Stat 500 - Homework 2 (Solutions)

Part A.

1. We first remove observations associated with negative values of the variable `experience`:

```
> library(faraway)
> data(uswages)
> newdata <- subset(uswages, uswages$exper >= 0)
```

Now, we regress `weekly wages` onto `years of education` and `experience`. By default `R` always includes an intercept.

```
> fit <- lm(wage ~ educ + exper, data=newdata)
> summary(fit)

Call:
lm(formula = wage ~ educ + exper, data = newdata)

Residuals:
    Min      1Q  Median      3Q     Max
-1014.7  -235.2   -52.1   150.1  7249.2

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -239.1146    50.7111  -4.715 2.58e-06 ***
educ          51.8654     3.3423  15.518  < 2e-16 ***
exper          9.3287     0.7602  12.271  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 426.8 on 1964 degrees of freedom
Multiple R-squared:  0.1348,Adjusted R-squared:  0.1339
F-statistic:   153 on 2 and 1964 DF,  p-value: < 2.2e-16
```

2. Our linear model explains 13.48 % of the variation in the response. Note that only if the model contains an intercept `R` outputs the correct value of the coefficient of determination. This is because only with intercept the variance decomposition relation holds. What happens if you do not include the intercept?

3. The case number of the largest residual is 1550, the value of his residual is 7249.174.

```
> which.max(fit$res) # case number (index)
15387
 1550
> fit$res[which.max(fit$res)] # value of max. residual
   15387
7249.174
```

4. The mean of the residuals is $-1.381535 \times 10^{-15} \approx 0$, while the median of the residuals is -52.14337. This suggests that the (empirical) distribution of the residuals is skewed to the right.

```
> mean(fit$res)
[1] -1.381535e-15
> median(fit$res)
[1] -52.14337
```

5. This is an exercise in how to interpret the estimated coefficients of a linear model. Possible answers are: "*Based on the linear model we predict for two people with the same education and one year difference in experience a wage difference of $9.33.*" Or: "*Our linear model predicts that an increase of one year in experience results,* ceteris paribus, *in an increase of weekly wage by $9.33.*"

6. The correlation between fitted values and residuals is $6.35678 \times 10^{-17} \approx 0$. In geometric terms this means that the vectors of fitted values and residuals are orthogonal to each other, i.e. the vectors $X'\hat{\beta}$ and $\hat{\epsilon} = Y - X'\hat{\beta}$ from a right angle. Based on plot of residuals versus fitted values in Figure 1 do you think that the linear regression is a good model?

```
> cor(fit$fitted, fit$res)
[1] 6.35678e-17
> plot(fit$fitted, fit$res, xlab="Fitted", ylab="Residuals")
> abline(h=0) # add horizontal line at zero
```

Part B.
1. To compute $\hat{\beta} = (X'X)^{-1}XY$ we use the following code:

```
> set.seed(1504) # initialize random number generator to get reproducible results
> X <- cbind(rep(1, 10),c(2,-1,3,3,2,1,0,0,-1,0), c(-2,-2,-2,3,3,3,0,0,0,1))
> beta0 <- c(1,-1,2)
> sigma <- 1
> y <- X%*%beta0 + rnorm(10, 0, sigma)
> solve(t(X)%*%X)%*%t(X)%*%y
           [,1]
[1,]  1.0623948
[2,] -0.9453115
[3,]  2.2332188
```

2. The population ("true") variance of $\hat{\beta}$ is $\sigma^2(X'X)^{-1}$, i.e.

```
> sigma^2*solve(t(X)%*%X)
             [,1]         [,2]         [,3]
[1,]  0.139180672 -0.042016807 -0.003413866
[2,] -0.042016807  0.050420168 -0.008403361
[3,] -0.003413866 -0.008403361  0.027442227
```

3. An unbiased estimate for σ^2 is given by $\frac{1}{7}\sum_{i=1}^{10}(y_i - x_i'\hat{\beta})^2$, i.e.

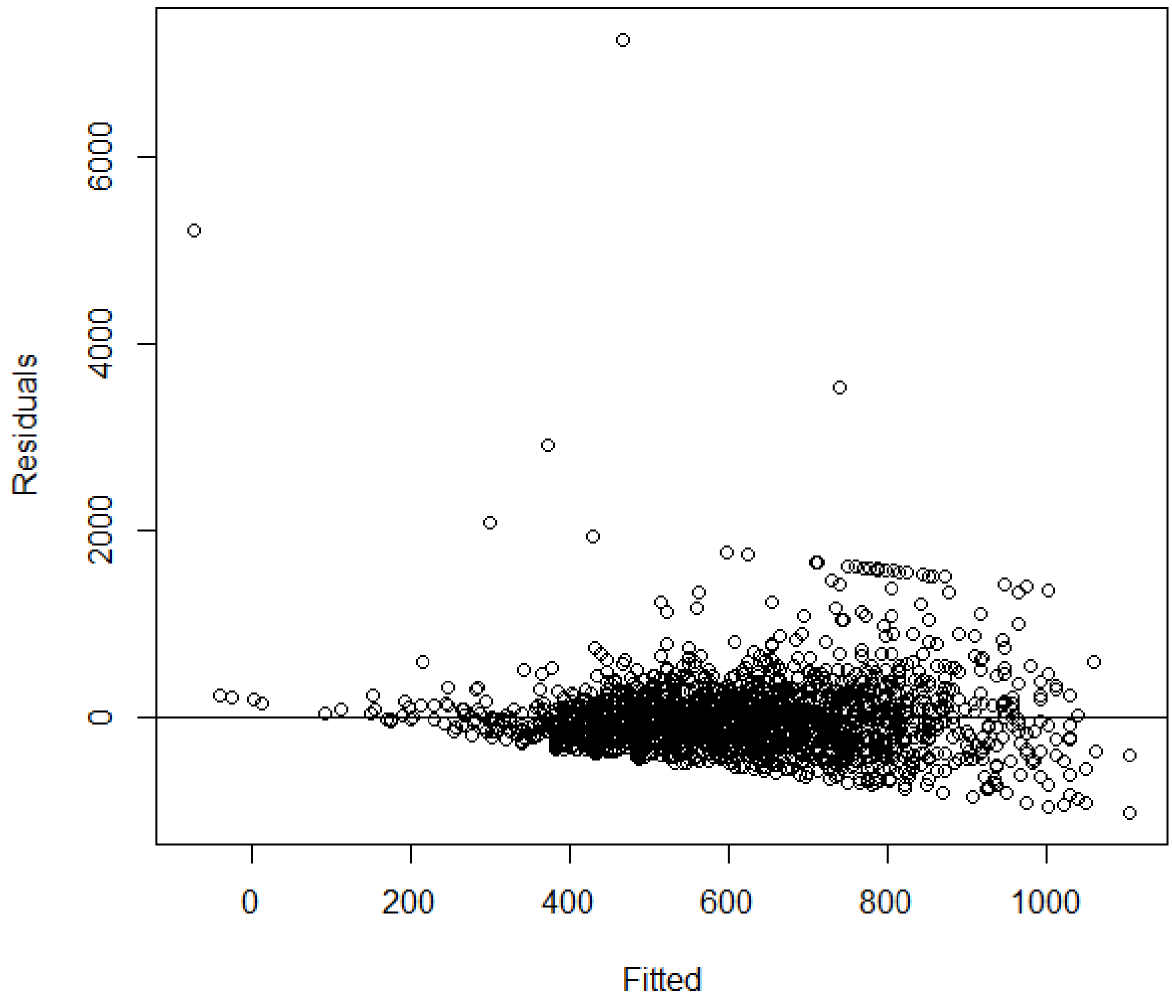

Figure 1: Residuals versus Fitted Values

```
> fitted <- y - X%*%beta
> sigma2_hat <- sum(fitted^2)/(length(fitted)-3)
> sigma2_hat
[1] 1.887114
```

4. & 5. We solve questions 4 and 5 together in one loop but comment separately on the results.

```
> B <- matrix(NA, ncol=3, nrow=1000)
> S <- matrix(NA, ncol=1, nrow=1000)
> for (i in 1:1000) {
+   y <- X%*%beta0 + rnorm(10, 0, sigma)
```

```
+   B[i,] <- solve(t(X)%*%X)%*%t(X)%*%y
+   fitted <- y - X%*%B[i,]
+   S[i] <- sum(fitted^2)/(length(fitted) -3)
+ }
> var(B[,1]) # variance of beta_1 etc...
[1] 0.1486725
> var(B[,2])
[1] 0.05159052
> var(B[,3])
[1] 0.0284923
```

From above output we learn that the estimates of the variances for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ match the population variances in question 2 quite well. Moreover, the histograms of the estimates are centered around the true values of β:

```
> hist(B[,1], main=expression(paste("Histogram of ", beta[1])), xlab=expression(hat(beta)[1]))
> hist(B[,2], main=expression(paste("Histogram of ", beta[2])), xlab=expression(hat(beta)[2]))
> hist(B[,3], main=expression(paste("Histogram of ", beta[3])), xlab=expression(hat(beta)[3]))
> hist(S, main=expression(paste("Histogram of ", hat(sigma))), xlab=expression(hat(sigma)))
```

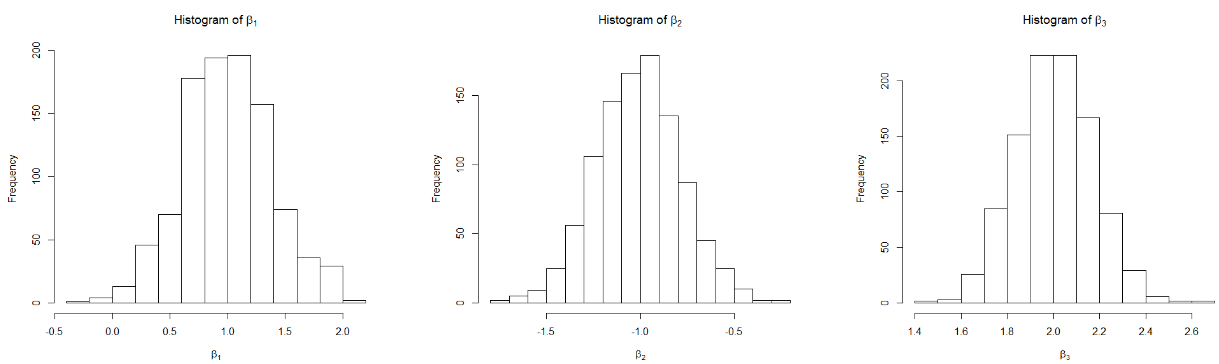

Figure 2: (a) Histogram of estimates for β_1, (b) Histogram of estimates for β_2, and (c) Histogram of estimates for β_3. Each histogram is based on 1000 simulations.

5. The mean of the estimates for σ^2 is also quite accurate:

```
> mean(S)
[1] 0.9958682
```

We can also compare the histogram of the estimates for σ^2 with the histogram of samples from the population distribution of estimates for σ^2:

```
> hist(S, main=expression(paste("Histogram of ", hat(sigma))), xlab=expression(hat(sigma)))
> chi2 <- rchisq(1000,7)
[1] 0.9982037
> hist(chi2/7, main="")
```

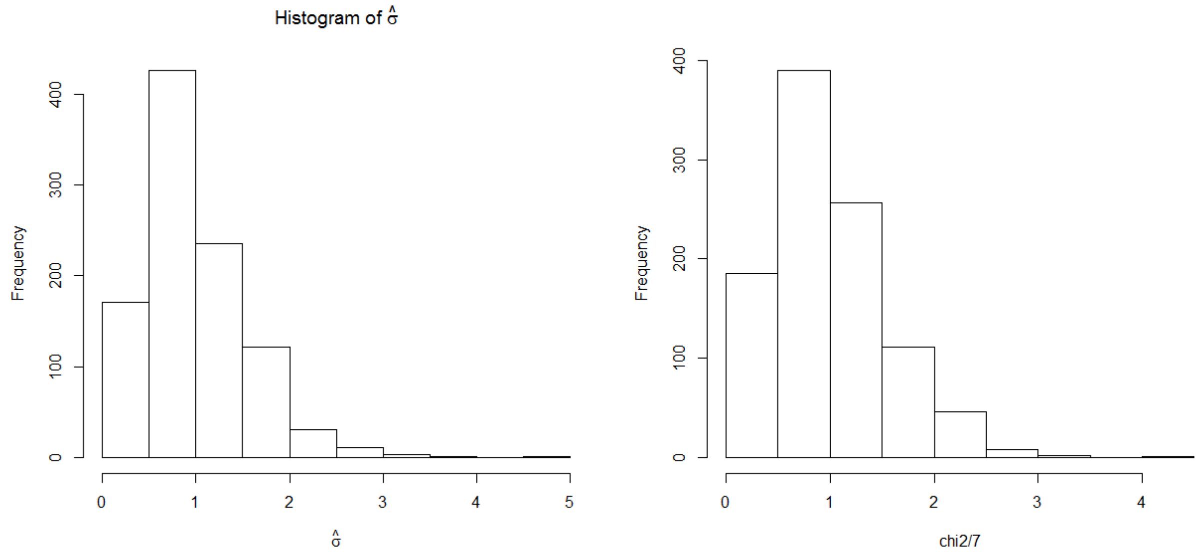

Figure 3: (a) Histogram of estimates for σ^2, (b) Histogram of samples from the population distribution of the estimate for σ^2 imates for β_2. Each histogram is based on 1000 simulations.

We see that the two histograms have the same centers of mass but that the histogram of the estimates for σ^2 is slightly more spread out.

7. We suggest to re-run the code with errors following the uniform distribution $U[-\sqrt{3}, \sqrt{3}]$. (Check for yourself that this distribution has indeed mean 0 and variance 1.)

```
> B <- matrix(NA, ncol=3, nrow=1000)
> S <- matrix(NA, ncol=1, nrow=1000)
> for (i in 1:1000) {
+   y <- X%*%beta0 + runif(10, -sqrt(3), sqrt(3))
+   B[i,] <- solve(t(X)%*%X)%*%t(X)%*%y
+   fitted <- y - X%*%B[i,]
+   S[i] <- sum(fitted^2)/(10-3)
+ }
>
> var(B[,1])
[1] 0.1296519
> var(B[,2])
[1] 0.04997595
> var(B[,3])
[1] 0.02816145
>
> hist(B[,1], main=expression(paste("Histogram of ", beta[1])), xlab=expression(hat(beta)[1]))
> hist(B[,2], main=expression(paste("Histogram of ", beta[2])), xlab=expression(hat(beta)[2]))
> hist(B[,3], main=expression(paste("Histogram of ", beta[3])), xlab=expression(hat(beta)[3]))
>
> hist(S, main=expression(paste("Histogram of ", hat(sigma))), xlab=expression(hat(sigma)))
> mean(S)
```

```
[1] 1.02071
```

We observe that neither the variances of the estimates of β nor the mean of the estimates of σ^2 are much affected by the change in the distribution of the error term. However, from Figure 4 we see that the variation of the estimates for β has increased (albeit only slightly). Notably, the histogram of the estimates of σ^2 looks now very different from the histogram based on the correct distribution depicted in Figure 3 (b) (note the change in the spread!).

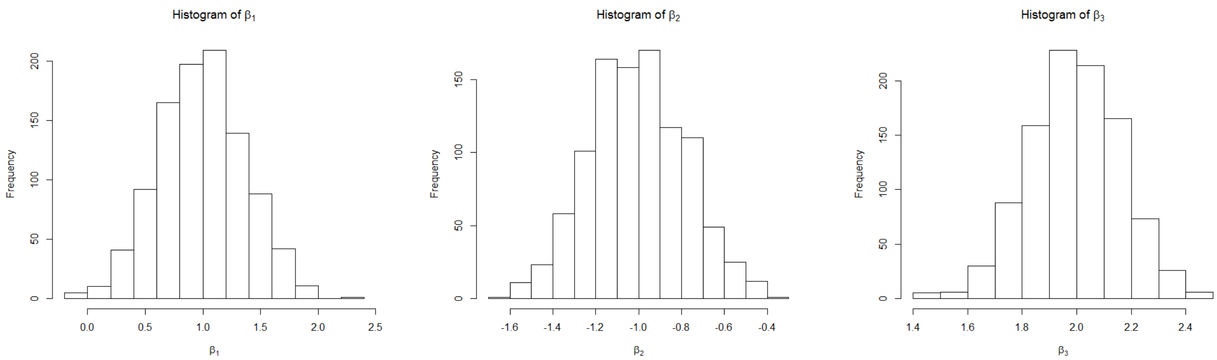

Figure 4: (a) Histogram of estimates for β_1, (b) Histogram of estimates for β_2, and (c) Histogram of estimates for β_3. Error distribution is the uniform distribution $U[-\sqrt{3}, \sqrt{3}]$. Each histogram is based on 1000 simulations.

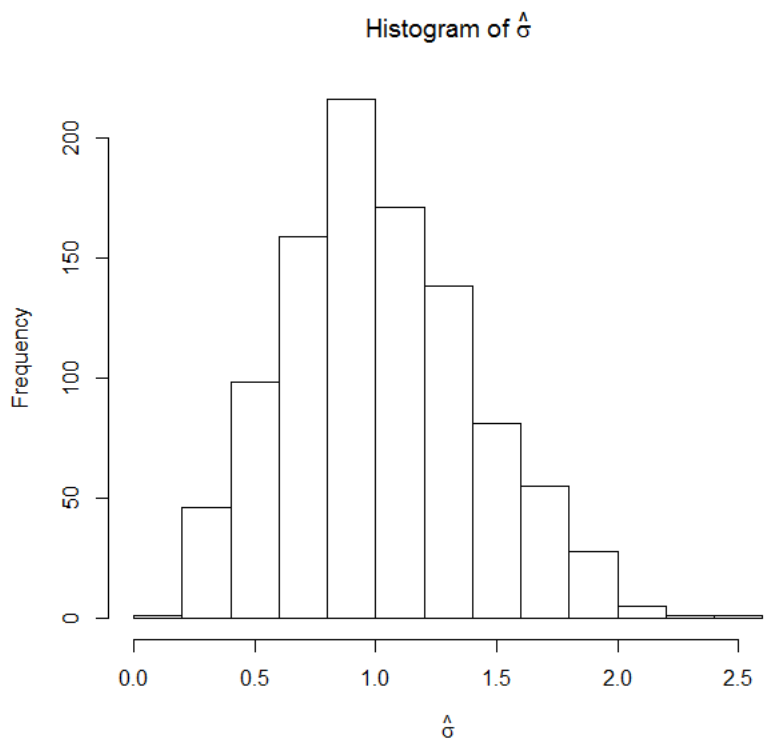

Figure 5: Residuals versus Fitted Values. Error distribution is the uniform distribution $U[-\sqrt{3}, \sqrt{3}]$. Histogram is based on 1000 simulations.