Simple Linear Regression

Inference

Prof. Maria Tackett

STA 210

[Click for PDF of slides]

Topics

Topics

- Conduct a hypothesis test for β_1

Topics

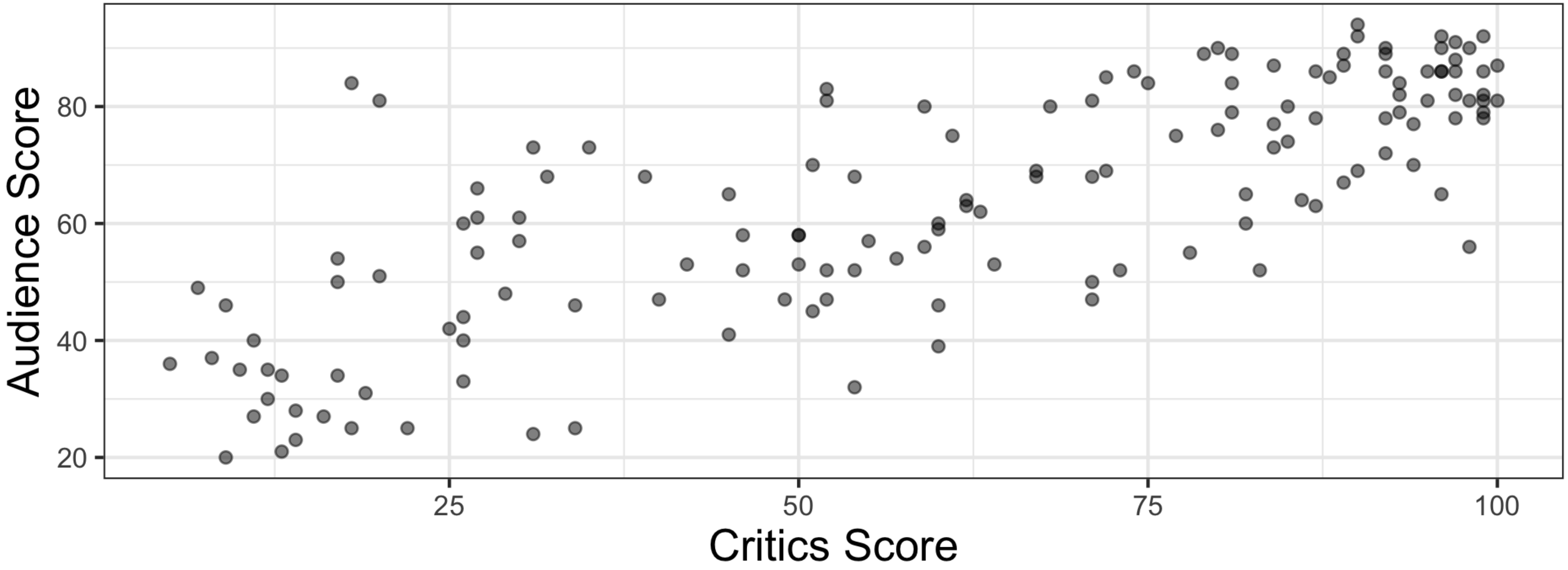
- Conduct a hypothesis test for β_1

- Calculate a confidence interval for β_1

STA 210

Movie ratings data

The data set contains the "Tomatometer" score (**`critics`**) and audience score (**`audience`**) for 146 movies rated on rottentomatoes.com.



The model

```
model <- lm(audience ~ critics, data = movie_scores)
```

```
model %>%
  tidy() %>%
  kable(format = "html", digits = 3)
```

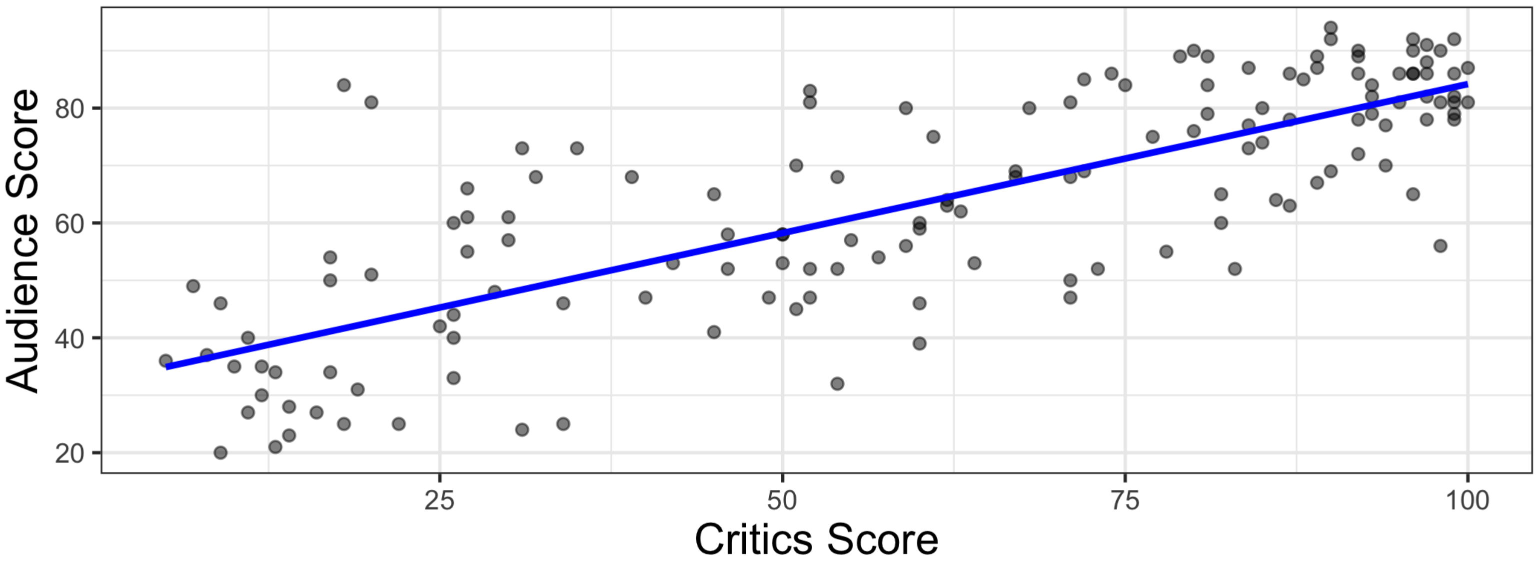
term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

STA 210

The model

$$\widehat{\text{audience}} = 32.316 + 0.519 \times \text{critics}$$

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0



Does the data provide sufficient evidence that β_1 is significantly different from 0?

Outline of a hypothesis test

Outline of a hypothesis test

1️⃣ State the hypotheses.

Outline of a hypothesis test

1️⃣ State the hypotheses.

2️⃣ Calculate the test statistic.

STA 210

Outline of a hypothesis test

1️⃣ State the hypotheses.

2️⃣ Calculate the test statistic.

3️⃣ Calculate the p-value.

Outline of a hypothesis test

1️⃣ State the hypotheses.

2️⃣ Calculate the test statistic.

3️⃣ Calculate the p-value.

4️⃣ State the conclusion.

1️⃣ State the hypotheses

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

STA 210

1️⃣ State the hypotheses

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

Null hypothesis

STA 210

1️⃣ State the hypotheses

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

Null hypothesis

Alternative hypothesis

STA 210

2️⃣ Calculate the test statistic

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$\text{test statistic} = \frac{\text{Estimate} - \text{Hypothesized}}{\text{Standard error}}$$

STA 210

2️⃣ Calculate the test statistic

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$t = \frac{\hat{\beta}_1 - 0}{SE_{\hat{\beta}_1}}$$

STA 210

2️⃣ Calculate the test statistic

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$t = \frac{\hat{\beta}_1 - 0}{SE_{\hat{\beta}_1}}$$

$$t = \frac{0.5187 - 0}{0.0345} = \mathbf{15.03}$$

STA 210

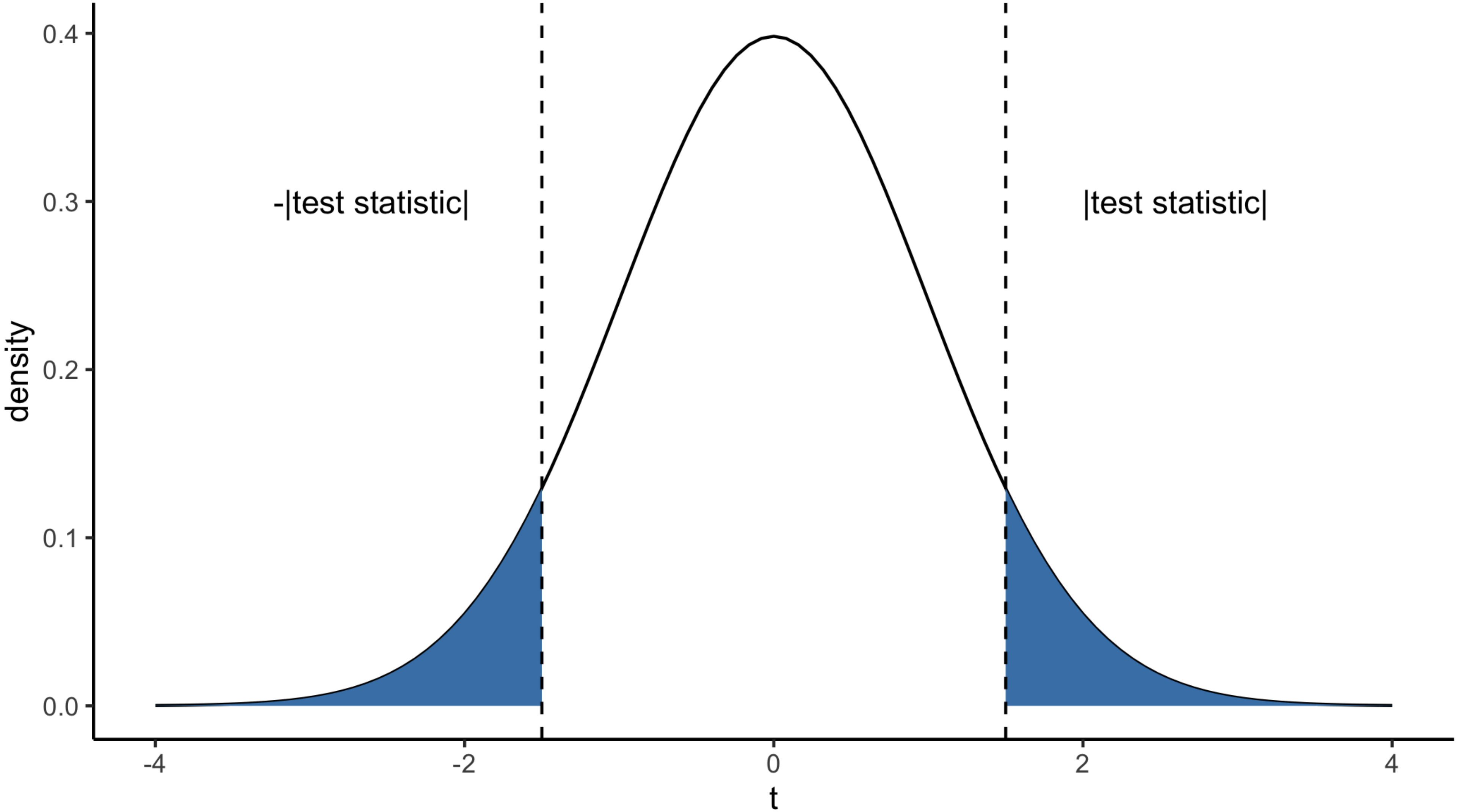
3️⃣ Calculate the p-value

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$\text{p-value} = P(|t| \geq |\text{test statistic}|)$$

Calculated from a t distribution with $n - 2$ degrees of freedom

3️⃣ Calculate the p-value



Understanding the p-value

Magnitude of p-value	Interpretation
p-value < 0.01	strong evidence against H_0
0.01 < p-value < 0.05	moderate evidence against H_0
0.05 < p-value < 0.1	weak evidence against H_0
p-value > 0.1	effectively no evidence against H_0

These are general guidelines. The strength of evidence depends on the context of the problem.

STA 210

4️⃣ State the conclusion

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

STA 210

4️⃣ State the conclusion

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

The data provide sufficient evidence that the population slope β_1 is different from 0.

There is a linear relationship between the critics score and audience score for movies on rottentomatoes.com.

What is a plausible range of values for the population slope β_1?

Confidence interval for β_1

$$\text{Estimate} \pm \ (\text{critical value}) \ \times \text{SE}$$

STA 210

Confidence interval for β_1

$$\text{Estimate} \pm \text{(critical value)} \times \text{SE}$$

$$\hat{\beta}_1 \pm t^* \times SE_{\hat{\beta}_1}$$

t^* is calculated from a t distribution with $n-2$ degrees of freedom

STA 210

Calculating the 95% CI for β_1

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$\hat{\beta}_1 = 0.519 \qquad t^* = 1.977 \qquad SE_{\hat{\beta}_1} = 0.035$$

STA 210

Calculating the 95% CI for β_1

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0

$$\hat{\beta}_1 = 0.519 \qquad t^* = 1.977 \qquad SE_{\hat{\beta}_1} = 0.035$$

$$0.519 \pm 1.977 \times 0.035$$

$$\mathbf{[0.450, 0.588]}$$

STA 210

Interpretation

$$[0.450, 0.588]$$

STA 210

Interpretation

$$[0.450, 0.588]$$

We are 95% confident that for every one point increase in the critics score, the audience score is predicted to increase on average between 0.450 and 0.588 points.

Recap

Recap

- Conducted a hypothesis test for β_1

Recap

- Conducted a hypothesis test for β_1

- Calculated a confidence interval for β_1