Assessing Model Accuracy

Nate Wells

Math 243: Stat Learning

September 9th, 2020

Outline

In today's class, we will...

- Discuss theoretical foundation for linear regression
- Assess accuracy of simple linear models
- Implement simple linear regression in R

Foundations

Linear Regression

- Suppose we have one or more predictors $(X_1, X_2, \ldots, X_p)$ and a *quantitative* respone variable Y, and that
$$Y = f(X_1, \ldots, X_p) + \epsilon$$

Linear Regression

- Suppose we have one or more predictors $(X_1, X_2, \ldots, X_p)$ and a *quantitative* respone variable Y, and that
$$Y = f(X_1, \ldots, X_p) + \epsilon$$

- The function f could theoretically take many forms. But the simplest form assumes f is a linear function:

$$f(x_1, x_2, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Linear Regression

- Suppose we have one or more predictors $(X_1, X_2, \ldots, X_p)$ and a *quantitative* respone variable Y, and that
$$Y = f(X_1, \ldots, X_p) + \epsilon$$

- The function f could theoretically take many forms. But the simplest form assumes f is a linear function:

$$f(x_1, x_2, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

(That is, the change in f is constant per unit change in any of the inputs.)

- If Y depends on only 1 predictor X, then the linear model reduces to

$$\hat{f}(x) = \beta_0 + \beta_1 x$$

Linear Regression

- Suppose we have one or more predictors $(X_1, X_2, \ldots, X_p)$ and a *quantitative* respone variable Y, and that
$$Y = f(X_1, \ldots, X_p) + \epsilon$$
- The function f could theoretically take many forms. But the simplest form assumes f is a linear function:
$$f(x_1, x_2, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$
(That is, the change in f is constant per unit change in any of the inputs.)
- If Y depends on only 1 predictor X, then the linear model reduces to
$$\hat{f}(x) = \beta_0 + \beta_1 x$$
- We'll use the Simple Linear Model (SLM) to build intuition about all linear models

Linear Regression

- Suppose we have one or more predictors $(X_1, X_2, \ldots, X_p)$ and a *quantitative* respone variable Y, and that
$$Y = f(X_1, \ldots, X_p) + \epsilon$$
- The function f could theoretically take many forms. But the simplest form assumes f is a linear function:
$$f(x_1, x_2, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$
(That is, the change in f is constant per unit change in any of the inputs.)
- If Y depends on only 1 predictor X, then the linear model reduces to
$$\hat{f}(x) = \beta_0 + \beta_1 x$$
- We'll use the Simple Linear Model (SLM) to build intuition about all linear models

Approximations and Estimates

- In reality, the relationship f between Y and $X_1, \ldots, X_p$ may not be linear

Approximations and Estimates

- In reality, the relationship f between Y and $X_1, \ldots, X_p$ may not be linear
- But many functions can be well-approximated by linear ones (especially when inputs are restricted to a small range)

Approximations and Estimates

- In reality, the relationship f between Y and $X_1, \ldots, X_p$ may not be linear
- But many functions can be well-approximated by linear ones (especially when inputs are restricted to a small range)
- But even if f is truly linear, we still have problems: We do not know the parameters of the linear model.

Approximations and Estimates

- In reality, the relationship f between Y and $X_1, \dots, X_p$ may not be linear
- But many functions can be well-approximated by linear ones (especially when inputs are restricted to a small range)
- But even if f is truly linear, we still have problems: We do not know the parameters of the linear model.
- Based on data, we estimate the parameters to create an estimated linear model

$$\hat{f} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$$

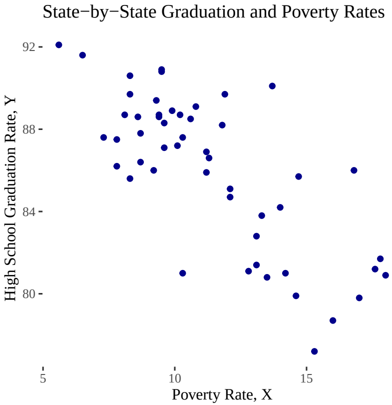
Approximations and Estimates

- In reality, the relationship f between Y and $X_1, \ldots, X_p$ may not be linear
- But many functions can be well-approximated by linear ones (especially when inputs are restricted to a small range)
- But even if f is truly linear, we still have problems: We do not know the parameters of the linear model.
- Based on data, we estimate the parameters to create an estimated linear model

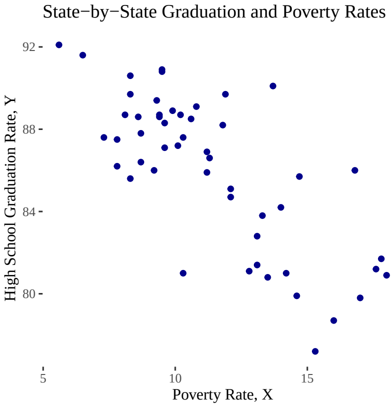
$$\hat{f} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$$

- So we are **estimating** an **approximation** to a relationship between response and predictors.

SLR Review

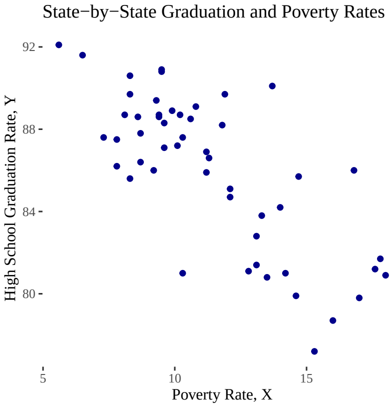

State−by−State Graduation and Poverty Rates
High School Graduation Rate, Y
92
88
84
80
5
10
15
Poverty Rate, X

SLR Review



- Suppose we want to model graduation rate Y as a function of poverty rate X

SLR Review



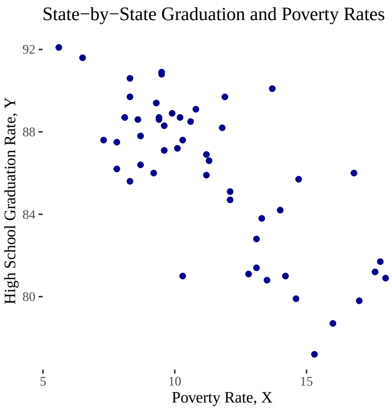
- Suppose we want to model graduation rate Y as a function of poverty rate X
- Let's assume a linear relationship

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

- Model (hand-fitted):

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X = 96.2 - 0.9X$$

SLR Review



- Suppose we want to model graduation rate Y as a function of poverty rate X
- Let's assume a linear relationship

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

- Model (hand-fitted):

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X = 96.2 - 0.9X$$

Residuals

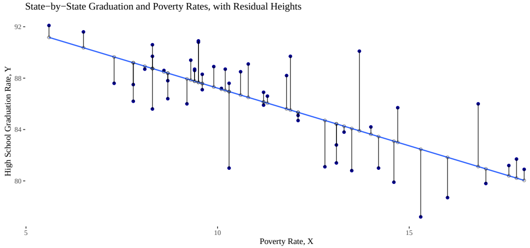
- **Residuals** are the leftover variation in the data after accounting for model fit.
- Each observation (x_i, y_i) has its own residual e_i, which is the difference between the observed (y_i) and predicted $(\hat{y}_i)$ value:

$$e_i = y_i - \hat{y}_i$$

Residuals

- **Residuals** are the leftover variation in the data after accounting for model fit.
- Each observation (x_i, y_i) has its own residual e_i, which is the difference between the observed (y_i) and predicted $(\hat{y}_i)$ value:

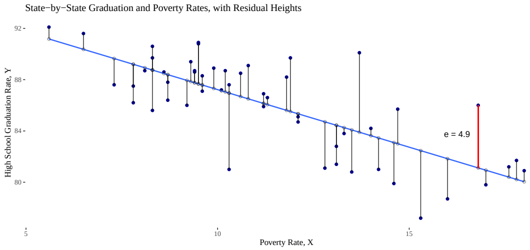
$$e_i = y_i - \hat{y}_i$$



Residuals

- **Residuals** are the leftover variation in the data after accounting for model fit.
- Each observation (x_i, y_i) has its own residual e_i, which is the difference between the observed (Y_i) and predicted $(\hat{y}_i)$ value:

$$e_i = y_i - \hat{y}_i$$

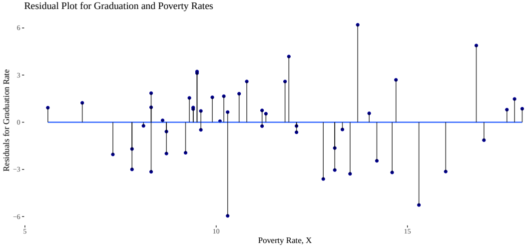


- D.C.'s residual is

$$e = y - \hat{y} = 86 - 81.1 = 4.9$$

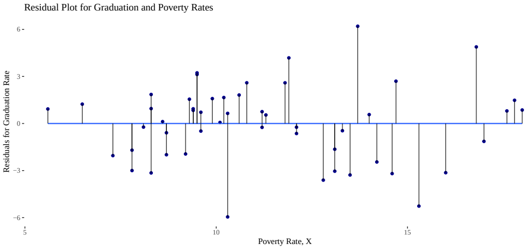
Residual Plot

- To visualize the degree of accuracy of a linear model, we use residual plots:



Residual Plot

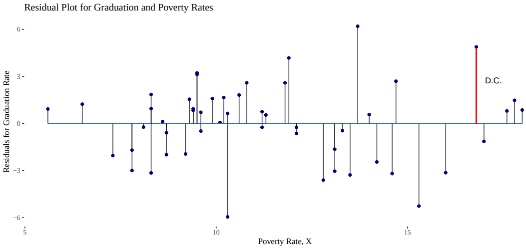
- To visualize the degree of accuracy of a linear model, we use residual plots:



- Points preserve original x-position, but with y-position equal to residual.

Residual Plot

- To visualize the degree of accuracy of a linear model, we use residual plots:



- Points preserve original x-position, but with y-position equal to residual.

Residual Sum of Squares

- Define the **Residual Sum of Squares** (RSS) as

$$\mathrm{RSS} = e_1^2 + \cdots + e_n^2$$

Residual Sum of Squares

- Define the **Residual Sum of Squares** (RSS) as

$$\mathrm{RSS} = e_1^2 + \cdots + e_n^2$$

Note that $\mathrm{RSS} = n\mathrm{MSE}$.

Residual Sum of Squares

- Define the **Residual Sum of Squares** (RSS) as

$$\mathrm{RSS} = e_1^2 + \cdots + e_n^2$$

  Note that $\mathrm{RSS} = n\mathrm{MSE}$.

- Using calculus, we can show that RSS is minimized when

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Assessing Accuracy

Statistical Inference

- **Goal**: Use *statistics* calculated from data to make estimates about unknown *parameters*

Statistical Inference

- **Goal**: Use *statistics* calculated from data to make estimates about unknown *parameters*
- **Parameters**: β_0, β_1

Statistical Inference

- **Goal**: Use *statistics* calculated from data to make estimates about unknown *parameters*
- **Parameters**: β_0, β_1
- **Statistics**: $\hat{\beta}_0$, $\hat{\beta}_1$

Statistical Inference

- **Goal**: Use *statistics* calculated from data to make estimates about unknown *parameters*
- **Parameters**: β_0, β_1
- **Statistics**: $\hat{\beta}_0$, $\hat{\beta}_1$
- **Tools**: confidence intervals, hypothesis tests

Statistical Inference

- **Goal**: Use *statistics* calculated from data to make estimates about unknown *parameters*
- **Parameters**: β_0, β_1
- **Statistics**: $\hat{\beta}_0$, $\hat{\beta}_1$
- **Tools**: confidence intervals, hypothesis tests
- **The Problems**: Our model will change if built using a different random sample. So in addition to estimates, we need to know about variability

The Confidene Interval

- Confidence Intervals give estimates **and** express an amount of uncertainty we have about those estimates

The Confidene Interval

- Confidence Intervals give estimates **and** express an amount of uncertainty we have about those estimates
- A *C*-level confidence interval for a parameter θ using the statistic $\hat{\theta}$ takes the form

$$\hat{\theta} \pm t_C^* \cdot \mathrm{SE}(\hat{\theta})$$

The Confidene Interval

- Confidence Intervals give estimates **and** express an amount of uncertainty we have about those estimates
- A C-level confidence interval for a parameter θ using the statistic $\hat{\theta}$ takes the form

$$\hat{\theta} \pm t_C^* \cdot \mathrm{SE}(\hat{\theta})$$

- Where t_C^* is the $1-(1-C)/2$ quantile for the sampling distribution of $\hat{\theta}$

The Confidene Interval

- Confidence Intervals give estimates **and** express an amount of uncertainty we have about those estimates
- A *C*-level confidence interval for a parameter θ using the statistic $\hat{\theta}$ takes the form

$$\hat{\theta} \pm t_C^* \cdot \mathrm{SE}(\hat{\theta})$$

- Where t_C^* is the $1 - (1 - C)/2$ quantile for the sampling distribution of $\hat{\theta}$
- And where $\mathrm{SE}(\hat{\theta})$ is the standard error of $\hat{\theta}$, or the standard deviation of the sampling distribution

Common Regression Assumptions

In order to safely use simple linear regression, use make use of these assumptions:

1. Y is related to x by a simple linear regression model.

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

Common Regression Assumptions

In order to safely use simple linear regression, use make use of these assumptions:

1. Y is related to x by a simple linear regression model.

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

2. The errors $e_1, e_2, \ldots, e_n$ are independent of one another.

Common Regression Assumptions

In order to safely use simple linear regression, use make use of these assumptions:

1. Y is related to x by a simple linear regression model.

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

2. The errors $e_1, e_2, \ldots, e_n$ are independent of one another.

3. The errors have a common variance $\mathrm{Var}(\epsilon) = \sigma^2$.

Common Regression Assumptions

In order to safely use simple linear regression, use make use of these assumptions:

1. Y is related to x by a simple linear regression model.

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

2. The errors $e_1, e_2, \ldots, e_n$ are independent of one another.
3. The errors have a common variance $\mathrm{Var}(\epsilon) = \sigma^2$.
4. The errors are normally distributed: $\epsilon \sim N(0, \sigma^2)$

The Sampling Distribution of $\hat{\beta}_1$

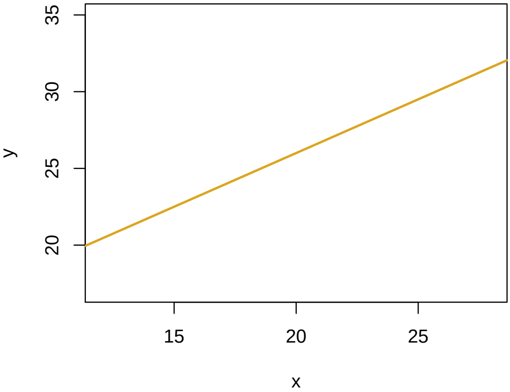
Assume the following true model:

$$f(x) = 12 + .7x; \epsilon \sim N(0, 4)$$

The Sampling Distribution of $\hat{\beta}_1$

Assume the following true model:

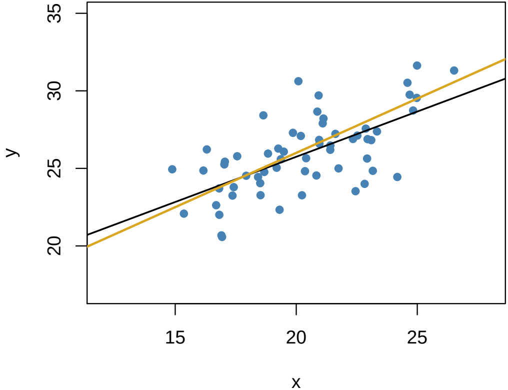
$$f(x) = 12 + .7x; \epsilon \sim N(0, 4)$$



The Sampling Distribution of $\hat{\beta}_1$

Assume the following true model:

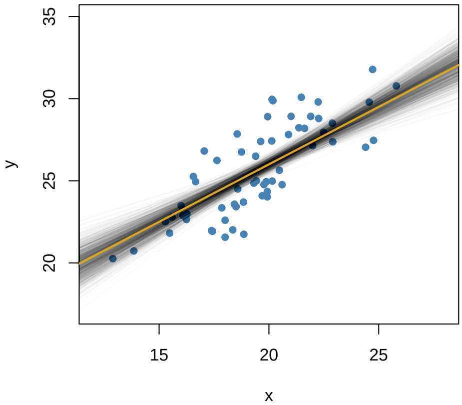
$$f(x) = 12 + .7x; \epsilon \sim N(0, 4)$$



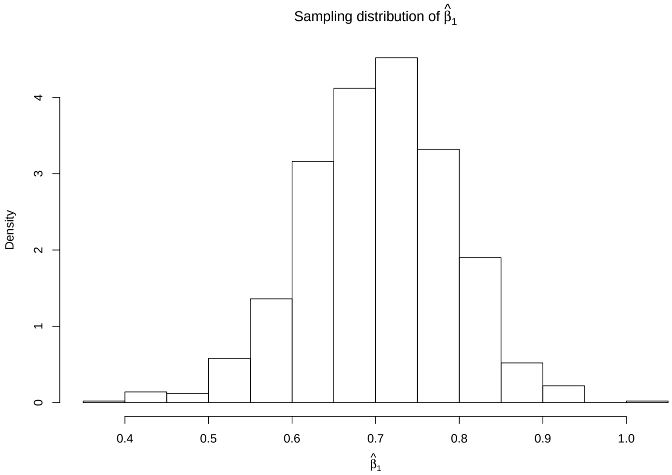
The Sampling Distribution of $\hat{\beta}_1$

Assume the following true model:

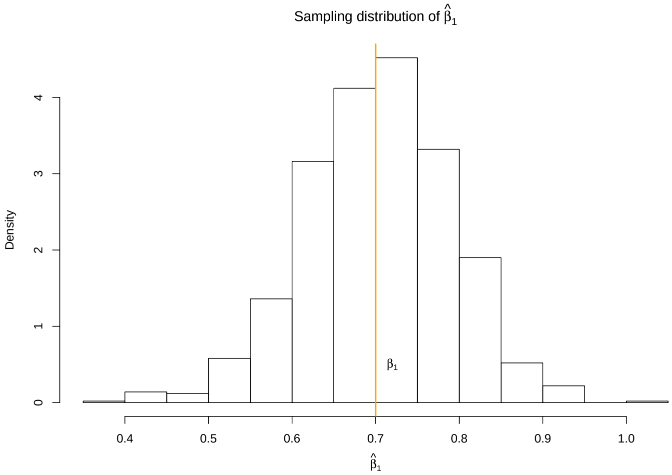
$$f(x) = 12 + .7x; \epsilon \sim N(0, 4)$$



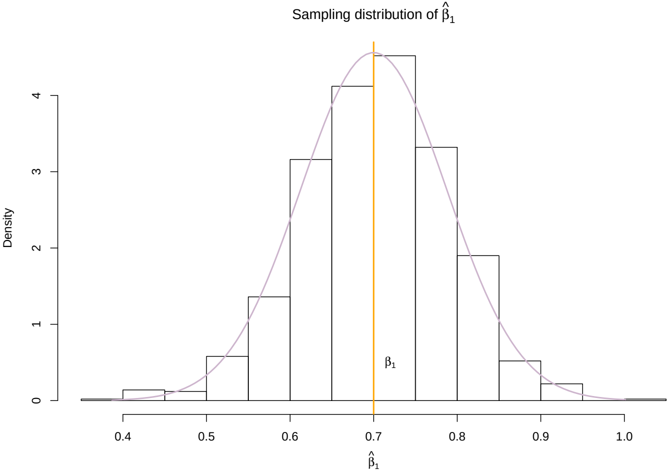
The Sampling Distribution of $\hat{\beta}_1$



The Sampling Distribution of $\hat{\beta}_1$



The Sampling Distribution of $\hat{\beta}_1$



The Sampling Distribution of $\hat{\beta}_1$

The Sampling Distribution has the following characteristics:

1. Centered at β_1, i.e. $E(\hat{\beta}_1) = \beta$.

The Sampling Distribution of $\hat{\beta}_1$

The Sampling Distribution has the following characteristics:

1. Centered at β_1, i.e. $E(\hat{\beta}_1) = \beta$.
2. $Var(\hat{\beta}_1) = \frac{\sigma^2}{SXX}$.
   - where $SXX = \sum_{i=1}^{n}(x_i - \bar{x})^2$

The Sampling Distribution of $\hat{\beta}_1$

The Sampling Distribution has the following characteristics:

1. Centered at β_1, i.e. $E(\hat{\beta}_1) = \beta$.
2. $Var(\hat{\beta}_1) = \frac{\sigma^2}{SXX}$.
   - where $SXX = \sum_{i=1}^{n}(x_i - \bar{x})^2$
3. $\hat{\beta}_1 | X \sim N(\beta_1, \frac{\sigma^2}{SXX})$.

Approximating the Sampling Dist. of $\hat{\beta}_1$

Our best guess of β_1 is $\hat{\beta}_1$. And since we have to estimate σ with $\hat{\sigma}^2 = RSS/n - 2$, the distribution isn't normal, but...

Approximating the Sampling Dist. of $\hat{\beta}_1$

Our best guess of β_1 is $\hat{\beta}_1$. And since we have to estimate σ with $\hat{\sigma}^2 = RSS/n - 2$, the distribution isn't normal, but...

T with n - 2 degrees of freedom.

Approximating the Sampling Dist. of $\hat{\beta}_1$

Our best guess of β_1 is $\hat{\beta}_1$. And since we have to estimate σ with $\hat{\sigma}^2 = RSS/n - 2$, the distribution isn't normal, but...

T with n - 2 degrees of freedom.

And we summarize that approximate sampling distribution using a CI:

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} * SE(\hat{\beta}_1)$$

where

$$SE(\hat{\beta}_1) = s/\sqrt{(}SXX)$$

Approximating the Sampling Dist. of $\hat{\beta}_1$

Our best guess of β_1 is $\hat{\beta}_1$. And since we have to estimate σ with $\hat{\sigma}^2 = RSS/n - 2$, the distribution isn't normal, but...

T with n - 2 degrees of freedom.

And we summarize that approximate sampling distribution using a CI:

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} * SE(\hat{\beta}_1)$$

where

$$SE(\hat{\beta}_1) = s/\sqrt{(}SXX)$$

Interpretation We are *95% confident* that the true slope between x and y lies between LB and UB.

Hypothesis test for $\hat{\beta}_1$

Suppose we are interested in testing the claim that the slope is zero.

$$H_0 : \beta_1^0 = 0 H_A : \beta_1^0 \neq 0$$

Hypothesis test for $\hat{\beta}_1$

Suppose we are interested in testing the claim that the slope is zero.

$$H_0 : \beta_1^0 = 0 H_A : \beta_1^0 \neq 0$$

We know that

$$T = \frac{\hat{\beta}_1 - \beta_1^0}{SE(\hat{\beta}_1)}$$

Hypothesis test for $\hat{\beta}_1$

Suppose we are interested in testing the claim that the slope is zero.

$$H_0 : \beta_1^0 = 0 H_A : \beta_1^0 \neq 0$$

We know that

$$T = \frac{\hat{\beta}_1 - \beta_1^0}{SE(\hat{\beta}_1)}$$

T will be t distributed with $n - 2$ degrees of freedom and with $SE(\hat{\beta}_1)$ calculated the same as in the CI.

Inference for $\hat{\beta}_0$

Often less interesting (but not always!). You use the t-distribution again but with a different SE.