $\alpha = 0.1$, we therefor reject the null hypothesis that the mean is equal to 150 in favour of the alternative hypothesis

Now if the significance level $\alpha = 0.05$ then the pvalue of 0.08719 is greater than the the the the significance level, than we would not reject H_0 in favour of H_1.

1.3 Part 3

1.4 Note

Reject H_0 if pvalue is less than significance level, and not reject H_0 otherwise. The significance level α is the probability of rejecting the null hypothesis when it is true. The pvale, is the probability of obtaining a result at least as extreme, given that the null hypothesis is true. The result is statistically significant, by the standard of the study when

$$pvalue < \alpha \tag{2}$$

References

[1] Petter J. Brockwell. Richard A. Davis *Introduction to Time Series and Forecasting.* Springer. Second edition. 2001