STAC67: Regression Analysis

Lecture 5

Thibault Randrianarisoa

September 23, 2026

University of Toronto Scarborough

Where we stopped last week

Everything so far used **only the first two moments** of the noise:

- $E(\epsilon_i) = 0$,
- $Var(\epsilon_i) = \sigma^2$,
- $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$.

We obtained the properties of least-square estimators:

- $\hat{\beta}_1 = \frac{S_{xY}}{S_{xx}} = \sum_{i=1}^{n} k_i Y_i$ with $k_i = \frac{x_i - \bar{x}}{S_{xx}}$ is linear in the Y_i,
- $E(\hat{\beta}_0) = \beta_0$, $E(\hat{\beta}_1) = \beta_1$: unbiased,
- $Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$ and $Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$
- **Gauss-Markov**: among all linear unbiased estimators, both have the smallest variance (BLUE)
- σ^2 is estimated without bias by $\hat{\sigma} = MSE = \frac{SSE}{n-2} = \frac{\sum_i e_i^2}{n-2}$, and

$$SE(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}, \quad SE(\hat{\beta}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

Inference about Regression Parameters

What we cannot do yet is make any probabilistic statement about the estimator.

Mean and variance do not say what is the probability $\hat{\beta}_j$ is to land within a given distance of β_j. Reminder: we can obtain **bounds** with Markov's inequality.

⚠ For that we need the **shape** of the distribution, which is what we assume today.

Inference about Regression Parameters

What we cannot do yet is make any probabilistic statement about the estimator. Mean and variance do not say what is the probability $\hat{\beta}_j$ is to land within a given distance of β_j. Reminder: we can obtain **bounds** with Markov's inequality.

⚠ For that we need the **shape** of the distribution, which is what we assume today.

So recall the simple Linear Model with **Normal Errors**

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

where

- β_0 and β_1 are parameters
- ϵ_i are i.i.d with normal distribution with mean 0 and variance, σ^2.

$$\epsilon_i \sim N(0, \sigma^2)$$

- Under this model,

$$Y_i | X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

Inference about Regression Parameters

- Inference about β_1

- σ is unknown, and so we can replace it with $\hat{\sigma}$. In this case, we have

$$T = \frac{\hat{\beta}_1 - \beta_1}{\left(\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right)} \sim t_{n-2}$$

- Review: the Student's t-distribution with k degrees of freedom can be defined as the random variable T

$$T = \frac{Z}{\sqrt{V/k}} \sim t_k,$$

Inference about Regression Parameters

- Inference about β_1

- σ is unknown, and so we can replace it with $\hat{\sigma}$. In this case, we have

$$T = \frac{\hat{\beta}_1 - \beta_1}{\left(\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right)} \sim t_{n-2}$$

- Review: the Student's t-distribution with k degrees of freedom can be defined as the random variable T

$$T = \frac{Z}{\sqrt{V/k}} \sim t_k,$$

where $Z \sim N(0,1), \quad V \sim \chi_k^2$, and Z and V are **independent**.

Inference about Regression Parameters

- Inference about β_1

- σ is unknown, and so we can replace it with $\hat{\sigma}$. In this case, we have

$$T = \frac{\hat{\beta}_1 - \beta_1}{\left(\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right)} \sim t_{n-2}$$

- Review: the Student's t-distribution with k degrees of freedom can be defined as the random variable T

$$T = \frac{Z}{\sqrt{V/k}} \sim t_k,$$

where $Z \sim N(0,1), \quad V \sim \chi^2_k,$ and Z and V are **independent**.

So to prove the student distribution we need three things: 1) a normal numerator, 2) a chi-square in the denominator, and 3) independence between them.

Sampling distribution of $\hat{\beta}_1$

How to prove $\dfrac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2}$?

Sampling distribution of $\hat{\beta}_1$

How to prove $\dfrac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2}$?

1. **Rewrite the ratio**, injecting σ^2 above and below:

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \underbrace{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/S_{xx}}}}_{\sim\mathcal{N}(0,1)} \Bigg/ \underbrace{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}_{\sqrt{\chi^2_{n-2}/(n-2)}\,?}$$

Sampling distribution of $\hat{\beta}_1$

How to prove $\dfrac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2}$?

1. **Rewrite the ratio**, injecting σ^2 above and below:

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \underbrace{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/S_{xx}}}}_{\sim\mathcal{N}(0,1)} \Bigg/ \underbrace{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}_{\sqrt{\chi^2_{n-2}/(n-2)}\,?}$$

2. We already know the numerator is standard normal

Sampling distribution of $\hat{\beta}_1$

How to prove $\dfrac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2}$?

1. **Rewrite the ratio**, injecting σ^2 above and below:

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \underbrace{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}}}_{\sim \mathcal{N}(0,1)} \Bigg/ \underbrace{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}_{\sqrt{\chi^2_{n-2}/(n-2)}\,?}$$

2. We already know the numerator is standard normal
3. **Admitted** The denominator is a chi-square over its degrees of freedom:

$$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{MSE}{\sigma^2} = \frac{V}{n-2}, \qquad \text{where } V = \frac{SSE}{\sigma^2} \sim \chi^2_{(n-2)}$$

Sampling distribution of $\hat{\beta}_1$

How to prove $\dfrac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2}$?

1. **Rewrite the ratio**, injecting σ^2 above and below:

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \underbrace{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2 / S_{xx}}}}_{\sim \mathcal{N}(0,1)} \Bigg/ \underbrace{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}}_{\sqrt{\chi^2_{n-2}/(n-2)}\,?}$$

2. We already know the numerator is standard normal
3. **Admitted** The denominator is a chi-square over its degrees of freedom:

$$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{MSE}{\sigma^2} = \frac{V}{n-2}, \qquad \text{where } V = \frac{SSE}{\sigma^2} \sim \chi^2_{(n-2)}$$

4. **Admitted** The two are independent. ($Cov(\hat{\beta}_1, e_i) = 0$, and under normality zero covariance gives independence.)

Inference about Regression Parameters

Theorem 3 (Sampling distributions)

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \sim t_{n-2}, \quad \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim t_{n-2}$$

Confidence Interval for β_0 and β_1:

Since $\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim t_{n-2},\ j = 0, 1$

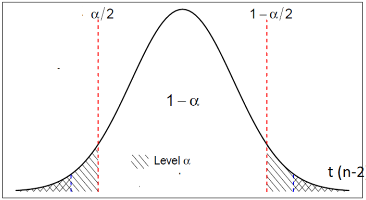
Theorem 3 (Sampling distributions)

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \sim t_{n-2}, \quad \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim t_{n-2}$$

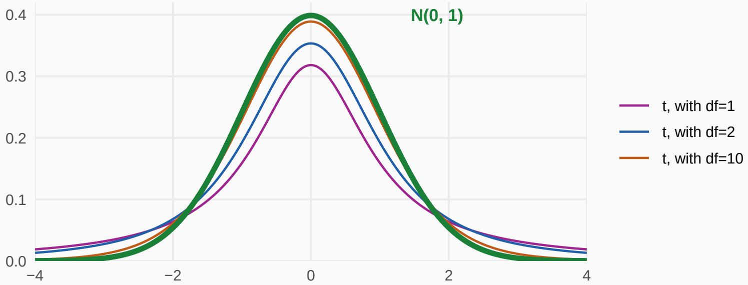
Confidence Interval for β_0 and β_1:

Since $\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim t_{n-2},\ j = 0, 1$

$$P\left(-t(1-\alpha/2; n-2) \leq \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \leq t(1-\alpha/2; n-2)\right) = 1 - \alpha$$



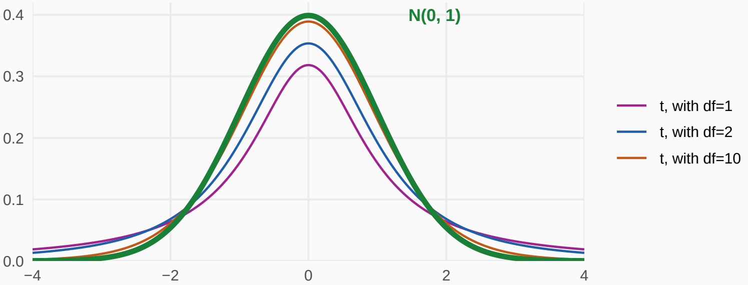
How far is t_{n-2} from the Gaussian distribution?



Heavier tails than the normal: larger quantiles are the price of estimating σ rather than knowing it.

The difference vanishes as the degrees of freedom grow.

How far is t_{n-2} from the Gaussian distribution?

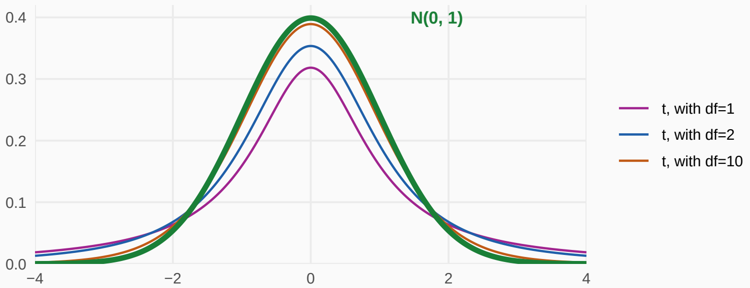


Heavier tails than the normal: larger quantiles are the price of estimating σ rather than knowing it.

The difference vanishes as the degrees of freedom grow.

ν	1	2	5	10	20	30	60	∞
$t(0.975;\ \nu)$	12.706	4.303	2.571	2.228	2.086	2.042	2.000	1.960

How far is t_{n-2} from the Gaussian distribution?



Heavier tails than the normal: larger quantiles are the price of estimating σ rather than knowing it.

The difference vanishes as the degrees of freedom grow.

ν	1	2	5	10	20	30	60	∞
$t(0.975;\ \nu)$	12.706	4.303	2.571	2.228	2.086	2.042	2.000	1.960

⚠ Be careful with small sample sizes as in this course! With $n = 24$, we use $t(0.975;\ 22) = 2.074$, and 1.96 would make the interval **6% too narrow**. This is why R always reports t intervals, whatever n is.

Confidence Intervals

$$\hat{\beta}_j \pm t(1-\alpha/2; n-2)SE(\hat{\beta}_j), \quad j = 0, 1$$

Why should we care about these confidence intervals?

Because they completely capture how much information in the data we have about the parameter.

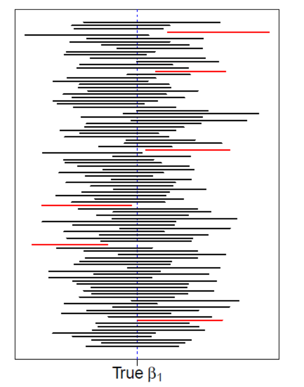
Confidence Intervals

$$\hat{\beta}_j \pm t(1-\alpha/2; n-2) SE(\hat{\beta}_j), \quad j = 0, 1$$

Why should we care about these confidence intervals?

Because they completely capture how much information in the data we have about the parameter.

The $1-\alpha$ confidence level, over repeated samples, the proportion of the intervals build this wat that contain β_j. The one you build from a particular dataset either does or does not.



Hypothesis test concerning β_1

1. Two-sided test: $H_0 : \beta_1 = \beta_1^0 \quad vs \quad H_a : \beta_1 \neq \beta_1^0$
2. Test statistics:

Hypothesis test concerning β_1

1. Two-sided test: $H_0 : \beta_1 = \beta_1^0 \quad vs \quad H_a : \beta_1 \neq \beta_1^0$
2. Test statistics:

$$t^* = \frac{\hat{\beta}_1 - \beta_1^0}{SE(\hat{\beta}_1)} \overset{H_0}{\sim} t_{n-2}$$

Hypothesis test concerning β_1

1. Two-sided test: $H_0 : \beta_1 = \beta_1^0 \quad vs \quad H_a : \beta_1 \neq \beta_1^0$

2. Test statistics:

$$t^* = \frac{\hat{\beta}_1 - \beta_1^0}{SE(\hat{\beta}_1)} \overset{H_0}{\sim} t_{n-2}$$

3. **Decision Rule**:

   - P-value approach: we can compute

     $$\text{p-value} = 2 \times P(t_{n-2} \geq |t^*|)$$

     If p-value $< \alpha$, **reject** H_0; otherwise **do not reject** H_0,

     where α is the level of significance, or probability of type I error, $\alpha =$ Pr(Reject H_0 | H_0 is true).

Hypothesis test concerning β_1

1. Two-sided test: $H_0 : \beta_1 = \beta_1^0 \quad vs \quad H_a : \beta_1 \neq \beta_1^0$

2. Test statistics:

$$t^* = \frac{\hat{\beta}_1 - \beta_1^0}{SE(\hat{\beta}_1)} \overset{H_0}{\sim} t_{n-2}$$

3. **Decision Rule**:

   - P-value approach: we can compute

     p-value $= 2 \times P(t_{n-2} \geq |t^*|)$

     If p-value $< \alpha$, **reject** H_0; otherwise **do not reject** H_0,

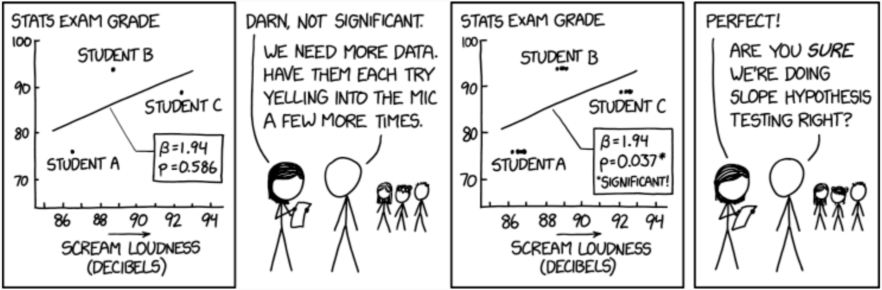
     where α is the level of significance, or probability of type I error, $\alpha =$ Pr(Reject H_0 | H_0 is true).

   - Critical value approach:

     If $|t^*| > t(1 - \alpha/2; n - 2)$, **reject** H_0.

     If $|t^*| \leq t(1 - \alpha/2; n - 2)$, **do not reject** H_0.

On the importance of assumptions



The estimate never moves: $\hat{\beta}_1 = 1.94$ in both panels. Only the p-value does, because the extra points are repeated measurements on the same three students and add more confidence in the estimates.

Exercise (Crime Rate)

Let's go back to the example from last week.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	20517.5999	3277.64269	6.259865	0.00e+00
X	-170.5752	41.57433	-4.102897	9.57e-05

- We want to test whether or not there is a linear association between crime rate and percentage of high school graduates, using a t test with $\alpha = 0.01$. State the hypotheses, decision rule, and conclusion.

Exercise (Crime Rate)

- Estimate β_1 with a 99 % confidence interval. Interpret the interval estimate.