HW5

Andrew Shao

2024-10-27

(a)

In the `happy` dataset there are 39 observations and 5 variables, with each observation corresponding to the measurements for a single University of Chicago MBA student. The response variable in this analysis is `happy`, a subjective measurement of happiness on a 10 point scale with 10 being the happiest. `money` measures the students' family incomes in thousands of dollars, `sex` is a binary indicator variable describing whether participants were satisfied with their sexual activity, and `love` is a categorical variable where 1 indicates the student is lonely, 2 indicates the student has secure relationships and 3 indicates a deep feeling of belonging and caring. `work` is a 5 point scale where 1 is no job and 5 is a great job. `happy` does not appear to be normally distributed and looks skewed to the left. No two variables have a correlation value greater than 0.9.

(b)

The fitted model is:

$$happy = -0.072 + 0.0096 \cdot money - 0.149 \cdot sex + 1.919 \cdot love + 0.476 \cdot work$$

The residual standard error is 1.058 on 34 df.
The multiple R-squared is 0.7102 and the adjusted R-squared is 0.6761.
The F-statistic p-value is $9.364 \cdot 10^{-9}$.
The two variables significant at the 0.05 level are `love` and `work`.

The model has rather high R-squared values with a very small F-statistic p-value which indicates it explains most of the variance in happiness and the model is statistically significant and thus useful to predict happiness using the input variables. Additionally, the Q-Q plot residuals seem to generally follow the line which suggests that the residuals are indeed normally distributed. However, there seems to be a slight curve in the fitted vs residuals plot which suggests a possible non-linear relationship. Also, the scale-location plot suggests that variance isn't constant. The residuals vs leverage plot also suggests that point 36 could be an outlier but it seems to be on the border of being influential.

(c)

The plot of the square root of the absolute residuals versus the fitted values shows a non-constant spread, with some potential increase in variability as fitted values increase. The Breusch-Pagan test yielded a p-value of 0.03069, which is below the 5% significance level, suggesting that the homoscedasticity assumption is violated. This indicates heteroscedasticity, meaning the model's residual variance is not constant.
The Q-Q plot of residuals generally follows the line, indicating approximate normality. The histogram of residuals shows a roughly symmetric bell shape, supporting this assumption. The Shapiro-Wilk normality test has a p-value of 0.8749, which is well above 0.05, indicating no significant deviation from normality.
The leverage points identified (indices 2, 6, 7, and 10) have high leverage, which means these observations have the potential to unduly influence the model's predictions.
The largest absolute studentized residual is -3.276 for observation 36 and no studentized residuals exceed the Bonferroni critical value of 3.510854, suggesting no observations are significant outliers at the 5% significance

level.
No observations exceed the threshold for Cook's distance, suggesting no highly influential points.

(d)

To address the nonconstant variance issue I would suggest doing weighted least squares approach. The other assumptions/diagnostics are fine so I don't think anything else is necessary. The final model from 20 rounds of WLS is:

$$happy = 0.3135 + 0.0058 \cdot money - 0.1328 \cdot sex + 1.8671 \cdot love + 0.4698 \cdot work$$

(e)

Kathy's final model was as follows:

$$\begin{aligned}
happy = & -33.52 + 0.228 \cdot money + 25.74 \cdot sex + 17.23 \cdot love + 5.71 \cdot work \\
& - 0.140 \cdot (money \times sex) - 0.110 \cdot (money \times love) + 0.009 \cdot (money \times work) \\
& - 11.69 \cdot (sex \times love) - 4.97 \cdot (sex \times work) - 2.41 \cdot (love \times work) \\
& + 0.069 \cdot (money \times sex \times love) + 2.07 \cdot (sex \times love \times work)
\end{aligned}$$

The residual standard error is 0.8535 on 26 df.
The multiple R-squared is 0.8559 and the adjusted R-squared is 0.7894.
The F-statistic p-value is $4.832 \cdot 10^{-8}$.
The variables significant at the 0.05 level are `money`, `sex`, `love`, `work`, and the interaction terms between; `money` and `love`, `sex` and `love`, `sex` and `work`, `love` and `work`, and `sex`, `love`, and 'work.
This model seems to be better at predicting happiness than Jack's model as it has higher R-squared values with a very significant F-statistic p-value and also has a lot more significant variables. However, with this many variables the model could be too complex and overfitting as a result. Validation on testing data is necessary.

(f)

The plot of square root of absolute residuals versus fitted values shows a fairly random scatter, suggesting no strong evidence of heteroscedasticity. The Breusch-Pagan test returned a p-value of 0.551, which is well above 0.05, indicating no significant evidence of heteroscedasticity.
The Q-Q plot of residuals follows the 45-degree line fairly closely, suggesting approximate normality, with minor deviations at the tails. The histogram of residuals shows a roughly symmetric shape, supporting this assumption. The Shapiro-Wilk test has a p-value of 0.3524 and the Durbin-Watson test gives a p-value of 0.3092, indicating not enough evidence to suggest the residuals aren't normally distributed.
Observations with high leverage were identified (indices 1, 15, 19, and 36), indicating these observations have potential to influence the model substantially.
The largest absolute studentized residual is 2.77 for observation 11 and no studentized residuals exceed the Bonferroni critical value of 3.510854, meaning no observations are identified as significant outliers.
Observation 36 is an influential point, so it should be carefully considered as it may significantly impact model estimates.

(g)

The F-statistic p-value is 0.01005 which is significant and suggests that Kathy's model provides a significantly better fit to the data than Jack's model, as it captures additional variance through its interaction terms.
Kathy's model has a significantly lower training error mean and variance compared to Jack's model, suggesting it fits the training data much better. This makes sense, since Kathy's model is more complex and should capture more interactions among predictors. Jack's model has a lower testing error compared to Kathy's model, both in terms of mean and variance. This suggests that Kathy's model is overfitting the data, resulting

in poor performance on new data and shows inability to generalize.

The t-test for training error yields a highly significant p-value, indicating that the difference in training errors between Jack's and Kathy's models is statistically significant. The t-test for testing errors results in a p-value of 0.05492, which is close to 0.05 but not significant, suggesting that the difference in testing errors is not statistically significant. The Wilcoxon test for training errors gives an extremely small p-value, suggesting a significant difference in training error between the models. The Wilcoxon test yields a p-value of 0.1696, indicating no statistically significant difference in testing error between the models.

Overall I would prefer Jack's model as it seems to perform better when generalizing to new data. However, it moreso depends on the purpose of the model and whether it needs to be capable of generalization. I would suggest to experiment with (e.g. cross-validation or LASSO) adding more terms/complexity to the model to see if there is a better middle point between the complexity of the two models to optimize testing performance and complexity.

(h)

My advice would be to focus on relationships (`love`) and finding a fulfilling job (`work`) in that order, since these are the most significant and highest magnitude variables within the model. However, since the sample size of the data isn't huge and the sample is a very specific demographic (University of Chicago MBA students) it may be inappropriate to apply the results to the general population or other demographics which aren't similar to the sample.

Appendix

(a)

```
### Read the data
library(faraway)
```

```
## Warning: package 'faraway' was built under R version 4.3.3
```

```
head(teengamb);
```

```
##   sex status income verbal gamble
## 1   1     51   2.00      8    0.0
## 2   1     28   2.50      8    0.0
## 3   1     37   2.00      6    0.0
## 4   1     28   7.00      4    7.3
## 5   1     65   2.00      8   19.6
## 6   1     61   3.47      6    0.1
```

```
## (a) Please modify the R codes in HW#4 for  Exploratory data analysis
dim(happy)
```

```
## [1] 39  5
```

```
summary(happy)
```

```
##      happy            money             sex              love      
##  Min.   : 2.000   Min.   :  0.00   Min.   :0.0000   Min.   :1.000  
##  1st Qu.: 5.000   1st Qu.: 42.50   1st Qu.:0.0000   1st Qu.:2.000  
##  Median : 7.000   Median : 50.00   Median :1.0000   Median :3.000  
##  Mean   : 6.744   Mean   : 62.15   Mean   :0.6923   Mean   :2.462  
##  3rd Qu.: 8.000   3rd Qu.: 78.00   3rd Qu.:1.0000   3rd Qu.:3.000  
##  Max.   :10.000   Max.   :175.00   Max.   :1.0000   Max.   :3.000  
##       work      
##  Min.   :1.000  
```

```
##  1st Qu.:3.000
##  Median :4.000
##  Mean   :3.359
##  3rd Qu.:4.000
##  Max.   :5.000
## histogram
hist(happy$happy)
```

Histogram of happy$happy
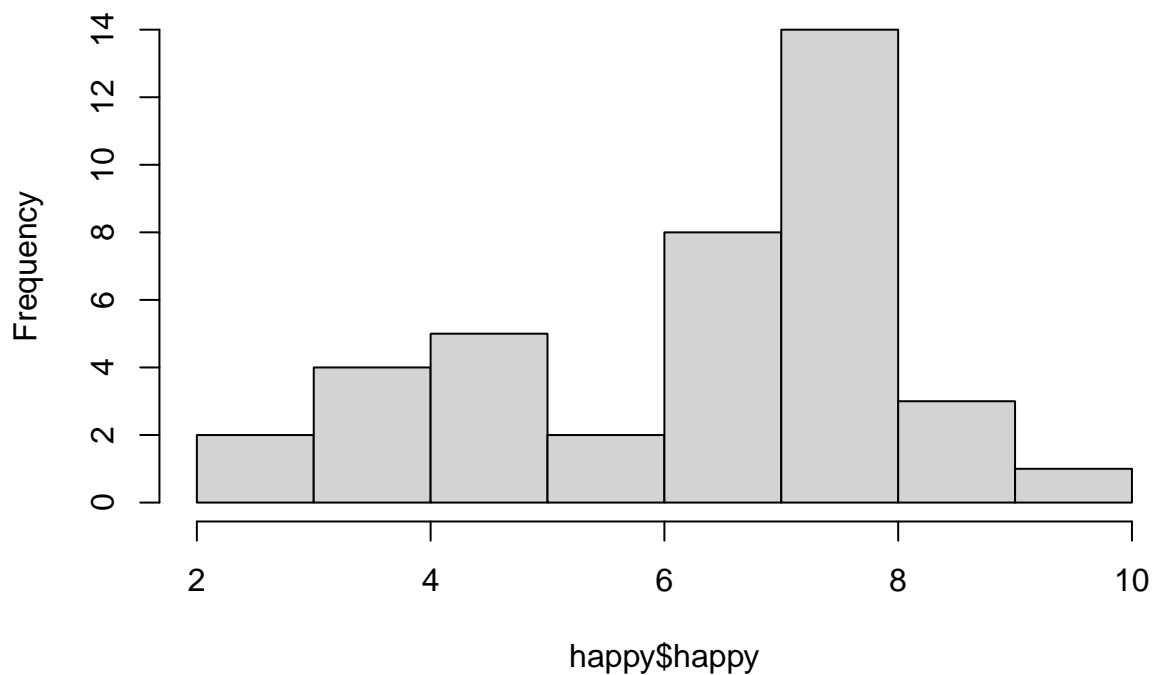

```
## plot
library(lattice)
```

```
## Warning: package 'lattice' was built under R version 4.3.3
```

```
##
## Attaching package: 'lattice'
```
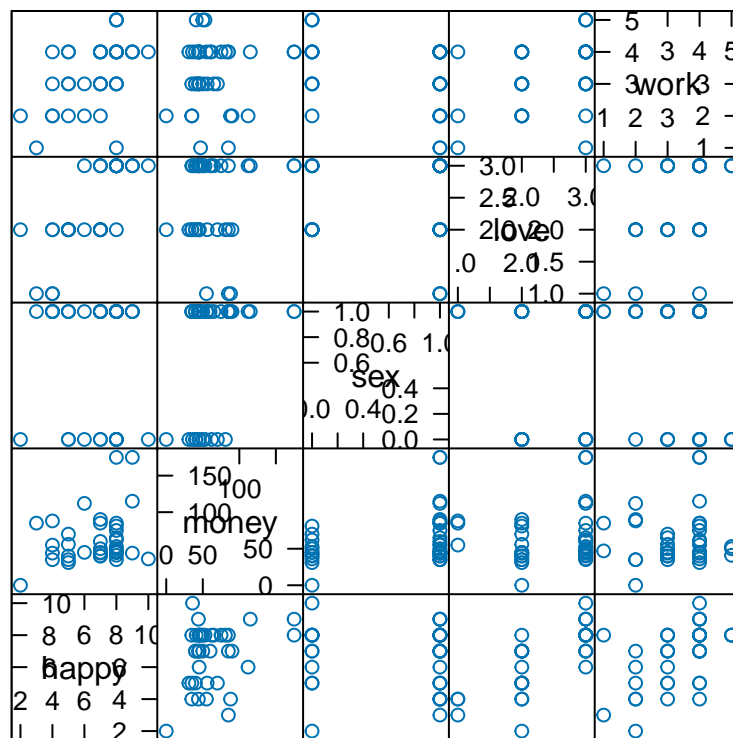
```
## The following object is masked from 'package:faraway':
##
##     melanoma
```

```
splom(happy)
```

Scatter Plot Matrix

```
## correlation matrix
round(cor(happy),2)
```

```
##       happy money   sex love  work
## happy  1.00  0.27 -0.03 0.78  0.54
## money  0.27  1.00  0.31 0.13  0.07
## sex   -0.03  0.31  1.00 0.05 -0.32
## love   0.78  0.13  0.05 1.00  0.39
## work   0.54  0.07 -0.32 0.39  1.00
```
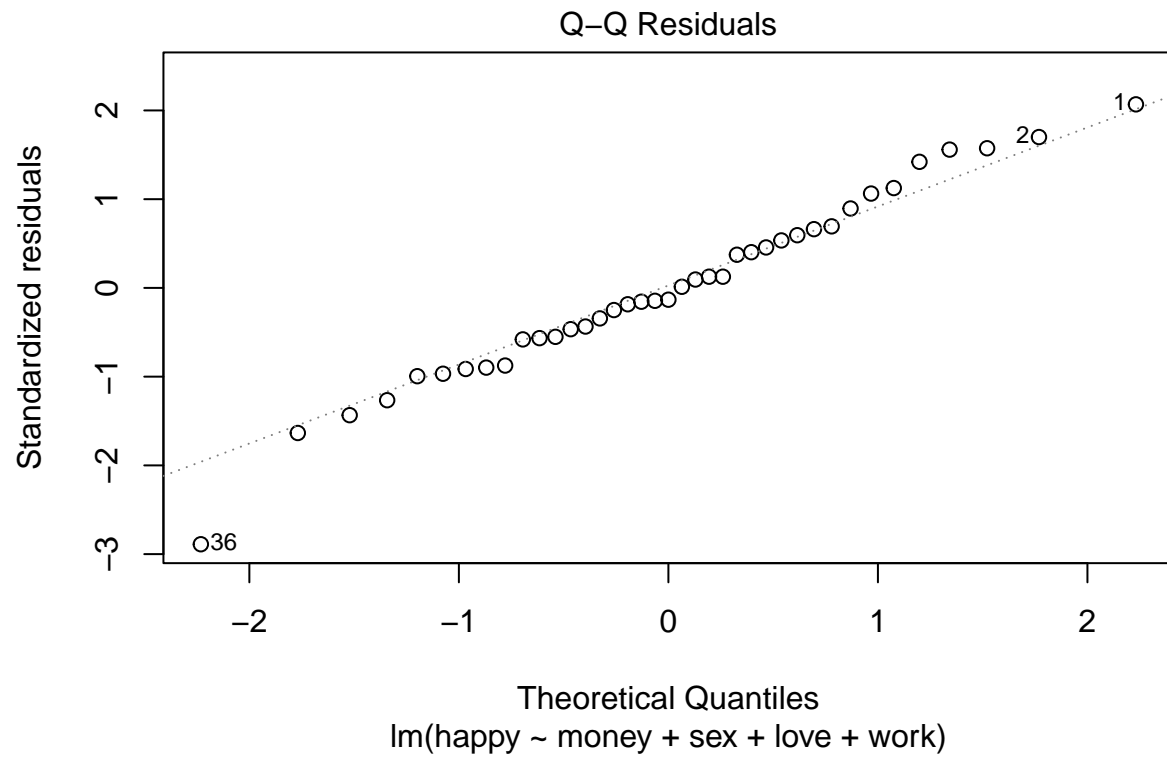
(b)

```
### (b) Jack's model
Jackmod <- lm(happy ~ money + sex + love + work, data = happy)
summary(Jackmod)
```
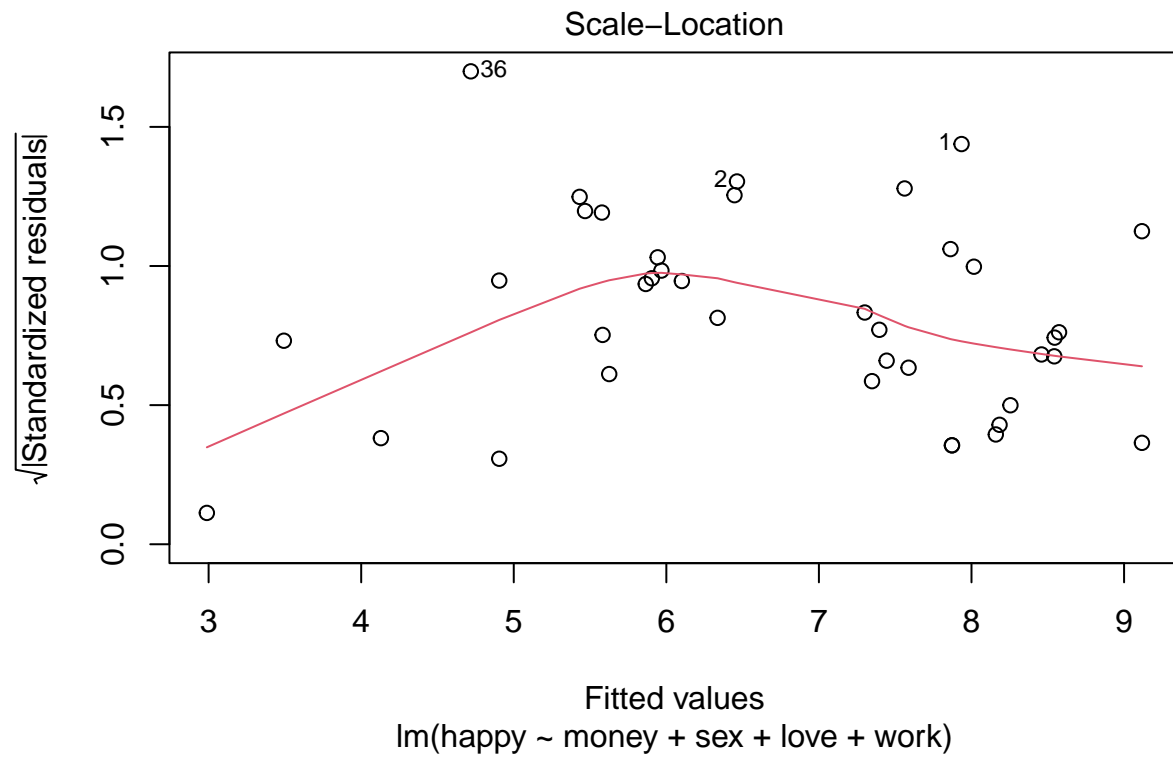
```
##
## Call:
## lm(formula = happy ~ money + sex + love + work, data = happy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7186 -0.5779 -0.1172  0.6340  2.0651
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

5

```
## (Intercept) -0.072081    0.852543   -0.085    0.9331
## money        0.009578    0.005213    1.837    0.0749 .
## sex         -0.149008    0.418525   -0.356    0.7240
## love         1.919279    0.295451    6.496 1.97e-07 ***
## work         0.476079    0.199389    2.388    0.0227 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.058 on 34 degrees of freedom
## Multiple R-squared:  0.7102, Adjusted R-squared:  0.6761
## F-statistic: 20.83 on 4 and 34 DF,  p-value: 9.364e-09
```

```
plot(Jackmod)
```

Residuals vs Fitted

Fitted values
lm(happy ~ money + sex + love + work)

Q–Q Residuals

Theoretical Quantiles
lm(happy ~ money + sex + love + work)

Scale–Location

√|Standardized residuals|

Fitted values
lm(happy ~ money + sex + love + work)

Residuals vs Leverage

lm(happy ~ money + sex + love + work)

(c)

```
### (c) Model diagnostic for Jack's model
## 1. Check the constant variance assumption for the errors
plot(fitted(Jackmod),sqrt(abs(residuals(Jackmod))), xlab="Fitted",
     ylab=expression(sqrt(hat(epsilon))))
```
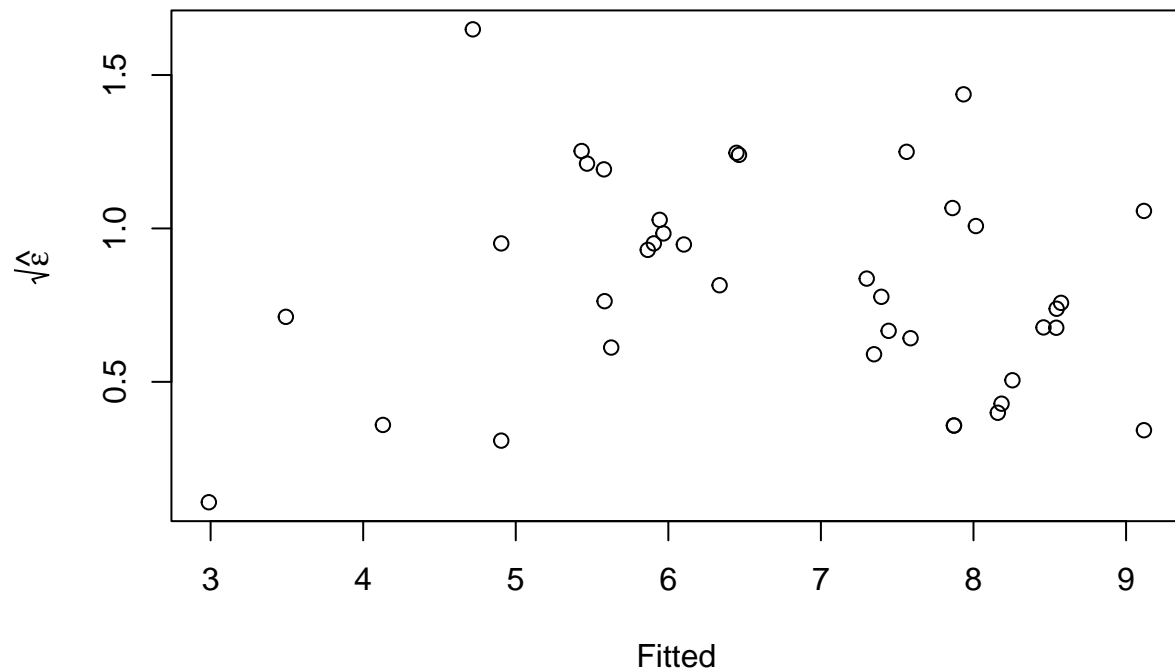
```r
library("lmtest")
```

```
## Warning: package 'lmtest' was built under R version 4.3.2
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
bptest(Jackmod)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  Jackmod
## BP = 10.658, df = 4, p-value = 0.03069
```
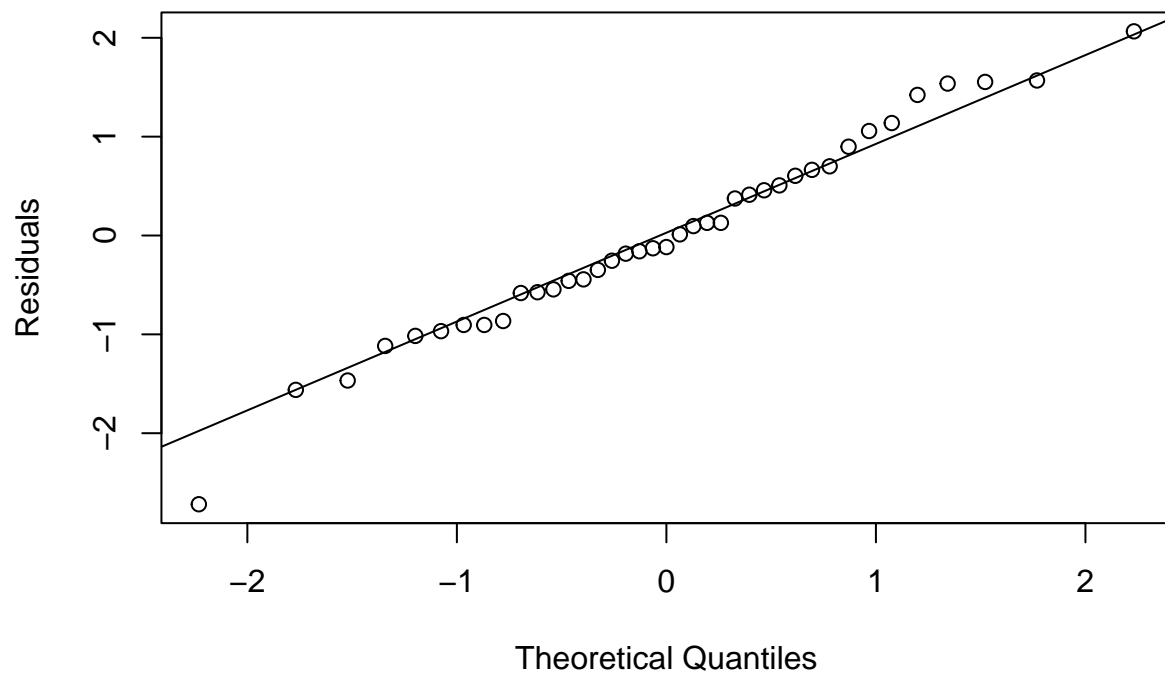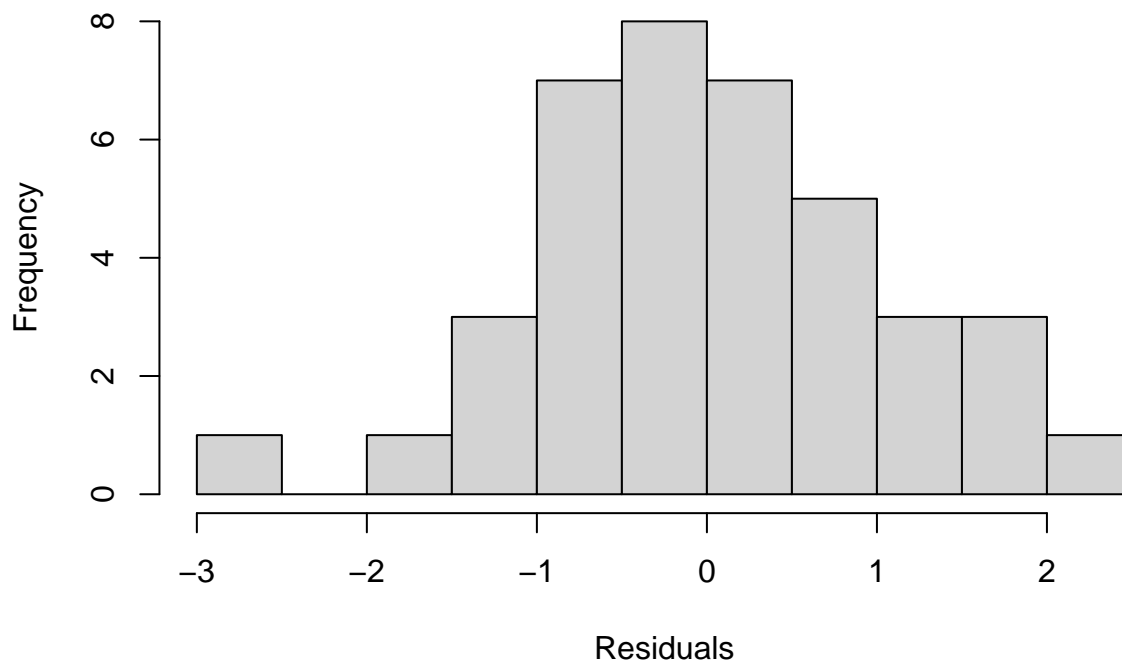
```r
## 2.  Check the normality assumption
qqnorm(residuals(Jackmod),ylab="Residuals",main="")
qqline(residuals(Jackmod))
```

```
hist(residuals(Jackmod),xlab="Residuals",main="")
```

```
shapiro.test(residuals(Jackmod))
```

```
##
##   Shapiro-Wilk normality test
##
## data:  residuals(Jackmod)
## W = 0.98507, p-value = 0.8749
```

```
dwtest(Jackmod)
```

```
##
##   Durbin-Watson test
##
## data:  Jackmod
## DW = 1.5411, p-value = 0.06815
## alternative hypothesis: true autocorrelation is greater than 0
```

```
## 3. Check for large leverage points
hatv <- hatvalues(Jackmod)
head(hatv)
```

```
##          1          2          3          4          5          6
## 0.11033549 0.27025953 0.12718610 0.09136326 0.19914885 0.30443392
```

```
n <- nrow(happy)
p <- sum(hatv)  ## there are p=5 variables including intercept
as.numeric(which(hatv>2*p/n))
```

```
## [1]  2  6  7 10
```

```r
## 4. Check for outliers
# largest studentized residuals in the absolute value
stud <- rstudent(Jackmod)
stud[which.max(abs(stud))]
```

```
##        36
## -3.276013
```

```r
# Compute Bonferroni critical value, qt(1-alpha/(2n), n-p)
p <- 5 # there are p=5 beta values including intercept
qt(1- 0.05/(n*2),n-p)
```

```
## [1] 3.510854
```

```r
# Get all outliers (if it is empty, then no outliers)
which(abs(stud)>qt(1-0.05/(n*2),n-p))
```

```
## named integer(0)
```

```r
## 5. Check for influential points (if it is empty, then no influential points)
cook <- cooks.distance(Jackmod)
which(cook > qf(0.5, p, n-p))
```

```
## named integer(0)
```

(d)

```r
loop = 20;
happy.wls <- lm(happy ~ money + sex + love + work, data = happy)
w = rep(1, dim(happy)[1]);
result <- NULL;
for (i in 1:loop){
 sdfit  <- lm( abs(resid(happy.wls)) ~  money + sex + love + work, data = happy);
 w  <- 1/ ( fitted(sdfit) )^2;
 happy.wls <- lm(happy ~ money + sex + love + work, data = happy, weights = w);
 result <- rbind(result, coef(happy.wls));
}
result
```

```
##        (Intercept)       money        sex      love      work
##  [1,]    0.2166680 0.006788997 -0.1606195 1.896653 0.4622175
##  [2,]    0.2934960 0.005960597 -0.1407222 1.874298 0.4679040
##  [3,]    0.3093835 0.005797024 -0.1348329 1.868650 0.4693756
##  [4,]    0.3126483 0.005760431 -0.1332852 1.867439 0.4697045
##  [5,]    0.3133341 0.005752347 -0.1329228 1.867176 0.4697792
##  [6,]    0.3134815 0.005750577 -0.1328418 1.867119 0.4697958
##  [7,]    0.3135135 0.005750190 -0.1328240 1.867107 0.4697994
##  [8,]    0.3135205 0.005750106 -0.1328201 1.867104 0.4698002
##  [9,]    0.3135220 0.005750087 -0.1328192 1.867104 0.4698004
## [10,]    0.3135224 0.005750083 -0.1328191 1.867103 0.4698004
## [11,]    0.3135224 0.005750082 -0.1328190 1.867103 0.4698004
## [12,]    0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
## [13,]    0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
## [14,]    0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
```
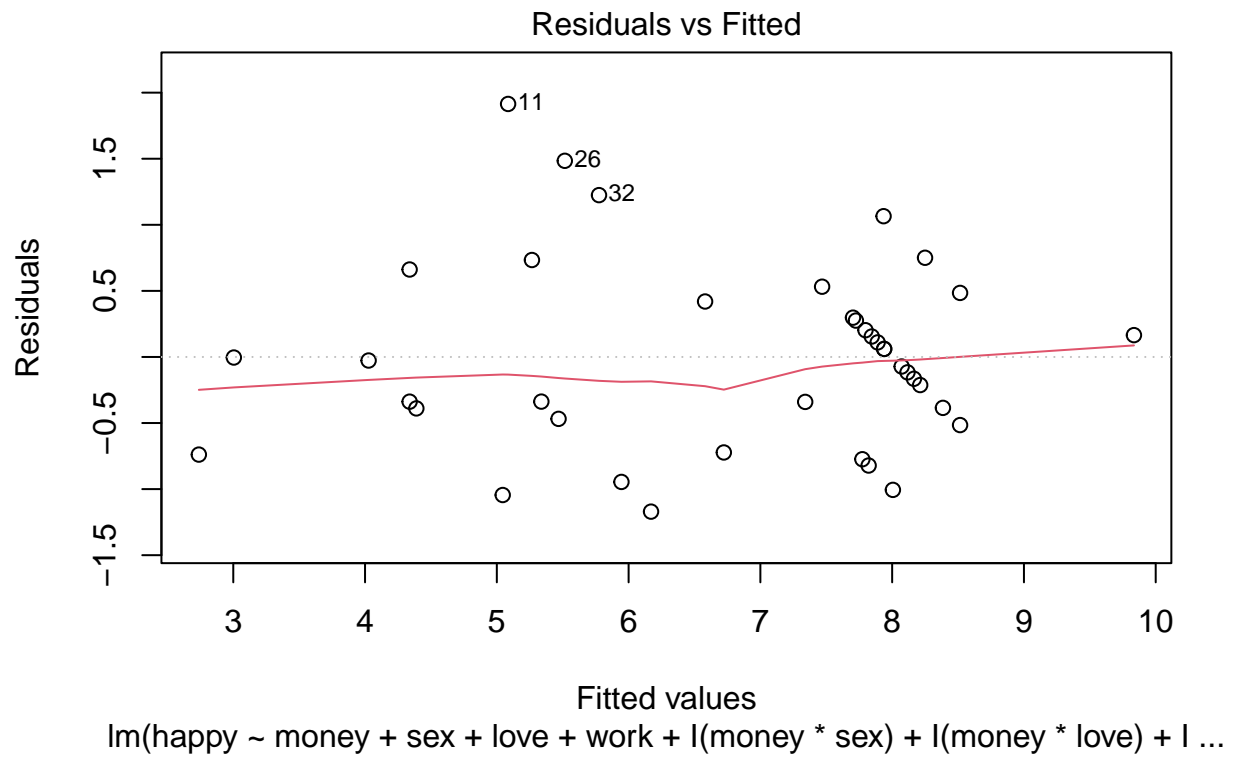
```
## [15,]   0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
## [16,]   0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
## [17,]   0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
## [18,]   0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
## [19,]   0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
## [20,]   0.3135225 0.005750082 -0.1328190 1.867103 0.4698004
```
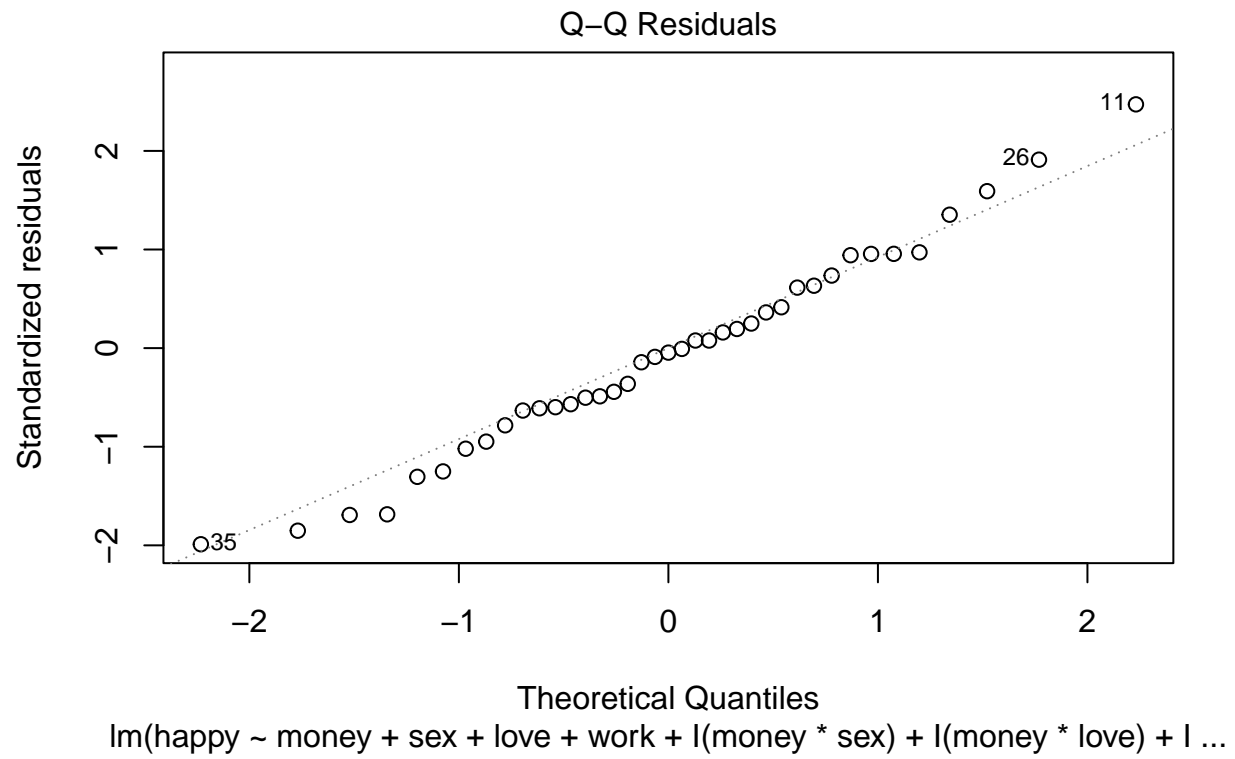
(e)

```
### (e) Kathy's model
# it can also be derived as follows
# Kathymod.full <- lm(happy ~ money*sex*love*work, data = happy)
# Kathymod <- step(Kathymod.full);
Kathymod <- lm(happy ~ money + sex + love + work + I(money*sex) +
               I(money*love) + I(money*work) + I(sex*love) + I(sex*work) + I(love*work)
               ↪
               + I(money*sex*love) + I(sex*love*work), data = happy)
summary(Kathymod);
```
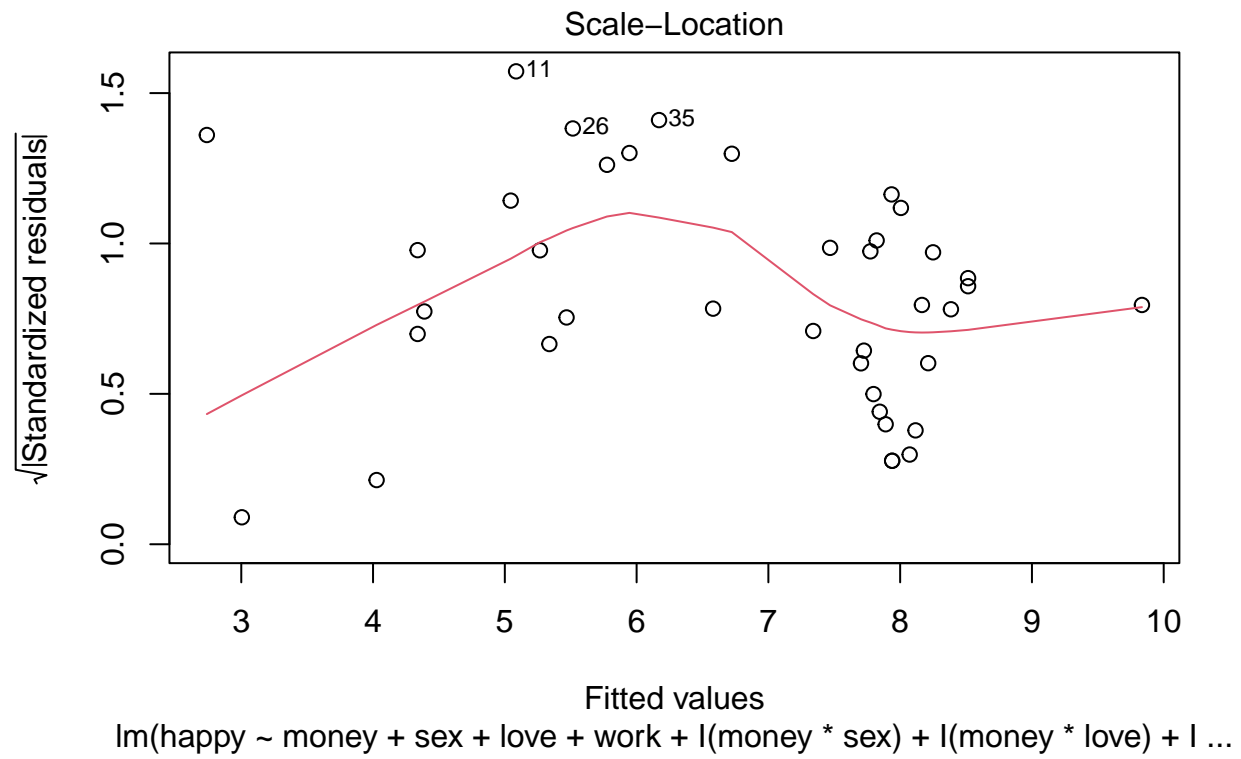
```
##
## Call:
## lm(formula = happy ~ money + sex + love + work + I(money * sex) +
##     I(money * love) + I(money * work) + I(sex * love) + I(sex *
##     work) + I(love * work) + I(money * sex * love) + I(sex *
##     love * work), data = happy)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -1.17034 -0.42886 -0.02696  0.35843  1.91431
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -33.521928   9.756408  -3.436 0.001995 **
## money                 0.227726   0.096778   2.353 0.026462 *
## sex                  25.737309  10.456628   2.461 0.020794 *
## love                 17.229592   4.582430   3.760 0.000872 ***
## work                  5.712229   2.293095   2.491 0.019447 *
## I(money * sex)       -0.139834   0.102817  -1.360 0.185495
## I(money * love)      -0.109560   0.045345  -2.416 0.023008 *
## I(money * work)       0.009183   0.006402   1.434 0.163394
## I(sex * love)       -11.692553   4.730753  -2.472 0.020319 *
## I(sex * work)        -4.967474   2.411944  -2.060 0.049587 *
## I(love * work)       -2.405741   0.967122  -2.488 0.019603 *
## I(money * sex * love) 0.069496   0.046456   1.496 0.146705
## I(sex * love * work)  2.066936   0.998452   2.070 0.048508 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8535 on 26 degrees of freedom
## Multiple R-squared:  0.8559, Adjusted R-squared:  0.7894
## F-statistic: 12.87 on 12 and 26 DF,  p-value: 4.832e-08
```

```
plot(Kathymod)
```

Residuals vs Fitted

Residuals

Fitted values
lm(happy ~ money + sex + love + work + I(money * sex) + I(money * love) + I ...

Q–Q Residuals

Standardized residuals

Theoretical Quantiles
lm(happy ~ money + sex + love + work + I(money * sex) + I(money * love) + I ...

Scale–Location

√|Standardized residuals|

Fitted values
lm(happy ~ money + sex + love + work + I(money * sex) + I(money * love) + I ...

Residuals vs Leverage

lm(happy ~ money + sex + love + work + I(money * sex) + I(money * love) + I ...
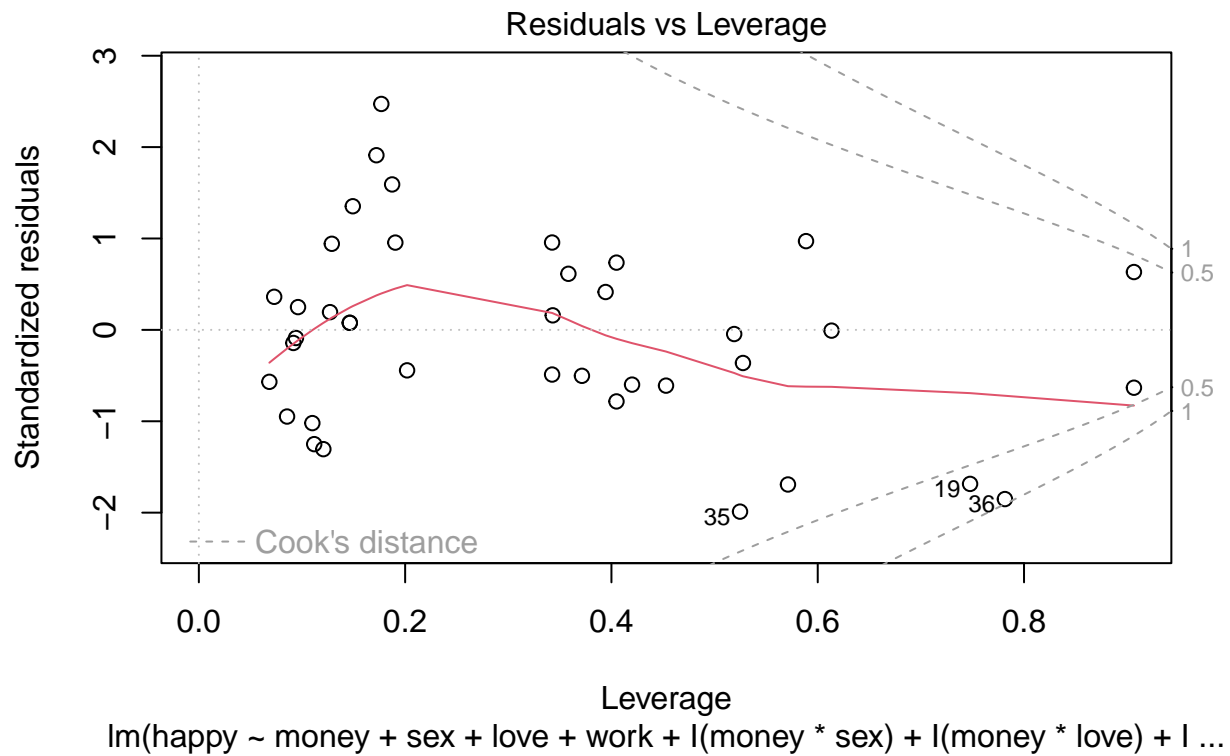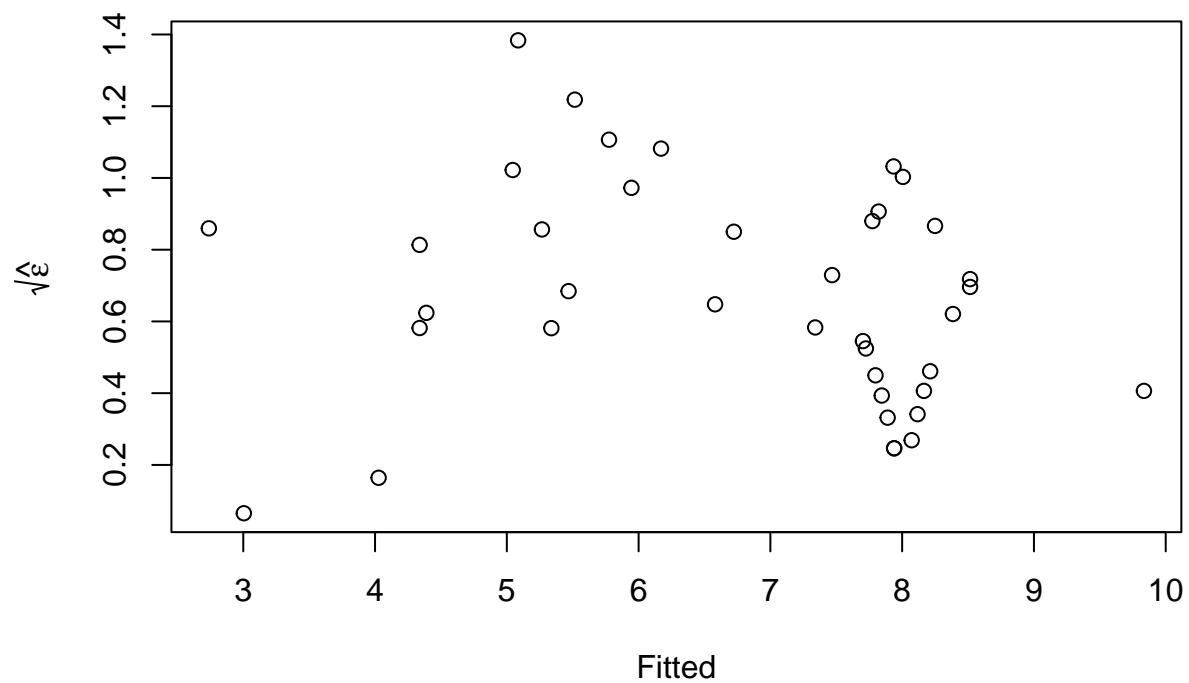
(f)

```
## 1. Check the constant variance assumption for the errors
plot(fitted(Kathymod),sqrt(abs(residuals(Kathymod))), xlab="Fitted",
     ylab=expression(sqrt(hat(epsilon))))
```
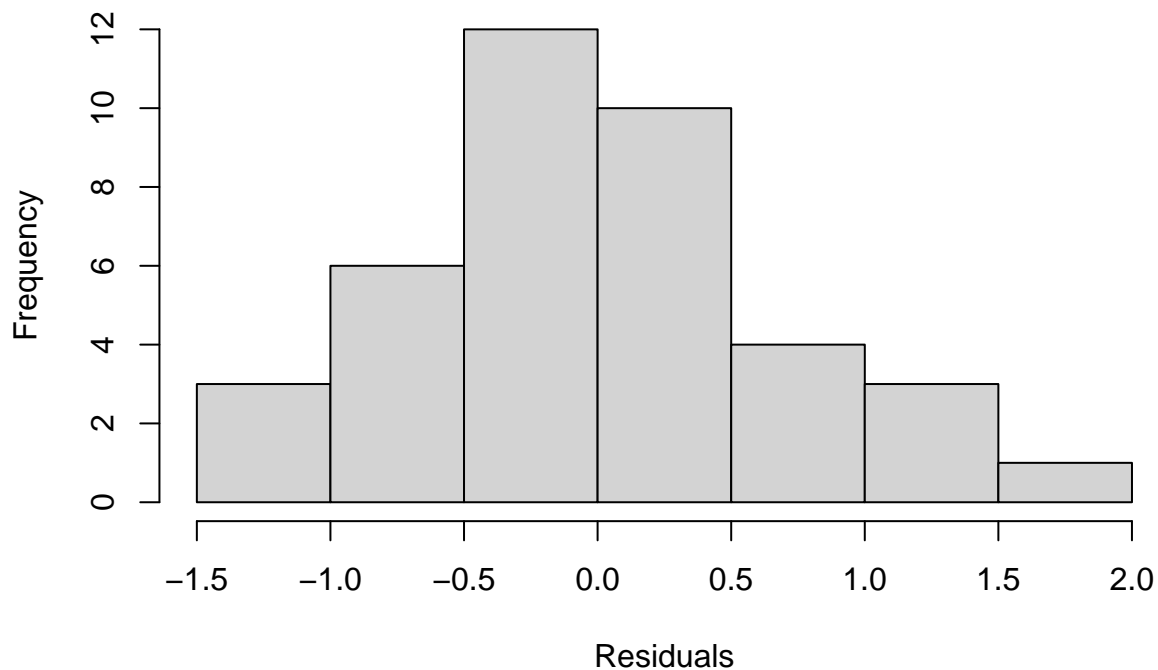
```
library("lmtest")
bptest(Kathymod)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  Kathymod
## BP = 10.744, df = 12, p-value = 0.551
```

```
## 2.  Check the normality assumption
qqnorm(residuals(Kathymod),ylab="Residuals",main="")
qqline(residuals(Kathymod))
```
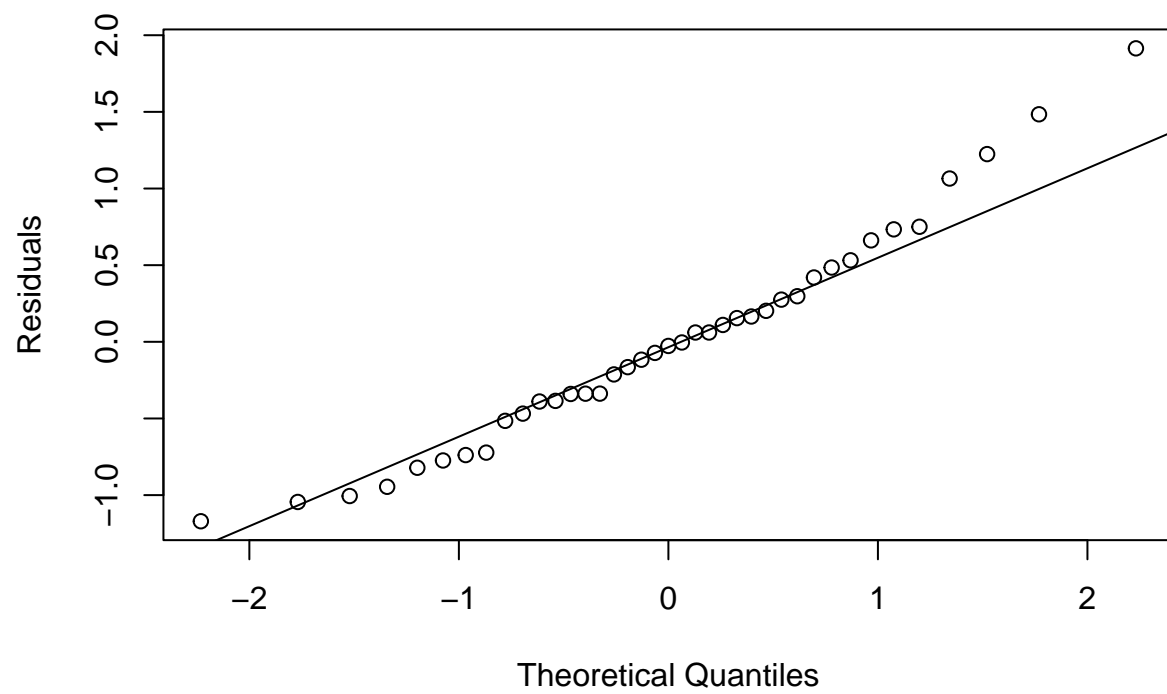
```
hist(residuals(Kathymod),xlab="Residuals",main="")
```

```r
shapiro.test(residuals(Kathymod))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(Kathymod)
## W = 0.96911, p-value = 0.3524
```

```r
dwtest(Kathymod)
```

```
##
##  Durbin-Watson test
##
## data:  Kathymod
## DW = 1.904, p-value = 0.3092
## alternative hypothesis: true autocorrelation is greater than 0
```

```r
## 3. Check for large leverage points
hatv <- hatvalues(Kathymod)
head(hatv)
```

```
##         1         2         3         4         5         6
## 0.9065843 0.5886568 0.3943581 0.1271706 0.4199954 0.4049010
```

```r
n <- nrow(happy)
p <- sum(hatv)  ## there are p=5 variables including intercept
as.numeric(which(hatv>2*p/n))
```

```
## [1]  1 15 19 36
```

```r
## 4. Check for outliers
# largest studentized residuals in the absolute value
stud <- rstudent(Kathymod)
stud[which.max(abs(stud))]
```

```
##       11
## 2.771763
```

```r
# Compute Bonferroni critical value, qt(1-alpha/(2n), n-p)
p <- 5 # there are p=5 beta values including intercept
qt(1- 0.05/(n*2),n-p)
```

```
## [1] 3.510854
```

```r
# Get all outliers (if it is empty, then no outliers)
which(abs(stud)>qt(1-0.05/(n*2),n-p))
```

```
## named integer(0)
```

```r
## 5. Check for influential points (if it is empty, then no influential points)
cook <- cooks.distance(Kathymod)
which(cook > qf(0.5, p, n-p))
```

```
## 36
## 36
```

(g)

```r
## (g) Comparison of these two models
##  (i) run ANOVA F-test
anova(Jackmod, Kathymod)
```

```
## Analysis of Variance Table
##
## Model 1: happy ~ money + sex + love + work
## Model 2: happy ~ money + sex + love + work + I(money * sex) + I(money *
##     love) + I(money * work) + I(sex * love) + I(sex * work) +
##     I(love * work) + I(money * sex * love) + I(sex * love * work)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     34 38.087
## 2     26 18.940  8    19.147 3.2856 0.01005 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##  (ii) please modify the cross-validation training/testing error codes in HW#4
##   to compare these two models, Jackmod and Kathymod, on your own
set.seed(2353);
B = 100;
TrainErr = NULL;  ## record all training errors
TestErr  = NULL;  ##record all testing errors
n <- nrow(happy)
n1 <- round(0.8 * n)
for (b in 1:B){
  ## with each loop, create training and testing subsets
```

```r
  flagtmp = sort(sample(1:n, n1));
  happy_train_tmp = happy[flagtmp, ];   ###"temporal" training subset
  happy_test_tmp = happy[-flagtmp, ];  ###"temporal" testing subset

  ###### Fit these two (2) models to the "temporal" training subset
  Jackmod_tmp <- lm(happy ~ money + sex + love + work, data = happy_train_tmp)
  Kathymod_tmp <- lm(happy ~ money + sex + love + work + I(money * sex) + I(money * love)
↪  + I(money * work) + I(sex * love) +
                      I(sex * work) + I(love * work) + I(money * sex * love) + I(sex *
                  ↪  love * work), data = happy_train_tmp)

  ## Save training errors of these two models for the "temporal" training subset
  TrainErr <- rbind(TrainErr, c(mean(Jackmod_tmp$residuals^2),
↪  mean(Kathymod_tmp$residuals^2)));

  ###### Prediction of these two (2) models to the "temporal" testing subset
  Pred_jack <- predict(Jackmod_tmp, newdata= happy_test_tmp[,2:5] );
  Pred_kathy <- predict(Kathymod_tmp, newdata= happy_test_tmp[,2:5] );
  TestErr <- rbind(TestErr, c(mean(Pred_jack - happy_test_tmp[,1])^2, mean(Pred_kathy -
↪  happy_test_tmp[,1])^2 ));
}
```

```
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
## Warning in predict.lm(Kathymod_tmp, newdata = happy_test_tmp[, 2:5]):
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```r
### Training Error Comaprison of two models
colnames(TrainErr) <- c("Jackmod", "Kathymod")
dim(TrainErr);
```

```
## [1] 100   2
```

```r
round(apply(TrainErr, 2, mean),4);
```

```
##  Jackmod Kathymod
##   0.9257   0.4152
```

```r
round(apply(TrainErr, 2, var),4);
```

```
##  Jackmod Kathymod
```

```
##    0.0194    0.0058
```

```r
### Testing  Error Comparisons of two models
colnames(TestErr) <- c("Jackmod", "Kathymod")
dim(TestErr);
```

```
## [1] 100    2
```

```r
round(apply(TestErr, 2, mean),4);
```

```
##   Jackmod Kathymod
##    0.1962   0.6005
```

```r
round(apply(TestErr, 2, var),4);
```

```
##   Jackmod Kathymod
##    0.0741   4.2672
```

```r
## Statistical difference testing to compare two models
##  We can adopt the paired two-sample test
##   since the same training/test subsets are used for fitting two models
## (a) First, use two-sample t-test for training and testing errors
t.test(TrainErr[,1],TrainErr[,2],paired=T)
```

```
##
##  Paired t-test
##
## data:  TrainErr[, 1] and TrainErr[, 2]
## t = 37.153, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   0.4831991 0.5377229
## sample estimates:
## mean difference
##        0.510461
```

```r
t.test(TestErr[,1],TestErr[,2],paired=T)
```

```
##
##  Paired t-test
##
## data:  TestErr[, 1] and TestErr[, 2]
## t = -1.9425, df = 99, p-value = 0.05492
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   -0.817370709  0.008691932
## sample estimates:
## mean difference
##       -0.4043394
```

```r
## which model has smaller training error? which has smaller testing error?
## Hints: you can check the values
round(apply(TrainErr, 2, mean),4);
```

```
##   Jackmod Kathymod
##    0.9257   0.4152
```

```r
round(apply(TestErr, 2, mean),4);
```

```
##   Jackmod Kathymod
##    0.1962    0.6005
```

```r
## (b) Second, use nonparametric Wilcox test for training and testing errors
wilcox.test(TrainErr[,1],TrainErr[,2],paired=T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  TrainErr[, 1] and TrainErr[, 2]
## V = 5050, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(TestErr[,1],TestErr[,2],paired=T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  TestErr[, 1] and TestErr[, 2]
## V = 2125, p-value = 0.1696
## alternative hypothesis: true location shift is not equal to 0
```