QMB Exercise 2 - Estimation and Testing

Karin Gryzlak & Dino Nienhold

10 May, 2015

Introduction

The following report is based on the QMB Exercise 2 - Estimation and Testing. The task description pdf file is bis_ex2-EstimationTesting-20150429.pdf

Requirements

Please make sure that you the following packages loaded in your workspace.

```
library("dplyr")
library("ggplot2")
library("ggExtra")
library("gridExtra")
```

Data Set

Please make sure you have the file housingrents.csv in the subdirectoy Data in your workspace.

```
housingrents <- read.csv("./Data/housingrents.csv",sep=";")
```

Data Processing

For analysis purposes it is necessary to convert the rooms and NRE variable to a factor. Furthermore a new variable rps (rent per square meter) is created

```
housingrents <- mutate(housingrents, rooms = factor(rooms),
  nre = factor(nre,levels=c(0,1),labels=c("no","yes")))
housingrents <- mutate(housingrents,rps = rent/area)
```

Task 1

...

Conclusion

...

Task 2

Task 2 checks normality of variable rent per square (rps). Additionally t-tests are conducted

a)

In this sub task the normal distribution of the variable rpf the NRE respectively non-NRE appartments are checked. Create two datasets as a first step:

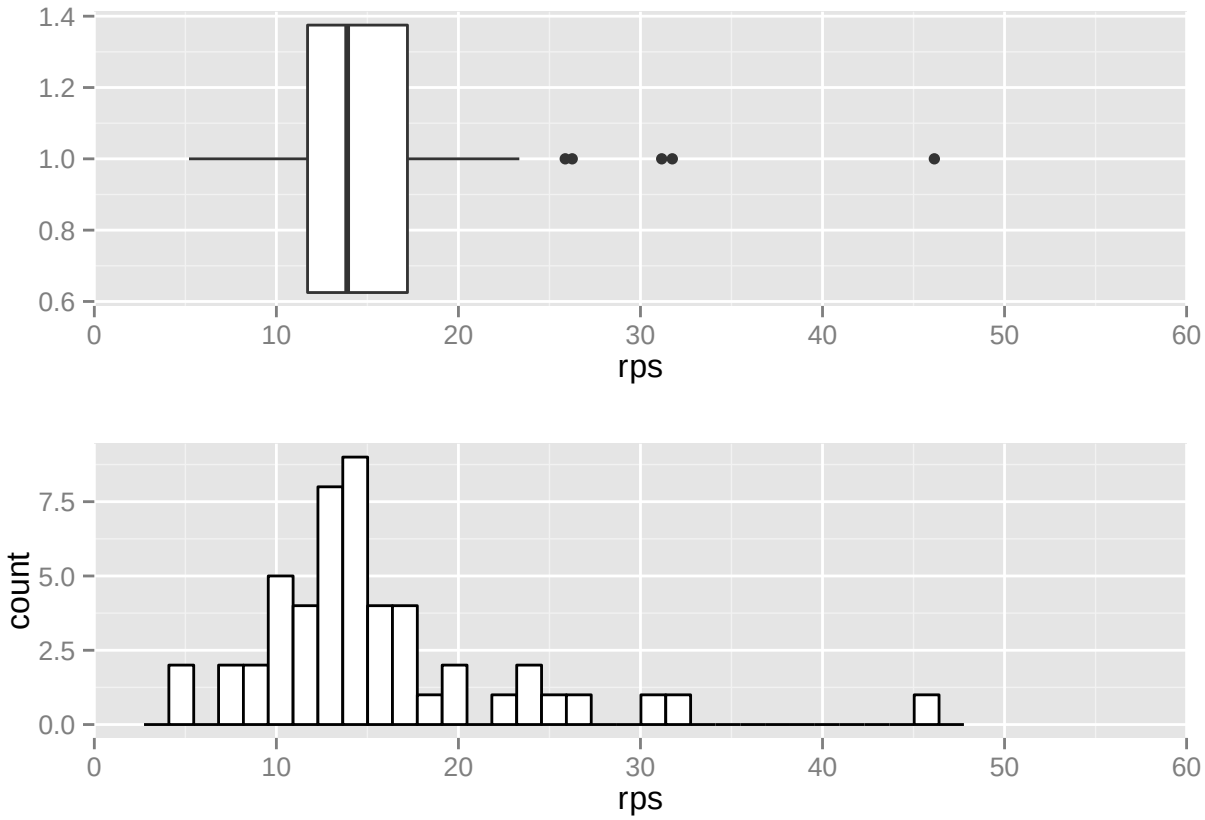
```
nrehousing <- filter(housingrents,nre=="yes")
nonnrehousing <- filter(housingrents,nre=="no")
```

Check normal distribution of rps variable for NRE appartments:

```
p1 <- qplot(x = 1, y = rps, data = nrehousing, xlab = "", geom = 'boxplot') +
        coord_flip(ylim=c(0,60))

p2 <- ggplot(nrehousing, aes(x = rps)) +
        geom_histogram(colour="black", fill="white") +
        coord_cartesian(xlim=c(0,60))

grid.arrange(p1, p2, widths = c(1, 2))
```



```
print(gg_qq(nrehousing$rps))
```

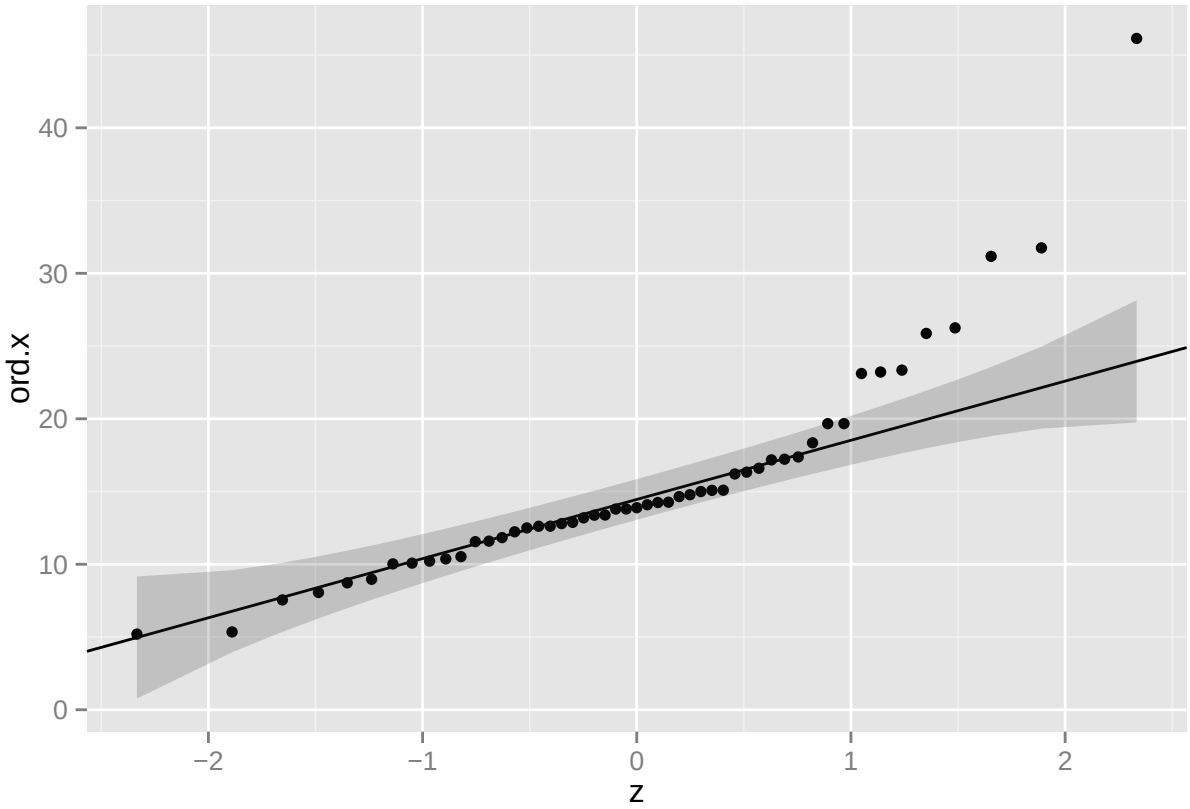


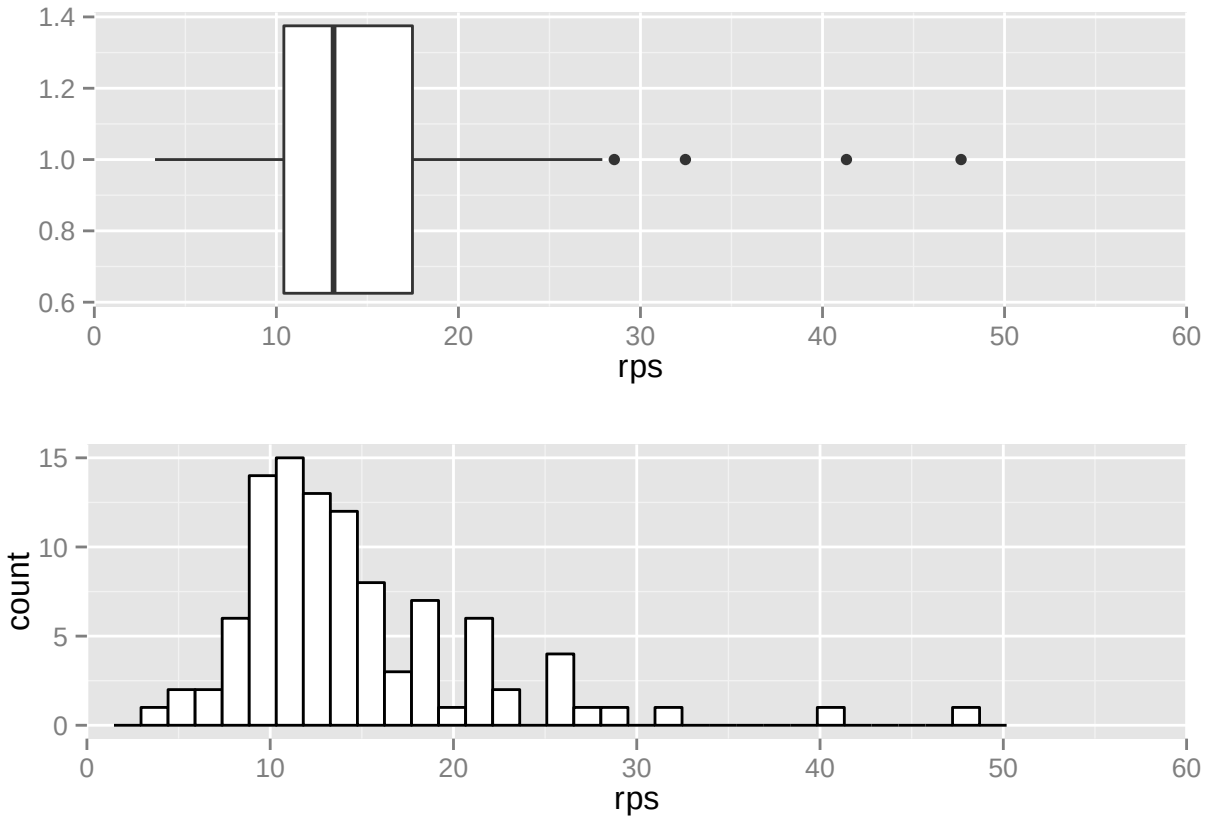
Figure 1:

Check normal distribution of rps variable for non-NRE appartments:

```
p1 <- qplot(x = 1, y = rps, data = nonnrehousing, xlab = "", geom = 'boxplot') +
        coord_flip(ylim=c(0,60))

p2 <- ggplot(nonnrehousing, aes(x = rps)) +
        geom_histogram(colour="black", fill="white") +
        coord_cartesian(xlim=c(0,60))

grid.arrange(p1, p2, widths = c(1, 2))
```



```
print(gg_qq(nonnrehousing$rps))
```

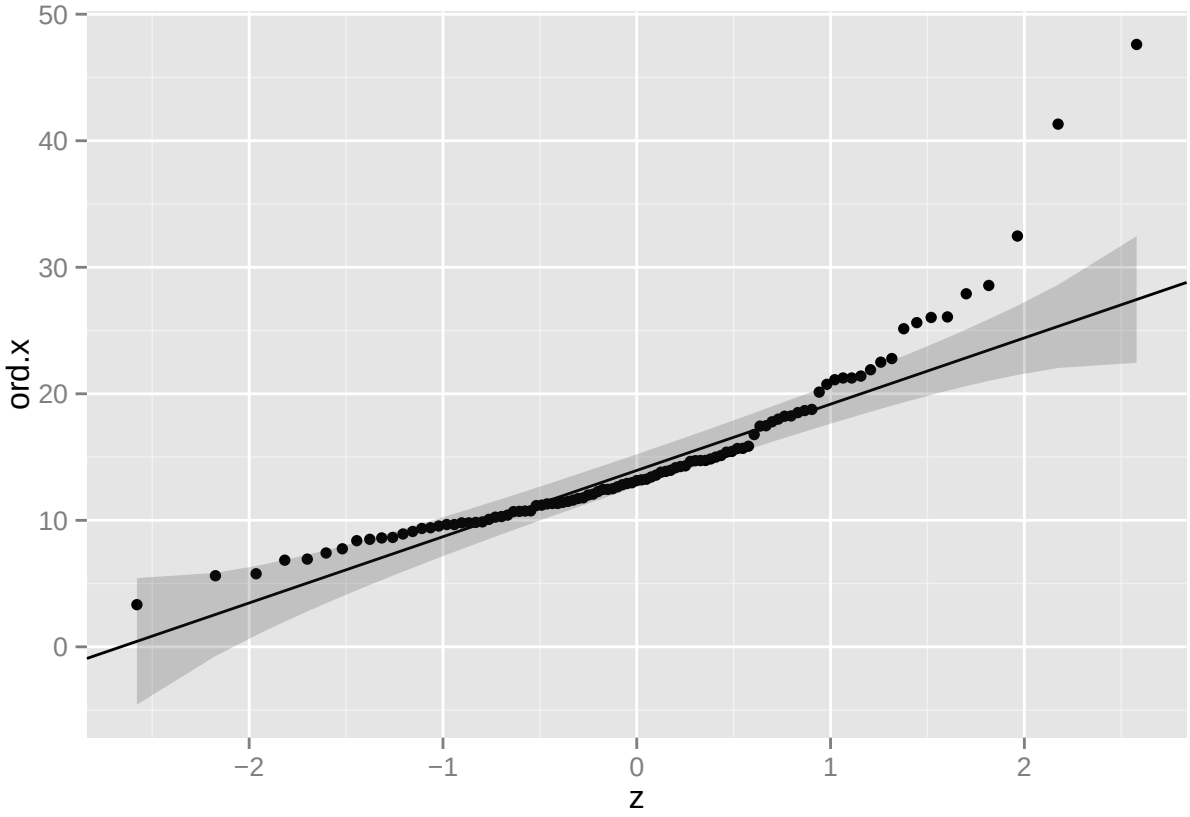


Figure 2:

Conclusion

...

b)

In this section a two-sided Student's t-test is conducted in order to check the following hypothesis:

H_0 = Mean diference is equal 0.

H_a = Mean diference is not equal 0.

```
t.test(housingrents$rps~housingrents$nre,alternative = "two.sided", mu=0, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  housingrents$rps by housingrents$nre
## t = -0.64441, df = 96.874, p-value = 0.5208
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.203661  1.633229
## sample estimates:
##  mean in group no mean in group yes
##          14.77912          15.56434
```

```
t.test(housingrents$rps~housingrents$nre,alternative = "two.sided", mu=0, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  housingrents$rps by housingrents$nre
## t = -0.65318, df = 150, p-value = 0.5146
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.160557  1.590124
## sample estimates:
##  mean in group no mean in group yes
##          14.77912          15.56434
```

c)

In this section a one-sided Student's t-test is conducted in order to check the following hypothesis:

H_0 = Mean diference is equal 0.

H_a = Mean diference is greater 0.

```
t.test(housingrents$rps~housingrents$nre,alternative = "greater", mu=0, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  housingrents$rps by housingrents$nre
## t = -0.64441, df = 96.874, p-value = 0.7396
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
```

```
##  -2.808839       Inf
## sample estimates:
##  mean in group no mean in group yes
##          14.77912          15.56434
```

Task 3

http://ww2.coastal.edu/kingw/statistics/R-tutorials/independ.html

```
housingrentTbl <- xtabs(~rooms+nre, data=housingrents)
housingrentChi <- chisq.test(housingrentTbl,simulate.p.value=TRUE)
housingrentChi
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  housingrentTbl
## X-squared = 26.749, df = NA, p-value = 0.0004998
```

```
housingrentChi$observed
```

```
##      nre
## rooms no yes
##     1 18   5
##     2 23   6
##     3 32   5
##     4 14  18
##     5  9  15
##     6  5   2
```

```
housingrentChi$expected
```

```
##      nre
## rooms        no       yes
##     1 15.282895  7.717105
##     2 19.269737  9.730263
##     3 24.585526 12.414474
##     4 21.263158 10.736842
##     5 15.947368  8.052632
##     6  4.651316  2.348684
```

```
housingrentChi$resid
```

```
##      nre
## rooms         no        yes
##     1  0.6950302 -0.9780910
##     2  0.8497704 -1.1958513
##     3  1.4953421 -2.1043413
##     4 -1.5751126  2.2165994
##     5 -1.7397058  2.4482255
##     6  0.1616756 -0.2275202
```

```r
resid(housingrentChi)
```

```
##      nre
## rooms         no        yes
##     1  0.6950302 -0.9780910
##     2  0.8497704 -1.1958513
##     3  1.4953421 -2.1043413
##     4 -1.5751126  2.2165994
##     5 -1.7397058  2.4482255
##     6  0.1616756 -0.2275202
```

Conclusion

...