STAT509: Simple Linear Regression

Peijie Hou

University of South Carolina

November 3, 2014

Introduction

- Many problems in engineering and the sciences involve a study or analysis of the relationship between two or more variables.
- For example, we want to study the displacement of a particle d_t and time t. Let d_0 be the displacement of the particle from the origin at time $t = 0$ and v be the velocity, then we have a **deterministic linear relationship** $d_t = d_0 + vt$. We say it is **deterministic** since the model predicts displacement perfectly.
- However, there are many situations where the relationship between variables is not deterministic.
- For example, the electrical energy consumption of a house (y) is related to the size of the house (x, in square feet), but it is unlikely to be a deterministic relationship.

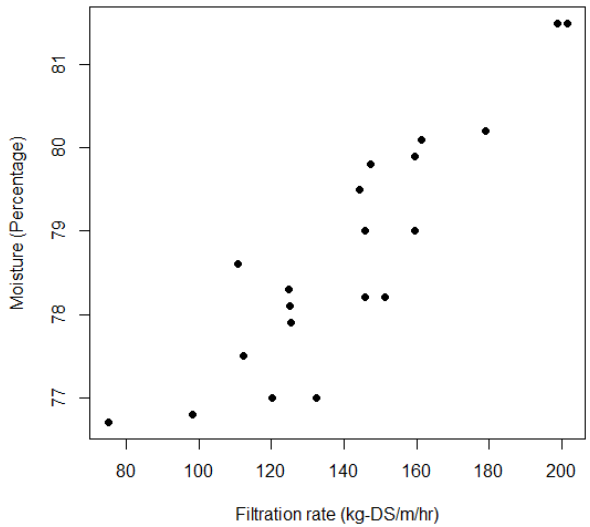
An Motivational Example

As part of a waste removal project, a new compression machine for processing sewage sludge is being studied. In particular, engineers are interested in the following variables:

Y = moisture control of compressed pellets (measured as a percent)

x = machine filtration rate (kg-DS/m/hr).

Engineers collect $n = 20$ observations of (x, Y); the data are displayed in the scatter diagram.



Introduction to Linear Regression

- No simple curve passed exactly through all the points.
- All the points scattered randomly around a straight line.
- It is reasonable to assume that the mean of the random variable Y is related to x by the following straight-line relationship:

$$E(Y) = \beta_0 + \beta_1 x$$

- **Regression coefficients:** β_0 (intercept), β_1 (slope)
- A probabilistic model is

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where ϵ is the **random error** term.

- We assume that $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2$
- We will call this model the **simple linear regression** model, because it has only one independent variable or **regressor**.

Properties of Simple Linear Regression

- β_0 quantifies the mean of Y when $x = 0$.
- β_1 quantifies the change in $E(Y)$ brought about by a one-unit change in x
- For the model $Y = \beta_0 + \beta_1 x + \epsilon$, we have

$$E(Y) = E(\beta_0 + \beta_1 x + \epsilon) = \beta_0 + \beta_1 x + E(\epsilon) = \beta_0 + \beta_1 x,$$

and

$$\mathrm{Var}\,(Y) = \mathrm{Var}\,(\beta_0 + \beta_1 x + \epsilon) = \mathrm{Var}\,(\epsilon) = \sigma^2.$$

Least squares estimation

- We want to fit a regression model, i.e, we would like to estimate the regression coefficients β_0 and β_1 using **least squares estimation**.
- Least squares says to choose the values β_0 and β_1 that minimize

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 x_i)]^2.$$

- Recall that we can minimize or maximize a multivariable function by taking the derivatives with respect to each arguments and set them to 0. So, taking partial derivative of $Q(\beta_0, \beta_1)$, we obtain

$$\begin{aligned} \frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_0} &= -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i) \stackrel{\text{set}}{=} 0 \\ \frac{\partial Q(\beta_0, \beta_1)}{\partial \beta_1} &= -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i) x_i \stackrel{\text{set}}{=} 0 \end{aligned}$$

Solution of LSE

- Solve above system of equations yields the **least squares estimators**

$$\begin{aligned}\hat{\beta}_0 &= \overline{Y} - \hat{\beta}_1 \overline{x} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{SS_{xy}}{SS_{xx}}.\end{aligned}$$

- In real life, it is rarely necessary to calculate $\hat{\beta}_0$ and $\hat{\beta}_1$ by hand. Let us look at how to use R to fit a regression model in the waste removal project example

```
#enter the data
filtration.rate=c(125.3,98.2,201.4,147.3,145.9,124.7,112.2,120.2,161.2,178.9,
                  159.5,145.8,75.1,151.4,144.2,125.0,198.8,132.5,159.6,110.7)
moisture=c(77.9,76.8,81.5,79.8,78.2,78.3,77.5,77.0,80.1,80.2,79.9,
                79.0,76.7,78.2,79.5,78.1,81.5,77.0,79.0,78.6)
# Fit the model
fit = lm(moisture~filtration.rate)
fit
Call:
lm(formula = moisture ~ filtration.rate)
Coefficients:
    (Intercept)  filtration.rate
       72.95855          0.04103
```

Solution of LSE

- From the output, we see that the least squares estimates are $\hat{\beta}_0 = 72.959$, and $\hat{\beta}_1 = 0.041$.
- Therefore, the equation of the least squares line that relates moisture percentage Y to the filtration rate x is

$$\widehat{Y} = 72.959 + 0.041x.$$

That is to say an estimate of expected moisture is given by

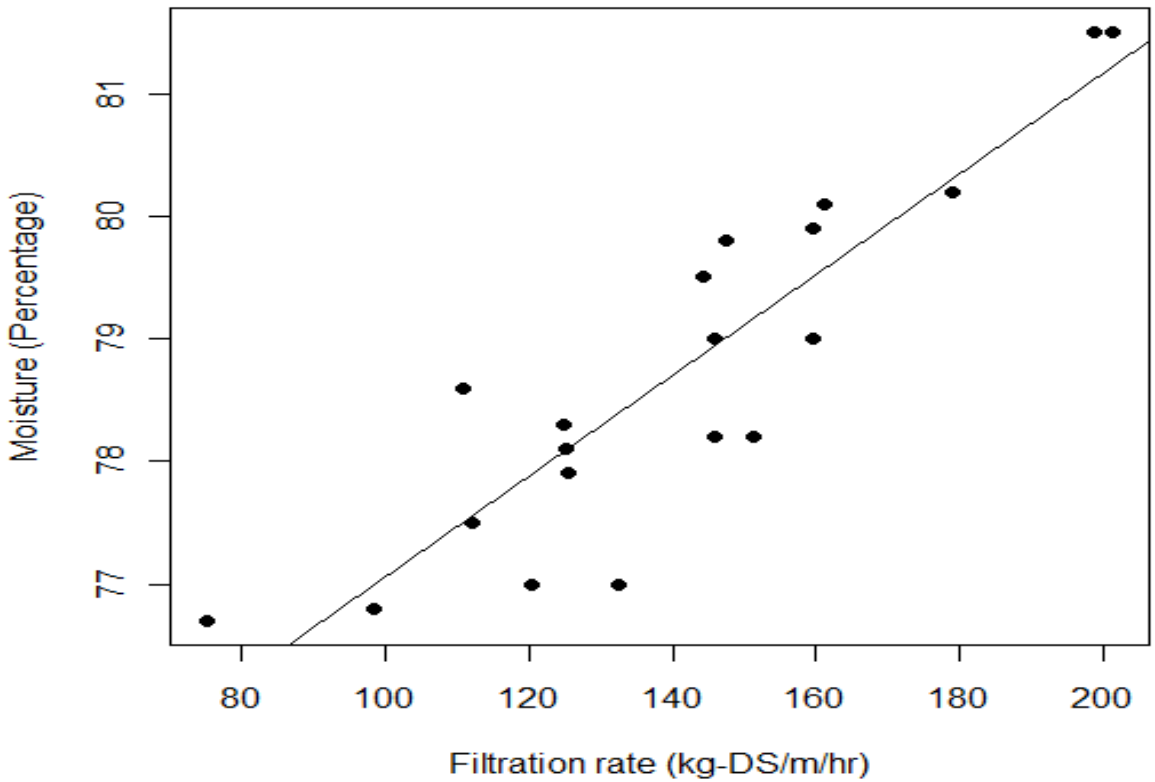
$$\widehat{\text{Moisture}} = 72.959 + 0.041\text{Filtration rate}.$$

- The least squares line is also called *prediction equation*. We can predict the mean response $E(Y)$ for any value of x. For example, when the filtration rate is $x = 150\text{kg}\cdot\text{ DS/m/hr}$, we would predict the mean moisture percentage to be

$$\widehat{Y}(150) = 72.959 + 0.041(150) = 79.109.$$

Scatter Plot with Least Squares Line

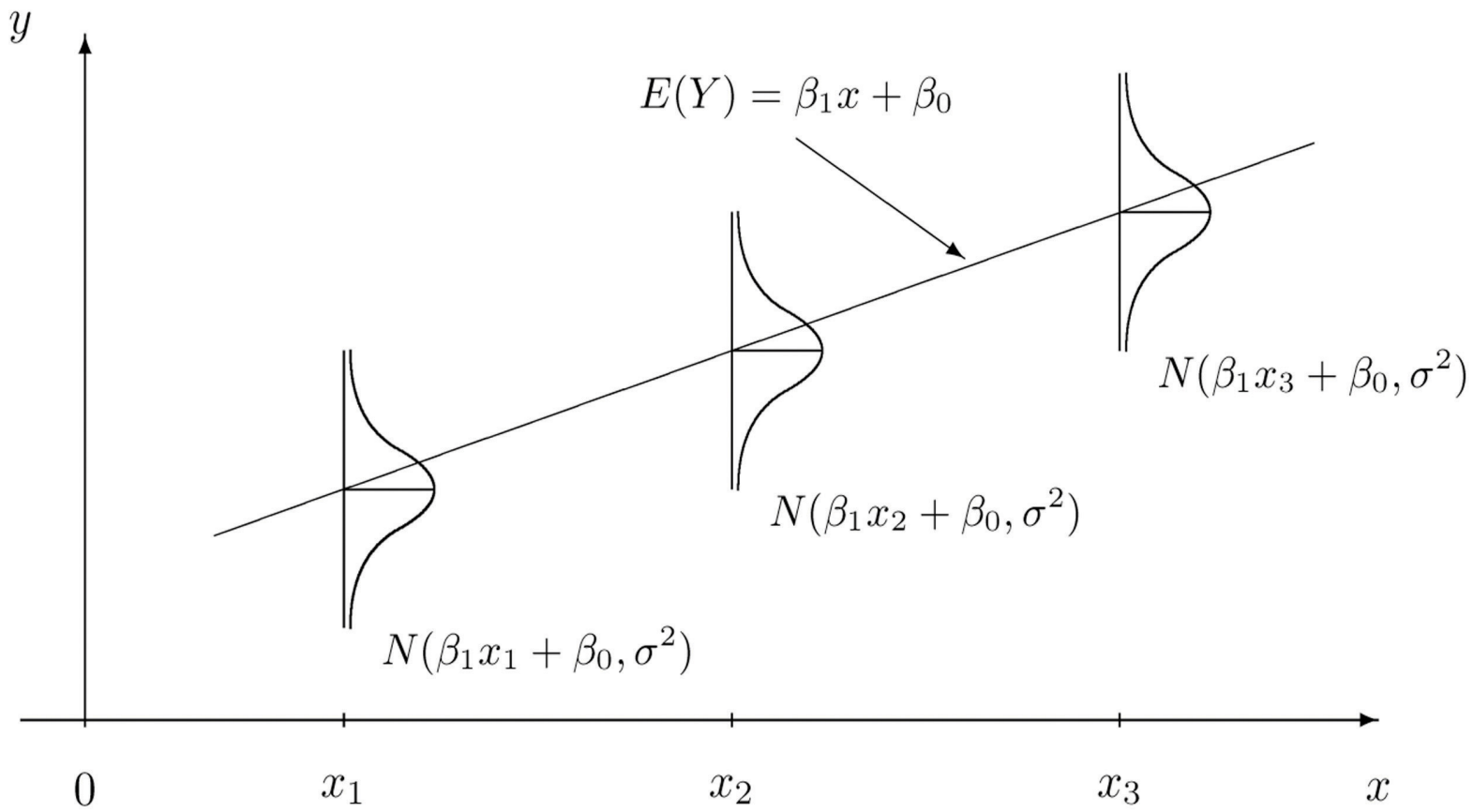
```
plot(filtration.rate,moisture,xlab = "Filtration rate (kg-DS/m/hr)",
      ylab = "Moisture (Percentage)",pch=16)
abline(fit)
```



Model Assumptions

- We have
$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$
We will assume the error term ϵ_i follows
  - $E(\epsilon_i) = 0$, for $i = 1, 2, \ldots, n$
  - $\mathrm{Var}(\epsilon_i) = \sigma^2$, for $i = 1, 2, \ldots, n$, i.e., the variance is constant
  - the random variable ϵ_i are independent
  - the random variable ϵ_i are normally distributed
- Those assumptions of the error terms can be summarized as
$$\epsilon_1, \epsilon_2, \ldots, \epsilon_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2),$$
where $i.i.d.$ stands for independent and identically distributed.
- Under the assumptions, it follows that
$$Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$
- We have three unknown but fixed parameters to estimate, namely, β_0, β_1, and σ^2.

Pictorial Illustration of Model Assumptions



Estimating σ^2

- We know we can use least squares method to estimate β_0 and β_1.
- The residuals $e_i = y_i - \hat{y}_i$ are used to obtain an estimator of σ^2. The sum of squares of the residuals, often called the **error sum of squares**, is

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

- It can be shown that the expected value of the error sum of squares is $E(SSE) = (n-2)\sigma^2$.
- Therefore an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{SSE}{n-2}$$

$\hat{\sigma}^2$ is also called mean squared error (**MSE**).

Properties of Least Squares Estimators

- Recall that

$$\begin{aligned}\hat{\beta}_0 &= \overline{Y} - \hat{\beta}_1 \overline{x} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{SS_{xy}}{SS_{xx}}.\end{aligned}$$

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of Y_i, so they are random variables and have their **sampling distributions.**
- It can be shown that

$$\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \left(\frac{1}{n} + \frac{\overline{x}^2}{SS_{xx}}\right)\sigma^2\right) \text{ and } \hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{SS_{xx}}\right)$$

- Note that both $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased.

Properties of Least Squares Estimators

- Since σ^2 is unknown, the **estimated standard error** of $\hat{\beta}_0$ and $\hat{\beta}_1$ are

$$se(\hat{\beta}_0) = \sqrt{\left(\frac{1}{n} + \frac{\overline{x}^2}{SS_{xx}}\right)\hat{\sigma}^2} \text{ and } se(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{SS_{xx}}}$$

where

$$\hat{\sigma}^2 = \frac{SSE}{n-2}$$

- We can use the standard errors to make hypothesis tests on β_0 and β_1.

Hypothesis Tests in Simple Linear Regression

- An important part of assessing the adequacy of a linear regression model is testing statistical hypotheses about the model parameters and constructing certain confidence intervals.
- In practice, inference for the slope parameter β_1 is of primary interest because of its connection to the independent variable x in the model.
- Inference for β_0 is less meaningful, unless one is explicitly interested in the mean of Y when $x = 0$. We will focus on inference on β_1.
- Under our model assumptions, the following sampling distribution arises:

$$t = \frac{\hat{\beta}_1 - \beta_1}{\widehat{se(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2/SS_{xx}}} \sim t(n-2)$$

Calculating $\hat{\sigma}^2$ in R

In R, `predict(fit)` gives the predicted value at each x_i, namely, $\hat{Y}(x_1), \hat{Y}(x_2), \ldots, \hat{Y}_{x_n}$.

```
> fit = lm(moisture~filtration.rate)
> fitted.values = predict(fit)
> residuals = moisture-fitted.values
> # Calculate MSE
> sum(residuals^2)/18
[1] 0.4426659
```

We have $\hat{\sigma}^2 = MSE = 0.443$.

Confidence Interval of $\hat{\beta}_1$

- The sampling distribution of $\hat{\beta}_1$ leads to the following $(1-\alpha)100\%$ confidence interval of β_1:

$$\underbrace{\hat{\beta}_1}_{\text{Point Estimate}} \pm \underbrace{t_{\alpha/2}}_{\text{Quantile}} \underbrace{\sqrt{\hat{\sigma}^2/SS_{xx}}}_{\text{standard error}}$$

- Note that this is two-sided confidence interval, which corresponds to the test $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$.
  - If '0'" is covered by this interval, we fail to reject H_0 at significance level of α. This suggests that Y and x are not linearly related.
  - If '0'" is not covered by this interval, we reject H_0 at significance level of α. This suggests that Y and x are linearly related.

Hypothesis Test for β_1

- Suppose we want to test β_1 equals to a certain value, say $\beta_{1,0}$, that is our interest is to test

$$H_0 : \beta_1 = \beta_{1,0} \text{ versus } H_a : \beta_1 \neq \beta_{1,0}$$

  where $\beta_{1,0}$ is often set to 0 (why?)
- The test statistic under the null is

$$t_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / SS_{xx}}} \sim t(n-2).$$

- The p-value of the test is $2P(T_{n-2} < -|t_0|)$, you can use R to find this probability. Remember that smaller p-value provide stronger evidence against H_0
- Let us look at removal project example...

Removal Project Example

We wish to test $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$.

```
fit = lm(moisture~filtration.rate)
summary(fit)

Call:
lm(formula = moisture ~ filtration.rate)

Residuals:
     Min       1Q   Median       3Q      Max
-1.39552 -0.27694  0.03548  0.42913  1.09901

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     72.958547   0.697528 104.596  < 2e-16 ***
filtration.rate  0.041034   0.004837   8.484 1.05e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6653 on 18 degrees of freedom
Multiple R-squared: 0.7999,     Adjusted R-squared: 0.7888
F-statistic: 71.97 on 1 and 18 DF,  p-value: 1.052e-07
```

What is your conclusion based the R output? Note that the residual standard error is $\sqrt{\hat{\sigma}^2} = \sqrt{MSE} = 0.6653$.

Analysis of Variance Approach to Test Significance of Regression

- (Analysis of Variance Identity) We decompose the total variability into

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

- We usually call $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ the error sum of squares and $SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ the regression sum of squares.
- Symbolically, we have

$$SSTO = SSR + SSE.$$

- We want to test $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.
- It can be shown that

$$E(SSE/(n-2)) = \sigma^2, \text{ and } E(SSR) = \sigma^2 + \beta_1^2 S_{xx}.$$

- If $H_0 : \beta_1 = 0$ is true, it can be shown that

$$F_0 = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F(1, n-2).$$

- We will reject H_0, if p-value is small.
- We can summarize these results in the ANOVA table.
- This overall test F test is equivalent to the t test approach for testing β_1.

ANOVA Table for Simple Linear Regression

Source of Variation	SS	df	MS	F
Regression	$\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$	1	$\frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
Error	$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$n-2$	$\frac{SSE}{n-2}$	
Total	$\sum_{i=1}^{n}(y_i - \overline{y})^2$	$n-1$		

Removal Project Example: F Test

We wish to test $H_0 : \beta_1 = 0$ against $H_a : \beta_1 \neq 0$ using the ANOVA approach. You can use `anova` command in `R`.

```
> # Fit the model
> fit = lm(moisture~filtration.rate)
> anova(fit)
Analysis of Variance Table

Response: moisture
                Df Sum Sq Mean Sq F value    Pr(>F)
filtration.rate  1 31.860  31.860  71.973 1.052e-07 ***
Residuals       18  7.968   0.443
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Again, we reject $H_0 : \beta_1 = 0$ at any reasonable α level and conclude that there is a strong evidence support $\beta_1 \neq 0$.

Confidence and prediction intervals for a given $x = x_0$ P.421 - 425

- Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

- We are interested in using the fitted model to learn about the response variable Y at a certain setting for the independent variable, say, $x = x_0$.
- Two potential goals:
  - Estimating the **mean response** of Y. This value is the mean of the following probability distribution

$$Y(x_0) \sim \mathcal{N}(\beta_0 + \beta_1 x_0, \sigma^2)$$

  - Predicting a **new response** Y, denoted by $Y^*(x_0)$. This value is a new outcome from

$$Y(x_0) \sim \mathcal{N}(\beta_0 + \beta_1 x_0, \sigma^2)$$

Confidence and prediction intervals for a given $x = x_0$ Cont'd

- *GOALS*: We would like to create $100(1-\alpha)\%$ intervals for the mean $E(Y|x_0)$ and for the new value $Y^*(x_0)$.
- The former is called a **confidence interval** and the latter is called a **prediction interval**.
- A $100(1-\alpha)\%$ **confidence interval** for the mean $E(Y|x_0)$ is

$$\widehat{Y}(x_0) \pm t_{n-2,\alpha/2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}\right]}$$

- A $100(1-\alpha)\%$ **prediction interval** for the new value $Y^*(x_0)$ is

$$\widehat{Y}(x_0) \pm t_{n-2,\alpha/2}\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}\right]}$$

Confidence and prediction intervals for a given $x = x_0$ Cont'd

- Note that the prediction interval is wider than the confidence interval!
- The length of the interval is smallest when $x_0 = \overline{x}$ and will get larger the farther x_0 is from $\overline{x}$ in either direction.
- Warning: It can be very dangerous to estimate $E(Y|x_0)$ or predict $Y^*(x_0)$ based on the fit of the model for values of x_0 outside the range of x values used in the experiment/study. This is called **extrapolation**.

Removal Project Example

In the removal Project example, suppose that we are interested in estimating $E(Y|x_0)$ and predicting a new $Y^*(x_0)$ when the filtration rate is $x_0 = 150$.

- $E(Y|x_0)$ denotes the mean moisture percentage for compressed pellets when the machine filtration rate is $x_0 = 150$.
- $Y^*(x_0)$ denotes a possible value of Y for a single run of the machine when the filtration rate is set at $x_0 = 150$.

Removal Project Example

- Confidence interval:

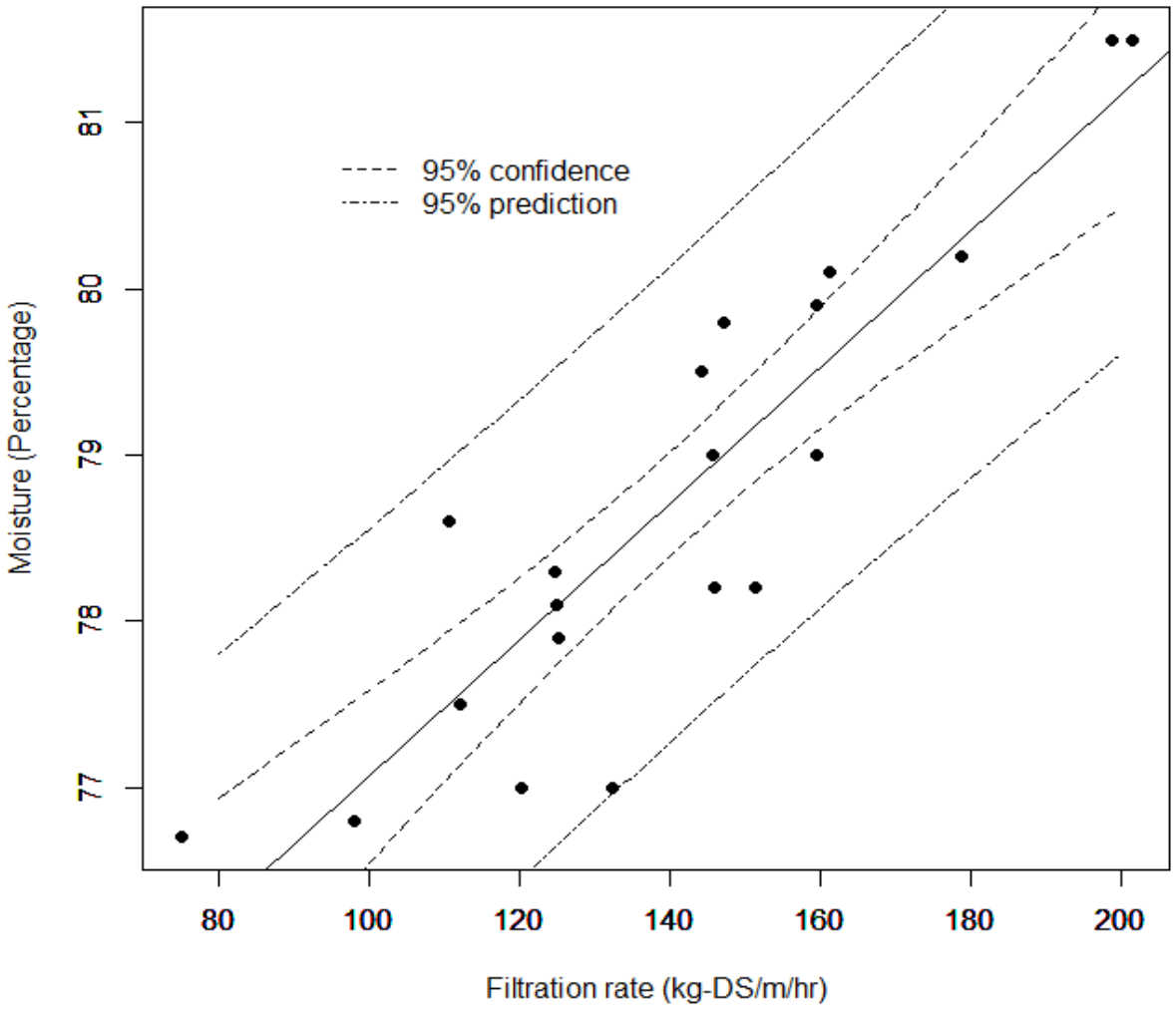
```
>  predict(fit,data.frame(filtration.rate=150),level=0.95,interval="confidence")
       fit      lwr      upr
1 79.11361 78.78765 79.43958
```

- Prediction interval:

```
>  predict(fit,data.frame(filtration.rate=150),level=0.95,interval="prediction")
       fit     lwr      upr
1 79.11361 77.6783 80.54893
```

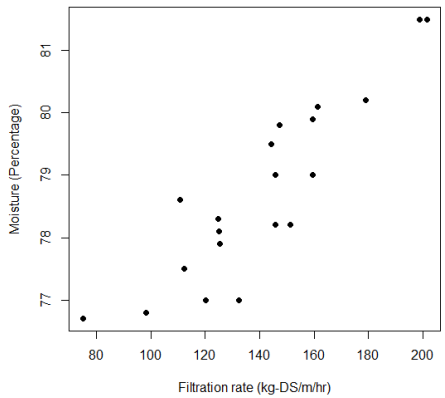
- Interpretation
  - A 95% confidence interval for $E(Y|x_0 = 150)$ is $(78.79, 79.44)$. When the filtration rate is $x_0 = 150$ kg-DS/m/hr, we are 95% confident that the **mean** moisture percentage is between 78.79 and 79.44 percent.
  - A 95 percent prediction interval for $Y^*(x_0 = 150)$ is $(77.68, 80.55)$. When the filtration rate is $x_0 = 150$ kg-DS/m/hr, we are 95% confident that the moisture percentage for a **single run** of the experiment will be between 77.68 and 80.55 percent.

Confidence Intervals Versus Prediction Intervals



Coefficient of Correlation

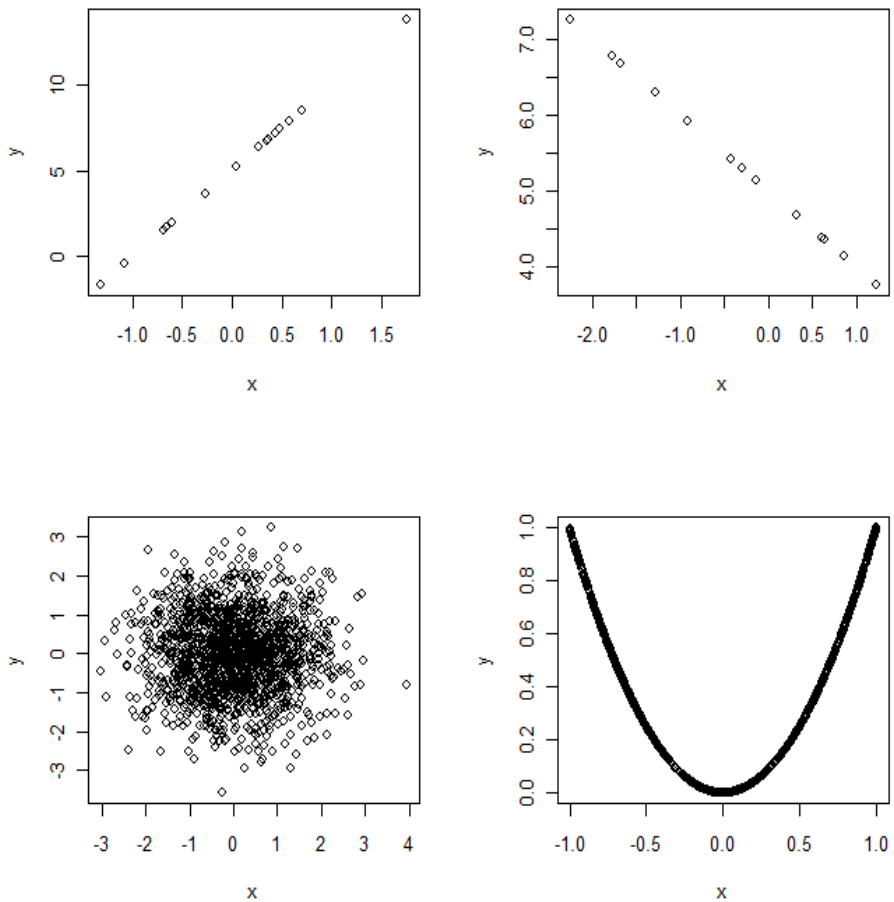
- **Correlation** measures the linear relationship between two quantitative variables.
- For example,



- To assign a numeric value: sample coefficient of correlation defined as

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})}} = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}.$$

The plot in the top left corner has $r = 1$; the plot in the top right corner has $r = -1$; the plot in the bottom left and right corner have $r \approx 0$;



Coefficient of Determination

- **Coefficient of Determination**, denoted by r^2, measures the contribution of x in the predicting of y.
- Recall that

$$SSTO = \sum_{i=1}^{n}(Y_i - \overline{Y})^2, SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

- If x makes no contribution to prediction of y, then $\beta_1 = 0$. In this case,

$$Y = \beta_0 + \epsilon.$$

  It can be shown that $\hat{Y}_i = \hat{\beta}_0 = \overline{Y}$, and $SSE = SSTO$.
- If x contribute to prediction of Y_i, then we expect $SSE << SSTO$. In other words, the independent variable x "explain" significant amount of variability among data.

Coefficient of Determination

- Intuitively, $SSTO$ is total sample variation around $\overline{Y}$, and SSE is unexplained sample variability after fitting regression line.
- Proportion of total sample variation explained by linear relationship:

$$\frac{\text{Explained Variability}}{\text{Total Variability}} = \frac{SSR}{SSTO}.$$

- Coefficient of determination is defined as

$$r^2 = \frac{SSTO - SSE}{SSTO} = \frac{SSR}{SSTO}.$$

- It can be shown that the coefficient of determination of a simple linear regression equals to the squared sample coefficient of correlation between x and Y.

Example: Removal Project Example

We can use command `cor` to calculate sample coefficient of correlation. The coefficient of determination r^2 is called `Multiple R-squared` in the summary of simple linear regression.

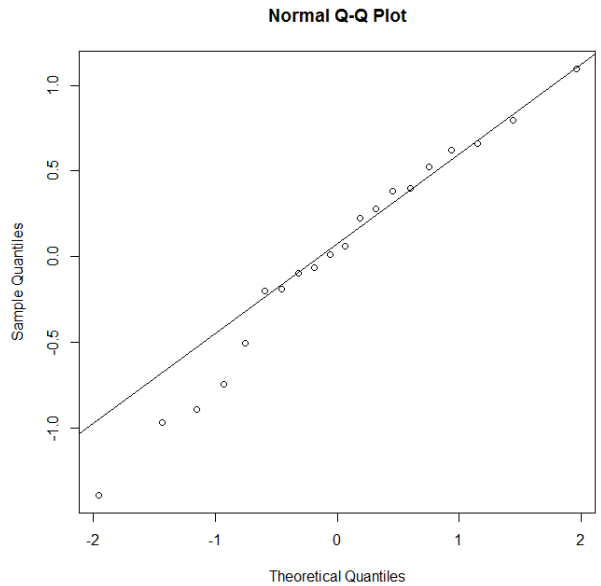
```
> cor(filtration.rate,moisture)
[1] 0.8943937
> fit<-lm(moisture~filtration.rate)
> summary(fit)
Call:
lm(formula = moisture ~ filtration.rate)
Residuals:
     Min       1Q   Median       3Q      Max
-1.39552 -0.27694  0.03548  0.42913  1.09901
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     72.958547   0.697528 104.596  < 2e-16 ***
filtration.rate  0.041034   0.004837   8.484 1.05e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6653 on 18 degrees of freedom
Multiple R-squared: 0.7999,     Adjusted R-squared: 0.7888
F-statistic: 71.97 on 1 and 18 DF,  p-value: 1.052e-07
> r<-cor(filtration.rate,moisture)
> r^2
[1] 0.7999401
```

Model Adequacy Checking: Residual Analysis P. 426

- The **residuals** from a regression model are $e_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, n.$, where y_i is an actual observation and $\hat{y}_i$ is the corresponding fitted value from the regression model.
- Analysis of the residuals is frequently helpful in checking the assumption that the errors are approximately normally distributed with constant variance, and in determining whether additional terms in the model would be useful.
- As an approximate check of normality, we can use apply the fat pencil test to the normal probability plot of residuals.
- Model checking is an important exercise because if the model assumptions are violated, then our analysis (and all subsequent interpretations) could be compromised.

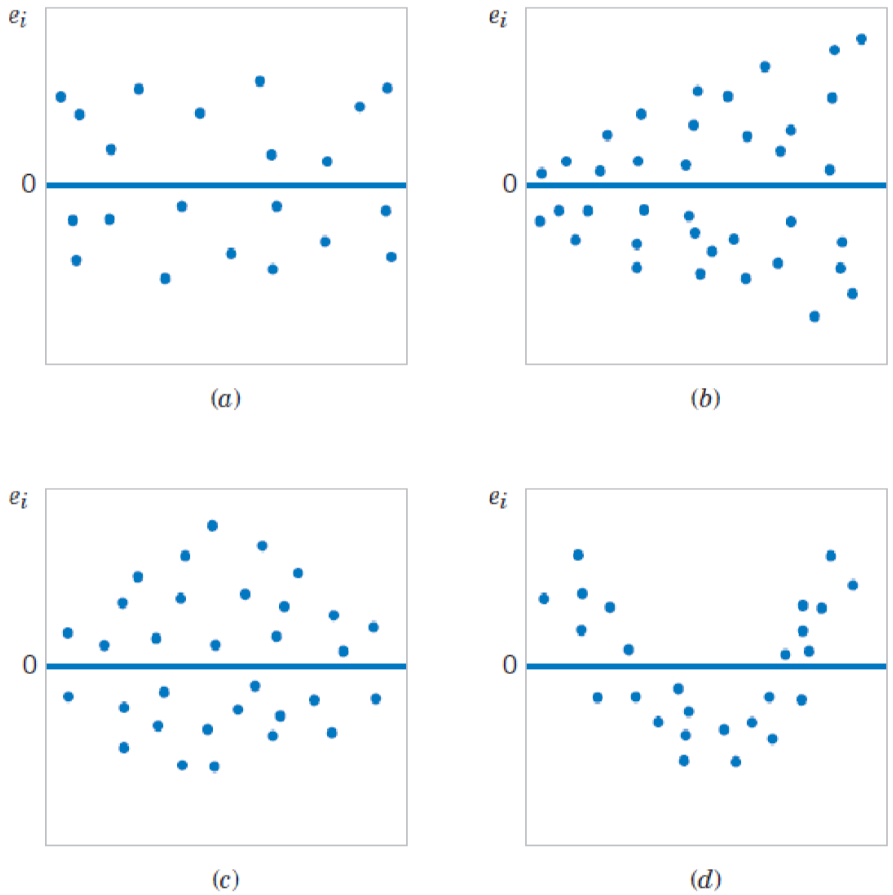
- Recall we have four assumptions on the error terms ϵ_i
  1. $E(\epsilon_i) = 0$, for $i = 1, 2, \ldots, n$
  2. $\text{Var}(\epsilon_i) = \sigma^2$, for $i = 1, 2, \ldots, n$, that is, variance is constant
  3. the random variables ϵ_i are independent
  4. the random variables ϵ_i are normally distributed.
- It is frequently helpful to plot the residuals (1) against $\hat{y}_i$, and (2) against the independent variable x.
- Q-Q plot for removal project

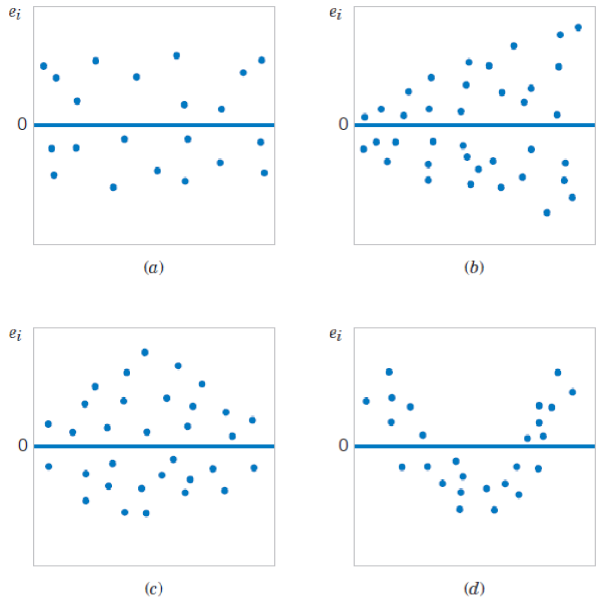
```
resid<-residuals(fit)
qqnorm(resid);qqline(resid)
```



Residual Plots

The residual plot is simply the scatterplot of residuals e_i's and predicted values. These graphs will usually look like one of the four general patterns shown below



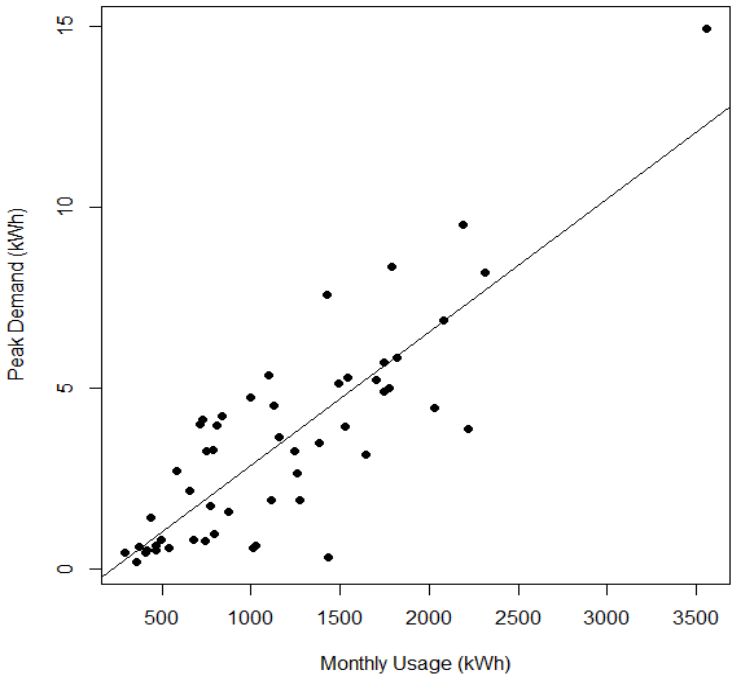


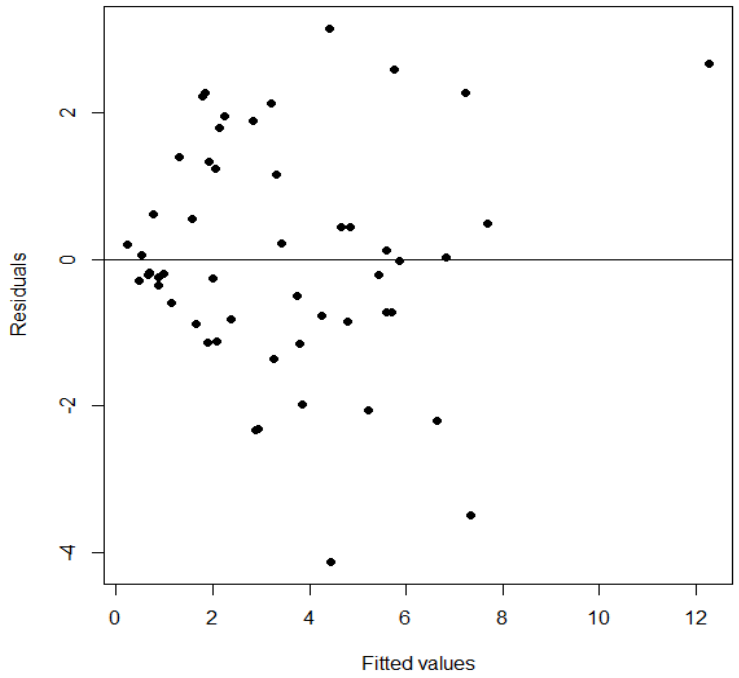
- Pattern (a) represents the ideal situation.
- Pattern (b) represents the cases where the variance of the observations may be increasing with the magnitude of y_i or x_i. Pattern (b) and (c) represents the unequal variance cases.
- Pattern (d) indicates the linear relationship between $E(Y_i)$ and x_i is not proper. We need to add higher order term, which requires multiple linear regression.

Example: Electricity Consumption

An electric company is interested in modeling peak hour electricity demand (Y) as a function of total monthly energy usage (x). This is important for planning purposes because the generating system must be large enough to meet the maximum demand imposed by customers.

```
electricity<-read.table(file.choose(),head=TRUE)
# Define variables
monthly.usage = electricity[,1]
peak.demand = electricity[,2]
# Fit the model
fit = lm(peak.demand ~ monthly.usage)


# Plots were constructed separately
# Scatterplot
plot(monthly.usage,peak.demand,xlab = "Monthly Usage (kWh)",
      ylab = "Peak Demand (kWh)", pch=16)
abline(fit)
# Residual plot
plot(fitted(fit),residuals(fit),pch=16,
      xlab="Fitted values",ylab="Residuals")
abline(h=0)
```



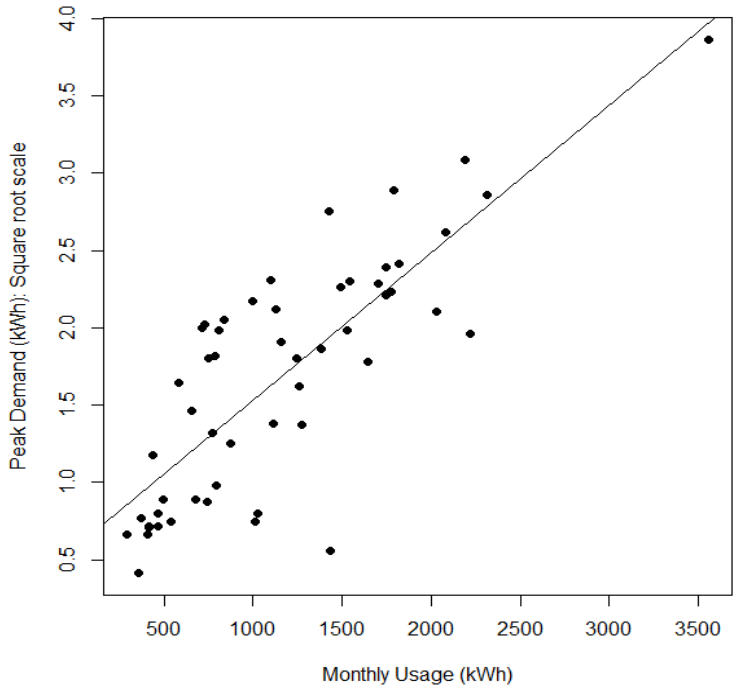


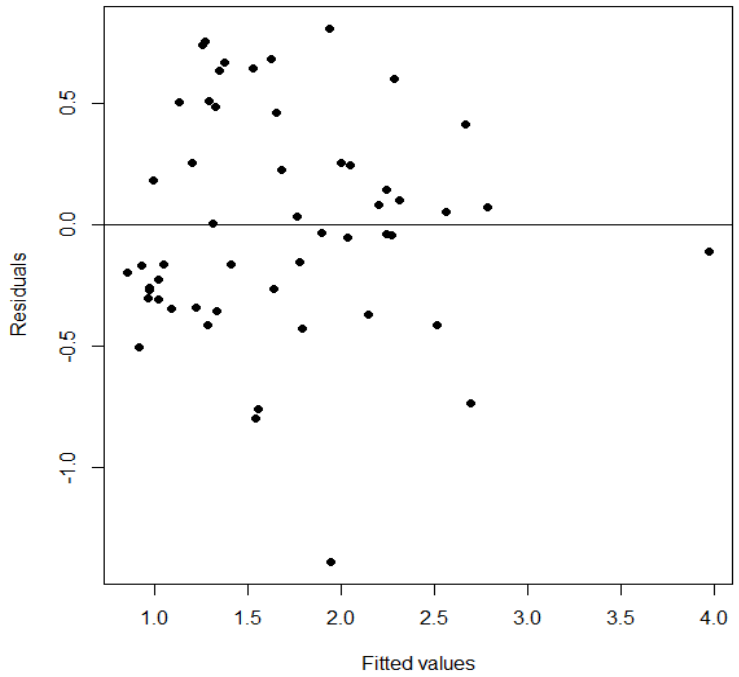
- The residual plot shows clearly a "megaphone" shape, which indicates that the equal variance assumption is violated.
- Widely used variance-stabilizing transformations include the use of $\sqrt{y}$, $\log y$, or $1/y$ as the response.
- Let us try $\sqrt{y}$ as the response.

Transforming the Response

You can use `sqrt(peak.demand)` in `R` to transform the response variable directly.

```
# Fit the transformed model
fit.2 = lm(sqrt(peak.demand) ~ monthly.usage)
fit.2

# Plots were constructed separately
# Scatterplot
plot(monthly.usage,sqrt(peak.demand),xlab = "Monthly Usage (kWh)",
      ylab = "Peak Demand (kWh): Square root scale", pch=16)
abline(fit.2)
# Residual plot
plot(fitted(fit.2),residuals(fit.2),pch=16,
      xlab="Fitted values",ylab="Residuals")
abline(h=0)
```





- The residual plot looks much better.
- Model interpretation: $\sqrt{Y_i} = \beta_0 + \beta_1 x_i + \epsilon_i$.