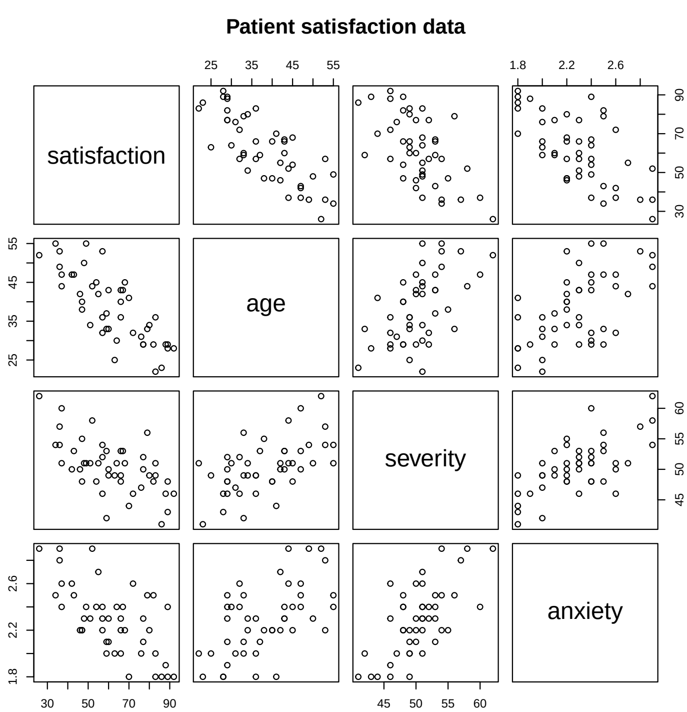
2.5.1: R - Multiple Predictors

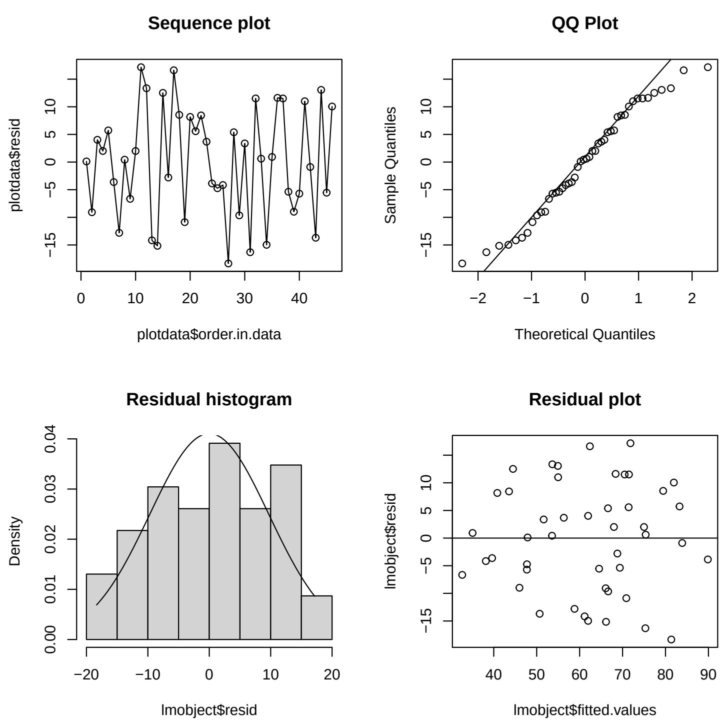
Stat 5100: Dr. Bean

Example: (Exercises 6.15-6.17) A hospital administrator is studying the relation between patient satisfaction (Y, an index) and a patient's age (X_1, in years), severity of illness (X_2, an index), and anxiety level (X_3, an index). Data are reported for 46 randomly selected patients. For all index variables, higher values indicate more (satisfaction, severity, anxiety).

```
# Input the data and take a look at the first few observations
library(stat5100)
data(patient)
head(patient)

##   satisfaction age severity anxiety
## 1           48  50       51     2.3
## 2           57  36       46     2.3
## 3           66  40       48     2.2
## 4           70  41       44     1.8
## 5           89  28       43     1.8
## 6           36  49       54     2.9

# Look at the scatterplot matrix
pairs( ~ satisfaction + age + severity + anxiety, data = patient,
       main = "Patient satisfaction data")
```

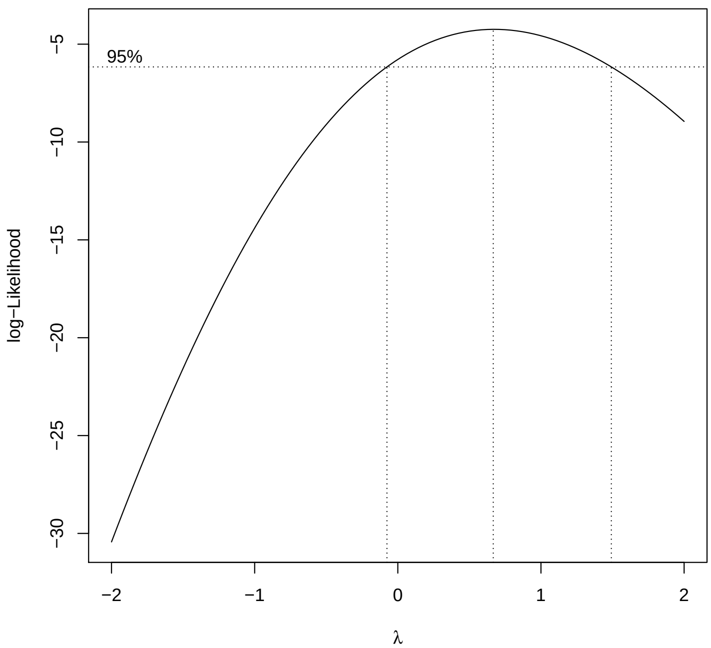


```r
# Fit a regression model
patient_lm <- lm(satisfaction ~ age + severity + anxiety, data = patient)
summary(patient_lm)

##
## Call:
## lm(formula = satisfaction ~ age + severity + anxiety, data = patient)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3524  -6.4230   0.5196   8.3715  17.1601
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
## age          -1.1416     0.2148  -5.315 3.81e-06 ***
## severity     -0.4420     0.4920  -0.898   0.3741
## anxiety     -13.4702     7.0997  -1.897   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.06 on 42 degrees of freedom
## Multiple R-squared:  0.6822,Adjusted R-squared:  0.6595
## F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10

# Check model assumptions
visual_assumptions(patient_lm)
```




```r
# Numerical assumptions
brown_forsythe_lm(patient_lm)

## [1] "Brown-forsythe test for constant variance in the residuals:"
```

```
## [1] "T-statistic: -0.1236, p-value: 0.9022"

cor_normality_lm(patient_lm)

## Correlation test of normality:
##                   resid expected_norm
## resid         1.0000000     0.9885077
## expected_norm 0.9885077     1.0000000
##
## Total observations: 46
## Make sure to consult with table B.6 for your final result.

# Joint 90% intervals for beta1, beta2, and beta3
# Because we don't care about the intercept term, we will set our confidence
# level to 0.86668 which corresponds to a 90% confidence interval for just the
# non-interept betas.
coefficient_confidence_lm(patient_lm, confidence = 0.86668, simul = TRUE)

## lower.est and upper.est are the 96.667% confidence limits.
## The Bonferroni adjustment for simultaneous confidence levels was made.
##                Estimate Std. Error    t value     Pr(>|t|)  lower.est
## (Intercept) 158.4912517 18.1258887  8.7439162 5.260955e-11 118.606836
## age          -1.1416118  0.2147988 -5.3147960 3.810252e-06  -1.614258
## severity     -0.4420043  0.4919657 -0.8984452 3.740702e-01  -1.524531
## anxiety     -13.4701632  7.0996608 -1.8972967 6.467813e-02 -29.092339
##               upper.est
## (Intercept) 198.3756674
## age          -0.6689661
## severity      0.6405229
## anxiety       2.1520128

# Simultaneous 90% prediction limits on two new patients (with Scheffe and Bonferroni)
# with profiles:
#       1. age = 35, severity = 45, anxiety = 2.2
#       2. age = 42, severity = 61, anxiety = 1.8
two_new_patients <- data.frame(age = c(35, 42),
                               severity = c(45, 61),
                               anxiety = c(2.2, 1.8))
simul_prediction_limits(patient_lm, two_new_patients, confidence = 0.90)

##   age severity anxiety     yhat se_yhat_pred  S_lower  S_upper  B_lower
## 1  35       45     2.2 69.01029     10.40495 46.05533 91.96524 48.01224
## 2  42       61     1.8 59.33500     12.67149 31.37972 87.29028 33.76291
##    B_upper
## 1 90.00833
## 2 84.90709

# Would we need a transformation?
library(MASS)
boxcox(patient_lm)
```



This plot above tells us that if we wanted to, we could try a square root transform on the response variable.