Chapter 4

SLR Model Assumptions

Chanwoo Yoo, Division of Advanced Engineering,
Korea National Open University

Contents

1. Background

1. LINE Assumptions

- Linear Function: The mean of the response, $E(Y_i)$, at each value of the predictor, x_i, is a linear function of the x_i.
- Independent: The errors, ϵ_i, are independent.
- Normally Distributed: The errors, ϵ_i, at each value of the predictor, x_i, are normally distributed.
- Equal Variances: The errors, ϵ_i, at each value of the predictor, x_i, have equal variances (denoted σ^2).

2. Consequences of Violation

- All tests and intervals are very sensitive to even minor departures from independence.
- All tests and intervals are sensitive to moderate departures from equal variance.

2. Consequences of Violation

- The hypothesis tests and confidence intervals for β_0 and β_1 are fairly "robust" against departures from normality.
- Prediction intervals are quite sensitive to departures from normality.

2. Residuals vs. Fits Plot

1. Residuals vs. Fits Plot

- Scatter plot of residuals on the y axis and fitted values (estimated responses) on the x axis.
- The plot is used to detect non-linearity, unequal error variances, and outliers.

2. Data: Alcohol Arm

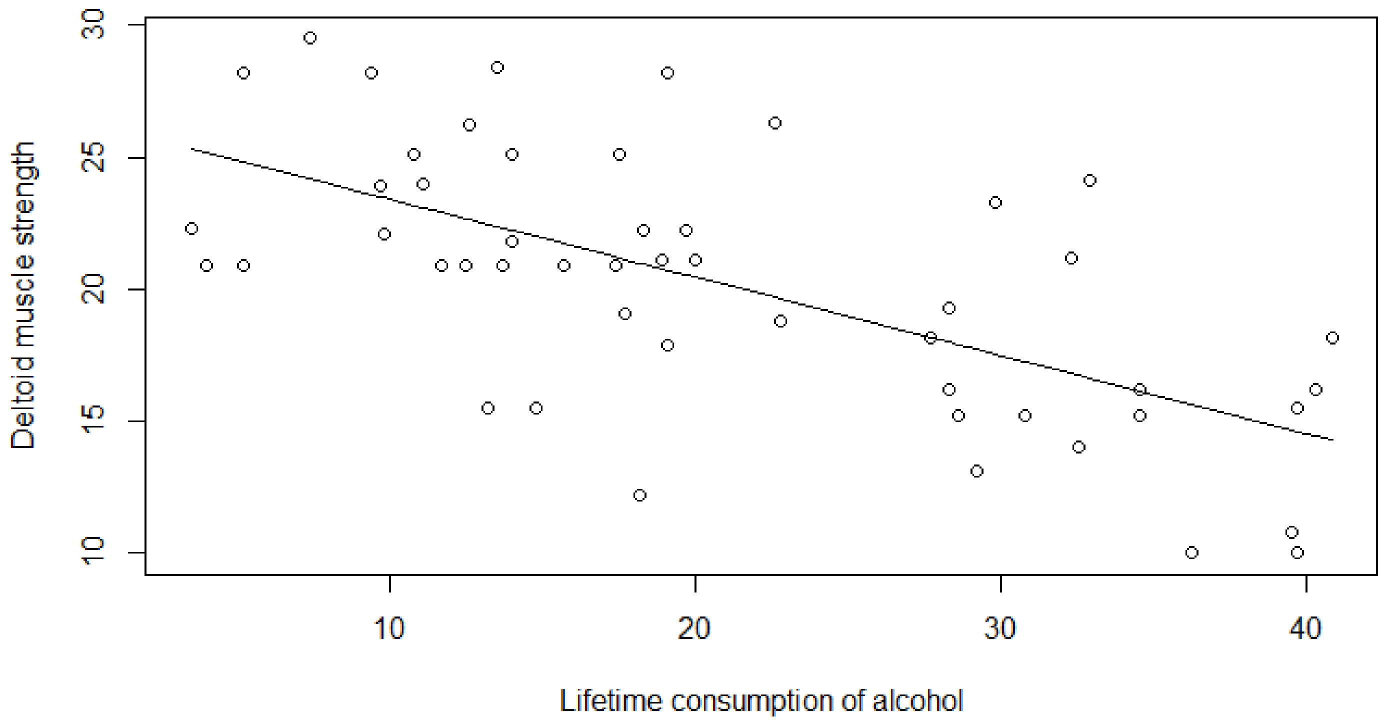
- Data: Alcohol Arm

```
alcoholarm <- read.table("alcoholarm.txt", header=T)
attach(alcoholarm)
model <- lm(strength ~ alcohol)
```

3. Regression Plot

```
plot(x=alcohol, y=strength,
     xlab="Lifetime consumption of alcohol", ylab="Deltoid
muscle strength",
     panel.last = lines(sort(alcohol),
fitted(model)[order(alcohol)]))
```

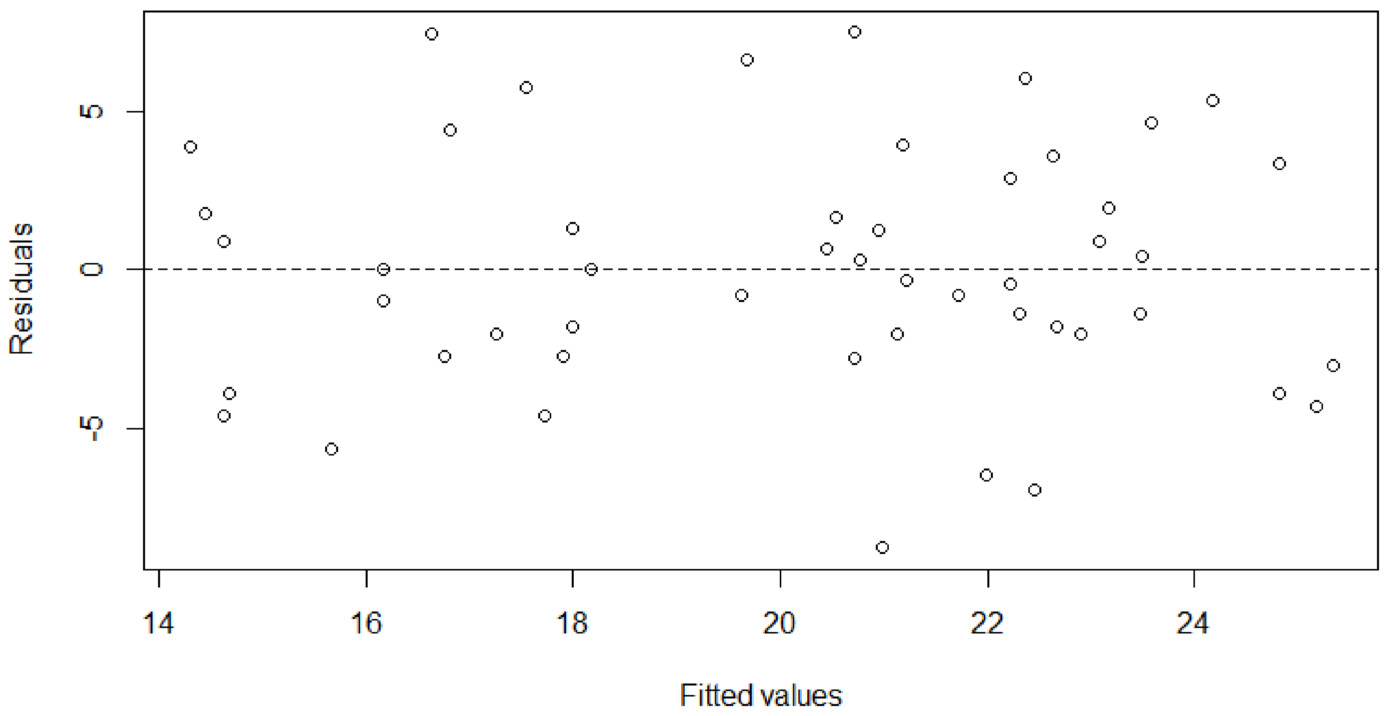
3. Regression Plot



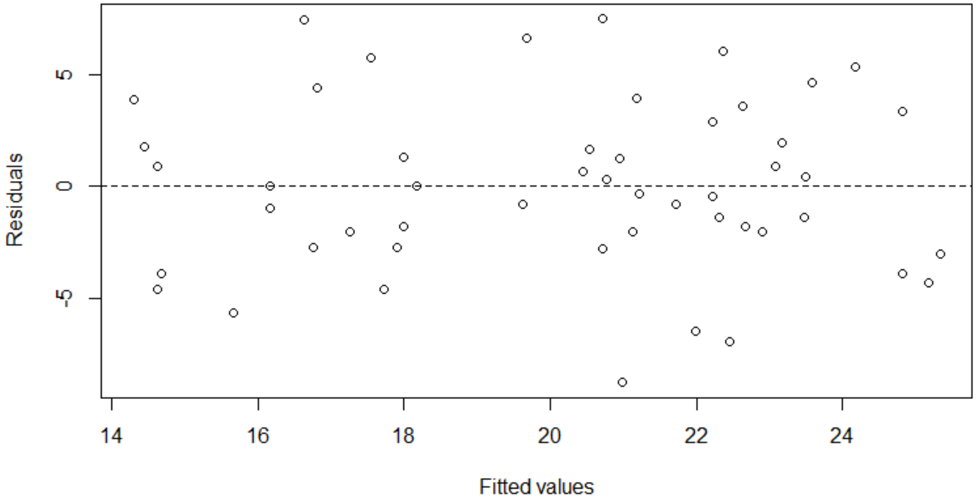
4. Residuals vs. Fits Plot

```
plot(x=fitted(model), y=residuals(model),
     xlab="Fitted values", ylab="Residuals",
     panel.last = abline(h=0, lty=2))
```

4. Residuals vs. Fits Plot

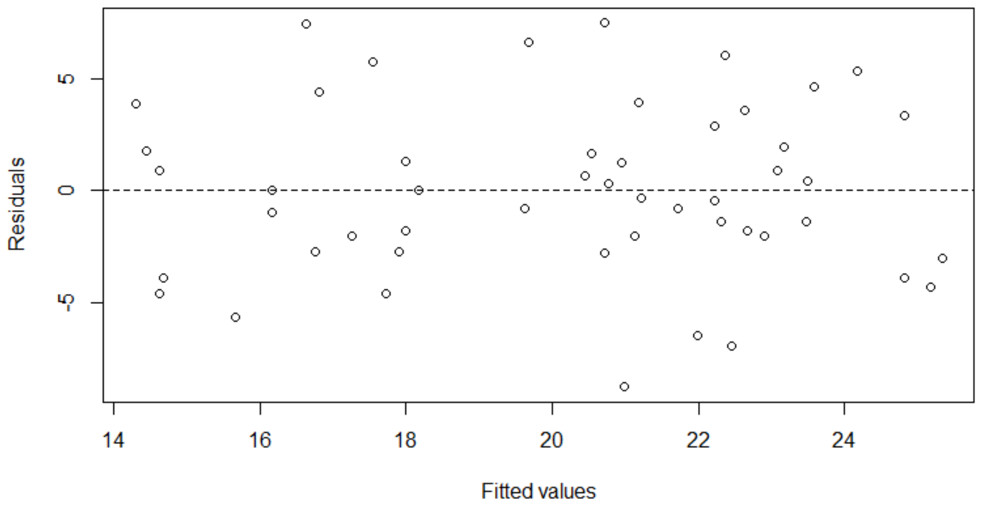


4. Residuals vs. Fits Plot



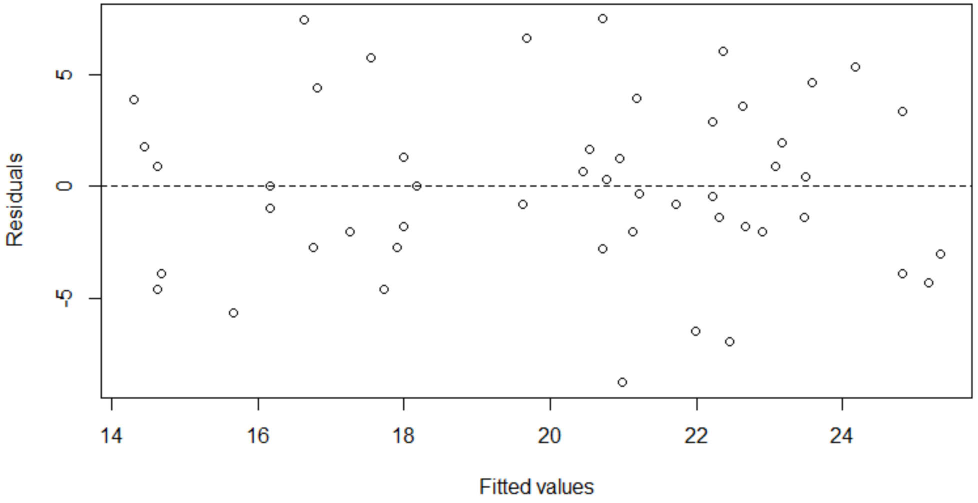
The residuals "bounce randomly" around the residual = 0 line. This suggests that the assumption that the relationship is linear is reasonable.

4. Residuals vs. Fits Plot



The residuals roughly form a "horizontal band" around the residual = 0 line. This suggests that the variances of the error terms are equal.

4. Residuals vs. Fits Plot



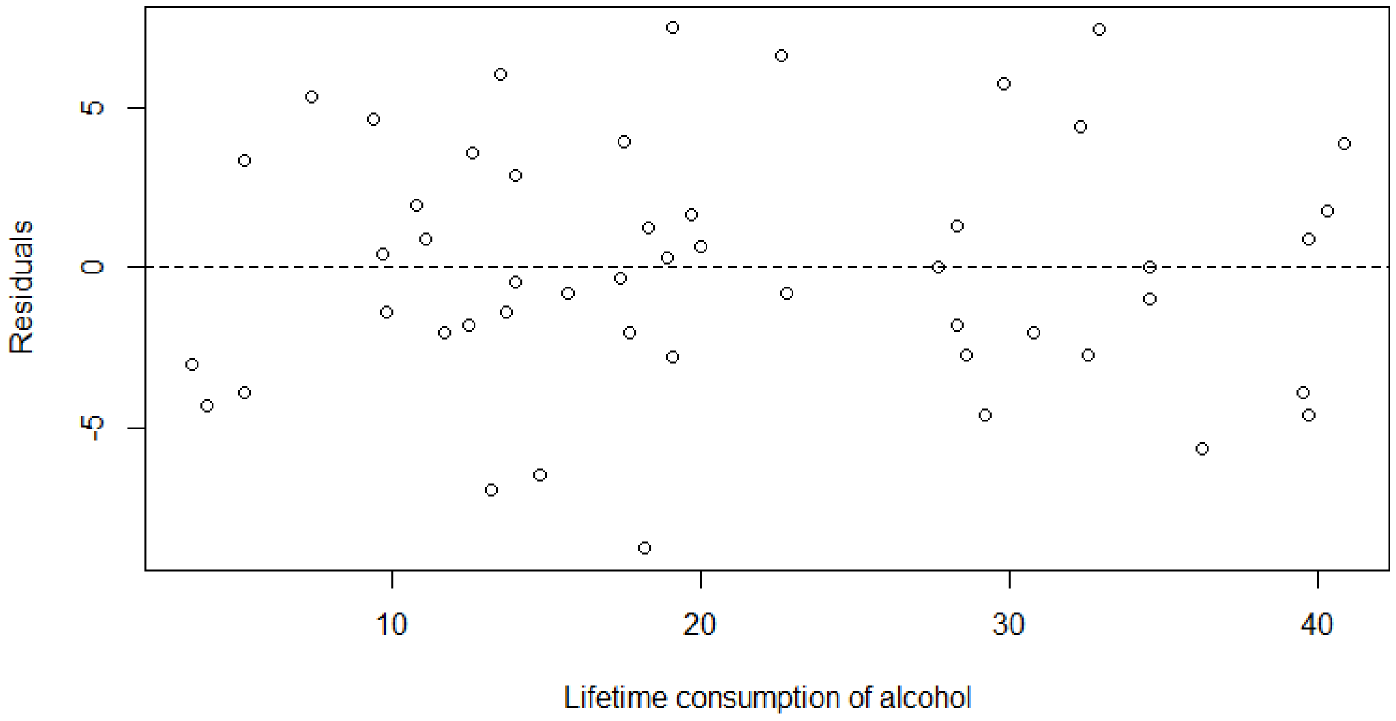
No one residual "stands out" from the basic random pattern of residuals. This suggests that there are no outliers.

3. Residuals vs. Predictor Plot

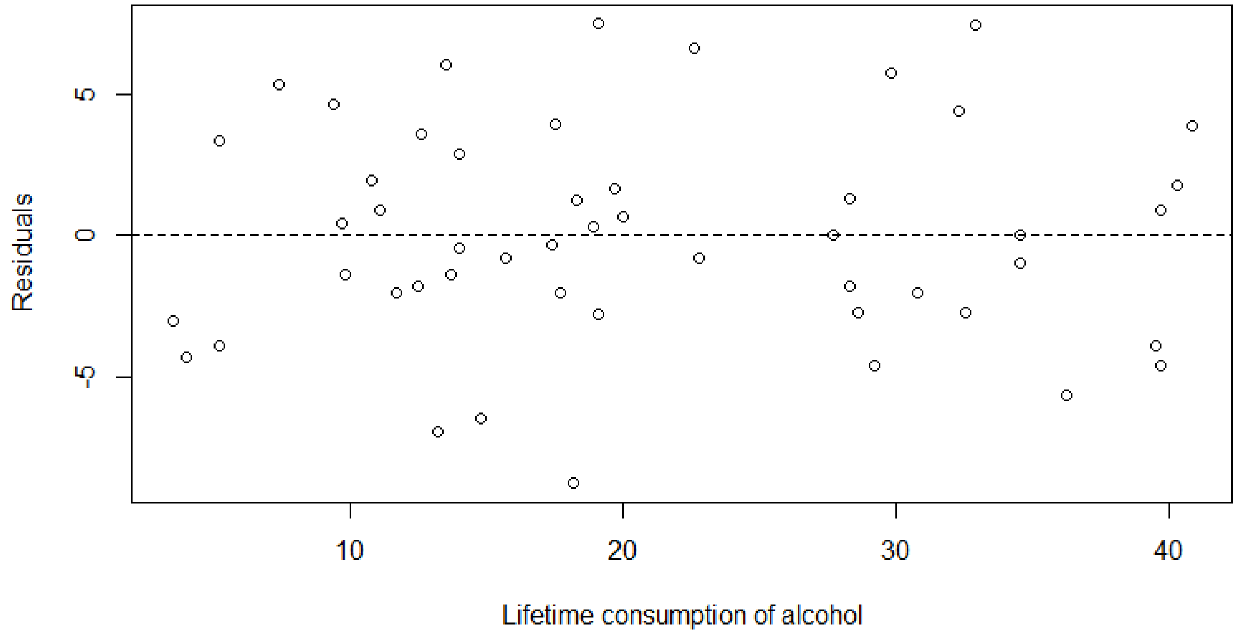
1. Residuals vs. Predictor Plot

```
plot(x=alcohol, y=residuals(model),
     xlab="Lifetime consumption of alcohol", ylab="Residuals",
     panel.last = abline(h=0, lty=2))
```

1. Residuals vs. Predictor Plot

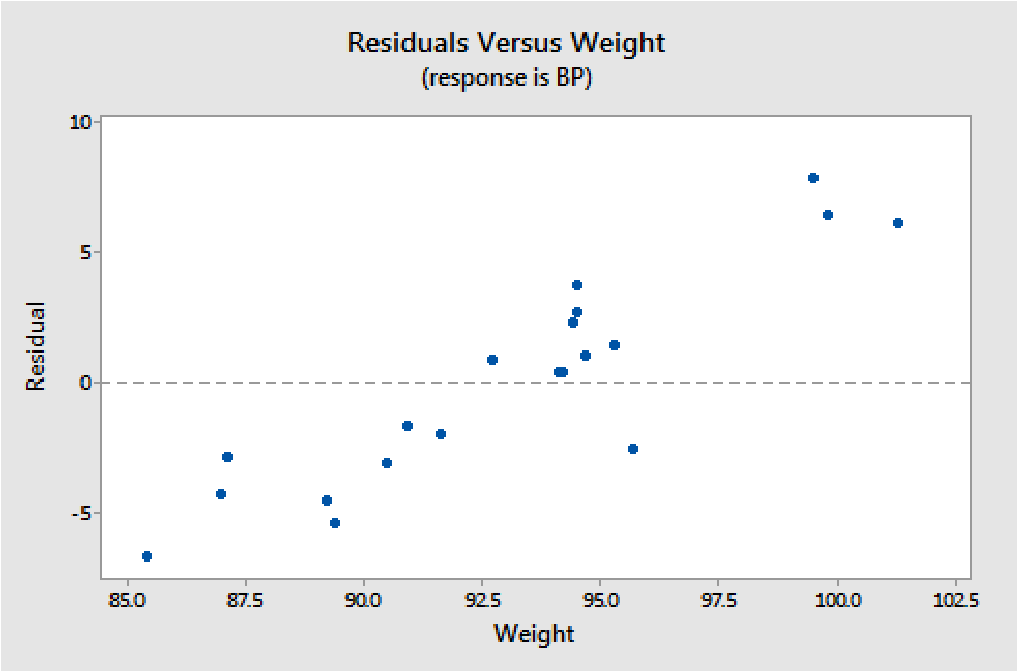


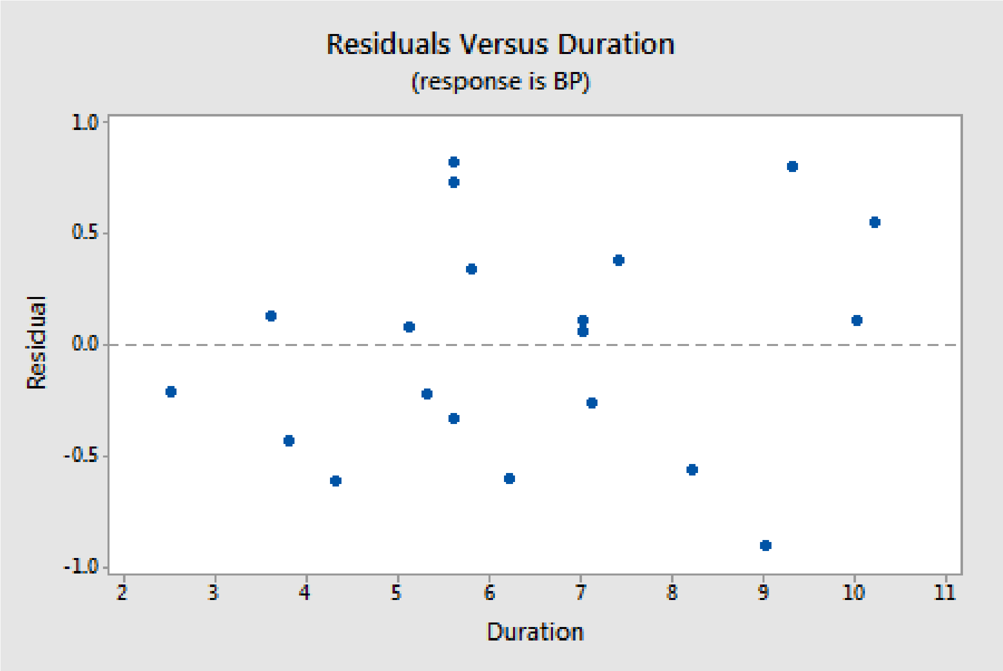
1. Residuals vs. Predictor Plot



For this example, the residuals vs. predictor plot is just a mirror image of the residuals vs. fits plot. The residuals vs. predictor plot offers no new information.

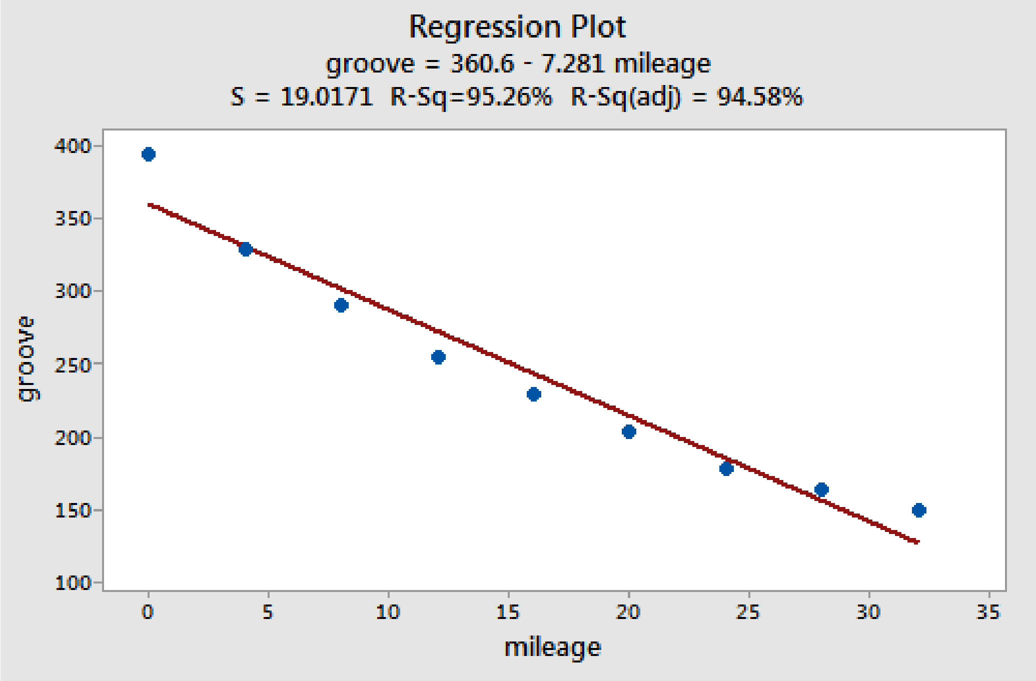
1. Residuals vs. Predictor Plot

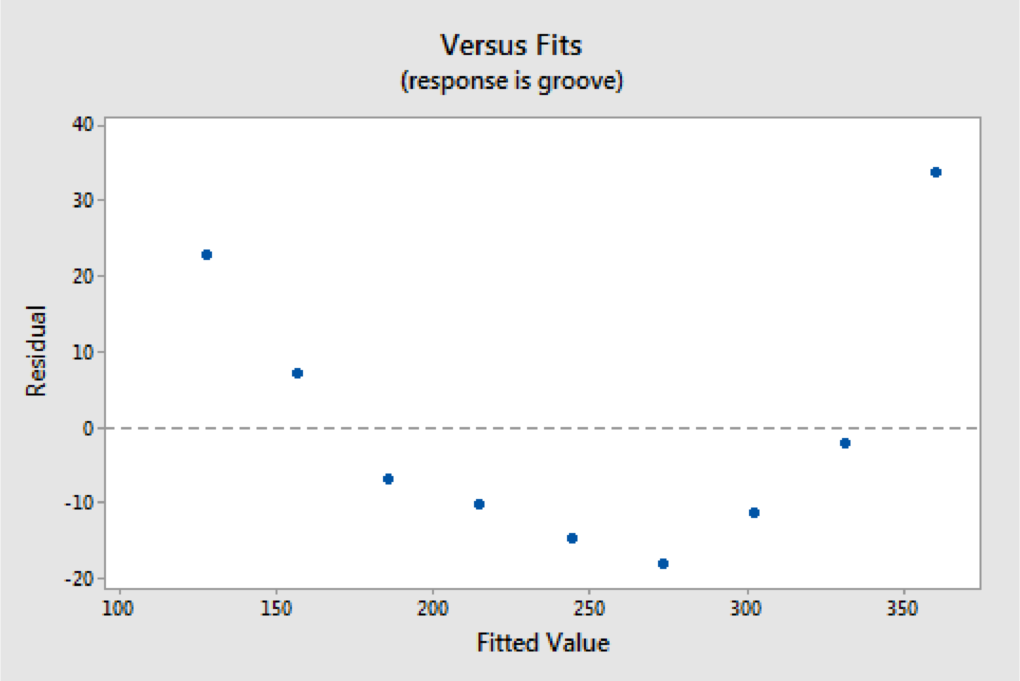




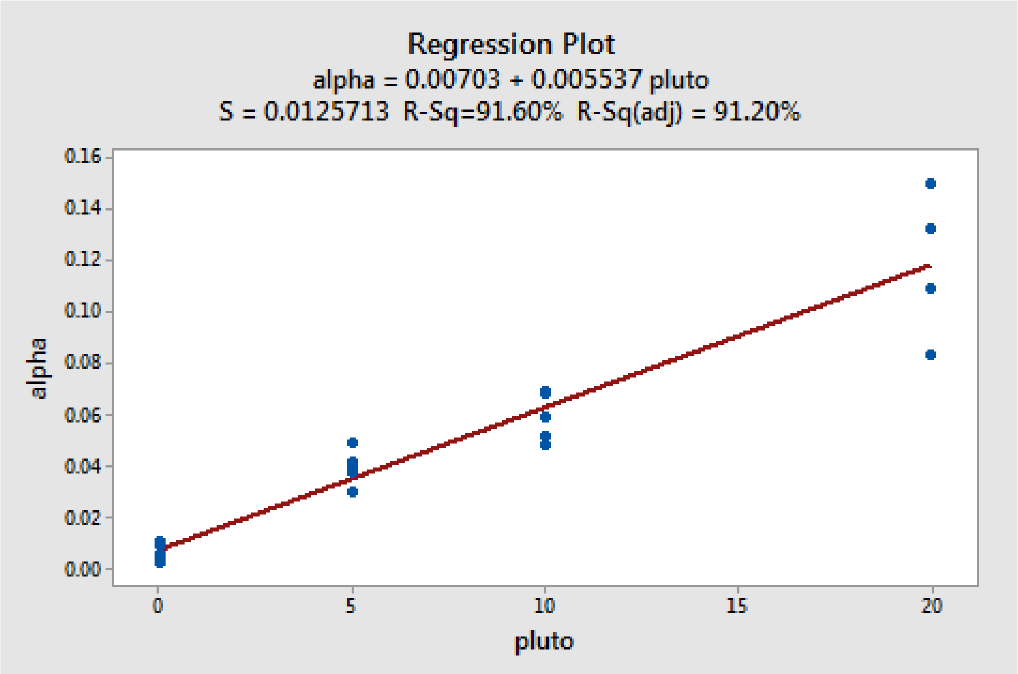
4. Identifying Specific Problems Using Residual Plots

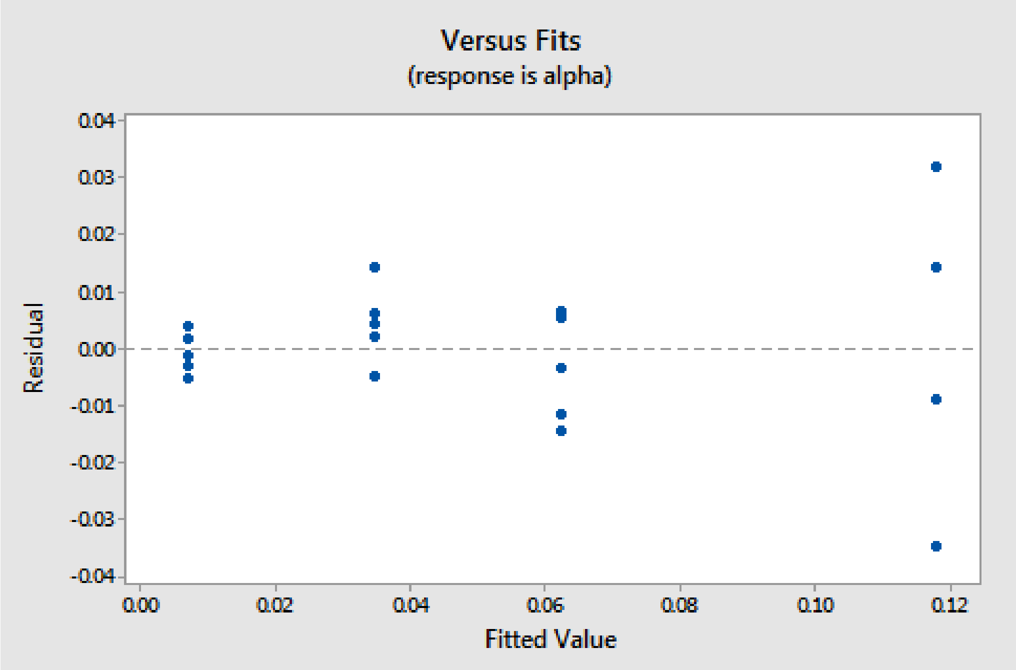
1. How does a non-linear regression function show up on a residual vs. fits plot?



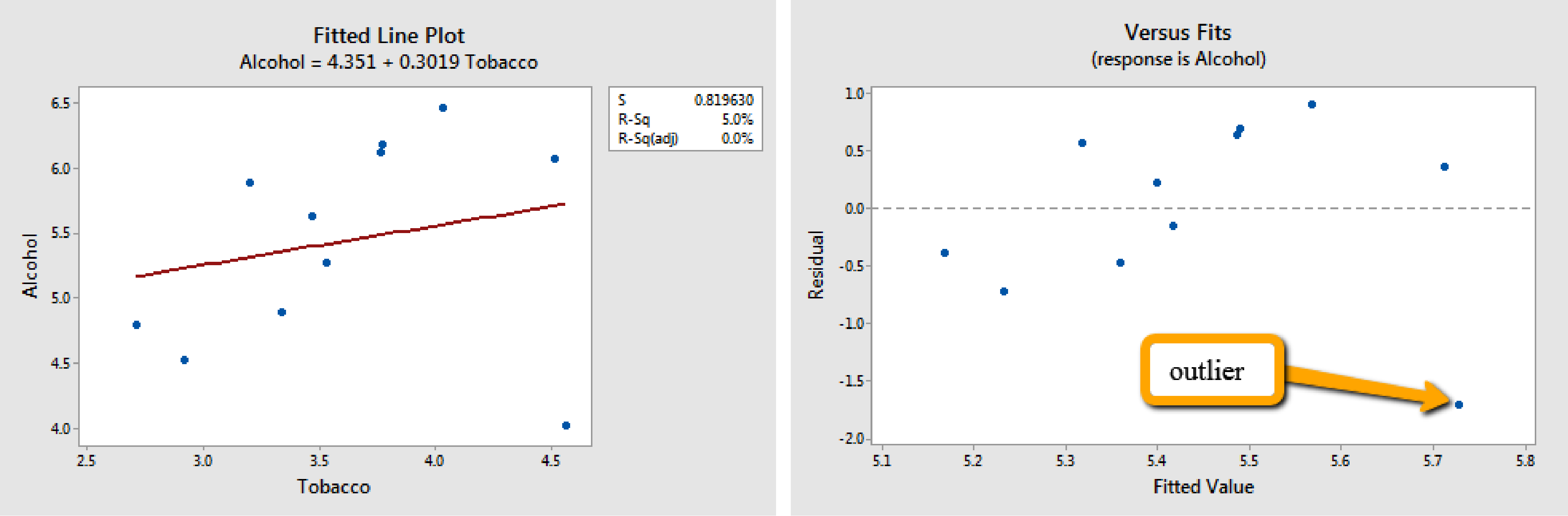


2. How does non-constant error variance show up on a residual vs. fits plot?

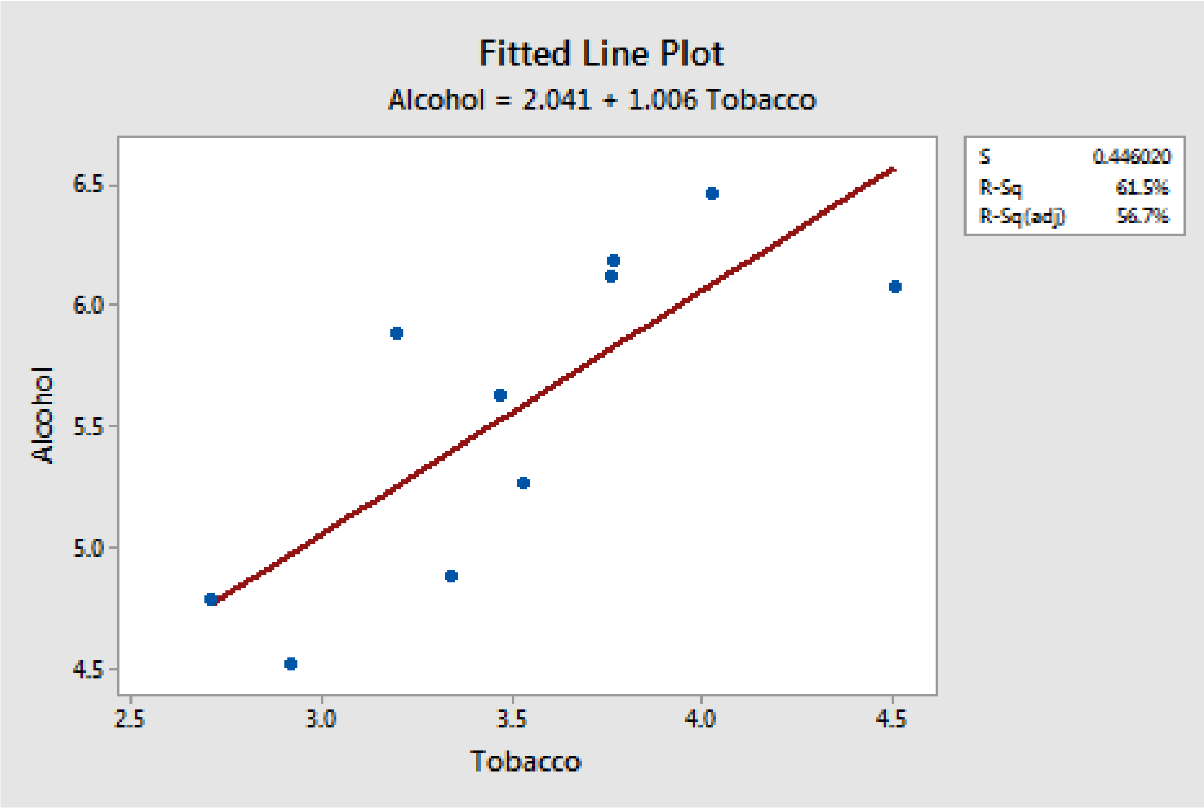




3. How does an outlier show up on a resudial vs. fits plot?

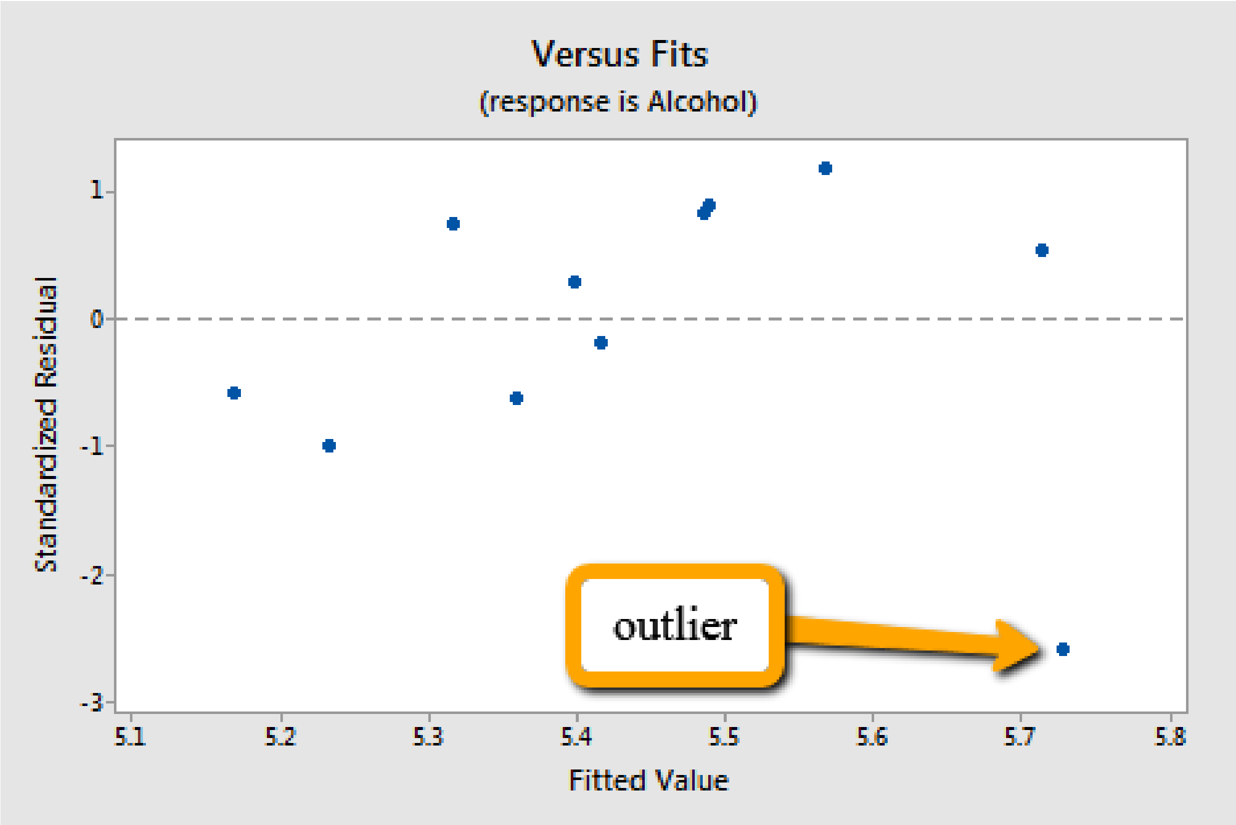


3. How does an outlier show up on a resudial vs. fits plot?



The r^2 value has jumped from 5% ("no-relationship") to 61.5% ("moderate relationship").

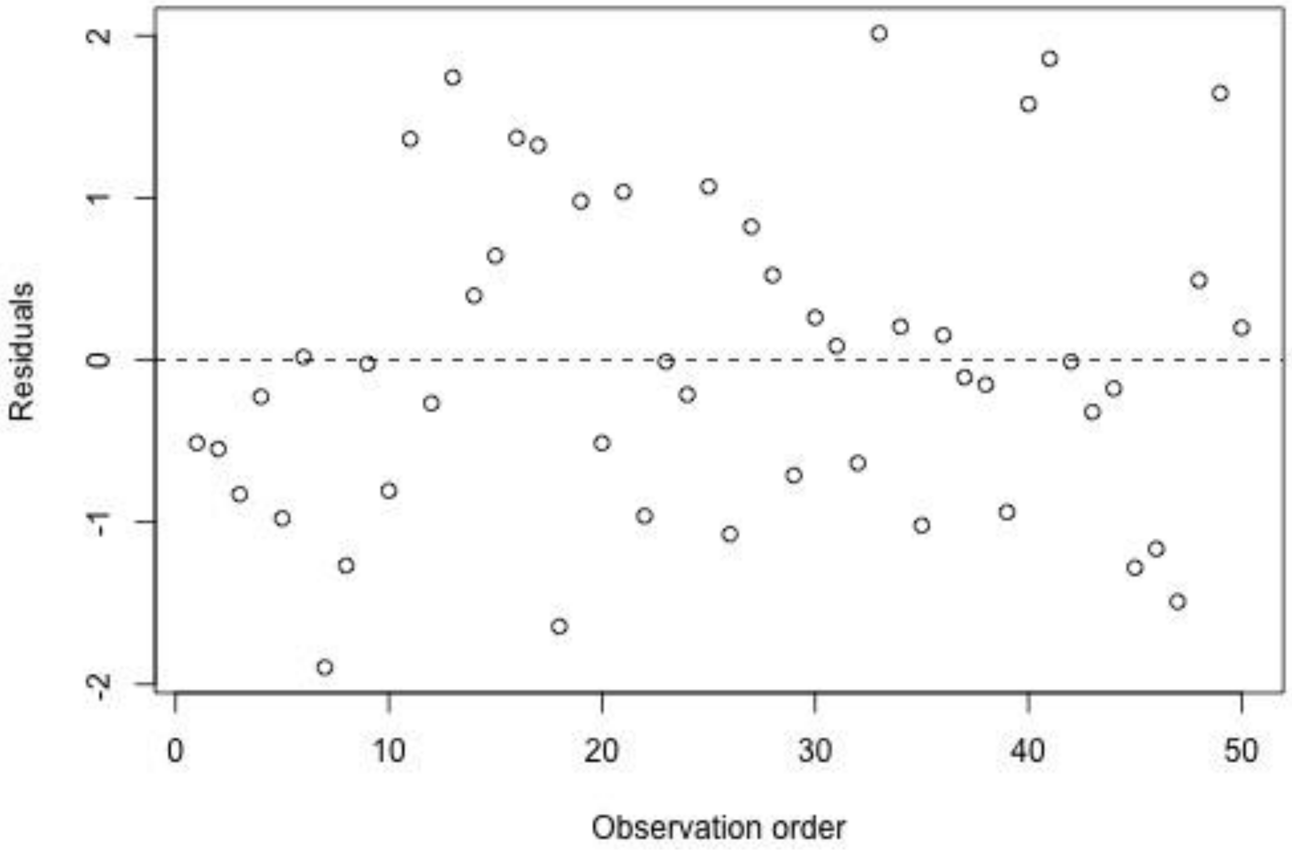
3. How does an outlier show up on a resudial vs. fits plot?



We can make the residuals "unitless" by dividing them by their standard deviation. In this way we create what are called "standardized residuals."

5. Residuals vs. Order Plot

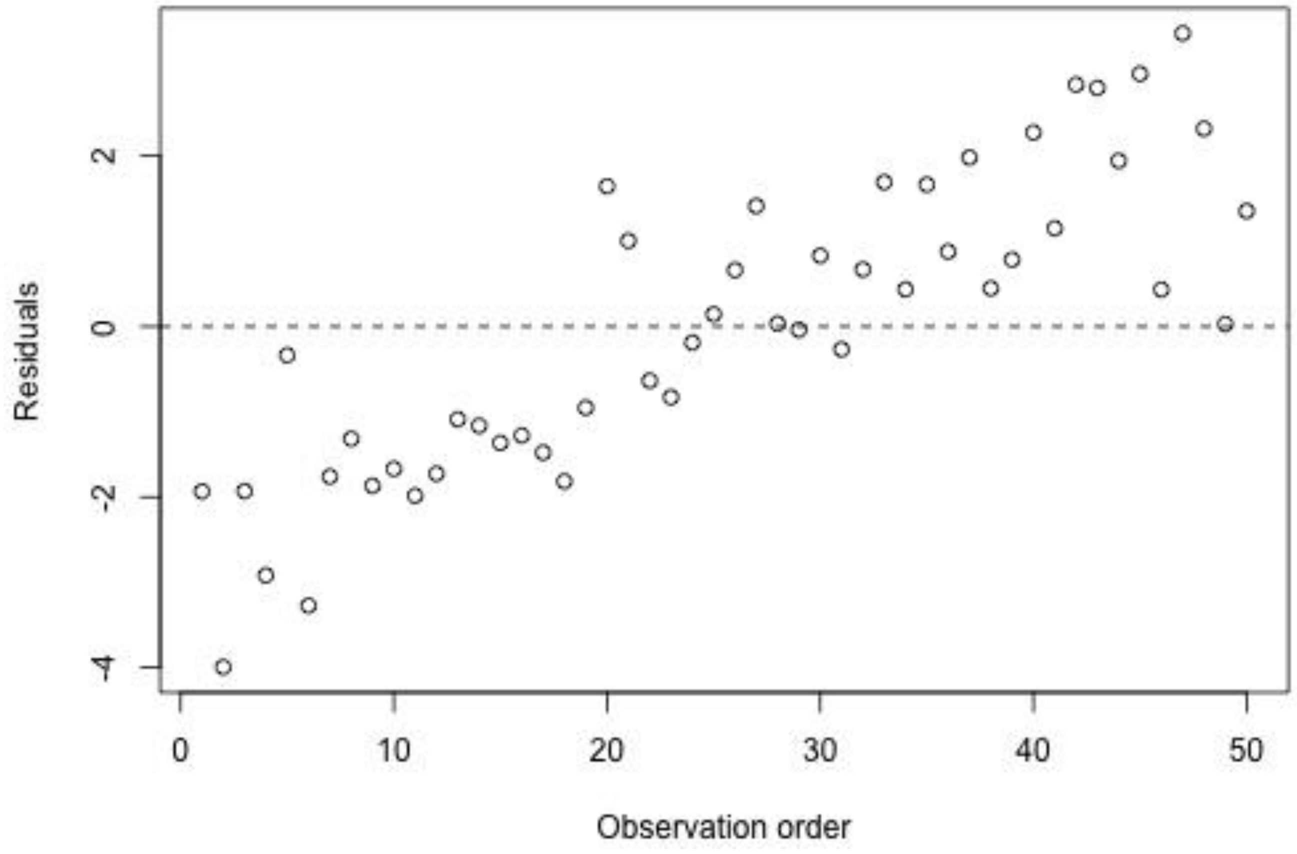
1. Residuals vs. Order Plot



1. Residuals vs. Order Plot

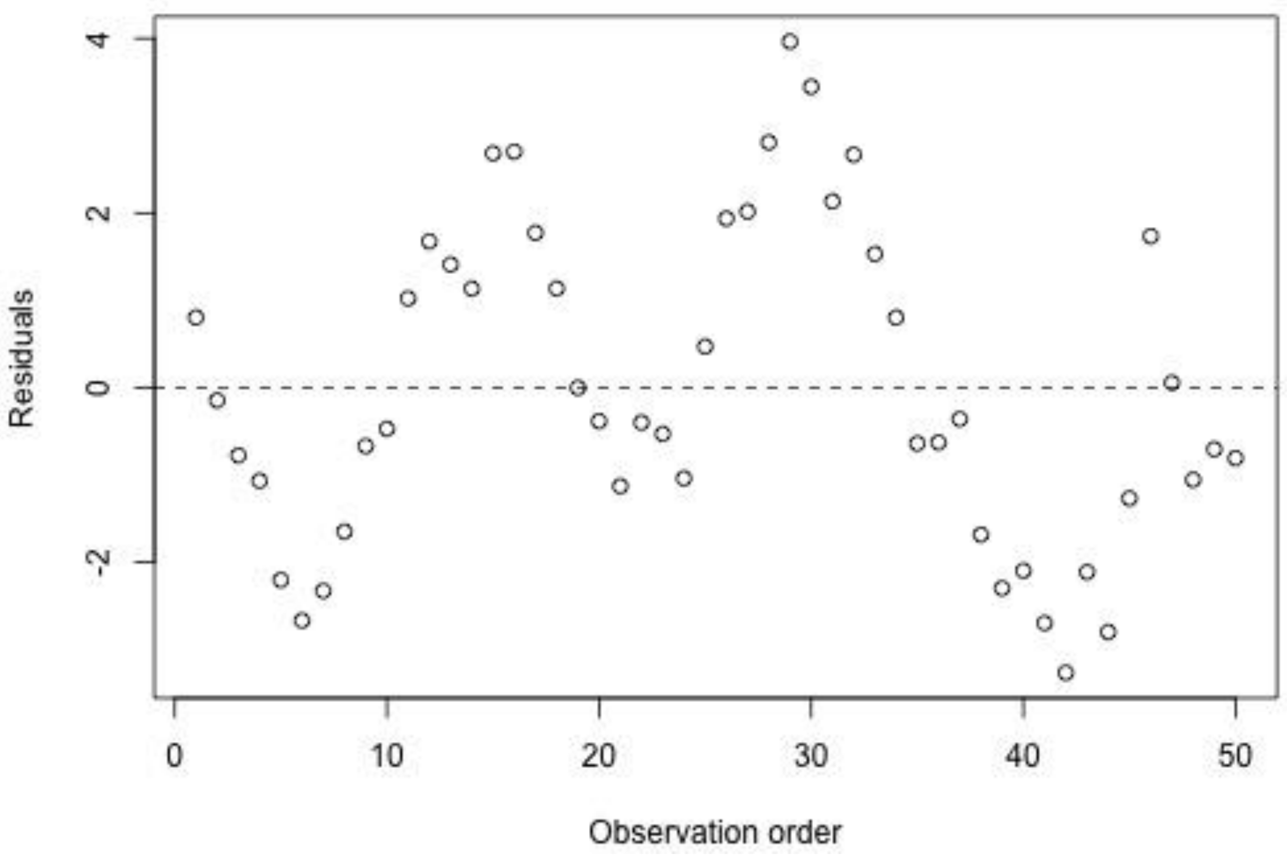
- A way of detecting a particular form of non-independence of the error terms, namely serial correlation.
- If the data are obtained in time (or space) sequence, a residuals vs. order plot helps to see if there is any correlation between the error terms that are near each other in the sequence.
- **The plot is only appropriate if you know the order in which the data were collected!**

2. A time trend



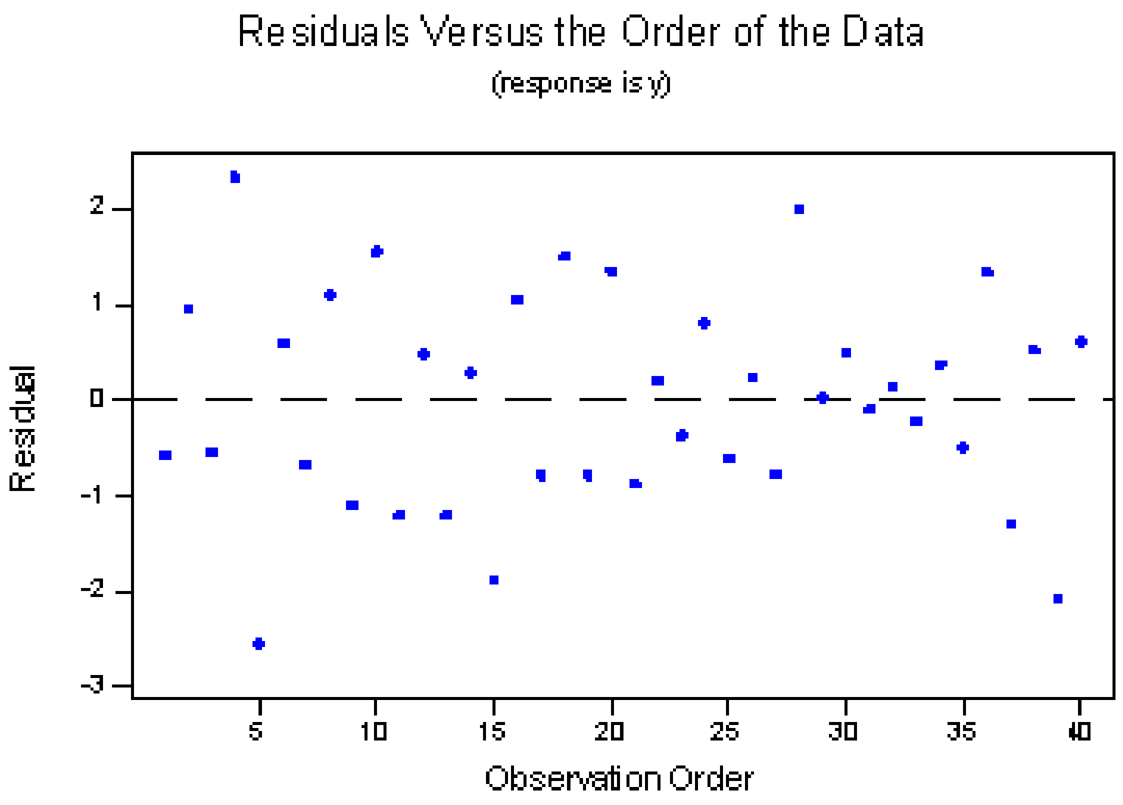
It might be a good idea to add the predictor "time" to the model.

3. Positive serial correlation



Time series modeling is more more appropriate than linear regression in this case.

4. Negative serial correlation



Time series modeling is more more appropriate than linear regression in this case.

5. Checking LINE Assumptions

- Linear Function → Residuals vs. Fits Plot
- Independent → Residuals vs. Order Plot
- Normally Distributed → ?
- Equal Variances → Residuals vs. Fits Plot

Next

Chapter 5
SLR Model Assumptions II