$$
\begin{aligned}
Var(\hat{\beta}_0) &= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sigma^2 \sum x_i^2}{n\sum(x_i - \bar{x})^2} \\
Var(\hat{\beta}_1) &= \frac{n\sigma^2}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \\
Cov(\hat{\beta}_0, \hat{\beta}_1) &= \frac{-\sigma^2 \sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{-\sigma^2 \bar{x}}{\sum(x_i - \bar{x})^2}
\end{aligned}
$$

Proof Let's calculate the variance of $\hat{\beta}_1$ first.

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}
$$

Therefore,

$$
\begin{aligned}
Var(\hat{\beta}_1) &= Var(\frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}) \\
&= \frac{1}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \sum_{i=1}^n (x_i - \bar{x})^2 Var(y_i) \\
&= \frac{1}{[\sum_{i=1}^n (x_i - \bar{x})^2]^2} \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2 \\
&= \frac{\sigma^2}{\sum(x_i - \bar{x})^2} = \frac{\sigma^2}{\sum x_i^2 - n\bar{x}^2}
\end{aligned}
$$

Lemma 1 Under the assumptions of linear regression model, $Cov(\bar{y}, \hat{\beta}_1) = 0$.

$$
\begin{aligned}
cov(\bar{y}, \hat{\beta}_1) &= cov(\bar{y}, \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}) \\
&= \frac{1}{\sum(x_i - \bar{x})^2} cov(\bar{y}, \sum(x_i - \bar{x})y_i) \\
&= \frac{1}{\sum(x_i - \bar{x})^2} cov(\frac{1}{n}\sum_j y_j, \sum_i (x_i - \bar{x})y_i) \\
&= \frac{1}{n\sum(x_i - \bar{x})^2} \sum_i \sum_j (x_i - \bar{x}) cov(y_i, y_j) \\
&= \frac{1}{n\sum(x_i - \bar{x})^2} \sum(x_i - \bar{x})\sigma^2 \\
&= 0
\end{aligned}
$$

To calculate the variance of $\hat{\beta}_0$, we will use the relationship between $\hat{\beta}_0$ and $\hat{\beta}_1$, i.e., $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

$$\begin{aligned}
Var(\hat{\beta}_0) &= Var(\bar{y} - \hat{\beta}_1 \bar{x}) \\
&= Var(\bar{y}) + Var(\hat{\beta}_1)\bar{x}^2 - 2cov(\bar{y}, \hat{\beta}_1 \bar{x}) \\
&= \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{\sum (x_i - \bar{x})^2} \\
&= \frac{\sigma^2}{n \sum (x_i - \bar{x})^2} [\sum (x_i - \bar{x})^2 + n\bar{x}^2] \\
&= \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}
\end{aligned}$$

Finally, let's calculate $Cov(\hat{\beta}_0, \hat{\beta}_1)$.

$$\begin{aligned}
cov(\hat{\beta}_0, \hat{\beta}_1) &= cov(\bar{y} - \bar{x}\hat{\beta}_1, \hat{\beta}_1) \\
&= cov(\bar{y}, \hat{\beta}_1) - \bar{x} var(\hat{\beta}_1) \\
&= \frac{-\sigma^2 \bar{x}}{\sum (x_i - \bar{x})^2}
\end{aligned}$$

Theorem 2 indicates that the variances depend on both the error variance σ^2 and x_i. Since x_i are known, we only need to estimate σ^2.

Note, the LSE of (β_0, β_1) st the same as its MLE. (homework).

3.1.3 Estimate of σ^2 and Inference of parameters

Definition Residual Sum of Squares (RSS)

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

An unbiased estimator of σ^2 is

$$s^2 = \frac{RSS}{n-2}.$$

The proof for the unbiasedness is a little bit complicated ...

$$\begin{aligned}
RSS &= \sum [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})]^2 \\
&= \sum (y_i - \bar{y})^2 + \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum (x_i - \bar{x})(y_i - \bar{y})
\end{aligned}$$

Thus $E[RSS] = ...(homework)$.

The variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ given in theorem 2 can be estimated by replacing σ^2 with s^2:

$$\begin{aligned} s^2_{\hat{\beta}_0} &= \frac{s^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\ s^2_{\hat{\beta}_1} &= \frac{ns^2}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \end{aligned}$$

Theorem 3 Under assumptions of linear model,

$$RRS/\sigma^2 \sim \chi^2_{n-2}$$

and RSS is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.

3.2 Statistical Inference

3.2.1 Inference of the Slope and Intercept

Theorem 4 Under assumptions of linear model,

$$T_j = \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \sim t_{n-2}$$

where $s_{\hat{\beta}_j} = \sqrt{\hat{Var}[\hat{\beta}_j]}$ is the standard error of $\hat{\beta}_j$ with $RRS/(n-2)$ plugged in for σ^2.

Proof

We will show the proof for $j = 1$. The proof for $j = 0$ is very similar.

$$\begin{aligned} T_1 &= \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \\ &= \frac{\hat{\beta}_1 - \beta_1}{\sqrt{var[\hat{\beta}_1]}} / \frac{s_{\hat{\beta}_1}}{\sqrt{var[\hat{\beta}_1]}} (***) \end{aligned}$$

It is clear that

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{var[\hat{\beta}_1]}} \sim N(0,1)$$

Note that

$$\begin{aligned}\frac{s_{\hat{\beta}_1}}{\sqrt{var[\hat{\beta}_1]}} &= \sqrt{\frac{\frac{s^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}{\frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}} \\ &= \sqrt{\frac{s^2}{\sigma^2}} \\ &= \sqrt{\frac{RSS/(n-2)}{\sigma^2}}\end{aligned}$$

We have already learned that $RSS/\sigma^2 \sim \chi^2_{n-2}$. In addition, RSS is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.

Let

$$U = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{var[\hat{\beta}_1]}}, V = \frac{RSS}{\sigma^2}$$

The above steps and (***) indicate that

- $T = \frac{U}{\sqrt{V/(n-2)}}$.
- $U \sim N(0,1)$, $V \sim \chi^2_{n-2}$.
- U and V are independent

By the definition of t-distribution, $T_1 \sim t_{n-2}$.

Hence $T \sim t_{n-2}$.

When the normality assumption is not assumed but n is large enough, T follows t_{n-2} approximately.

As a result of Theorem 4, a two-sided 95% confidence interval for β_j is

$$\hat{\beta}_j \pm t_{n-2,0.975} s_{\hat{\beta}_j}$$

It is often of interest to test whether or not there is a linear association between x and y. The hypotheses are

$$H_0 : \beta_1 = 0$$

vs

$$H_1 : \beta_1 \neq 0$$

Based on Theorem 4, we reject the null at significance level α if $T_i > t_{n-2,1-\alpha/2}$

```
#relationship between complaints and rating
s2=sum( (y-beta0.hat-beta1.hat*x)^2)/(n-2)
s=sqrt(s2)

#standard error of beta1.hat
se.beta1= s / sqrt(sum( (x-mean(x))^2 ))
se.beta1

#95% CI for beta1
c(beta1.hat- qt(0.975, df=n-2)*se.beta1, beta1.hat+ qt(0.975, df=n-2)*se.beta1)

#t-statistic to test for linear association
t.beta1=beta1.hat / se.beta1
t.beta1

#two-sided p-value
2*(1-pt(abs(t.beta1), df=n-2))
```

We estimated that one unit increase in complaint score is associated with 0.75 unit increase in overall rating. We are 95% confident that the slope is between 0.55 and 0.95.

At significance level $\alpha = 0.05$, we reject the null hypothesis and conclude that overall rating is linearly associated with complaint score.

3.2.2 Inference regarding $E[y|X = x_h] = \beta_0 + \beta_1 x_h$

A common use of regression model is to estimate the expected value of the outcome y conditional upon a given value of x.

The expectation of y given $x = x_h$ is $y_h = E[y|x = x_h] = \beta_0 + \beta_1 x_h$.

$$\begin{aligned}
\hat{y}_h &= \hat{\beta}_0 + \hat{\beta}_1 x_h \\
&= (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 x_h \\
&= \bar{y} + \hat{\beta}_1 (x_h - \bar{x})
\end{aligned}$$

The variance of $\hat{y}_h$

$$\begin{aligned}
var[\hat{y}_h] &= var[\hat{\beta}_0 + \hat{\beta}_1 x_h] \\
&= var(\bar{y} + \hat{\beta}_1(x_h - \bar{x}) \\
&= var(\bar{y}) + (x_h - \bar{x})^2 var(\hat{\beta}_1) + 2(x_h - \bar{x}) cov(\bar{y}, \hat{\beta}_1) \\
&= \sigma^2[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]
\end{aligned}$$

The distribution of $\hat{y}_h$

If $y_i's$ are independent and normally distributed, $\bar{y}$ and $\hat{\beta}_1$ are also normally distributed. Recall that $\hat{y}_h$ is a linear combination of the two. Thus, it also follows a normal distribution.

$$\hat{y}_h \sim N(\beta_0 + \beta_1 x_h, \sigma^2[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}])$$

Equivalently,

$$\frac{\hat{y}_h - \beta_0 + \beta_1 x_h}{\sqrt{\sigma^2[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]}} \sim N(0,1)$$

Usually σ^2 is unknown. Replace σ^2 with s^2, we have

$$\frac{\hat{y}_h - E[y_h]}{\sqrt{s^2[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]}} \sim t_{n-2}$$

Based on the distribution above, a $(1-\alpha)100\%$ CI for $E[y_h]$ is given by

$$\hat{y}_h \pm t_{n-2,1-\alpha/2} \sqrt{s^2[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]}$$

Please try the following R code to estimate and construct CI for $E[y|x_h = 60]$

```
#conditional inference
#inference regarding the mean rating when xh=60
xh=60
#estimate
yh=beta0.hat + xh*beta1.hat
yh
#standard error
se.yh=s*sqrt(1/n + (xh-mean(x))^2/sum((x-mean(x))^2) )
```

```
se.yh

#95% c.i.
c(yh- qt(0.975, df=n-2)*se.yh, yh+ qt(0.975, df=n-2)*se.yh)
```

You will find that the estimate is 59.7 and a 95% CI is from 56.7 to 62.6.

3.2.3 Prediction for a new observation

Consider the following situation. We fit a linear regression using a data set we have. We are told that there is a new observation (which is not one of the observations that were used to build the linear model). We know its x value but we don't know its y value. The goal is to predict the y value using the linear model we built.

(a). known $\beta_0, \beta_1, \sigma^2$

$$y_{h(new)} \sim N(\beta_0 + \beta_1 x_h, \sigma^2)$$

$$\frac{y_{h(new)} - (\beta_0 + \beta_1 x_h)}{\sigma} \sim N(0, 1)$$

A $(1-\alpha)100\%$ prediction interval for $y_{h(new)}$ is given by

$$(\beta_0 + \beta_1 x_h) \pm \sigma z_{1-\alpha/2}$$

(b). unknown β_0, β_1, known σ^2

Because the slope and interception are unknown, we need to consider the distribution of

$$y_{h(new)} - \hat{y}_h$$

(1)

$$E(y_{h(new)} - \hat{y}_h) = 0$$

(easy to verify)

(2)

$$var[y_{h(new)} - \hat{y}_h] = var[y_{h(new)}] + var[\hat{y}_h] = \sigma^2 + \sigma^2(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2})$$

(3) Under the assumptions we made (normal and independent errors),

$$y_{h(new)} - \hat{y}_h \sim N(0, \sigma^2 + \sigma^2(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}))$$

$$\frac{y_{h(new)} - \hat{y}_h}{\sqrt{\sigma^2 + \sigma^2(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2})}} \sim N(0,1)$$

A $(1-\alpha)100\%$ CI for $y_{h(new)}$ is

$$\hat{y}_h \pm z_{1-\alpha/2}\sqrt{\sigma^2[1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]}$$

(c). no parameters is known

$$\frac{y_{h(new)} - \hat{y}_h}{\sqrt{s^2 + s^2(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2})}} \sim t_{n-2}$$

A $(1-\alpha)100\%$ prediction interval for $y_{h(new)}$ is given by

$$\hat{y}_h \pm t_{n-2,1-\alpha/2}\sqrt{s^2[1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]}$$

Suppose that the complaint score of a company (which was not one of the observatoins in the original data) is 65. The following code shows how to make a prediction and construct a 95% CI.

```
##prediction
#suppose that the complaint score of a company (which was not surveyed) is 65
xh=60
#predicted value
yh=beta0.hat + xh*beta1.hat
yh
#standard error
se.yh=s*sqrt(1+1/n + (xh-mean(x))^2/sum((x-mean(x))^2) )
se.yh

#95% c.i.
c(yh- qt(0.975, df=n-2)*se.yh, yh+ qt(0.975, df=n-2)*se.yh)
```

So our prediction for the overall rating of the company is 59.7 and we are 95% confident that the true overall rating is between 45.0 and 74.3.

3.3 Residuals

Let $e_i = y_i - \hat{y}_i$, where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. e_i is called the residual for unit i, i.e., the difference between the observed and the fitted for unit i.

It is useful to look at other forms $e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i = (y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})$

(1)The sum of the residuals is zero: $\sum_{i=1}^n e_i = 0$. As a result, $\sum y_i = \sum \hat{y}_i$

Proof: $\sum e_i = \sum[(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})] = \sum (y_i - \bar{y}) - \hat{\beta}_1 \sum (x_i - \bar{x}) = 0$

(2) The sum of weighted residuals is zero when the residual in the ith trial is weighted by the level of the predictor variable in the ith observation:

$$\sum_{i=1}^n x_i e_i = 0$$

Note, the predictor vector and the residual vector are orthogonal, as their inner product is zero.

Proof:

$$\begin{aligned}
\sum x_i e_i &= \sum (x_i - \bar{x}) e_i \\
&= \sum x_i [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})] \\
&= (\sum x_i y_i - n\bar{x}\bar{y}) - \hat{\beta}_1 (\sum x_i^2 - n\bar{x}^2) \\
&= \sum (x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1 \sum (x_i - \bar{x})^2 \\
&= 0
\end{aligned}$$

(*) is true because

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

(3)A consequence of (1) and (2) is

$$\sum_{i=1}^n \hat{y}_i e_i = 0$$

So the vector of fitted values and the residuals are orthogonal.

Proof:

$$\sum_{i=1}^n e_i \hat{y}_i = \sum_{i=1}^n e_i (\hat{\beta}_0 + \hat{\beta}_1 x_i) = \hat{\beta}_0 \sum_{i=1}^n e_i + \hat{\beta}_1 \sum_{i=1}^n e_i x_i = 0$$

(4) The residual sum of squares (RSS)

$RSS = \sum e_i^2 = \sum (y_i - \bar{y})^2 - \sum (\hat{y}_i - \bar{y})^2$. Equivalently, we have $\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$

Proof:

$$\begin{aligned}
\sum e_i^2 &= \sum [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})]^2 \\
&= \sum (y_i - \bar{y})^2 + \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum (x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum (y_i - \bar{y})^2 + \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 - 2\hat{\beta}_1^2 \sum (x_i - \bar{x})^2 \\
&= \sum (y_i - \bar{y})^2 - \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 \\
&= \sum (y_i - \bar{y})^2 - \sum (\hat{y}_i - \bar{y})^2
\end{aligned}$$

The last step is true because

$$\hat{y}_i - \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x})$$

In fact, this is the basis for the ANOVA approach to linear regression:

$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 = SS_{Reg} + SS_{Error}$

3.4 ANOVA Approach to Linear Regression

The analysis of variance approach is based on partitioning of sums of squares and degrees of freedom associated with the response variable y.

The decomposition of total deviation:

$$y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y}$$

That is, total deviation = deviation around fitted regression line + deviation of fitted regression value around mean

The decomposition of $SSTO$:

$$SSTO = SS_{Reg} + SS_{Error}$$

where

$$\begin{aligned}
SSTO &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
SS_{Reg} &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2 = \sum (\hat{\beta}_1 x_i - \hat{\beta}_1 \bar{x})^2 \\
SS_{Error} &= \sum e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2
\end{aligned}$$