

Chapter 5

SLR Model Assumptions II

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

This work is a derivative of 'Regression Methods' by Iain Pardoe, Laura Simon and Derek Young, used under CC BY-NC.

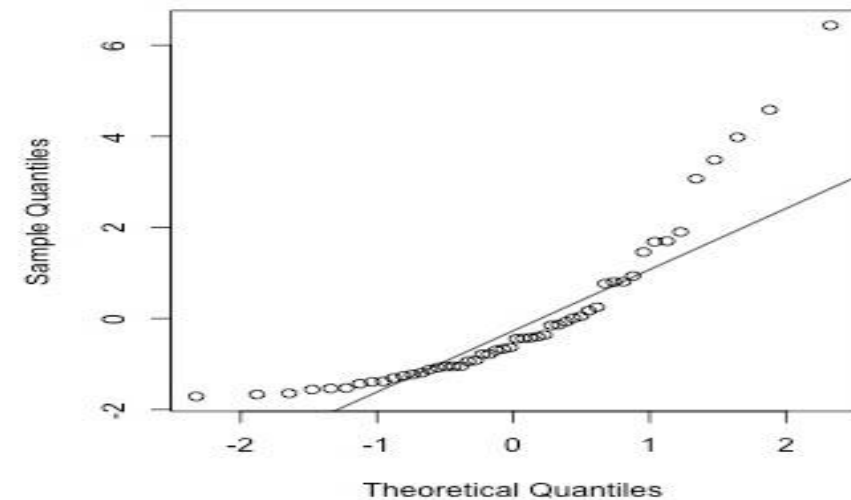
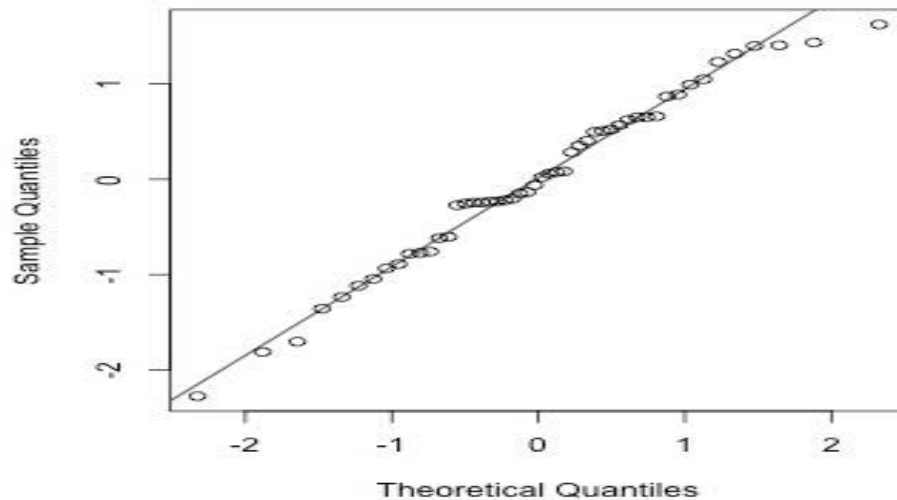
Contents

1. Normal Probability Plot of Residuals
2. Normal Probability Plots vs. Histograms
3. Assessing Linearity by Visual Inspection
4. Further Examples

1. Normal Probability Plot of Residuals

1. Normal Probability Plot of Residuals

- If the data follow a normal distribution, then a plot of the theoretical percentiles of the normal distribution versus the observed sample percentiles should be approximately linear.

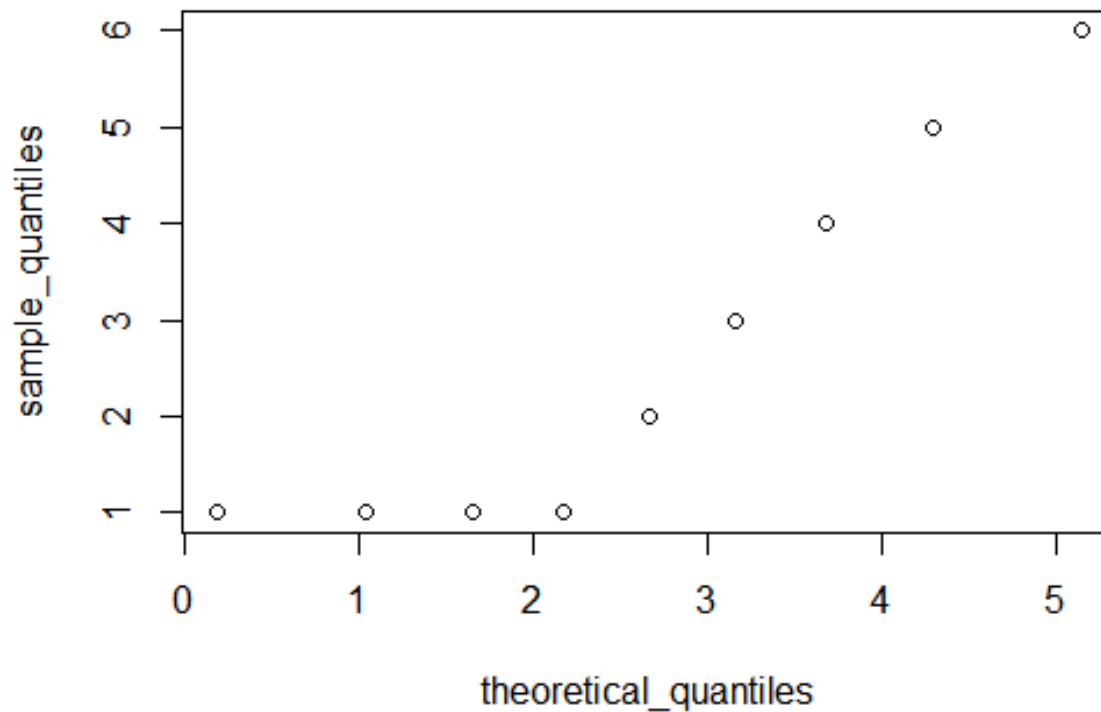


1. Normal Probability Plot of Residuals

```
> residual = c(1,1,1,1,2,3,4,5,6)
> p = seq(0.1,0.9,0.1)
> sample_quantiles = quantile(residual, p, type = 1)
> print(sample_quantiles)
10% 20% 30% 40% 50% 60% 70% 80% 90%
  1   1   1   1   2   3   4   5   6
> theoretical_quantiles = qnorm(p, mean = mean(residual), sd =
sd(residual))
> print(theoretical_quantiles)
[1] 0.1849527 1.0368742 1.6511694 2.1760621 2.6666667 3.1572712
[7] 3.6821639 4.2964592 5.1483806
```

1. Normal Probability Plot of Residuals

```
> qqplot(theoretical_quantiles, sample_quantiles)
```



1. Normal Probability Plot of Residuals

<pre>> std_residual = scale(residual) > print(std_residual)</pre> <pre> [,1] [1,] -0.8606630 [2,] -0.8606630 [3,] -0.8606630 [4,] -0.8606630 [5,] -0.3442652 [6,] 0.1721326 [7,] 0.6885304 [8,] 1.2049282</pre>	<pre>[9,] 1.7213259 attr(,"scaled:center") [1] 2.6666667 attr(,"scaled:scale") [1] 1.936492</pre>
---	--

1. Normal Probability Plot of Residuals

```
> p = seq(0.1,0.9,0.1)
> sample_quantiles = quantile(std_residual, p, type = 1)
> print(sample_quantiles)
```

10%	20%	30%	40%	50%
-0.8606630	-0.8606630	-0.8606630	-0.8606630	-0.3442652
0.1721326				
70%	80%	90%		
0.6885304	1.2049282	1.7213259		

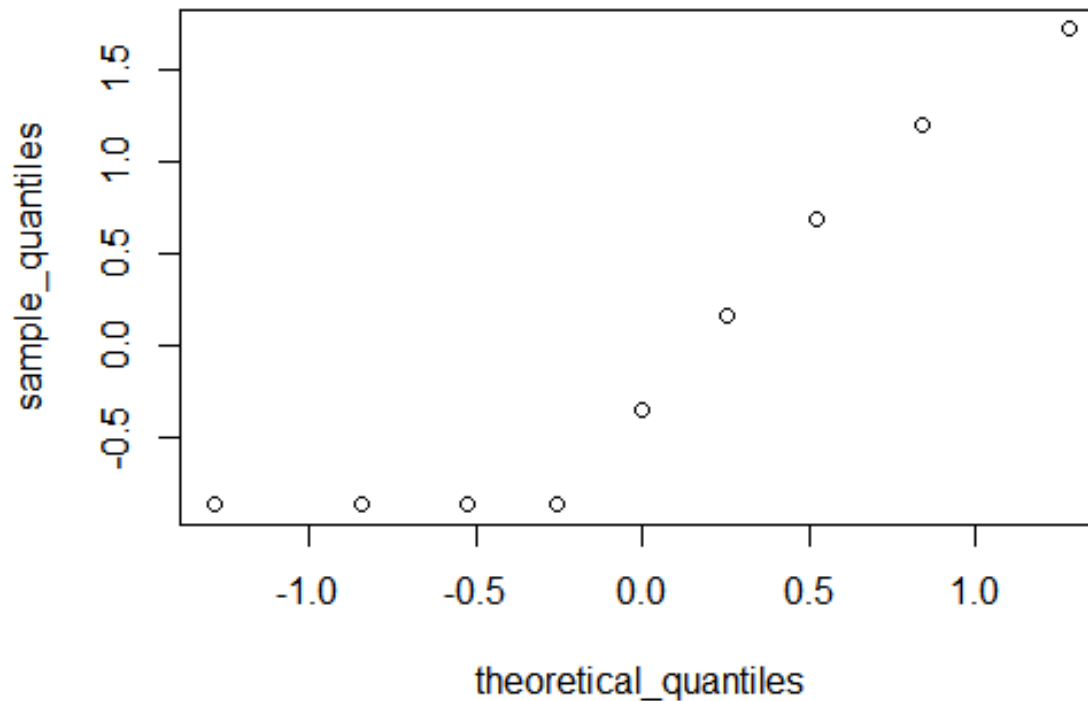
1. Normal Probability Plot of Residuals

```
> theoretical_quantiles = qnorm(p)
> print(theoretical_quantiles)
[1] -1.2815516 -0.8416212 -0.5244005 -0.2533471  0.0000000
0.2533471
[7]  0.5244005  0.8416212  1.2815516
```

<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867

1. Normal Probability Plot of Residuals

```
> qqplot(theoretical_quantiles, sample_quantiles)
```

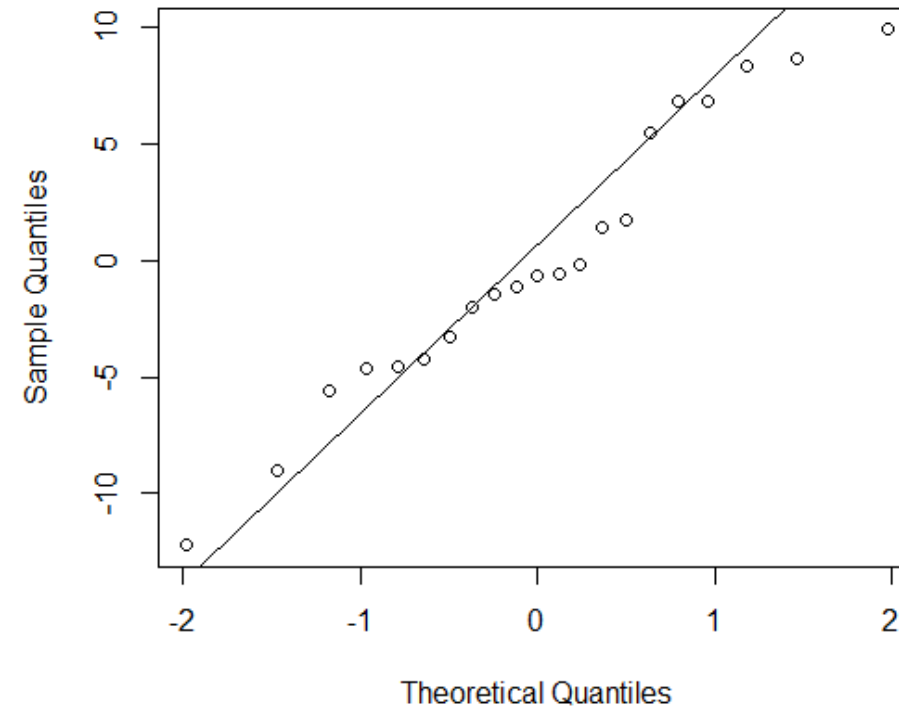
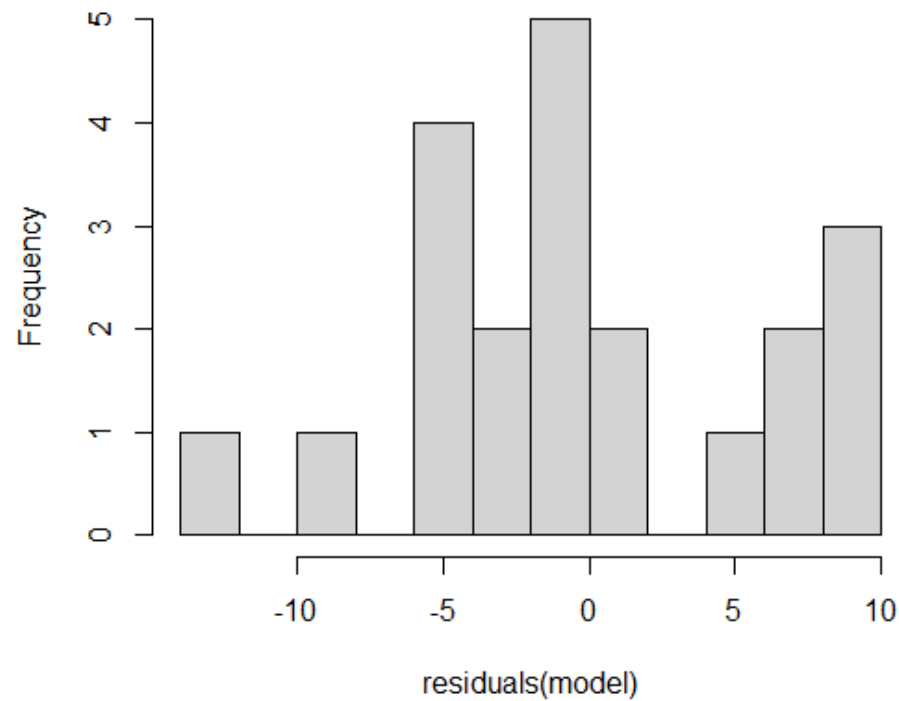


2. Data: Old Faithful

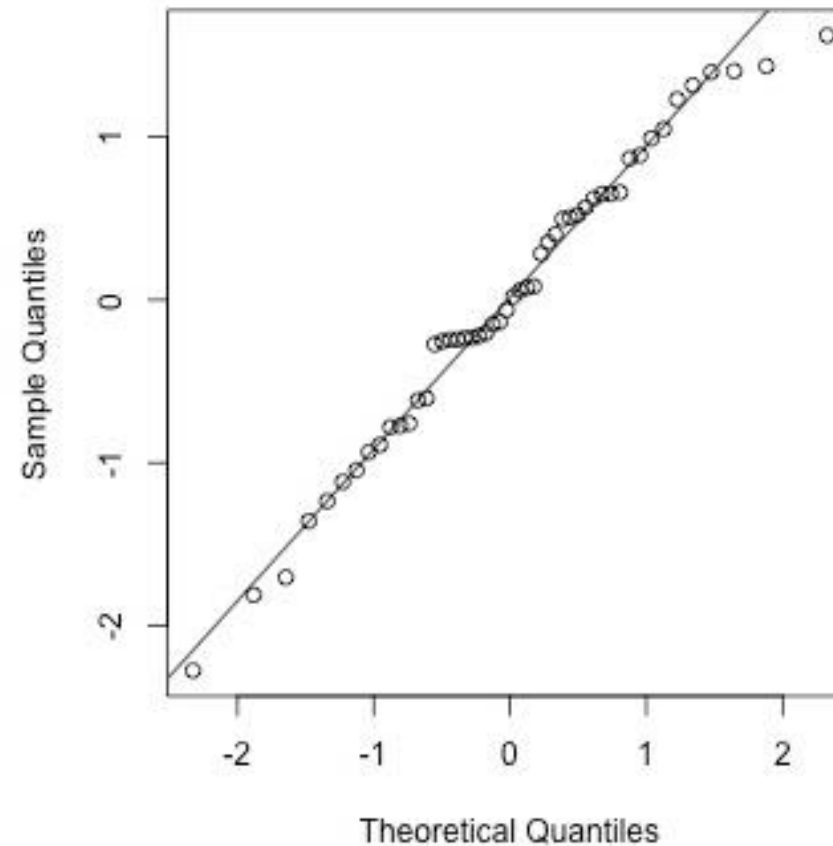
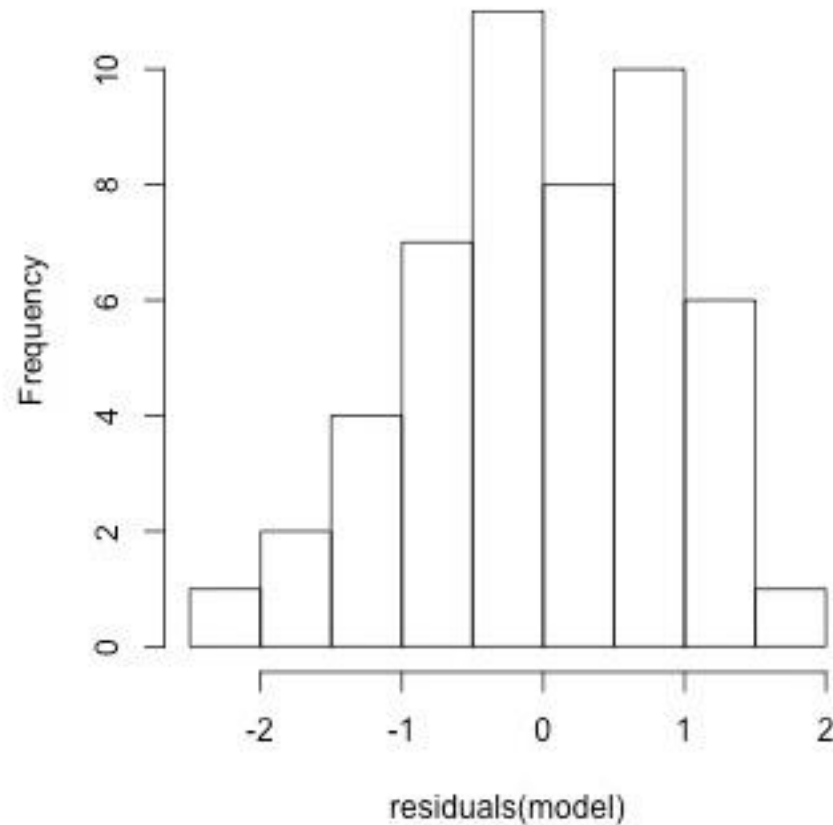
- Data: [Old Faithful](#)

```
oldfaithful <- read.table("oldfaithful.txt", header=T)
attach(oldfaithful)
model <- lm(next. ~ duration)
hist(residuals(model), main="", breaks=12)
qqnorm(residuals(model), main="")
qqline(residuals(model))
detach(oldfaithful)
```

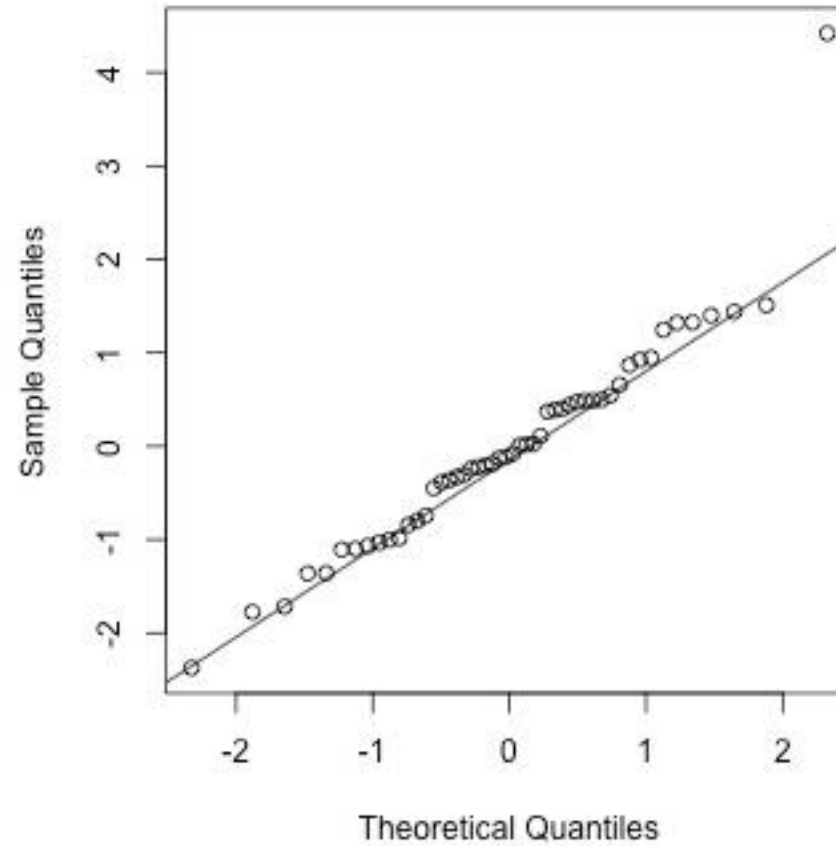
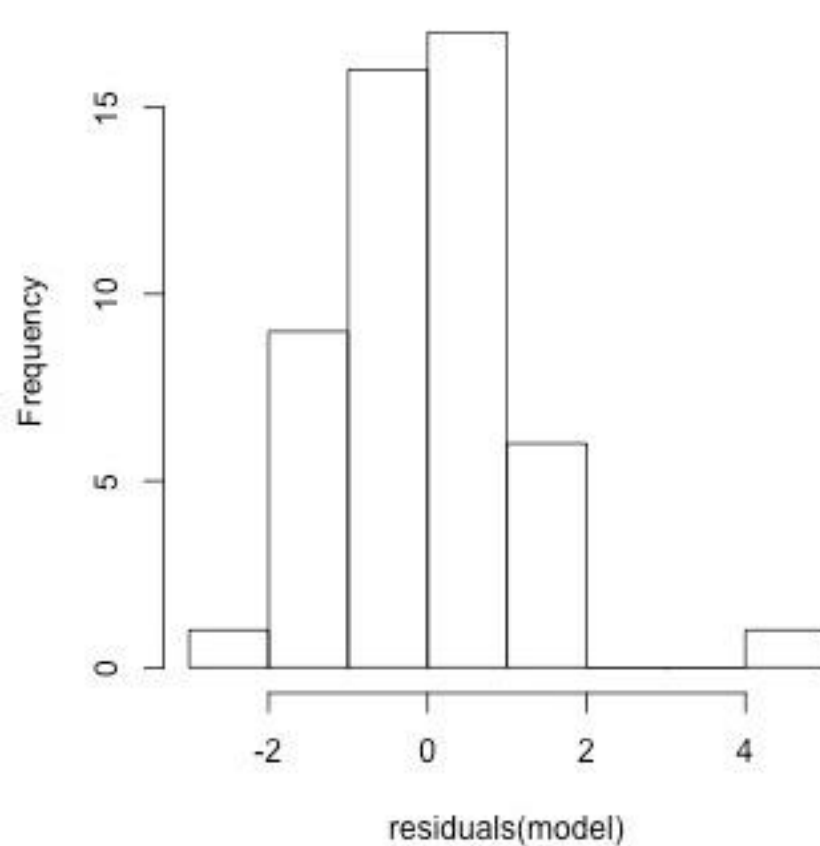
2. Data: Old Faithful



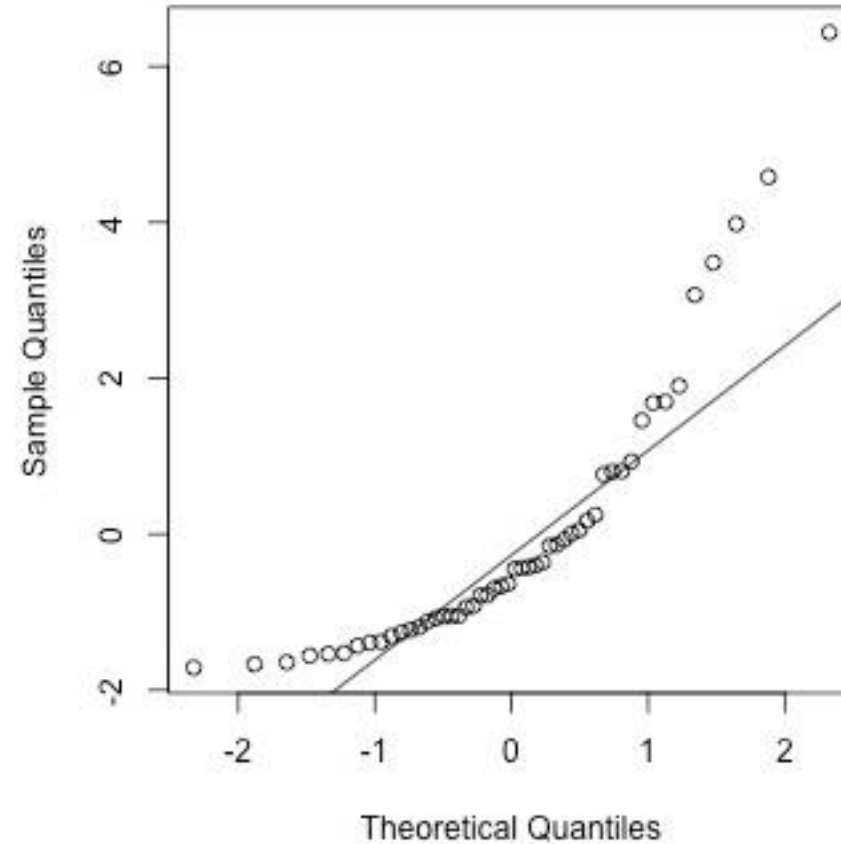
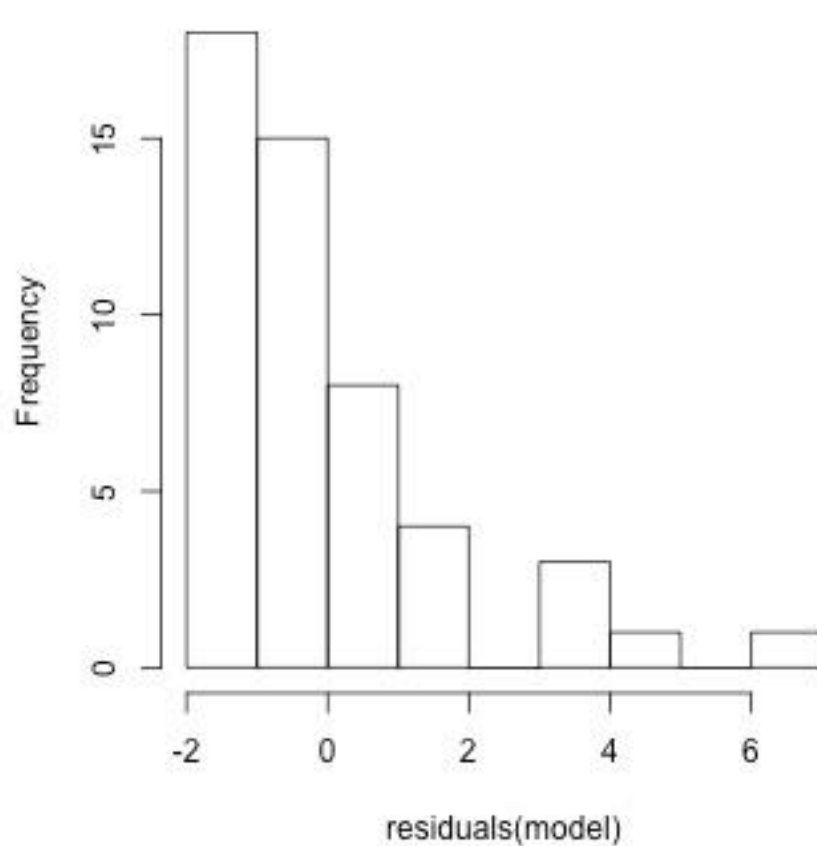
3. Normally distributed residuals



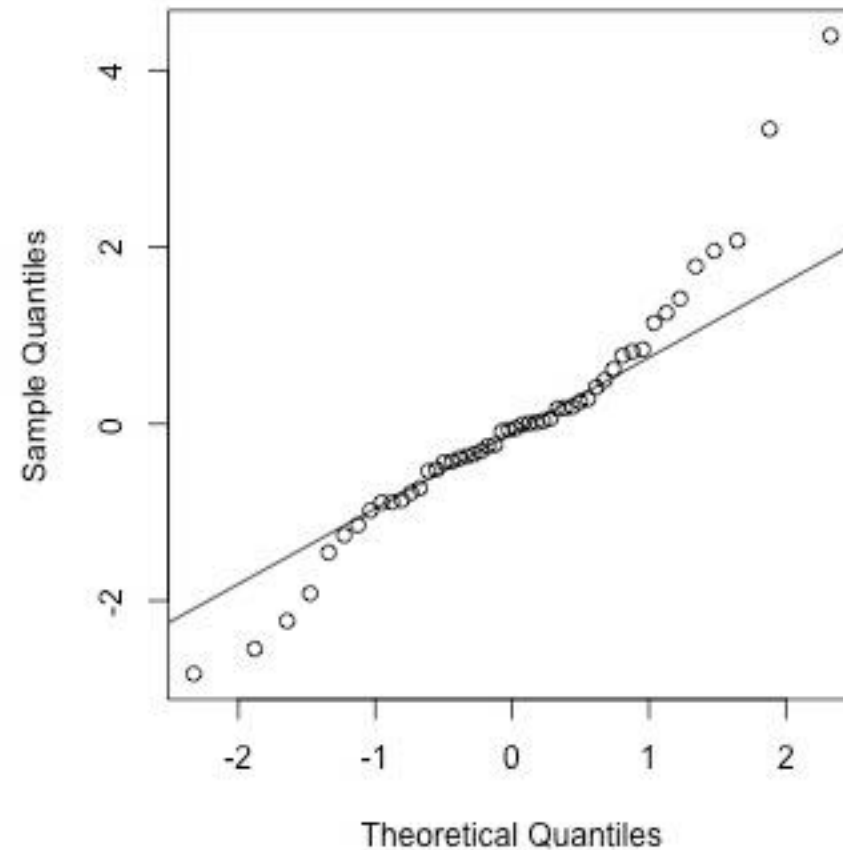
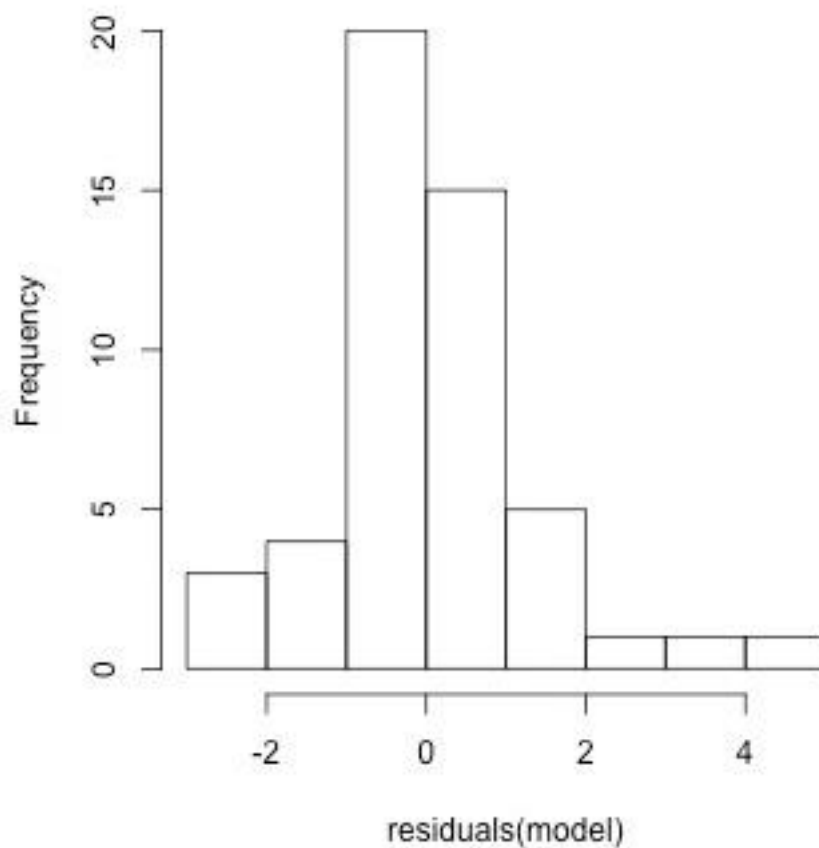
4. Normal residuals but with one outlier



5. Skewed residuals



6. Heavy-tailed residuals

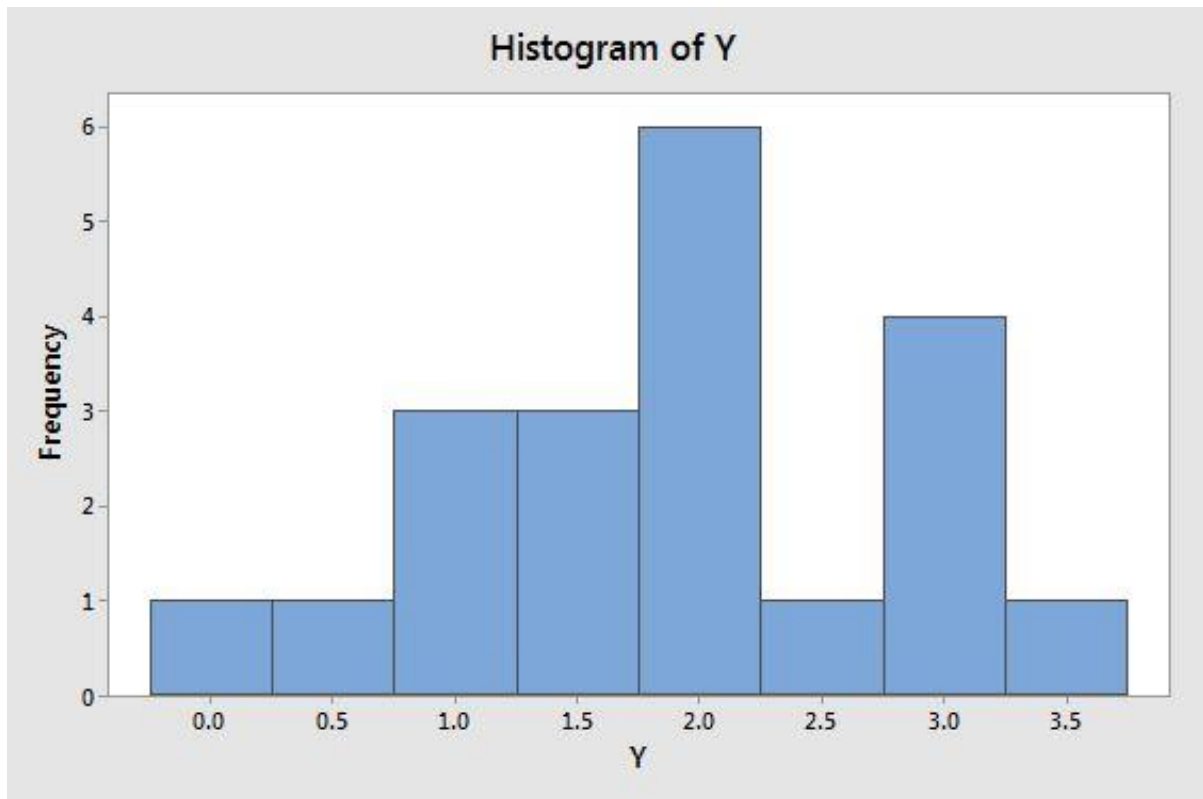


7. Checking LINE Assumptions

- Linear Function → Residuals vs. Fits Plot
- Independent → Residuals vs. Order Plot
- Normally Distributed → Normal Probability Plot of Residuals
- Equal Variances → Residuals vs. Fits Plot

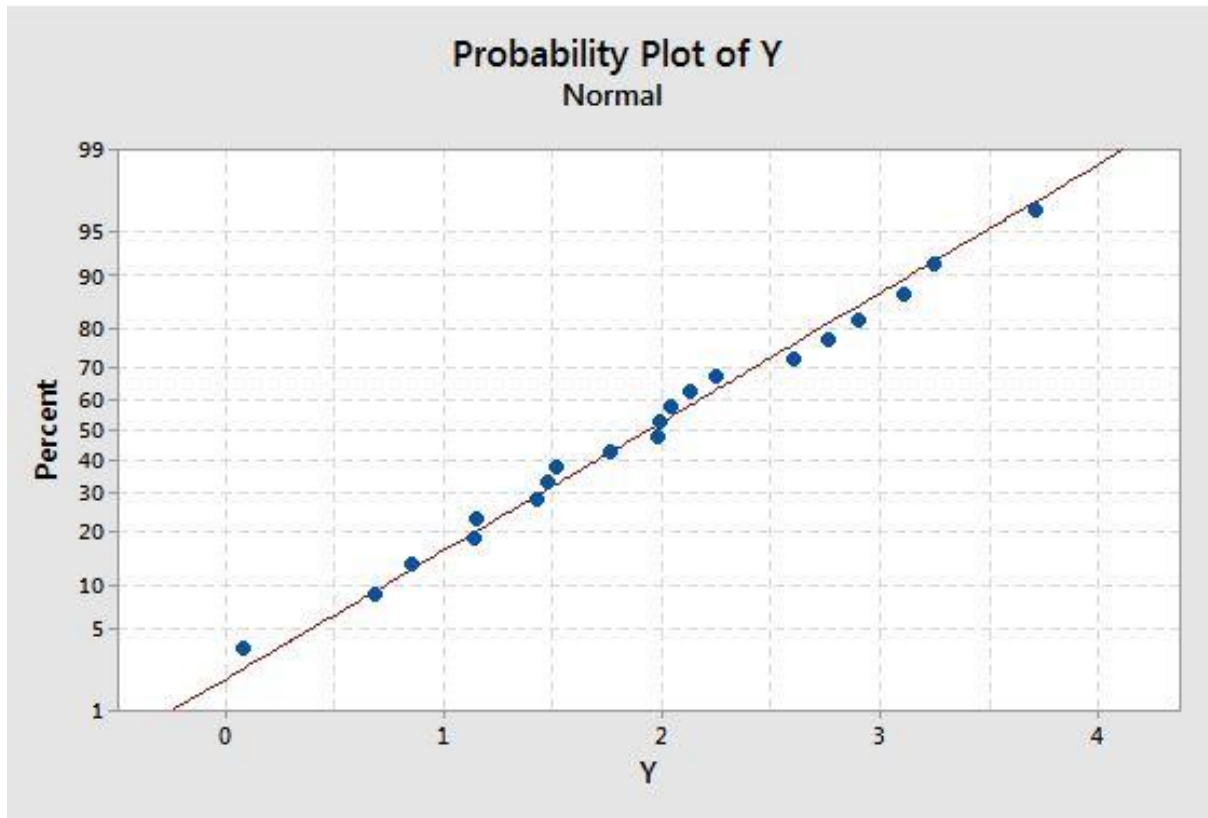
2. Normal Probability Plots vs. Histograms

1. Normal Probability Plots vs. Histograms



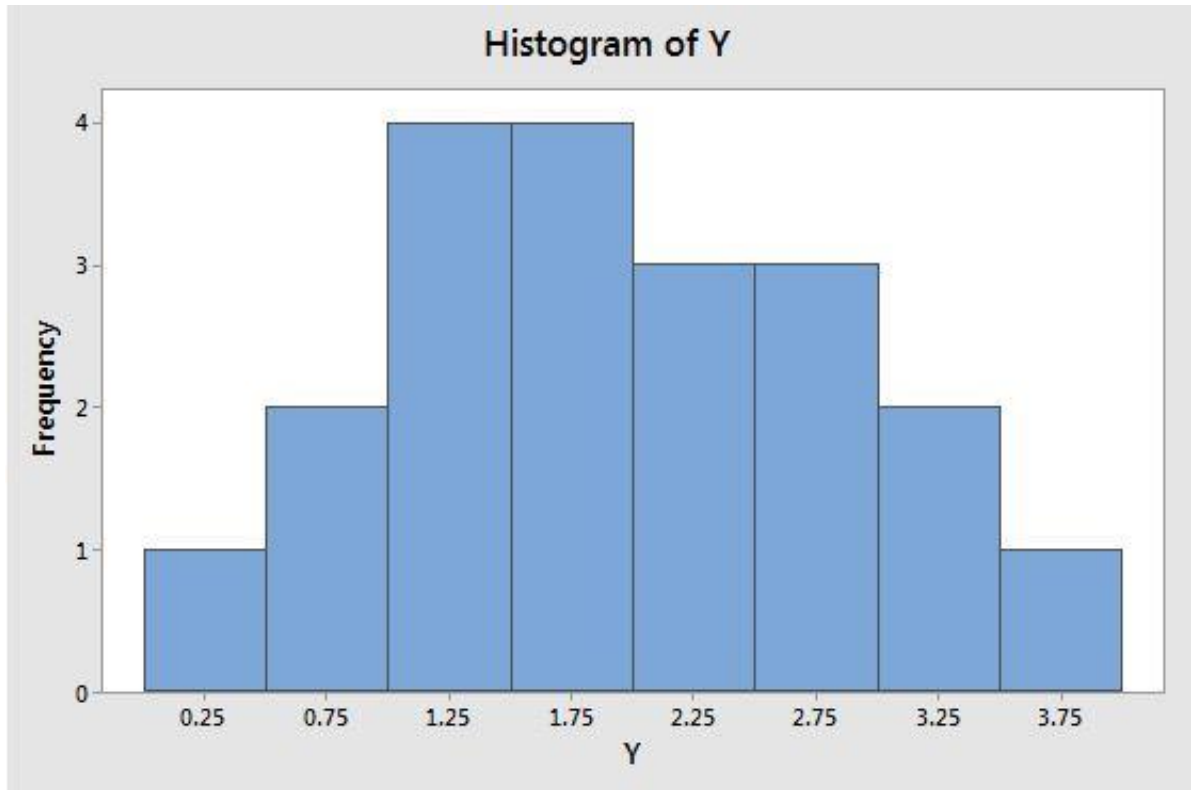
One problem is the sensitivity of a histogram to the choice of breakpoints for the bars – small changes can alter the visual impression quite drastically in some cases.

1. Normal Probability Plots vs. Histograms



By contrast, the normal probability plot is more straightforward and effective and it is generally easier to assess whether the points are close to the diagonal line.

1. Normal Probability Plots vs. Histograms



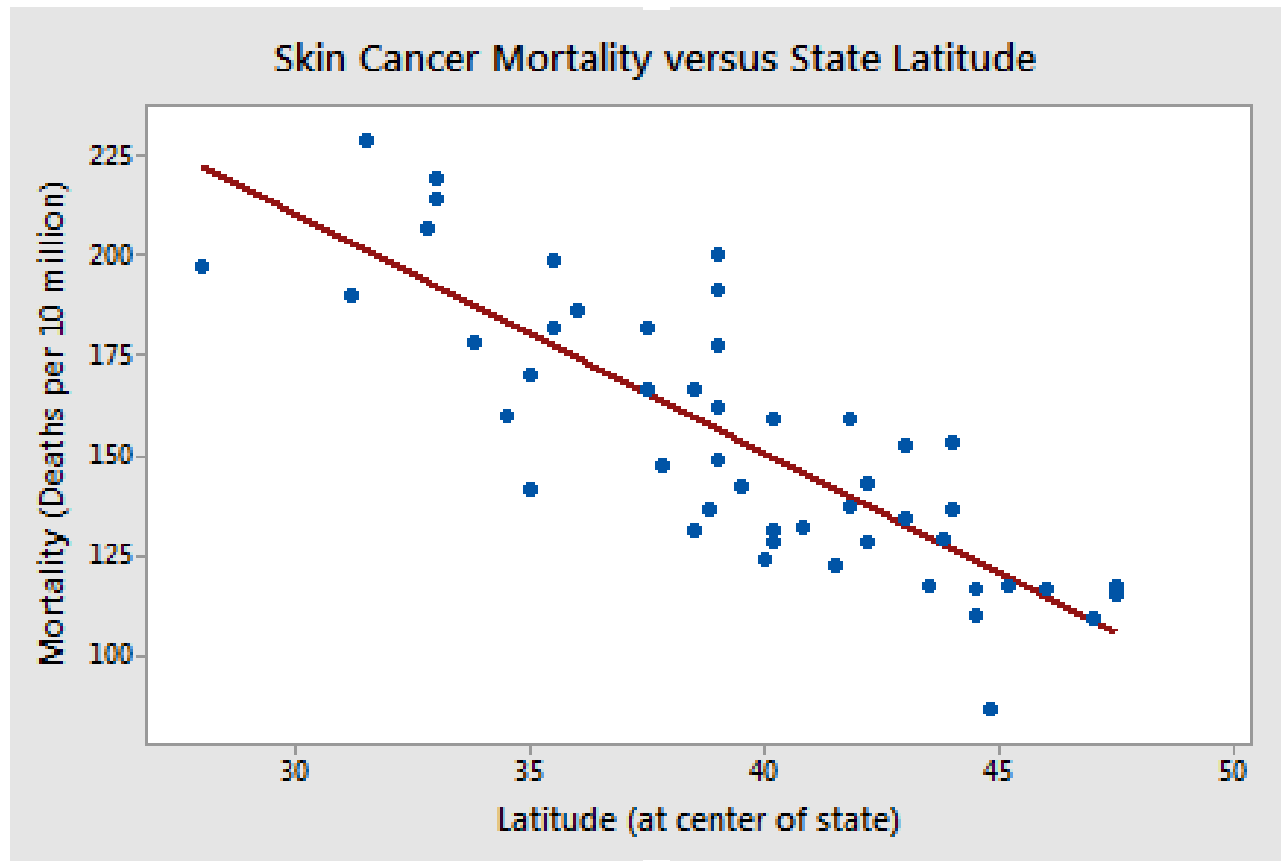
It is possible to create a more normal-looking histogram for these data by adjusting the breakpoints along the axis

3. Assessing Linearity by Visual Inspection

1. Assessing Linearity by Visual Inspection

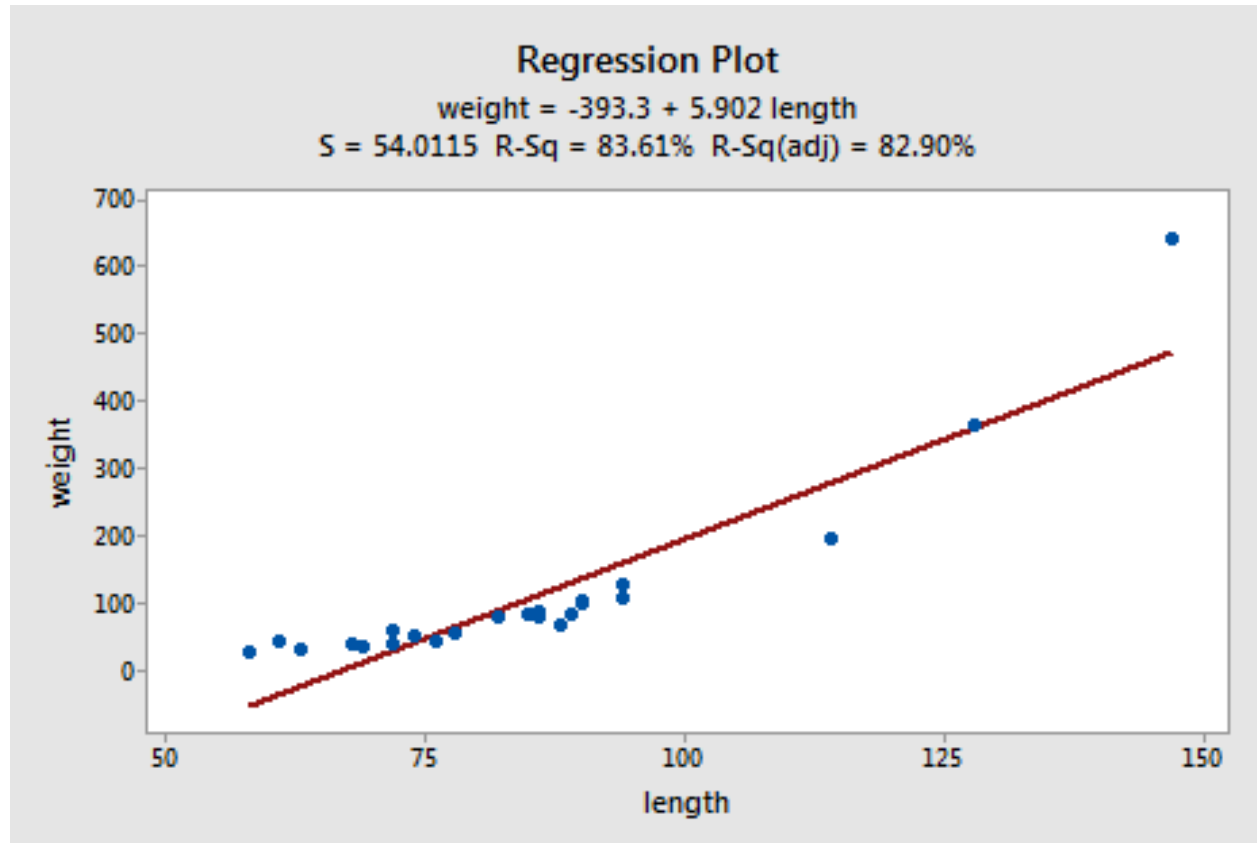
- The first simple linear regression model condition concerns linearity: the mean of the response at each predictor value should be a linear function of the predictor. The neat thing about simple linear regression — in which there is a response y and just one predictor x — is that we can get a good feel for this condition just by looking at a simple scatter plot.

1. Assessing Linearity by Visual Inspection



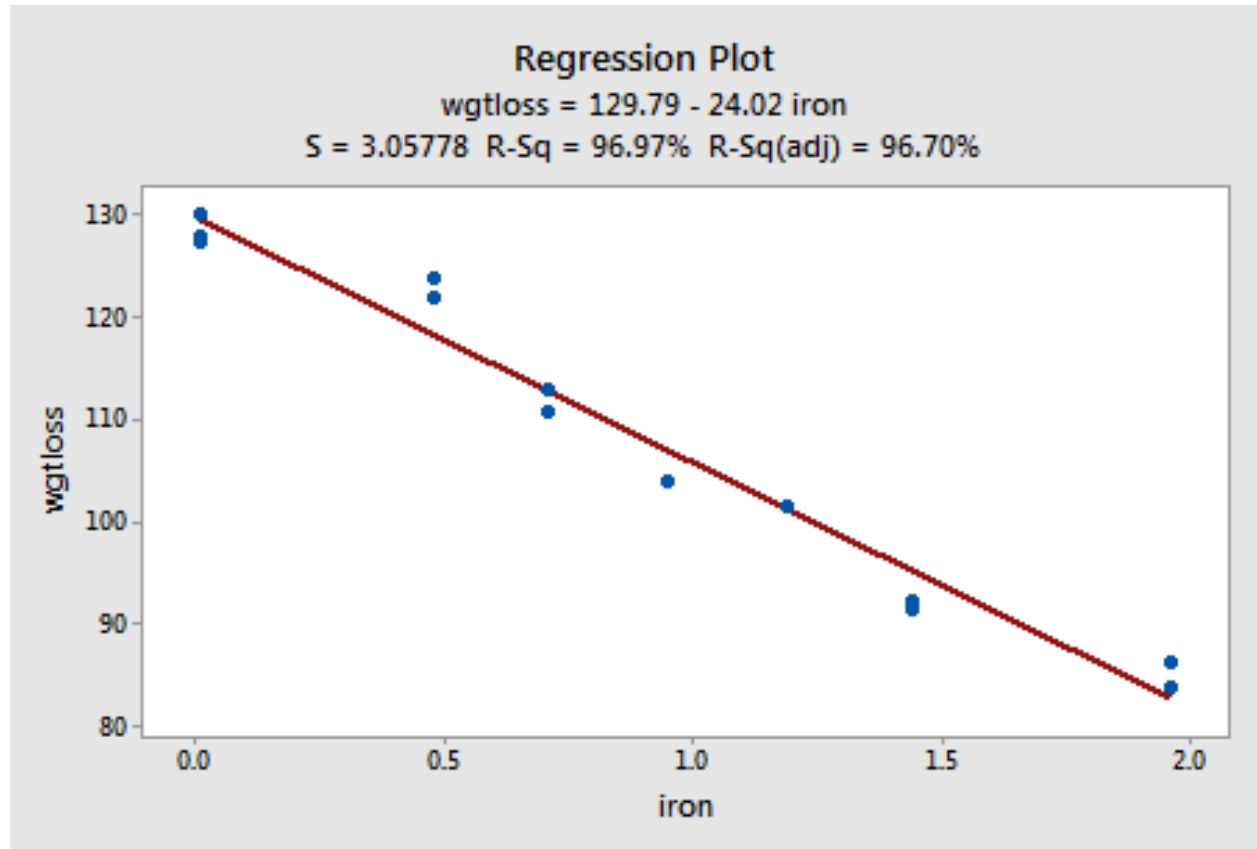
It appears as if the relationship between latitude and skin cancer mortality is indeed linear

1. Assessing Linearity by Visual Inspection



A curved function more adequately describes the trend.

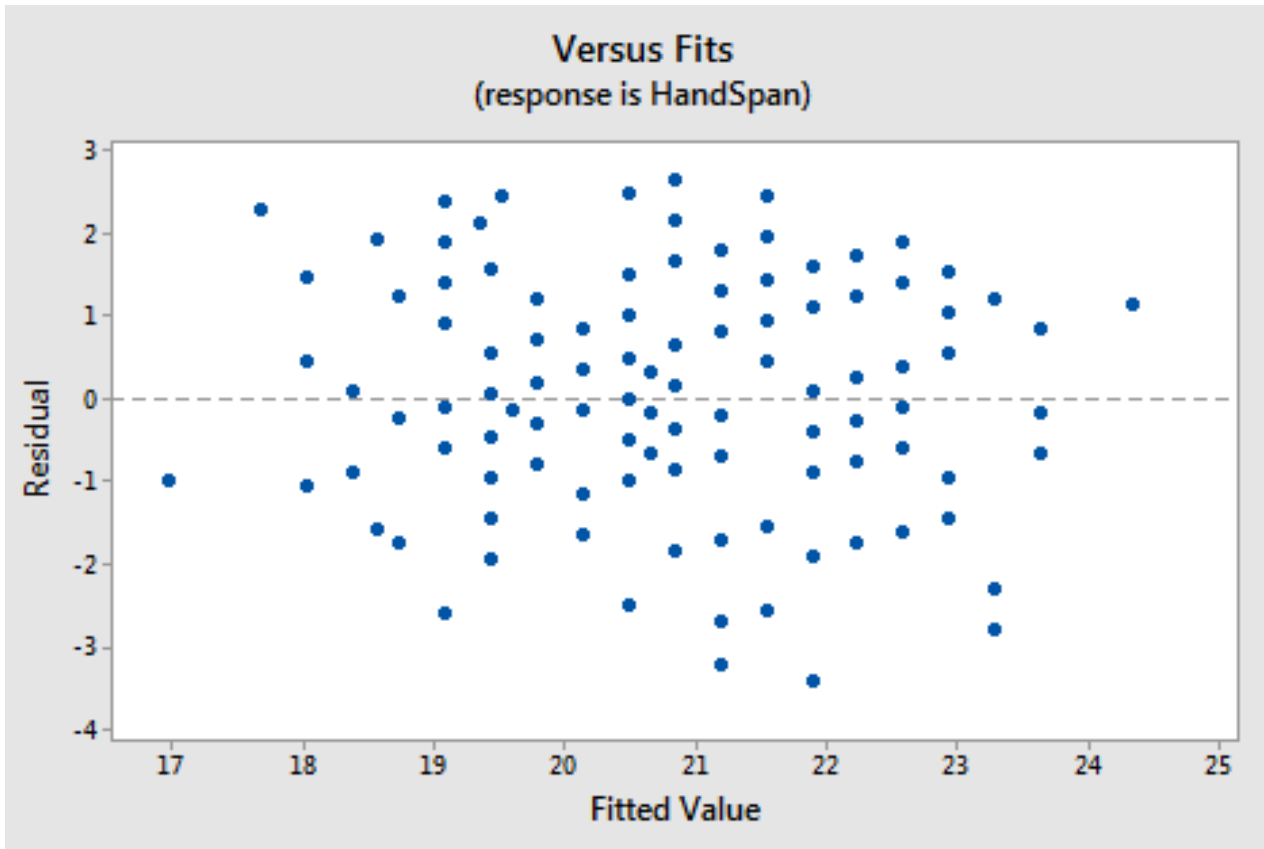
1. Assessing Linearity by Visual Inspection



A linear model would be adequate in describing the trend.

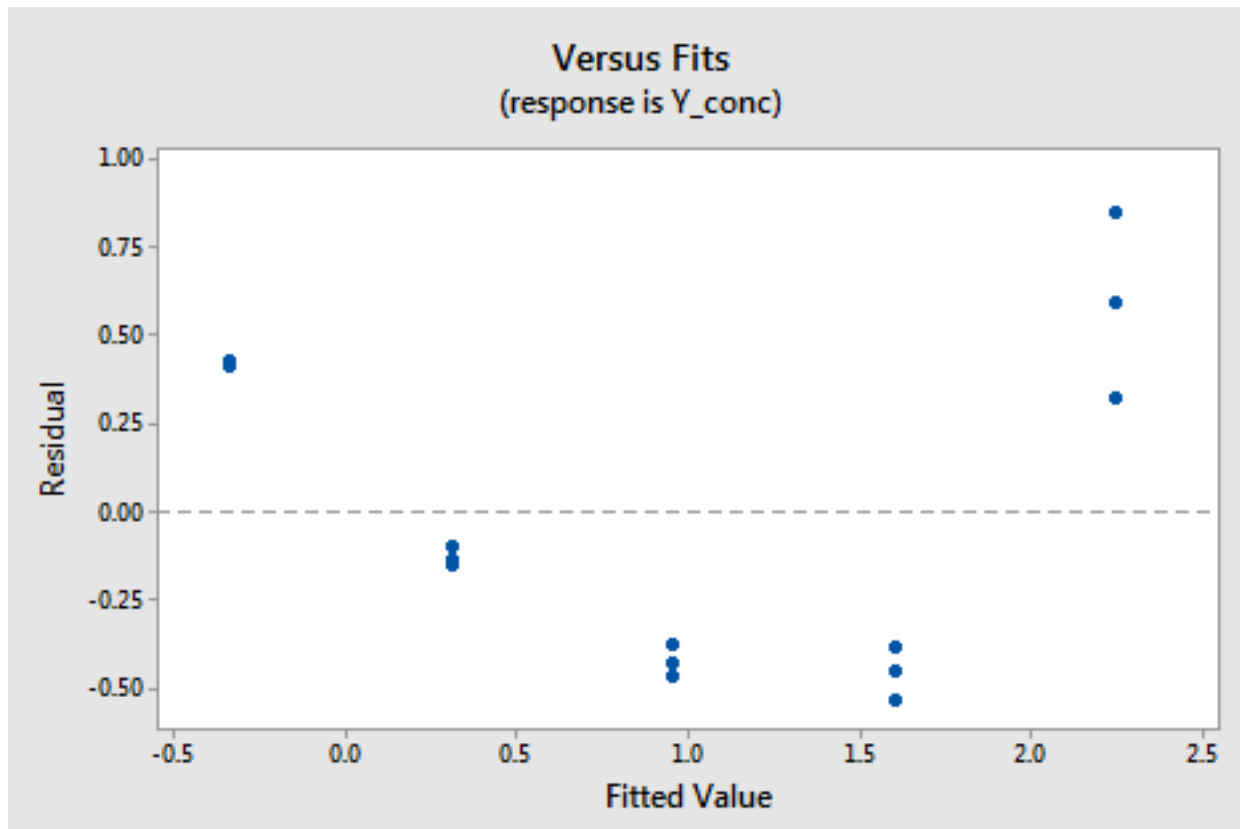
4. Further Examples

1. A Good Residual Plot



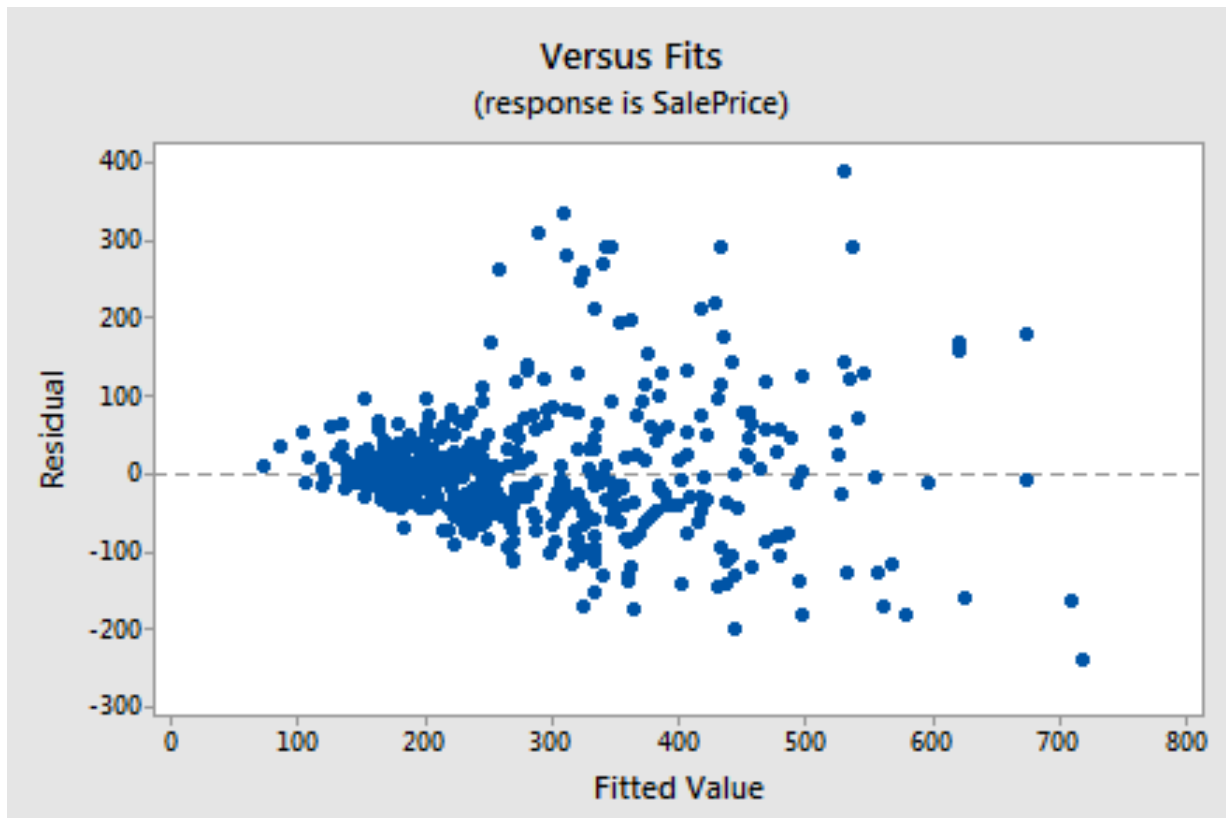
This plot looks good in that the variance is roughly the same all the way across and there are no worrisome patterns.

2. Residual Plot Resulting from Using the Wrong Model



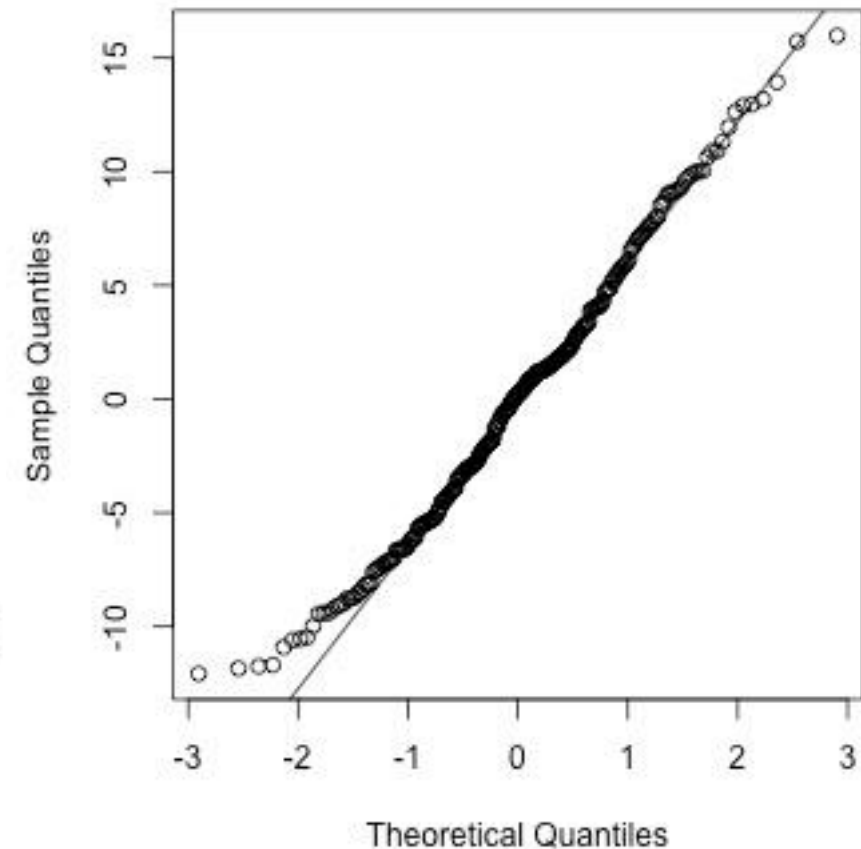
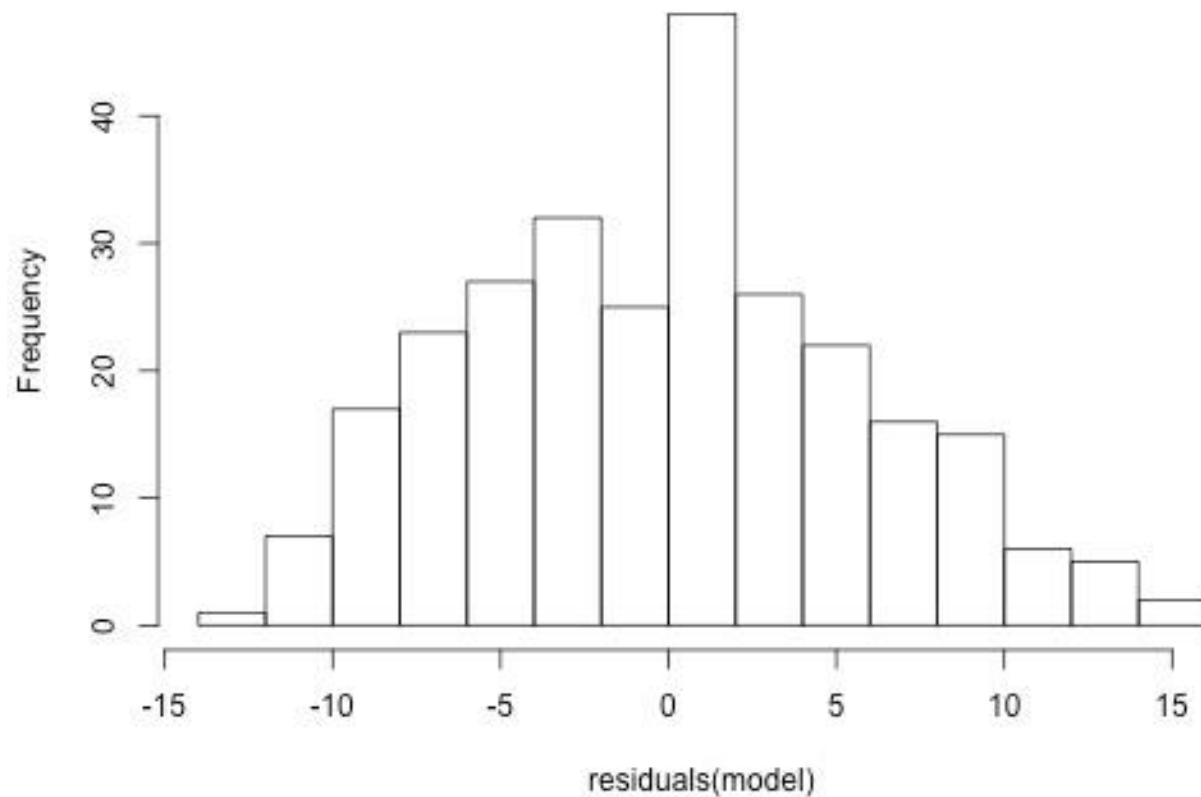
- The pattern is curved which indicates that the wrong type of equation was used.
- The variance increases as the fitted values increase.

3. Indications that Assumption of Constant Variance is Not Valid

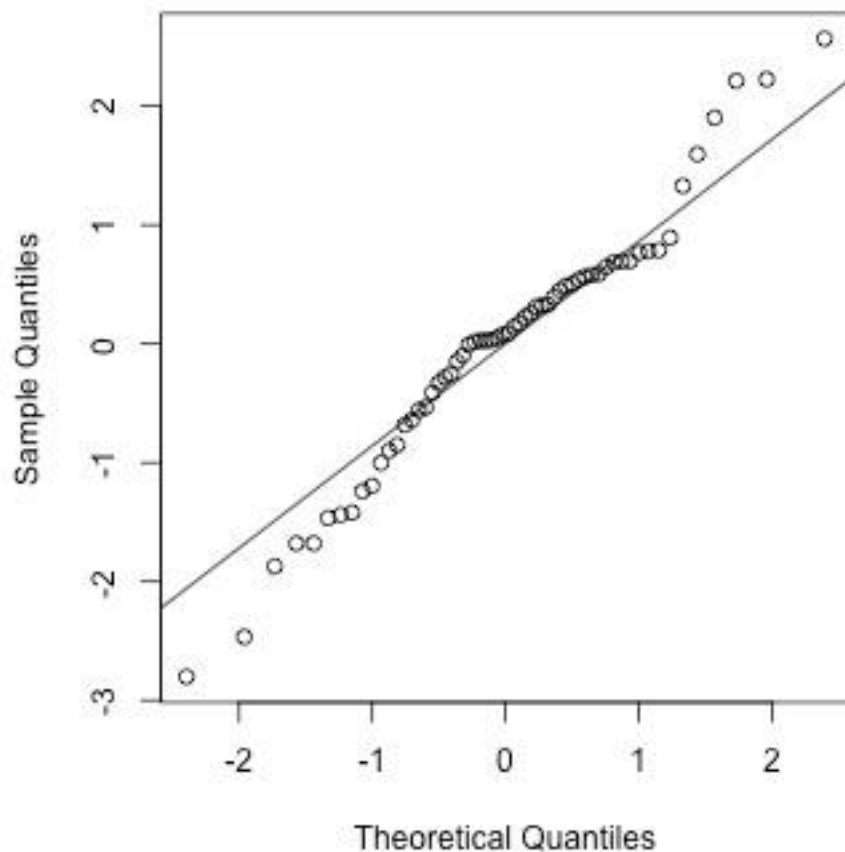


The residual variance increases as the fitted values increase.

4. Indications that Assumption of Normal Distribution for Errors is Valid



5. Indications that Assumption of Normal Distribution for Errors is Not Valid



The plot shows some deviation from the straight-line pattern indicating a distribution with heavier tails than a normal distribution.

Next

Chapter 6

Multiple Linear Regression