NCSU ST 503 HW 6

Probems 6.2,6.3,6.4,6.5,6.8 Faraway, Julian J. Linear Models with R, Second Edition Chapman & Hall / CRC Press.

Bruce Campbell

*24 September, 2017*

Complete exercises # 2 - 5 (parts a and b only for each) from Chapter 6. Complete exercise # 8 from Chapter 6.

6.2 Using the teengamb dataset, fit a model with gamble as the response and the other variables as predictors.

```
rm(list = ls())
data(teengamb, package = "faraway")
lm.fit <- lm(gamble ~ ., data = teengamb)
```

(a) Check the constant variance assumption for the errors.

```
plot(fitted(lm.fit), residuals(lm.fit), xlab = "Fitted", ylab = "Residuals")
abline(h = 0)
```
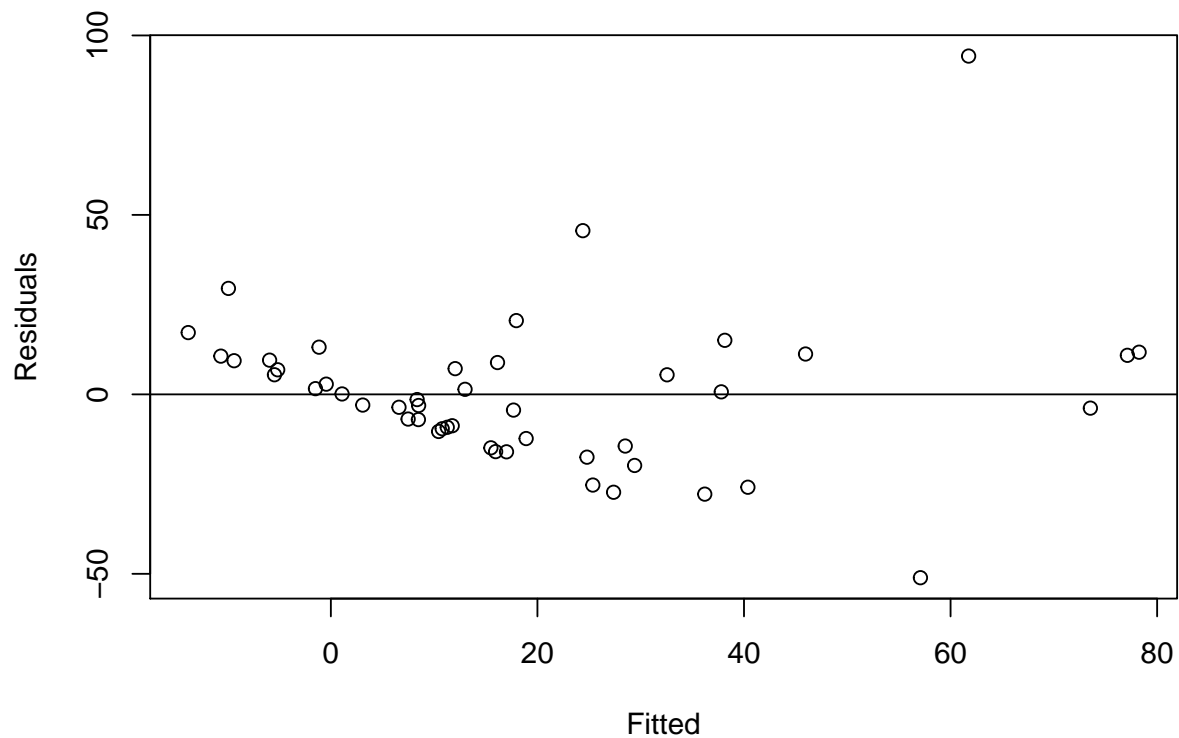
To check the assumption of constant variance we plot fitted values against the residuals - looking for any structure in the distribution of values about the theoretical mean value line $E[\epsilon] = 0$. There appears to be structure and heteroskedasticity in the plot. Below we plot the fitted values against the residuals for a model where the response has been transformed with the square root function. We see less structure and a more evenly distributed variance. We do see even with the transformed response evidence that the variance is not constant.

```
lm.fit <- lm(sqrt(gamble) ~ ., data = teengamb)
plot(fitted(lm.fit), residuals(lm.fit), xlab = "Fitted", ylab = "Residuals")
abline(h = 0)
```
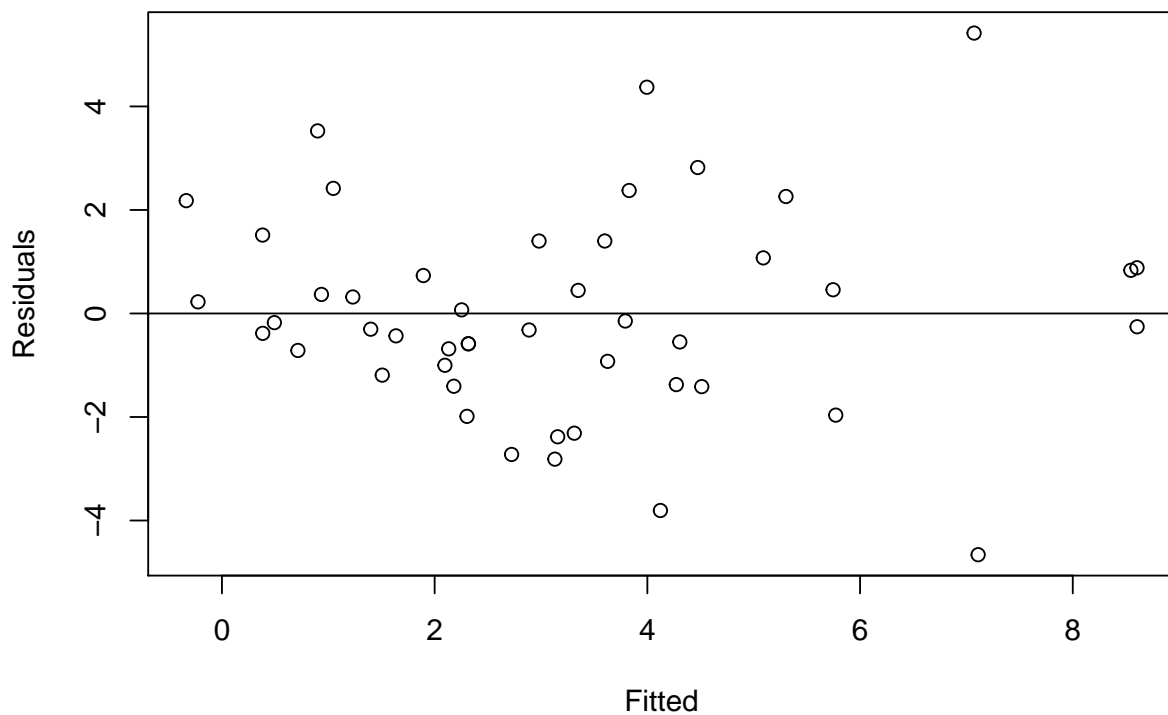
For reference we plot below what contant $N(0, 1)$ error over the same range of the response would look like for the same number of data points. We ran this a number of times to get a good idea of what contant variance looks like with this number of points. It's helpful to calibrate this way when evaluating whether variance is constant for a small and medium data sets.

```r
x <- runif(47, min = 0, max = 8.5)
y <- rnorm(47, mean = 0, sd = 1)
plot(x, y)
```
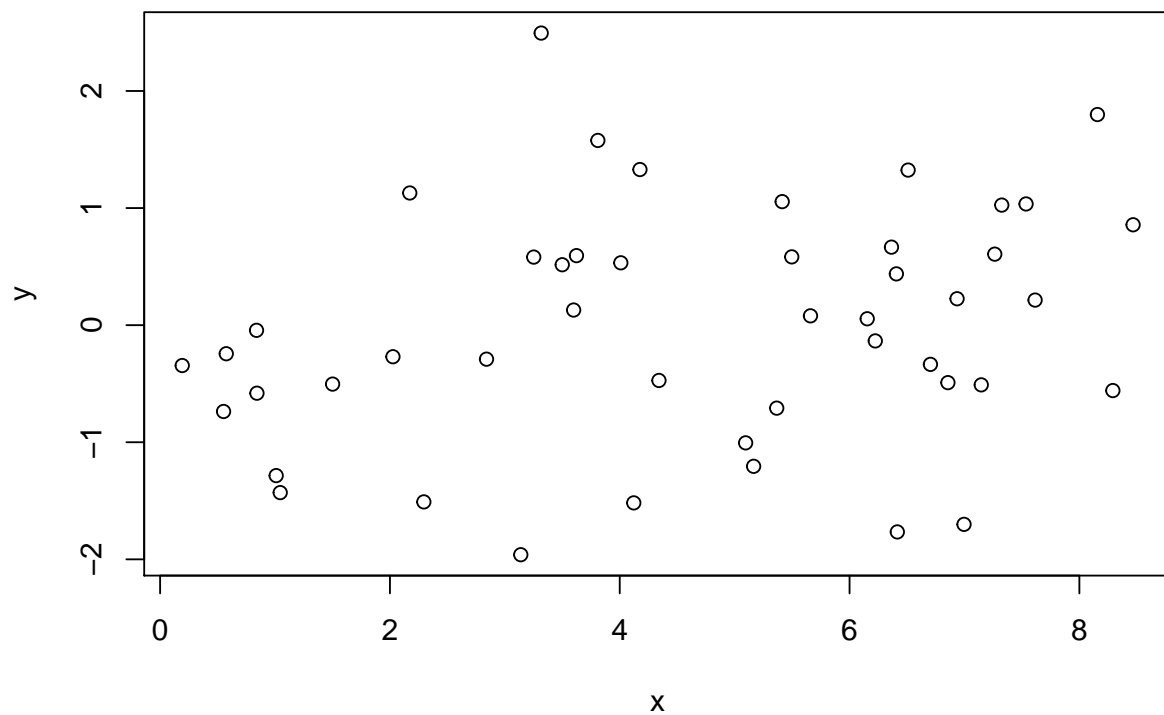
(b) Check the normality assumption.

```r
qqnorm(scale(residuals(lm.fit), center = TRUE, scale = TRUE), ylab = "Residuals",
    main = "Q-Q Plot of Standardized Residuals")
qqline(scale(residuals(lm.fit), center = TRUE, scale = TRUE))
```

Q–Q Plot of Standardized Residuals
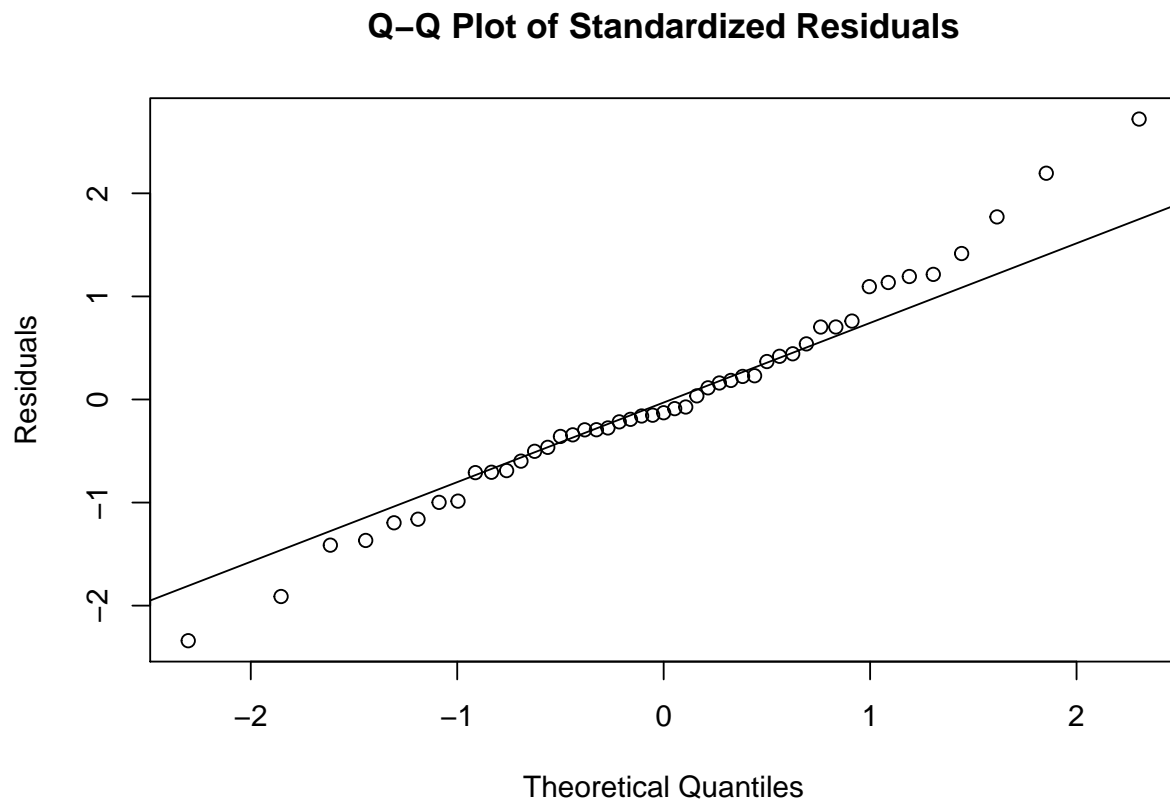
We see clear evidence of long tails in the distribution of the residuals.

6.3 For the prostate data, fit a model with lpsa as the response and the other variables as predictors.

```r
rm(list = ls())
data(prostate, package = "faraway")
lm.fit <- lm(lpsa ~ ., data = prostate)
```

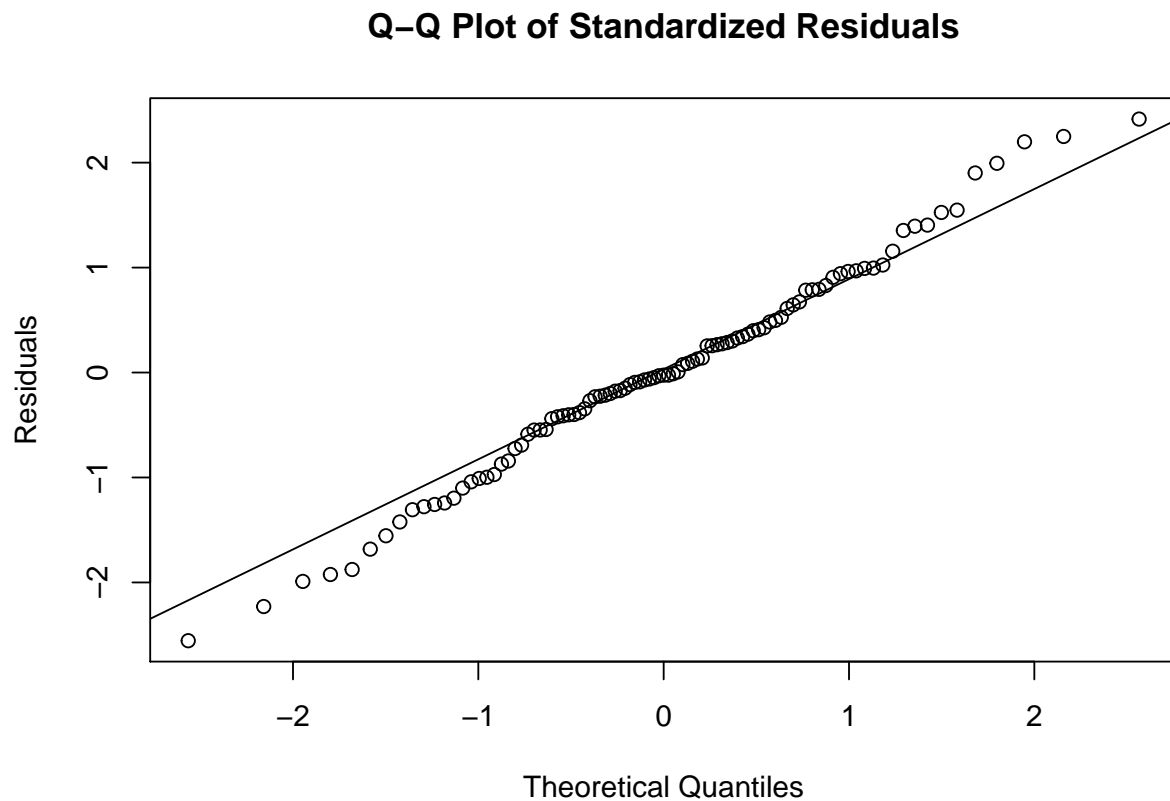**(a) Check the constant variance assumption for the errors.**

```r
plot(fitted(lm.fit), residuals(lm.fit), xlab = "Fitted", ylab = "Residuals")
abline(h = 0)
```

The variance of the residuals appears contant over the range of the fitted values. We're comfortable claiming homoskedasticity of residuals for this data set.

(b) Check the normality assumption.

```r
qqnorm(scale(residuals(lm.fit), center = TRUE, scale = TRUE), ylab = "Residuals",
    main = "Q-Q Plot of Standardized Residuals")
qqline(scale(residuals(lm.fit), center = TRUE, scale = TRUE))
```

Q–Q Plot of Standardized Residuals
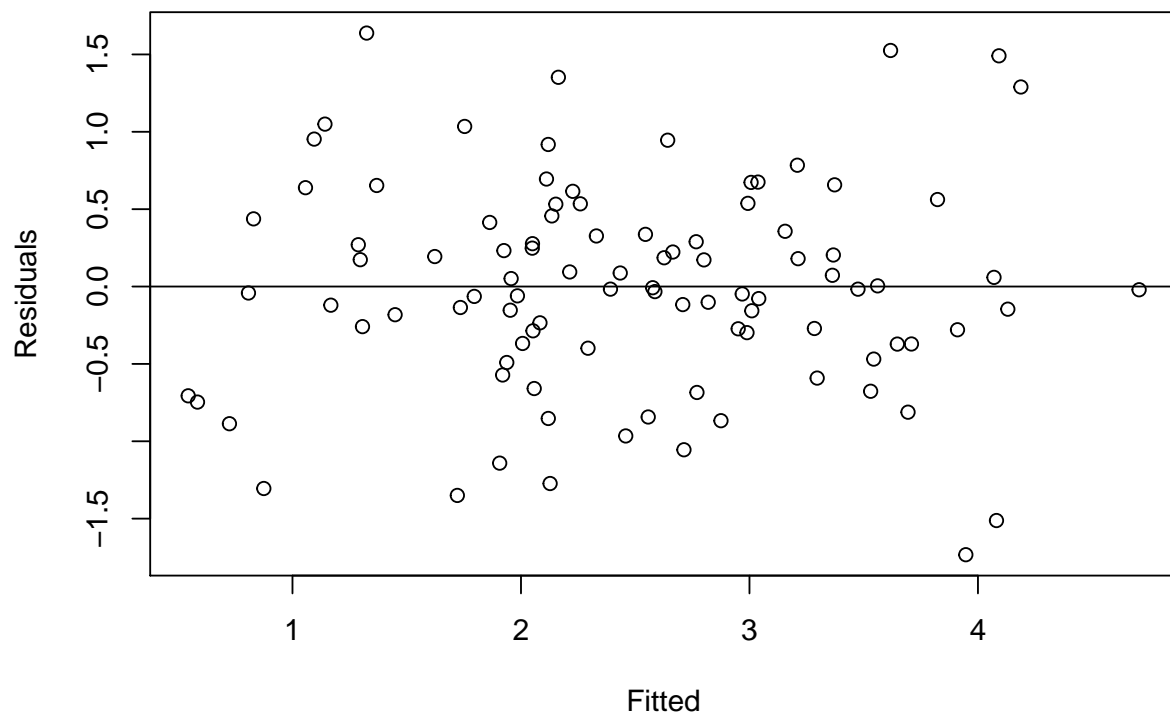
The studentized residuals appear to be lighly long tailed.

6.4 For the swiss data, fit a model with Fertility as the response and the other variables as predictors.

```
rm(list = ls())
data(swiss, package = "faraway")
lm.fit <- lm(Fertility ~ ., data = swiss)
```

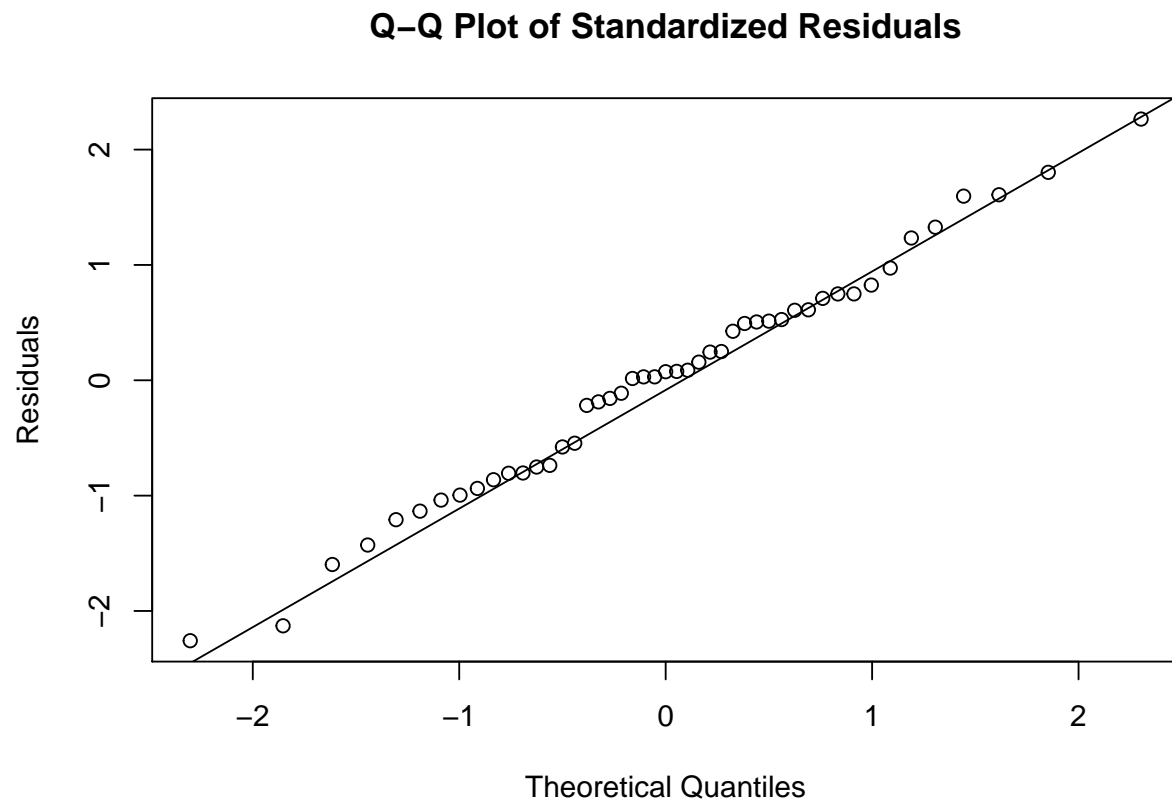**(a) Check the constant variance assumption for the errors.**

```
plot(fitted(lm.fit), residuals(lm.fit), xlab = "Fitted", ylab = "Residuals")
abline(h = 0)
```

The variance of the residuals appears contant over the range of the fitted values. We're comfortable claiming homoskedasticity of residuals for this data set.

(b) Check the normality assumption.

```r
qqnorm(scale(residuals(lm.fit), center = TRUE, scale = TRUE), ylab = "Residuals",
    main = "Q-Q Plot of Standardized Residuals")
qqline(scale(residuals(lm.fit), center = TRUE, scale = TRUE))
```
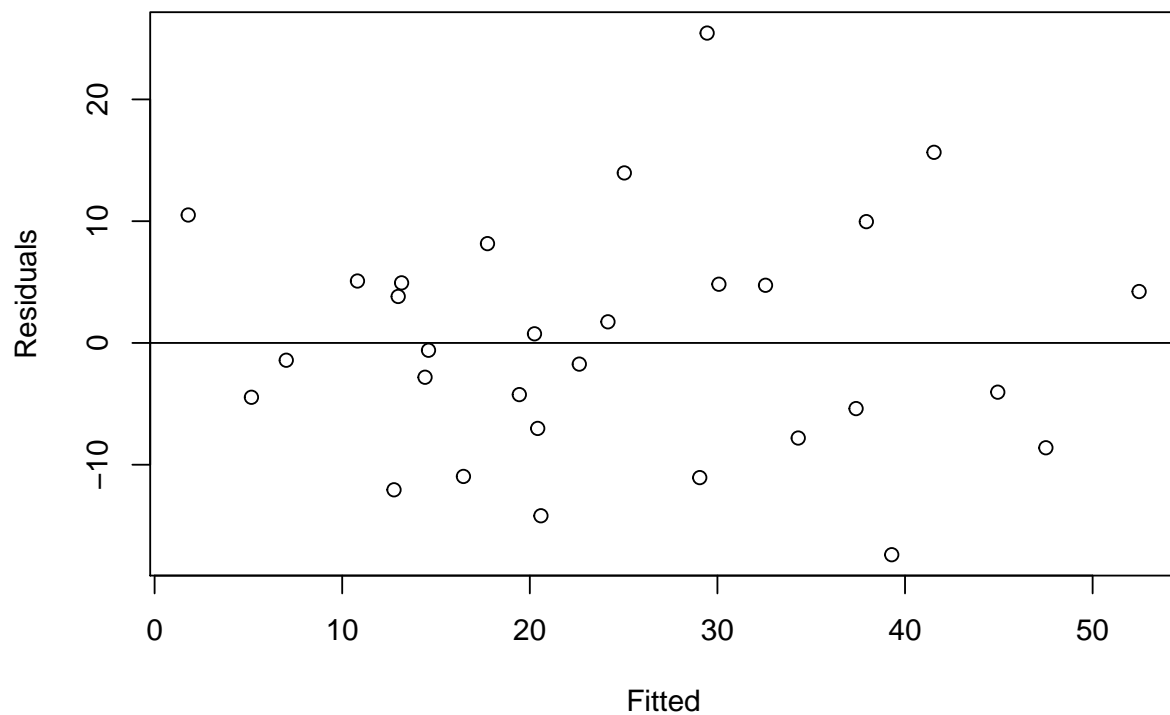
Q–Q Plot of Standardized Residuals
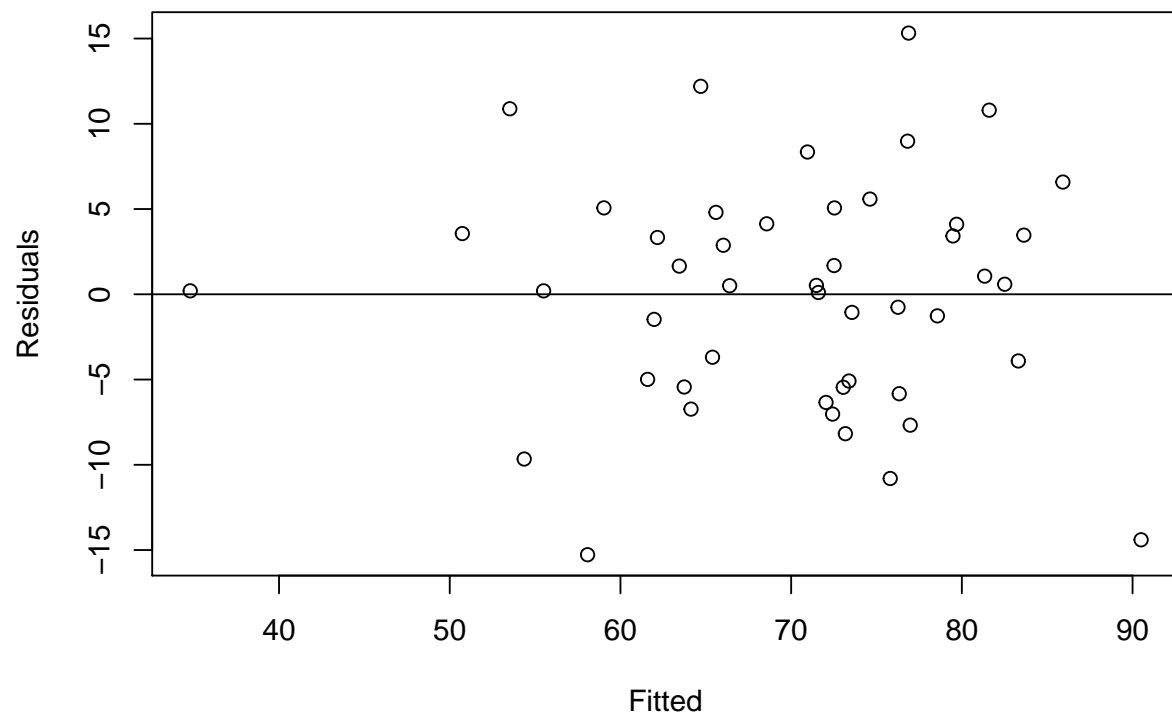
6.5 Using the cheddar data, fit a model with taste as the response and the other three variables as predictors.

```
rm(list = ls())
data(cheddar, package = "faraway")
lm.fit <- lm(taste ~ ., data = cheddar)
```

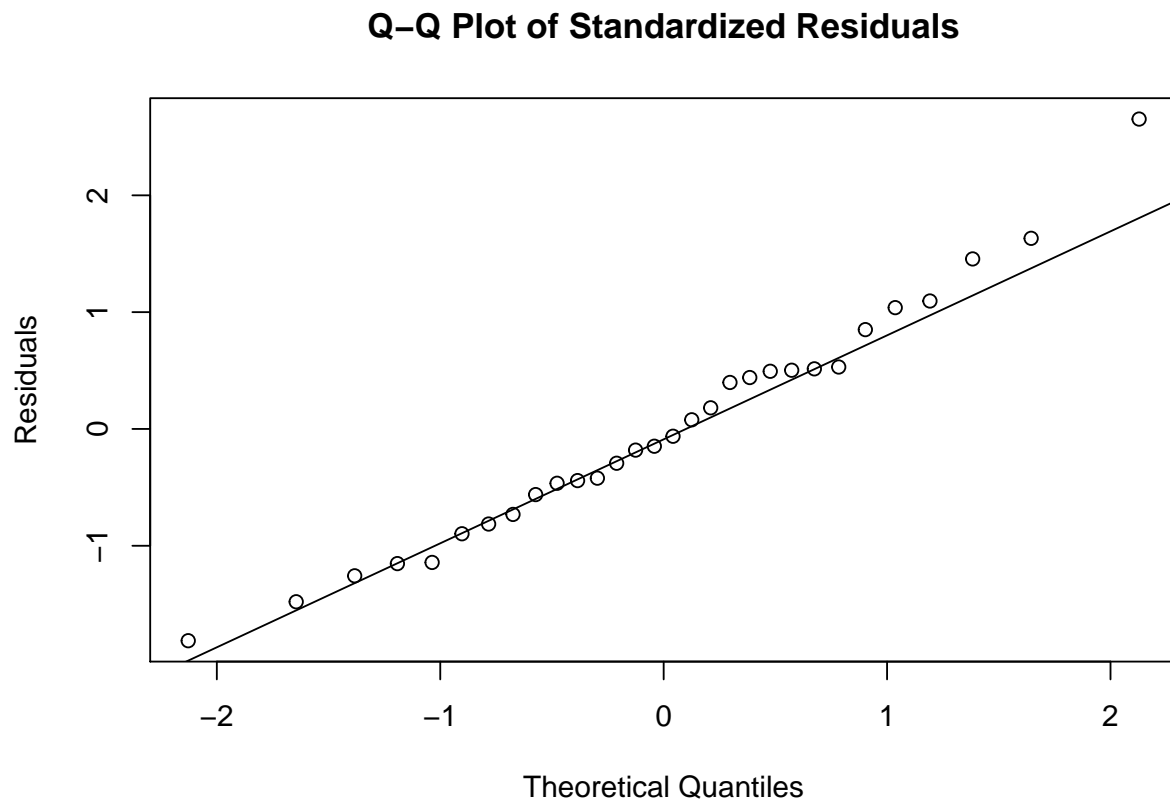(a) Check the constant variance assumption for the errors.

```
plot(fitted(lm.fit), residuals(lm.fit), xlab = "Fitted", ylab = "Residuals")
abline(h = 0)
```

The variance of the residuals appears contant over the range of the fitted values. We're comfortable claiming homoskedasticity of residuals for this data set.

(b) Check the normality assumption.

```r
qqnorm(scale(residuals(lm.fit), center = TRUE, scale = TRUE), ylab = "Residuals",
    main = "Q-Q Plot of Standardized Residuals")
qqline(scale(residuals(lm.fit), center = TRUE, scale = TRUE))
```
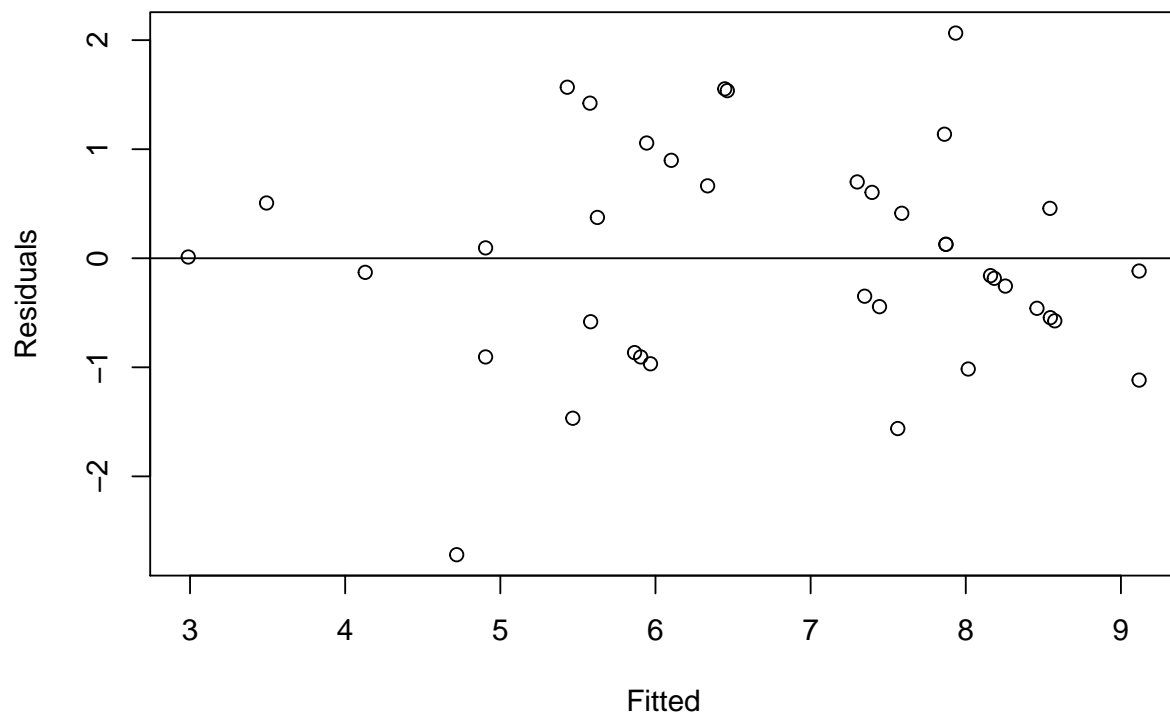
Q–Q Plot of Standardized Residuals

The standardized residuals appear to be normally distributed.

6.6 Using the happy data, fit a model with happy as the response and the other four variables as predictors.

```r
rm(list = ls())
data(happy, package = "faraway")
lm.fit <- lm(happy ~ ., data = happy)
```

**(a) Check the constant variance assumption for the errors.**

```r
plot(fitted(lm.fit), residuals(lm.fit), xlab = "Fitted", ylab = "Residuals")
abline(h = 0)
```

We see structure in the plot of the residuals. There is serial correlation in the residuals - which is a red flag for our model. The variance of the residuals appears slightly lower over the low end of the range of the response, and higher at the high end of the response. This is difficult to judge though since there ar eonly a few points at the low end of the range of the response.

(b) Check the normality assumption.

```r
qqnorm(scale(residuals(lm.fit), center = TRUE, scale = TRUE), ylab = "Residuals",
    main = "Q-Q Plot of Standardized Residuals")
qqline(scale(residuals(lm.fit), center = TRUE, scale = TRUE))
```

Q–Q Plot of Standardized Residuals
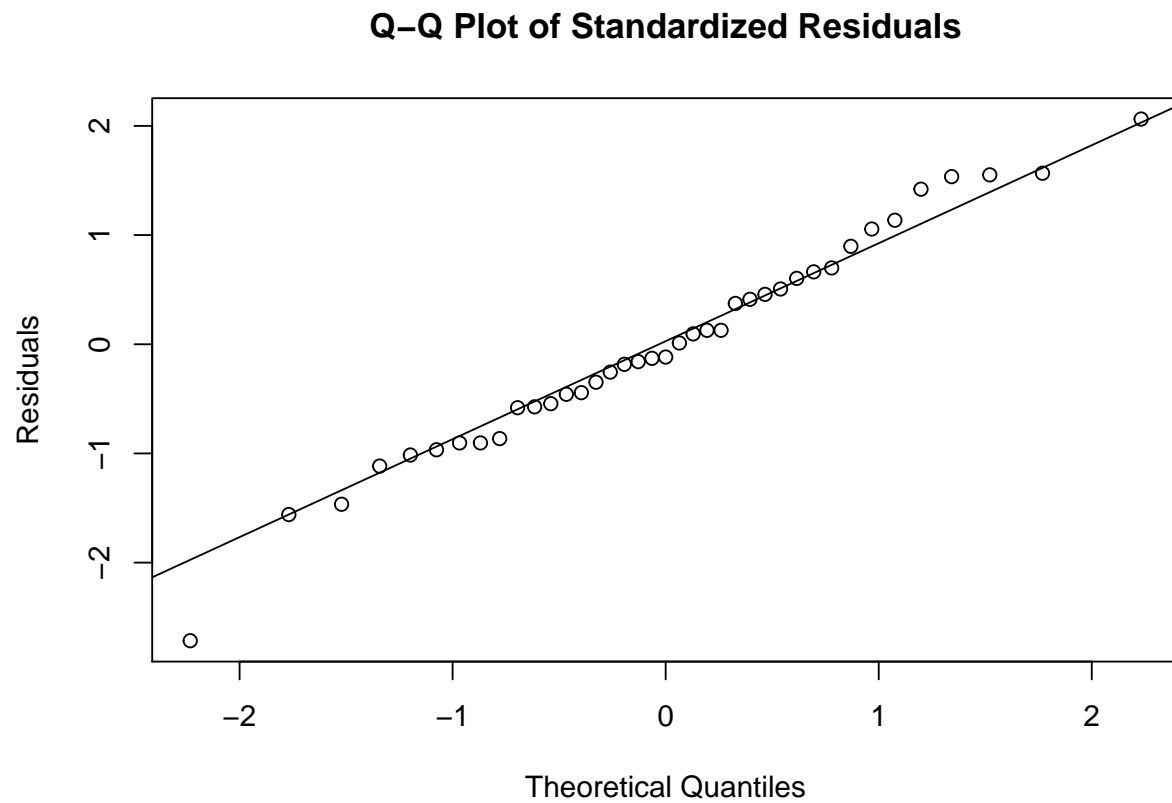
The standardized residuals appear to be normally distributed. This is interesting in light of the results from part a).