NMAI061-22-EX3

Matej Nemec

Hypothesis testing

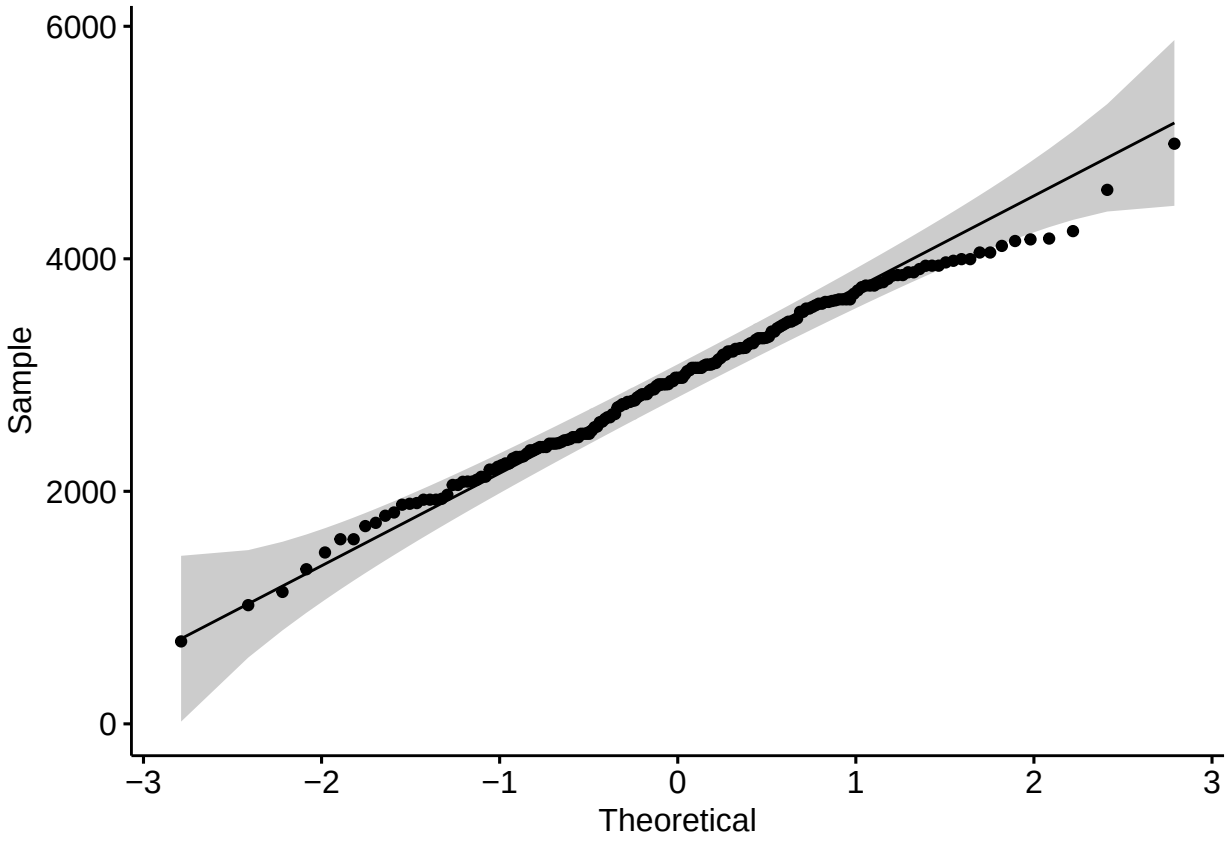
1)Load and check data

Our data come from publicly available dataset on risk factors associated with low infant birth weight. We want to know if there is a statistically significant relationship between smoking during pregnancy and reduced infant weight. (We suspect there is.) We should check the data for normality first lets examine the correlation of data distribution with corresponding normal distribution (parameters estimated from sample).

```
library(MASS)
library(ggpubr)
```

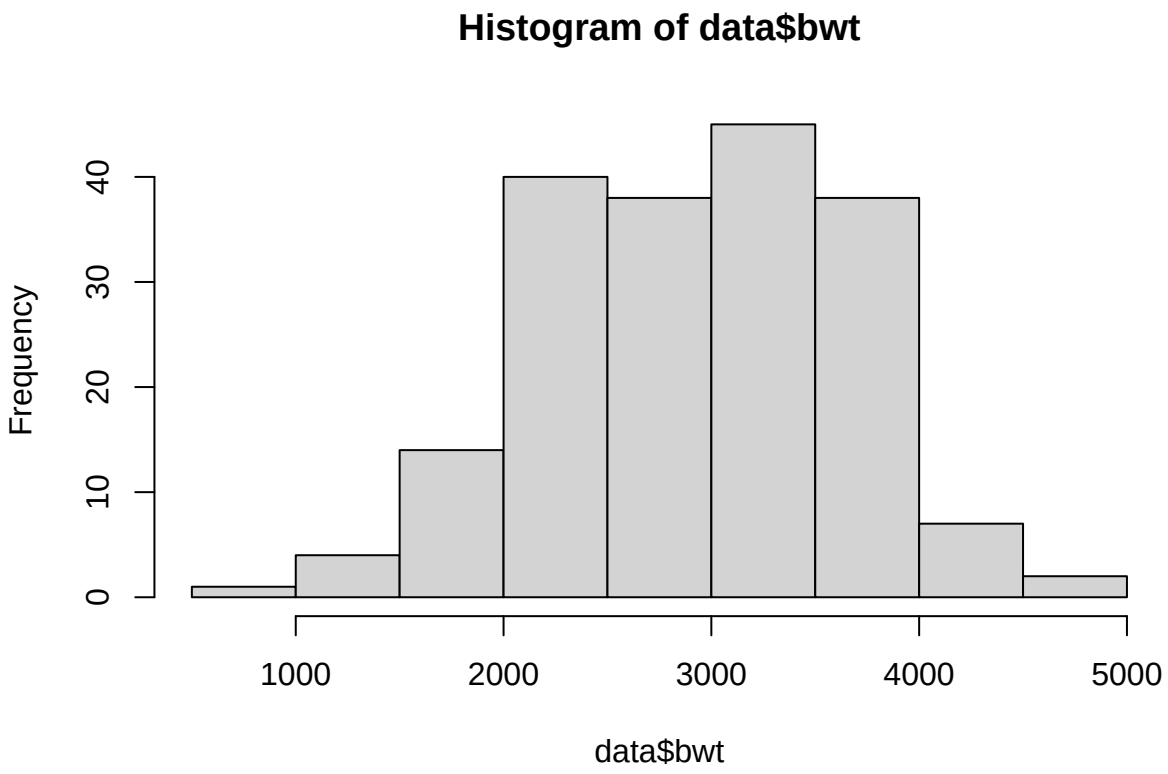
```
## Loading required package: ggplot2
```

```
data=birthwt
ggqqplot(birthwt$bwt)
```



Data distribution looks reasonably well correlated with normal distribution we can visualize the histogram.

```
hist(data$bwt)
```



Also looks normal enough. To be sure let's run Shapiro-Wilk normality test

```
shapiro.test(birthwt$bwt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  birthwt$bwt
## W = 0.99244, p-value = 0.4353
```

P-value of 0.4353 is large enough for us to say that any difference between the distribution of birth weights and normal distribution is by chance. Also there are no extreme outliers so we can run t-test.

3)Run t-test

```
t.test(bwt ~ smoke, data = data)
```

```
##
##  Welch Two Sample t-test
##
## data:  bwt by smoke
## t = 2.7299, df = 170.1, p-value = 0.007003
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
```

```
##   78.57486 488.97860
## sample estimates:
## mean in group 0 mean in group 1
##        3055.696        2771.919
```

We can see that difference in means for the two samples is very significant(p-value of ~0.007 and 95% CI: ~(79,489) and unlikely to be caused by chance. We can therefore claim with reasonable confidence that smoking during pregnancy negatively affects birth weight.