Case Study 3 - Group 4

Annika Janson h11829506
Jan Beck h11814291
Franz Uchatzi

29.11.2020

2 Descriptive statistics

Table 1: Summary statistics

Statistic	Mean	St. Dev.	Median	Min	Max
wage	988.475	406.512	937	115	3,078
hours	44.062	7.160	40	25	80
IQ	102.481	14.686	104	54	145
KWW	36.195	7.529	37	13	56
educ	13.680	2.231	13	9	18
exper	11.397	4.258	11	1	22
tenure	7.217	5.056	7	0	22
age	32.983	3.063	33	28	38
married	0.900	0.300	1	0	1
black	0.081	0.274	0	0	1
south	0.323	0.468	0	0	1
urban	0.719	0.450	1	0	1
sibs	2.846	2.241	2	0	14
brthord	2.178	1.488	2	1	10
meduc	10.828	2.823	12	0	18
feduc	10.273	3.288	11	0	18
lwage	6.814	0.412	6.843	4.745	8.032

2.1

The average wage is USD **988.48** and the median wage is USD **937**.

Histogram of wage
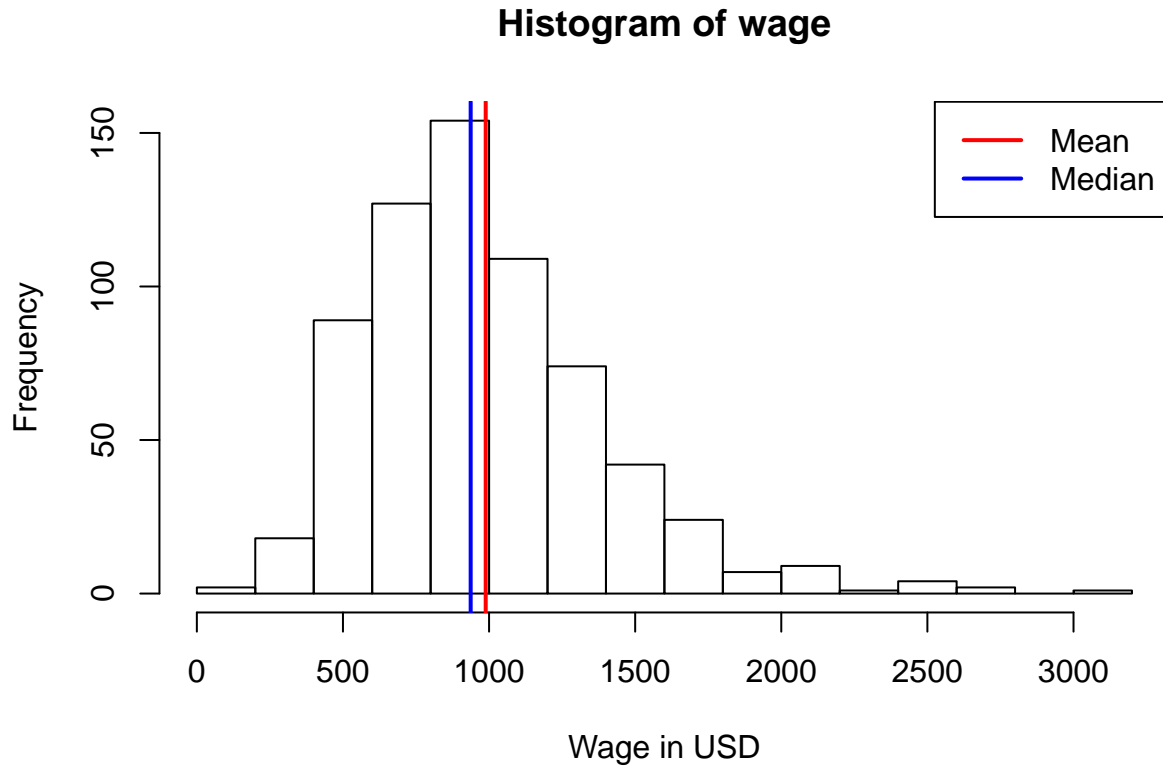
In the histogram we see that the distribution is right-skewed with a few observations exceeding USD 3000. Most observed values are concentrated around an interval of USD ±500 above and below the mean. Median and mean are fairly close to each other with the median being slightly higher due to the large outliers.

2.2

The proportion of workers working more than 40 hours a week is **42.081448%**.

2.3

The most common number of years of education among the workers is **270**.

2.4

No. The most frequent sibling pattern is having **1** sibling.

3 Data modelling

3.1

3.2

3.3

3.4

3.5.1

3.5.2

3.5.3

3.5.4

3.5.5

3.5.6

4 Simulation Study

4.1

```r
set.seed(1)

# our parameters according to spec
N1 <- 10
N2 <- 100
N3 <- 1000
beta0 <- -1
beta1 <- 0.2
mu <- 0
sigma <- sqrt(4)
minX = -3
maxX = 3

# model 1
x1 <- x <- runif(N1, min = minX, max = maxX)
u1 <- rnorm(N1, sd = sigma, mean = mu)
y1 <- beta0 + beta1*x1 + u1
lm1 <- lm(y1 ~ x) # using x instead of x1 to show as one row in stargazer output

# model 2
x2 <- x <- runif(N2, min = minX, max = maxX)
u2 <- rnorm(N2, sd = sigma, mean = mu)
y2 <- beta0 + beta1*x2 + u2
lm2 <- lm(y2 ~ x) # using x instead of x2 to show as one row in stargazer output

# model 3
x3 <- x <- runif(N3, min = minX, max = maxX)
u3 <- rnorm(N3, sd = sigma, mean = mu)
y3 <- beta0 + beta1*x3 + u3
lm3 <- lm(y3 ~ x) # using x instead of x3 to show as one row in stargazer output
```
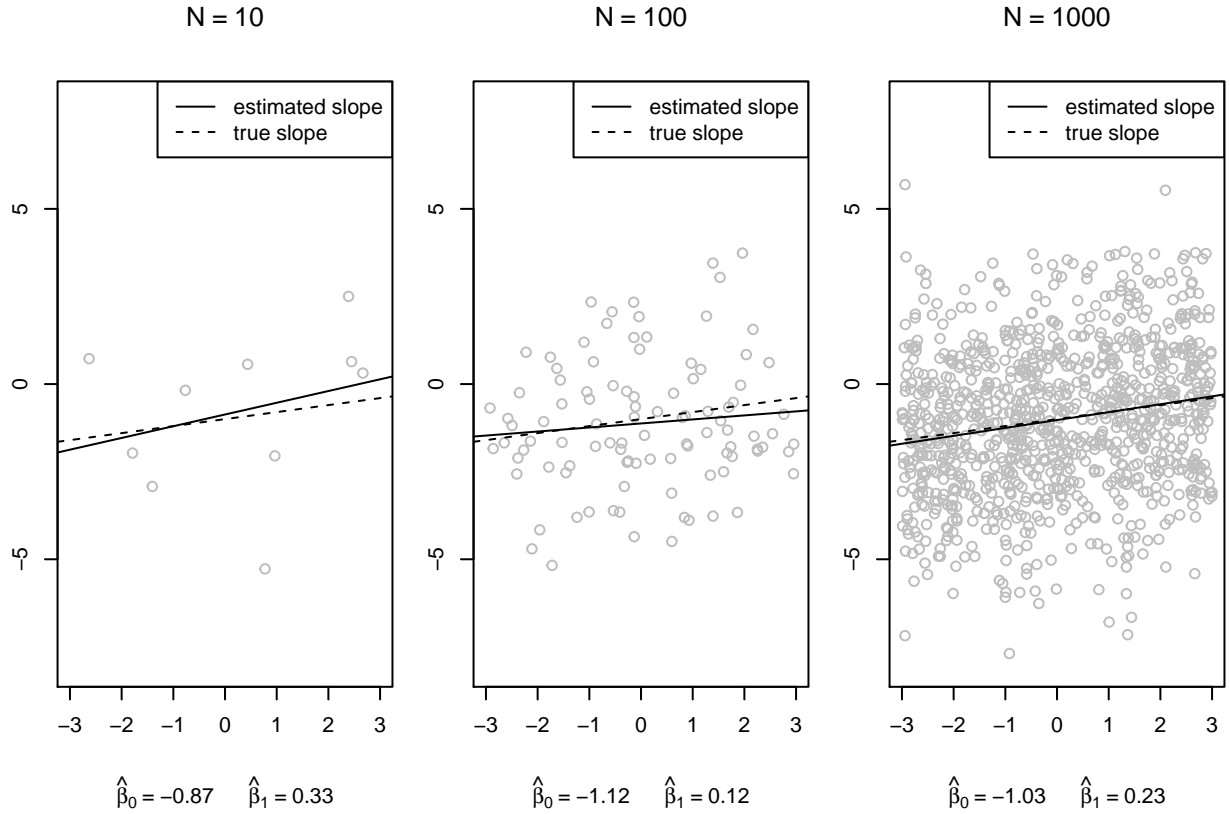
Table 2: Model comparison

	Dependent variable:		
	y1	y2	y3
	(1)	(2)	(3)
x	0.3346334	0.1150630	0.2255274***
	(0.4061975)	(0.1183332)	(0.0370705)
Constant	−0.8683543	−1.1244970***	−1.0266640***
	(0.7405519)	(0.1856205)	(0.0651777)
Observations	10	100	1,000
R^2	0.0782007	0.0095557	0.0357600
Adjusted R^2	-0.0370242	-0.0005509	0.0347938
Residual Std. Error	2.3079310 (df = 8)	1.8542300 (df = 98)	2.0608280 (df = 998)
F Statistic	0.6786786 (df = 1; 8)	0.9454931 (df = 1; 98)	37.0120100*** (df = 1; 998)

Note: $^*p<0.1;$ $^{**}p<0.05;$ $^{***}p<0.01$

N = 10   **N = 100**   **N = 1000**

$\hat{\beta}_0 = -0.87$ $\hat{\beta}_1 = 0.33$ $\hat{\beta}_0 = -1.12$ $\hat{\beta}_1 = 0.12$ $\hat{\beta}_0 = -1.03$ $\hat{\beta}_1 = 0.23$

4.2

4.3