Linear Regression Models

Segment 1 – Simple Linear Regression Model

Topic 4 – Accuracy of the Coefficient Estimates

Sudarsan N.S. Acharya (sudarsan.acharya@manipal.edu)

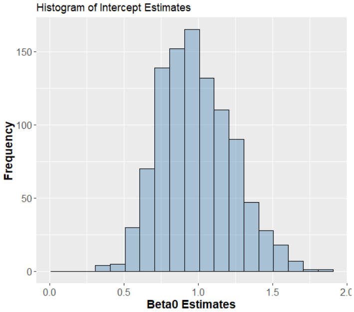
Topics

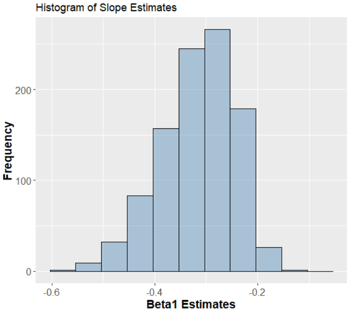
1. Accuracy of the Coefficient Estimates: Standard Errors
2. Accuracy of the Coefficient Estimates: Confidence Intervals
3. Accuracy of the Coefficient Estimates: Hypothesis Tests
4. Summary

Accuracy of the Coefficient Estimates: Standard Errors

Accuracy of the Coefficient Estimates: Standard Errors

How can we assess the accuracy of the SLRM coefficient estimates $\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$, and $\hat{\beta}_1 = \dfrac{s_{xy}}{s_{xx}}$ derived from a dataset?





Accuracy of the Coefficient Estimates: Standard Errors - Continued

Accuracy of the Coefficient Estimates: Standard Errors - Continued

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard deviations of the estimators are also referred to as their standard errors:

Accuracy of the Coefficient Estimates: Standard Errors - Continued

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- Standard deviations of the estimators are also referred to as their **standard errors**: $SE[\hat{\beta}_0] = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$,

Accuracy of the Coefficient Estimates: Standard Errors - Continued

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}} \right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with

Accuracy of the Coefficient Estimates: Standard Errors - Continued

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with $\sigma^2 \approx \frac{1}{n-2} \underbrace{\sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)}\right)^2}_{RSS}$.

Accuracy of the Coefficient Estimates: Standard Errors - Continued

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with $\sigma^2 \approx \frac{1}{n-2} \underbrace{\sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)}\right)^2}_{RSS}$.
- Note that the above are *in-sample* estimates;

Accuracy of the Coefficient Estimates: Standard Errors - Continued

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with $\sigma^2 \approx \frac{1}{n-2} \underbrace{\sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)}\right)^2}_{RSS}$.
- Note that the above are *in-sample* estimates; that is, calculated using the dataset.

Accuracy of the Coefficient Estimates: Standard Errors - Continued

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with $\sigma^2 \approx \frac{1}{n-2}\underbrace{\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2}_{RSS}$.
- Note that the above are *in-sample* estimates; that is, calculated using the dataset.
- Random error term $\varepsilon^{(i)}$ is assumed to have

Accuracy of the Coefficient Estimates: Standard Errors - Continued

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with

$$\sigma^2 \approx \frac{1}{n-2} \underbrace{\sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)}\right)^2}_{RSS}.$$

- Note that the above are *in-sample* estimates; that is, calculated using the dataset.
- Random error term $\varepsilon^{(i)}$ is assumed to have zero mean,

Accuracy of the Coefficient Estimates: Standard Errors - Continued

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with

$$\sigma^2 \approx \frac{1}{n-2} \underbrace{\sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)}\right)^2}_{RSS}.$$

- Note that the above are *in-sample* estimates; that is, calculated using the dataset.
- Random error term $\varepsilon^{(i)}$ is assumed to have zero mean, constant variance, and

Accuracy of the Coefficient Estimates: Standard Errors - Continued

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard deviations of the estimators are also referred to as their standard errors: $SE[\hat{\beta}_0] = \sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}_n^2}{s_{xx}}\right)}$, and $SE[\hat{\beta}_1] = \sqrt{\frac{\sigma^2}{s_{xx}}}$ with $\sigma^2 \approx \frac{1}{n-2}\underbrace{\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2}_{RSS}$.
- Note that the above are *in-sample* estimates; that is, calculated using the dataset.
- Random error term $\varepsilon^{(i)}$ is assumed to have zero mean, constant variance, and uncorrelated across the samples in the dataset.

Accuracy of the Coefficient Estimates: Confidence Intervals

Accuracy of the Coefficient Estimates: Confidence Intervals

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.

Accuracy of the Coefficient Estimates: Confidence Intervals

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.
- A 95% CI for β_0 is

Accuracy of the Coefficient Estimates: Confidence Intervals

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.
- A 95% CI for β_0 is $\left[\hat{\beta}_0 - 1.96 \times SE(\hat{\beta}_0),\ \hat{\beta}_0 + 1.96 \times SE(\hat{\beta}_0)\right]$.

Accuracy of the Coefficient Estimates: Confidence Intervals

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.
- A 95% CI for β_0 is $\left[\hat{\beta}_0 - 1.96 \times SE(\hat{\beta}_0),\ \hat{\beta}_0 + 1.96 \times SE(\hat{\beta}_0)\right]$.
- How to interpret this?

Accuracy of the Coefficient Estimates: Confidence Intervals

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.
- A 95% CI for β_0 is $\left[\hat{\beta}_0 - 1.96 \times SE(\hat{\beta}_0),\ \hat{\beta}_0 + 1.96 \times SE(\hat{\beta}_0)\right]$.
- How to interpret this? Each dataset gives us one such CI;

Accuracy of the Coefficient Estimates: Confidence Intervals

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.
- A 95% CI for β_0 is $\left[\hat{\beta}_0 - 1.96 \times SE(\hat{\beta}_0),\ \hat{\beta}_0 + 1.96 \times SE(\hat{\beta}_0)\right]$.
- How to interpret this? Each dataset gives us one such CI; approximately 95% of those intervals will contain the true population parameter β_0.

Accuracy of the Coefficient Estimates: Confidence Intervals

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.
- A 95% CI for β_0 is $\left[\hat{\beta}_0 - 1.96 \times SE(\hat{\beta}_0),\ \hat{\beta}_0 + 1.96 \times SE(\hat{\beta}_0)\right]$.
- How to interpret this? Each dataset gives us one such CI; approximately 95% of those intervals will contain the true population parameter β_0.
- In practice, however, we have one dataset and one CI that comes out of it.

Accuracy of the Coefficient Estimates: Confidence Intervals

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- The standard errors can be extended to calculate confidence intervals (CI) for population coefficients β_0 and β_1.
- A 95% CI for β_0 is $\left[\hat{\beta}_0 - 1.96 \times SE(\hat{\beta}_0),\ \hat{\beta}_0 + 1.96 \times SE(\hat{\beta}_0)\right]$.
- How to interpret this? Each dataset gives us one such CI; approximately 95% of those intervals will contain the true population parameter β_0.
- In practice, however, we have one dataset and one CI that comes out of it.
- For calculating CI as above, additional assumption on the random error term $\varepsilon^{(i)}$ that it is *normally distributed* is needed.

Accuracy of the Coefficient Estimates: Hypothesis Tests

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor:

Accuracy of the Coefficient Estimates: Hypothesis Tests

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- We want to check if there is indeed a relationship between mpg and hp;

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- We want to check if there is indeed a relationship between mpg and hp; how close should $\hat{\beta}_1$ be equal to 0 to conclude that there is no relationship?

Accuracy of the Coefficient Estimates: Hypothesis Tests

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- We want to check if there is indeed a relationship between mpg and hp; how close should $\hat{\beta}_1$ be equal to 0 to conclude that there is no relationship?
- We consider a *null hypothesis* $\beta_1 = 0$;

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- We want to check if there is indeed a relationship between mpg and hp; how close should $\hat{\beta}_1$ be equal to 0 to conclude that there is no relationship?
- We consider a *null hypothesis* $\beta_1 = 0$; that is, we hypothesize that there is no relationship between mpg and hp.

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- We want to check if there is indeed a relationship between mpg and hp; how close should $\hat{\beta}_1$ be equal to 0 to conclude that there is no relationship?
- We consider a *null hypothesis* $\beta_1 = 0$; that is, we hypothesize that there is no relationship between mpg and hp.

- *Assuming that the null hypothesis is true,*

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- We want to check if there is indeed a relationship between mpg and hp; how close should $\hat{\beta}_1$ be equal to 0 to conclude that there is no relationship?
- We consider a *null hypothesis* $\beta_1 = 0$; that is, we hypothesize that there is no relationship between mpg and hp.
- *Assuming that the null hypothesis is true,* $T = \frac{\hat{\beta}_1 - E[\hat{\beta}_1]}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \underbrace{\beta_1}_{=0}}{SE(\hat{\beta}_1)}$

Accuracy of the Coefficient Estimates: Hypothesis Tests

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Standard errors can be used to perform hypothesis tests on the coefficient estimates.
- Suppose that we build an SLRM using the *mtcars* dataset with mpg as the response and hp as the predictor: $\widehat{mpg} = \hat{\beta}_0 + \hat{\beta}_1 hp$.
- We want to check if there is indeed a relationship between mpg and hp; how close should $\hat{\beta}_1$ be equal to 0 to conclude that there is no relationship?
- We consider a *null hypothesis* $\beta_1 = 0$; that is, we hypothesize that there is no relationship between mpg and hp.
- *Assuming that the null hypothesis is true,* $T = \frac{\hat{\beta}_1 - E[\hat{\beta}_1]}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \underbrace{\beta_1}_{=0}}{SE(\hat{\beta}_1)}$ follows a t-distribution with $n-2$ degrees of freedom.

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

- From a particular dataset, we have the estimate $\hat{\beta}_1$ and the realized value of T denoted as t.

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- From a particular dataset, we have the estimate $\hat{\beta}_1$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, which is the probability of observing a realization of T that is more extreme than what we observed from the dataset given that the null hypothesis is true.

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

- From a particular dataset, we have the estimate $\hat{\beta}_1$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, which is the probability of observing a realization of T that is more extreme than what we observed from the dataset given that the null hypothesis is true.
- This probability is referred to as the p-value;

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

MANIPAL ACADEMY of HIGHER EDUCATION (Institution of Eminence Deemed to be University)

- From a particular dataset, we have the estimate $\hat{\beta}_1$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, which is the probability of observing a realization of T that is more extreme than what we observed from the dataset given that the null hypothesis is true.
- This probability is referred to as the p-value; we reject the null hypothesis if the p-value is smaller than a threshold, typically 0.05.

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

- From a particular dataset, we have the estimate $\hat{\beta}_1$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, which is the probability of observing a realization of T that is more extreme than what we observed from the dataset given that the null hypothesis is true.
- This probability is referred to as the p-value; we reject the null hypothesis if the p-value is smaller than a threshold, typically 0.05.
- Rejecting the null hypothesis is equivalent to accepting the alternative hypothesis that $\beta_1 \neq 0$, and therefore mpg and hp are indeed related.

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

- From a particular dataset, we have the estimate $\hat{\beta}_1$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, which is the probability of observing a realization of T that is more extreme than what we observed from the dataset given that the null hypothesis is true.
- This probability is referred to as the p-value; we reject the null hypothesis if the p-value is smaller than a threshold, typically 0.05.
- Rejecting the null hypothesis is equivalent to accepting the alternative hypothesis that $\beta_1 \neq 0$, and therefore mpg and hp are indeed related.
- If the p-value is greater than the threshold, we fail to reject the null hypothesis;

Accuracy of the Coefficient Estimates: Hypothesis Tests - Cotinued

- From a particular dataset, we have the estimate $\hat{\beta}_1$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, which is the probability of observing a realization of T that is more extreme than what we observed from the dataset given that the null hypothesis is true.
- This probability is referred to as the p-value; we reject the null hypothesis if the p-value is smaller than a threshold, typically 0.05.
- Rejecting the null hypothesis is equivalent to accepting the alternative hypothesis that $\beta_1 \neq 0$, and therefore mpg and hp are indeed related.
- If the p-value is greater than the threshold, we fail to reject the null hypothesis; this means that there possibly is no relationship between mpg and hp.

Summary

Summary

Different ways to assess the accuracy of coefficient estimates using standard errors, confidence intervals, and hypothesis tests.