

# HW4 - Lucas Fellmeth, Sven Bergmann

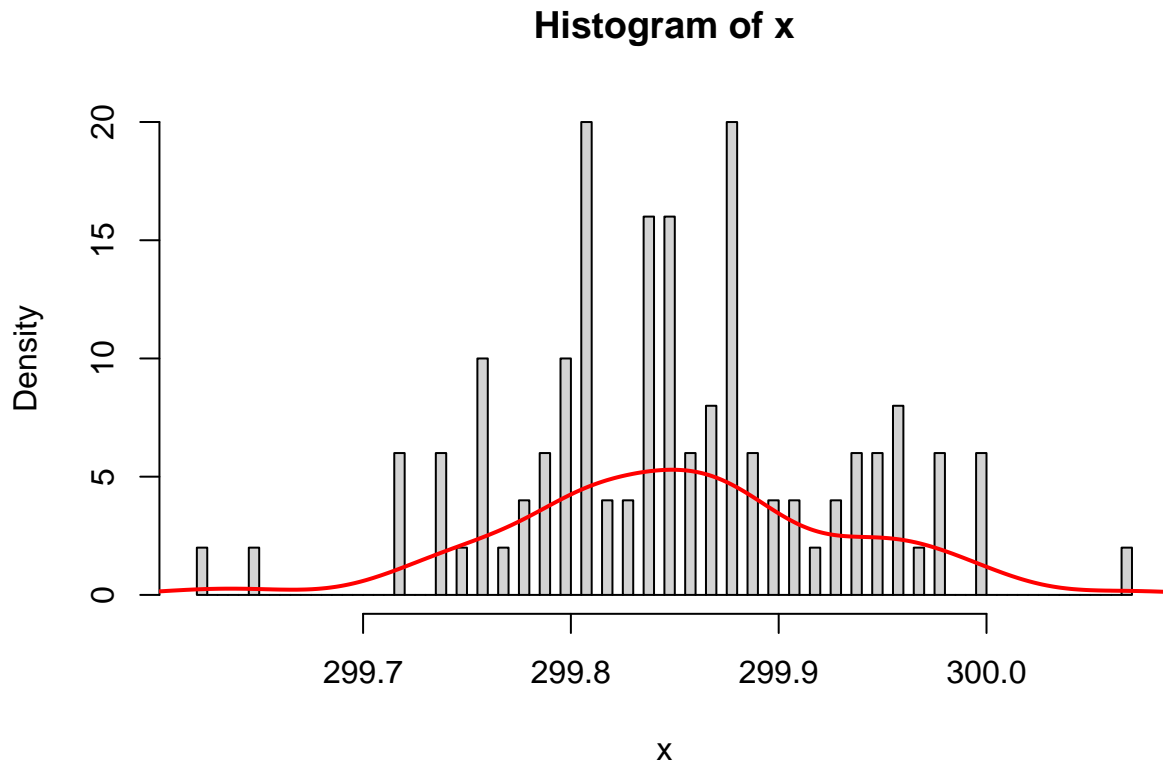
2023-11-01

## Chapter 11

### Exercise 5

Recall Exercise 6.3 based on 100 measurements of the speed of light in air. In that chapter, we tested the data for normality. Use the same data to construct a density estimator that you feel gives the best visual display of the information provided by the data. What parameters did you choose? The data can be downloaded from <http://www.itl.nist.gov/div898/strd/univ/data/Michelson.dat>

```
x <- as.numeric(read.delim2("Michelson.dat.txt"))[[1]]
hist(x, freq = F, nclass = 100)
kde <- density(x, bw = "SJ")
lines(kde$x, kde$y, col = "red", lwd = 2)
```



## Chapter 12

### Exercise 1

Using robust regression, find the intercept and slope  $\tilde{\beta}_0$  and  $\tilde{\beta}_1$  for each of the four data sets of Anscombe (1973) from p. 244. Plot the ordinary least-squares regression along with the rank regression estimator of slope. Contrast these with one of the other robust regression techniques. For which set does  $\tilde{\beta}_1$  differ the most from its LS counterpart  $\hat{\beta}_1 = 0.5$ ? Note that in the fourth set, 10 out of 11 Xs are equal, so one should use

$$S_{ij} = (Y_j - Y_i)/(X_j - X_i + \epsilon)$$

to avoid dividing by 0. After finding  $\tilde{\beta}_0$  and  $\tilde{\beta}_1$ , are they different than  $\hat{\beta}_0$  and  $\hat{\beta}_1$ ? Is the hypothesis  $H_0 : \beta_1 = 1/2$  rejected in a robust test against the alternative  $H_1 : \beta_1 < 1/2$ , for data set 3? Note here  $\beta_{10} = 1/2$ .

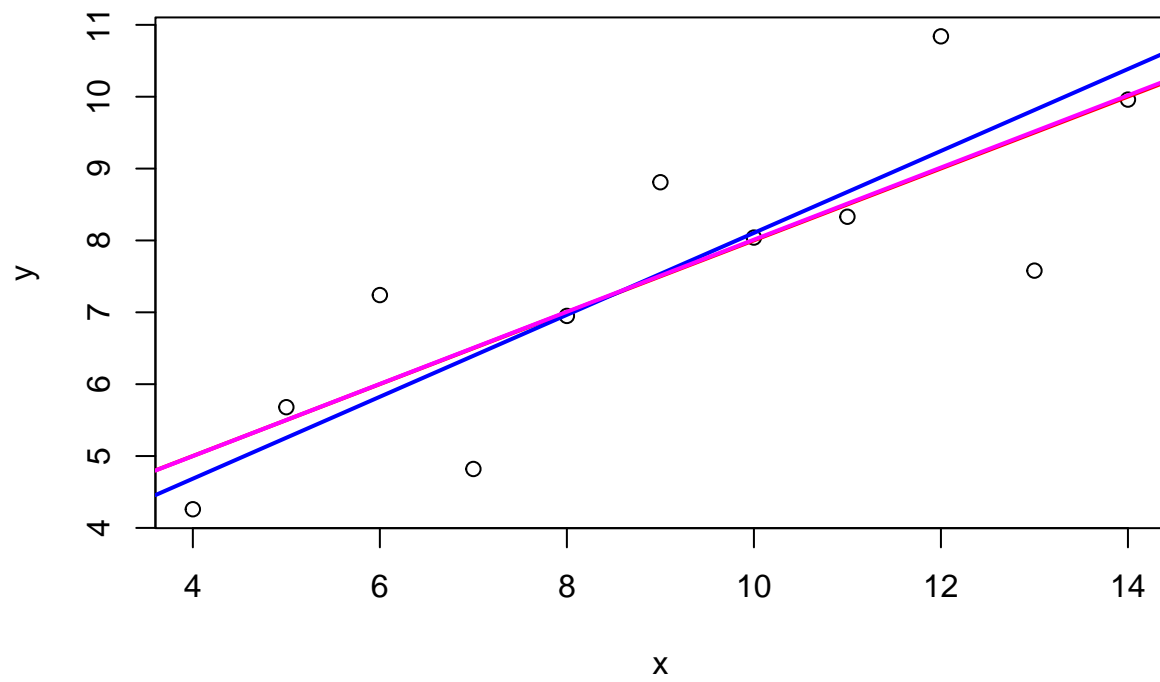
```
library(robustbase)
library(MASS)
library(L1pack)
```

```
## Loading required package: fastmatrix
```

```
data <- read.csv("anscombe.csv")
data$dataset <- factor(data$dataset)
data1 <- subset(data, subset = data$dataset == "I")
data2 <- subset(data, subset = data$dataset == "II")
data3 <- subset(data, subset = data$dataset == "III")
data4 <- subset(data, subset = data$dataset == "IV")
datasets <- list(data1, data2, data3, data4)
```

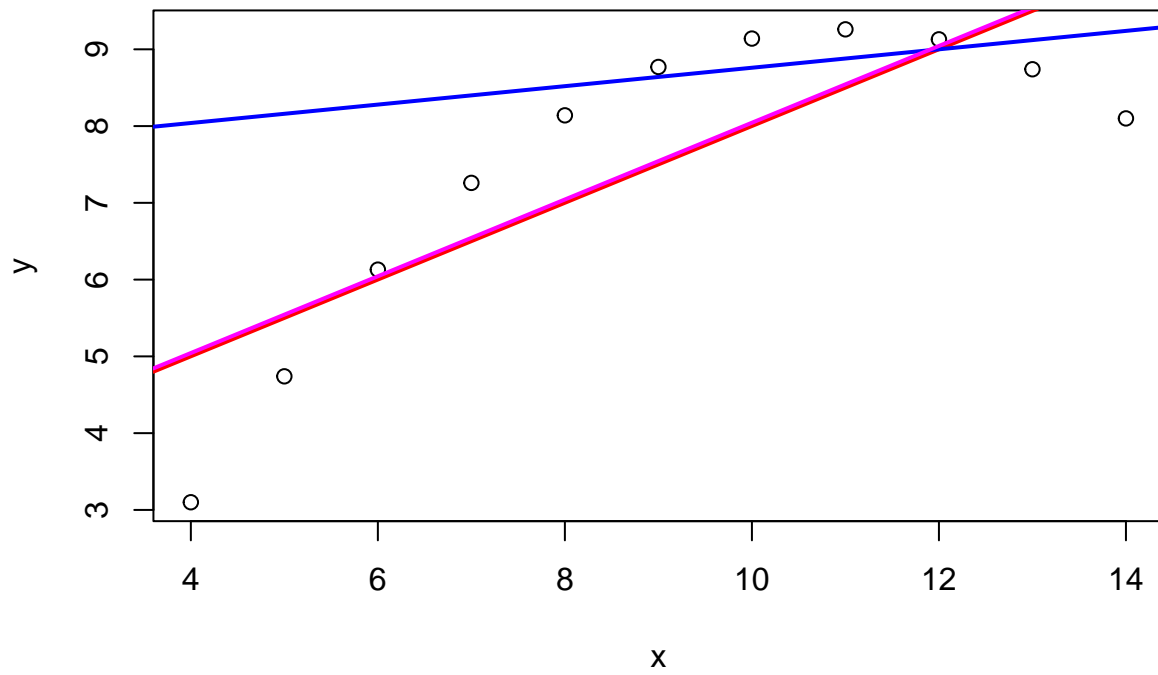
```
for (dataset in datasets) {
  plot(y ~ x, data = dataset, main = dataset[[1]])
  lmfit <- lm(y ~ x, data = dataset)
  abline(coef = coefficients(lmfit), col = "red", lwd = 2)
  lmedfit <- lmsreg(y ~ x, data = dataset)
  abline(coef = coefficients(lmedfit), col = "blue", lwd = 2)
  robfit <- rlm(y ~ x, data = dataset, method = "MM", psi = psi.bisquare, init = coefficients(lmedfit))
  abline(coef = coefficients(robfit), col = "magenta", lwd = 2)
  cat("Coefficients for dataset", dataset[[1]][1], ":\n")
  print(coefficients(lmfit))
  print(coefficients(lmedfit))
  print(coefficients(robfit))
}
```

I



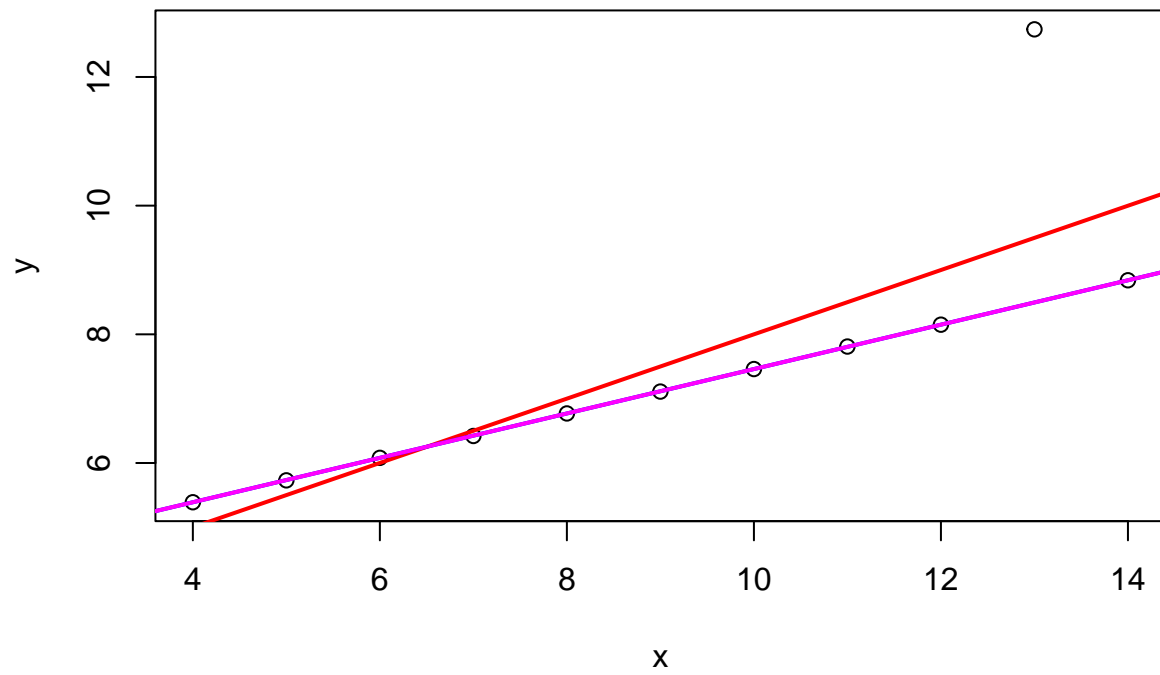
```
## Coefficients for dataset 1 :  
## (Intercept)      x  
## 3.0000909 0.5000909  
## (Intercept)      x  
## 2.405 0.570  
## (Intercept)      x  
## 2.992274 0.501876
```

II



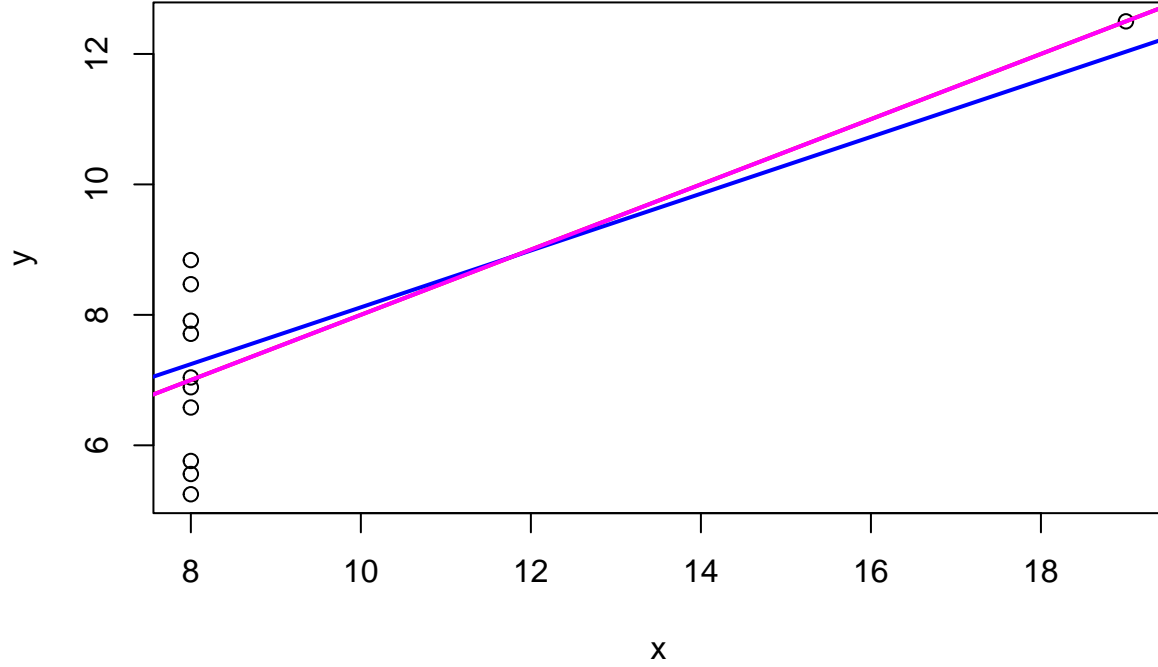
```
## Coefficients for dataset 2 :  
## (Intercept)      x  
## 3.000909      0.500000  
## (Intercept)      x  
## 7.56          0.12  
## (Intercept)      x  
## 3.0435347      0.4999972
```

III



```
## Coefficients for dataset 3 :  
## (Intercept)      x  
## 3.0024545 0.4997273  
## (Intercept)      x  
## 4.010 0.345  
## (Intercept)      x  
## 4.010 0.345
```

## IV



```
## Coefficients for dataset 4 :
## (Intercept)      x
##  3.0017273    0.4999091
## (Intercept)      x
##  3.7613636    0.4354545
## (Intercept)      x
##  3.0008522    0.4999551
```

For dataset 1  $\tilde{\beta}_1$  differs the at most 0.070 from the LS counterpart  $\hat{\beta}_1 = 0.5$ . This difference is attained with the least median of squares estimator.

For dataset 2  $\tilde{\beta}_1$  differs at most 0.38 from the LS counterpart  $\hat{\beta}_1 = 0.5$ . This difference is attained with the least median of squares estimator.

For dataset 3  $\tilde{\beta}_1$  differs at most 0.155 from the LS counterpart  $\hat{\beta}_1 = 0.5$ . This difference is attained with the least median of squares estimator & the M estimator using Tukey's bisquare function.

For dataset 4  $\tilde{\beta}_1$  differs at most 0.10 from the LS counterpart  $\hat{\beta}_1 = 0.5$ . This difference is attained with the least median of squares estimator.

Dataset 3 has the highest difference between  $\tilde{\beta}_1$  and  $\hat{\beta}_1 = 0.5$  with a difference of 0.155.

$\tilde{\beta}_0$  and  $\tilde{\beta}_1$  are different than  $\hat{\beta}_0$  and  $\hat{\beta}_1$  for two of the three estimators in dataset 3.

Is the hypothesis  $H_0 : \beta_1 = 1/2$  rejected in a robust test against the alternative  $H_1 : \beta_1 < 1/2$ , for data set 3?

```

n <- length(data3$x)
x <- data3$x
y <- data3$y
beta_1 <- 0.5
beta_10 <- 0.5
U <- y - beta_10 * x
rho_hat <- beta_1 * sqrt(sum(rank(x)^2 - mean(rank(x))))/sqrt(sum(rank(U)^2 - mean(rank(U))))
p <- 1 - pnorm(rho_hat * sqrt(n - 1))
p

```

```
## [1] 0.05692315
```

We accept the hypothesis  $H_0 : \beta_1 = 1/2$  since  $p > 0.05$