MVE055 2017 Lecture 14

Marco Longfils

Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg

Wednesday 11th October, 2017

The problem

- The Body Mass Index (BMI) is a commonly used measure of malnutrition/obesity.
- BMI $= \frac{\text{weight in Kg}}{\text{height in meters}^2}$;
- Weight statuses vs BMI:

BMI	Status
0-18.5	Underweight
18.5-24.9	Normal weight
25-29.9	Overweight
$\geq$30	Obese

- Many health risks (diabetes, heart diseases,...) are different for these categories.
- Before starting, we can see that BMI might be an incorrect measure, as it does not take into account mass composition (fat vs. muscles.)

Questions

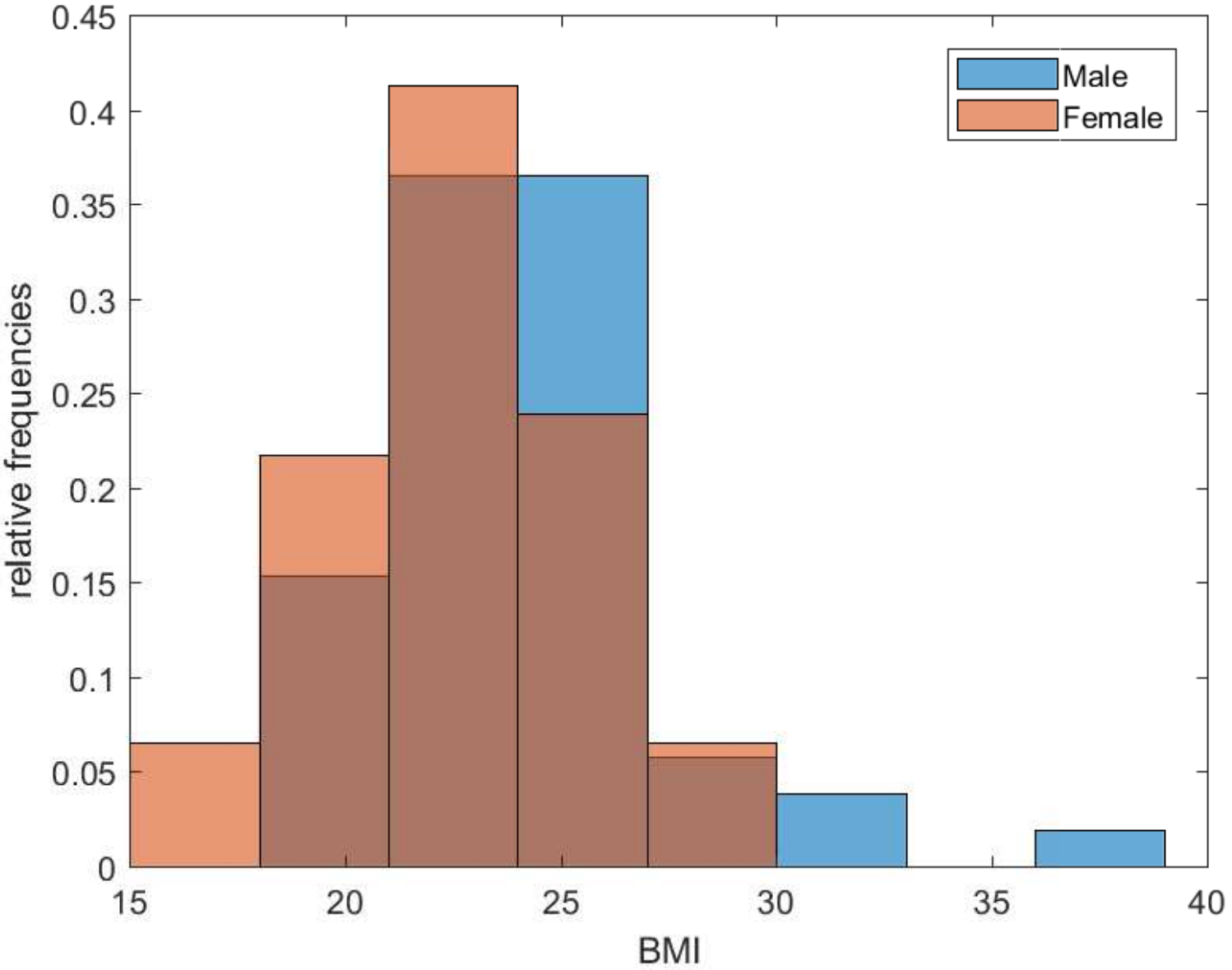
We would like to investigate/asnwer the following questions:

- Is there a difference in the distribution of the BMI of men and women?
- It is expected that changes in the metabolism and lifestyle cause on average an increase of the BMI.

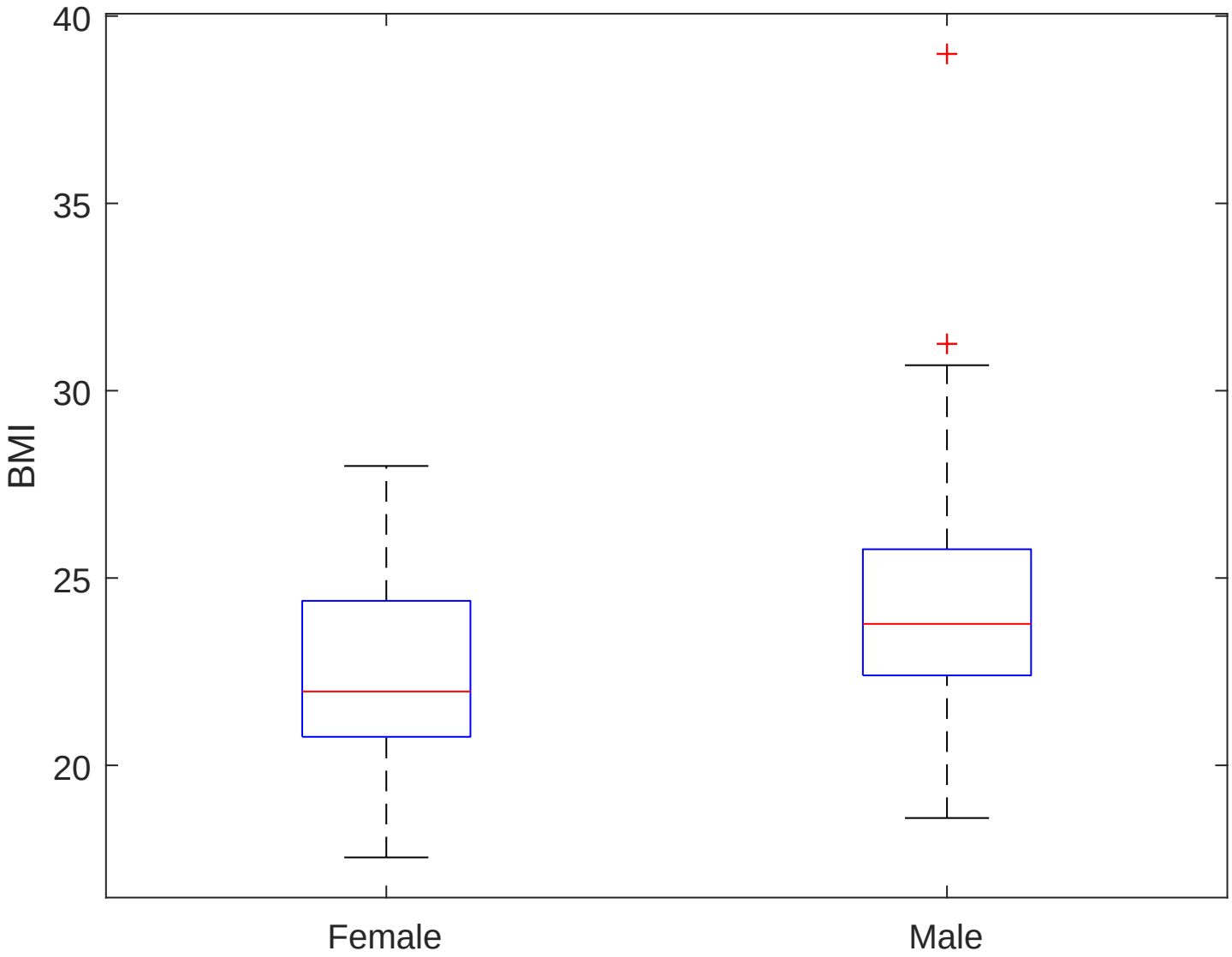
Data collection

- 100 persons have filled in an anonymous internet-based survey with questions about: height, weight, gender, age, health esteem, number of exercise and sitting hours, number of cigarettes per day;
- Missing/erroneous entries are presents: e.g. an age value 0f 186 years and gender not specified. As they are few compared to the sample size, we can simply ignore. (no trend in missingness can be seen)

Data visualization



Boxplot

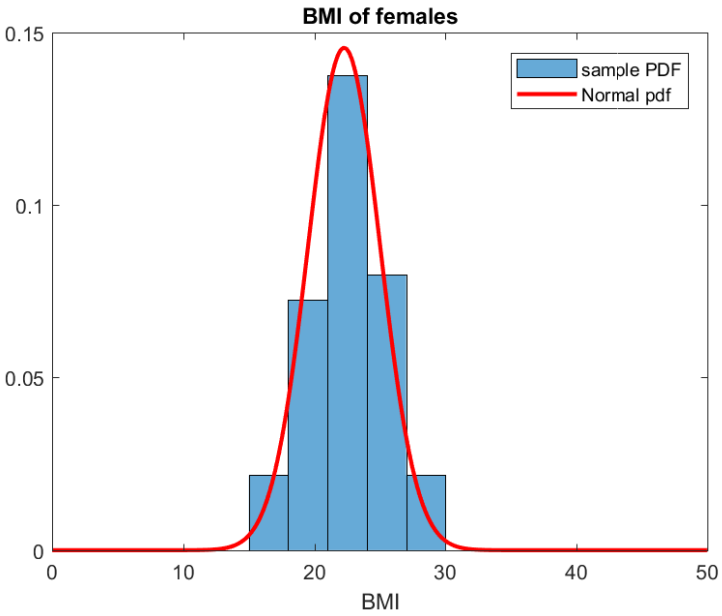


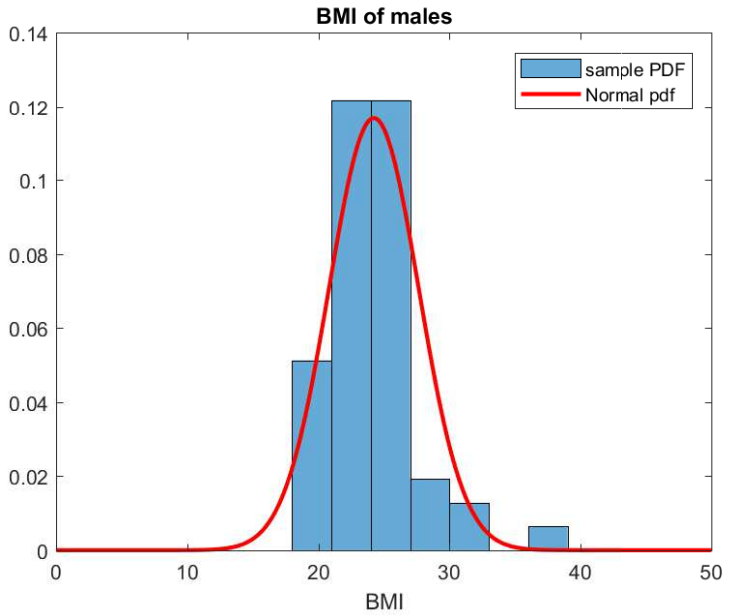
Comparing the mean BMI of men vs. women

To compare the mean of two populations we would need to use either:

1. two sample (unpaired) T test, with the following assumptions:
   - Normality;
   - equal/inequal variances;
2. non-parametric test (Wilcoxon: next lecture).

Normality


BMI of females
sample PDF
Normal pdf
0.15
0.1
0.05
0
0
10
20
30
40
50
BMI


BMI of males
sample PDF
Normal pdf
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
10
20
30
40
50
BMI

Variances

- We perform an F test to check if the variances are equal.
- p-value: 0.133 $\rightarrow$ we assume equal variances
- **The F test is higly sensitive to deviation from normality, thus I really should not have used it in this case!!!**
- We perform then a T test with equal variances and obtain a p-value of 0.0025;
- Non-parametric tests are the most suitable method to use. In this case a Wilcoxon test gives a p-value of 0.0036.

BMI vs. Age

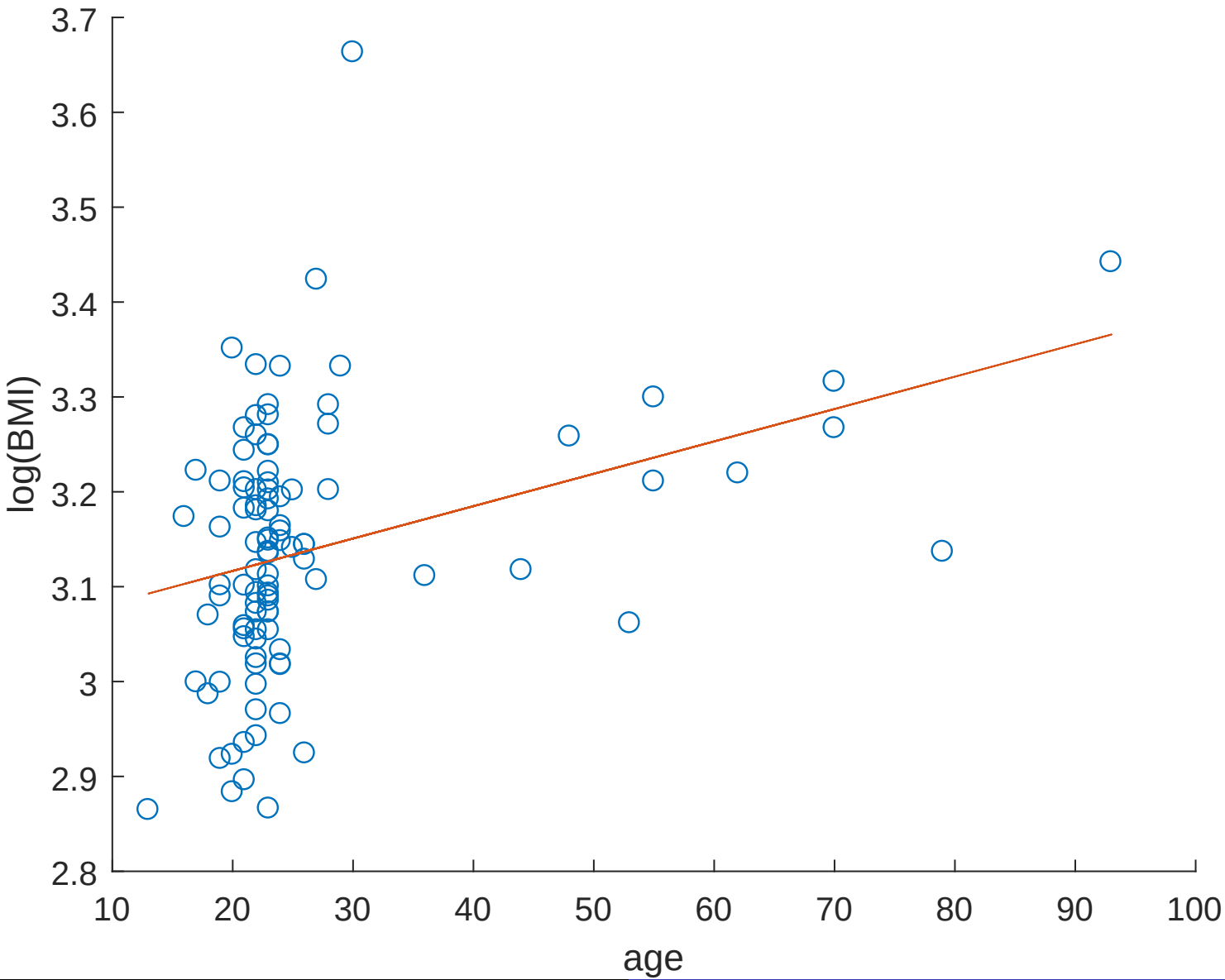
- We have noticed that the distribution of BMI is not normal;
- Consider log BMI

Estimated Coefficients:

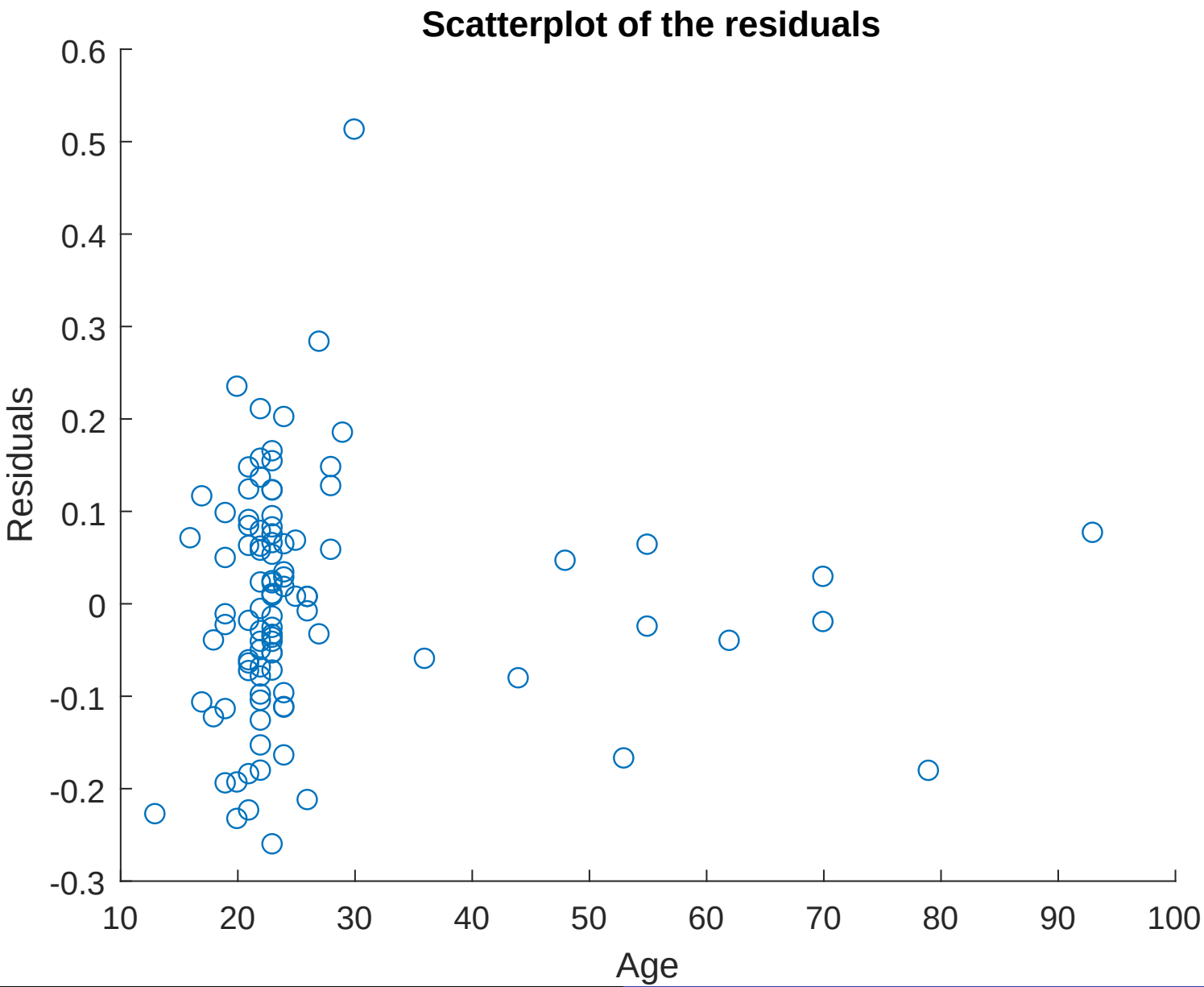
	Estimate	SE	tStat	pValue
(Intercept)	3.0484	0.028267	107.84	8.1802e-103
x1	0.0034126	0.000949	3.596	0.00051037

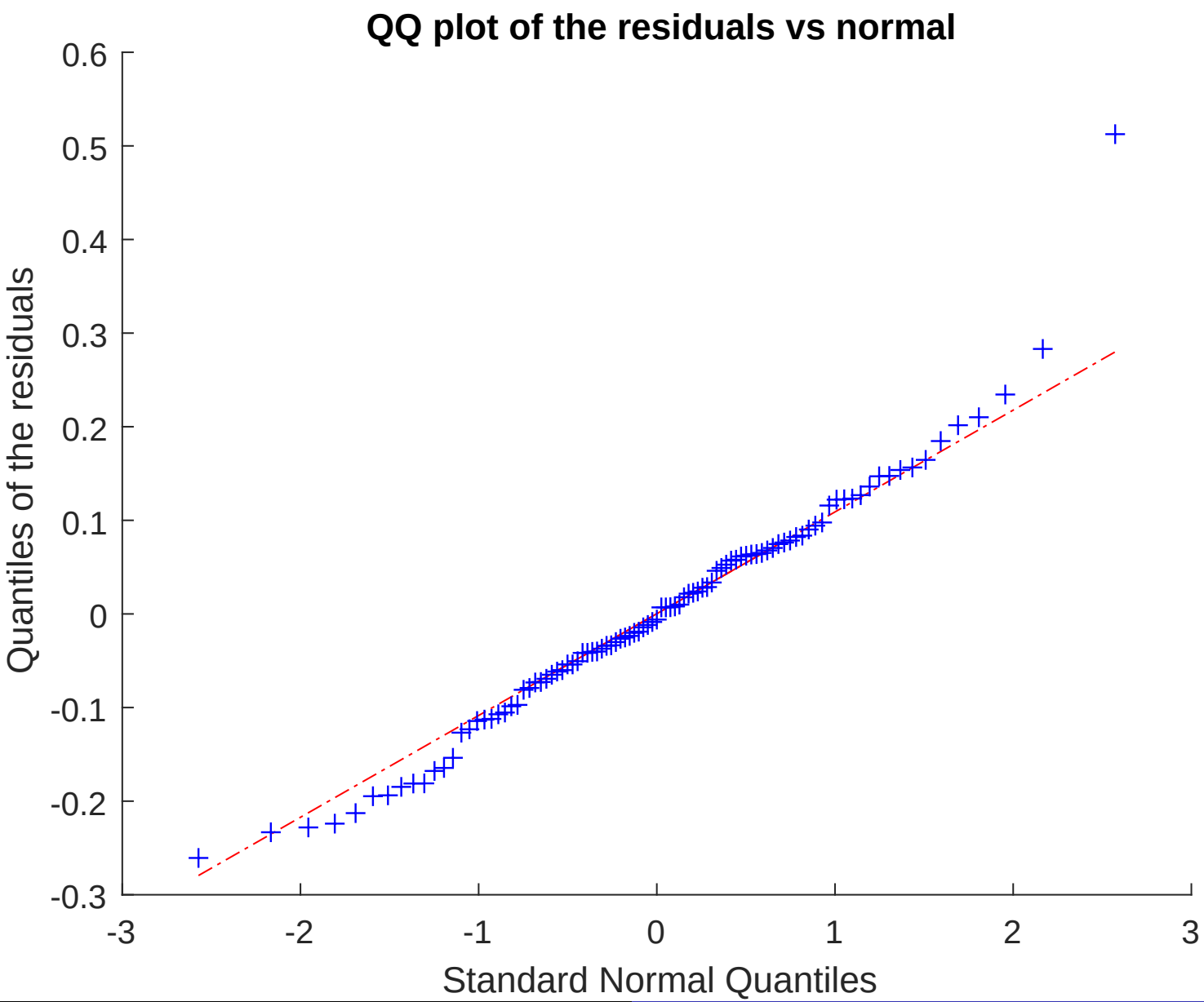
Number of observations: 99, Error degrees of freedom: 97
Root Mean Squared Error: 0.126
R-squared: 0.118, Adjusted R-Squared 0.109
F-statistic vs. constant model: 12.9, p-value = 0.00051

diagnostic



diagnostic


Scatterplot of the residuals
Residuals
Age
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
10
20
30
40
50
60
70
80
90
100


QQ plot of the residuals vs normal
Quantiles of the residuals
Standard Normal Quantiles
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-3
-2
-1
0
1
2
3