We determined that (given some assumptions): $\hat{\beta}_1$ is distributed $\mathcal{N}\left(\beta_1, \frac{Var(u)}{\sum_i (x_i - \bar{x})^2}\right)$. $Var(u)$ is unknown since u_i is unobservable, so we have to approximate it using the regression residuals e_i. Because of this, we also use the t-distribution, which is similar to the Normal distribution, but with slightly fatter tails.

We define "standard errors" as our estimate of the standard deviation of the regression coefficient. The formula for a simple regression standard error of $\hat{\beta}_1$ is:

$$\sqrt{\frac{\sum_i (u_i^2)}{(n-2)\sum_i (x_i - \bar{x})^2}}$$

.

1. **We would like our standard errors to be as small as possible so we can** increase the precision of our estimates. If we increased the number of observations (assuming all else is held equal), will our standard errors increase or decrease?

They should decrease. You can see this as "n" is in the denominator of the equation above.

2. **All else held equal, if the sample variance of x_i decreased, should we** expect the standard errors would increase or decrease?

   For the sake of building intuition, we will use the following study:

   There are two studies designed to find the effect of a blood pressure medication on health outcomes.

   Study A: 9 people take the placebo, 1 person takes the medication, so $X = (0,0,0,0,0,0,0,0,0,1)$

   Study B: 5 people take the placebo, 5 people take the medication, so $X = (0,0,0,0,0,1,1,1,1,1)$.

   Calculate the sample variance of X in both studies (sample variance $= \frac{\sum_i (x_i - \bar{x})^2}{n-1}$). Which study will yield a more confident estimate of the effect, and which study has a lower $Var(X)$?

Var. Study A: $\bar{x} = \frac{1}{10}$ & $n = 10 \rightarrow \frac{1}{10-1} \cdot \underbrace{9(0-1/10)^2}_{\text{From 0's}} + \underbrace{1(1-1/10)^2}_{\text{From 1's}} = \frac{9(0.01) + 0.81}{9} = 0.10$

Var. Study B: $\bar{x} = \frac{1}{2}$ & $n = 10 \rightarrow \frac{1}{10-1} \cdot \underbrace{5(0-1/2)^2}_{\text{From 0's}} + \underbrace{5(1-1/2)^2}_{\text{From 1's}} = \frac{5(0.25)+5(0.25)}{9} \approx 0.278$

Study B has more variation in treatment assignment so it will be more precise $\rightarrow$ Smaller Standard Errors

Study A has lower sample variance

3. **Inference**

Say you run the following model and have the data below:

$$\text{Exam Score}_i = \beta_0 + \beta_1 \text{Hours Studied}_i + u_i$$

Student	X_i Hours Studied	Y_i Exam Score	$\hat{y}_i$	$\hat{u}_i$
1	2	65	65.72	-0.72
2	3	70	68.90	1.10
3	5	75	75.27	-0.27
4	6	78	78.45	-0.45
5	8	85	84.82	0.18

You run the regression and receive the estimates: $\hat{\beta}_0 = 59.35$ and $\hat{\beta}_1 = 3.184$

(a) Recall we can estimate the variance of the error term σ^2 using $\frac{\sum_i \hat{u}_i^2}{n-k}$ where n is the number of observations and k are the number of regression parameters. What is σ^2 for this model?

$\hat{u}_i$ given by $y_i - \hat{y}_i$ & $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ (Values above) ; n = # of students = 5

k = # of parameters = 2

$$s^2 = \frac{\sum_i \hat{u}_i^2}{n-k} = \frac{2.0362}{3} = 0.6787$$

(b) Calculate the Standard Error of $\hat{\beta}_1$

$$SE(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum(x_i - \bar{x})^2}} = \sqrt{\frac{0.6787}{22.8}} = 0.1725$$

(c) Find the 95% Confidence Interval for $\hat{\beta}_1$ using the t-value of 3.182

95% C.I. = $\hat{\beta}_1 \pm$ t-value $\cdot SE(\hat{\beta}_1)$

↳ $3.184 \pm 3.182 \cdot 0.1725 = 3.184 \pm 0.549$

↳ CI is $[2.64, 3.73]$

4. **If we decreased the variance of u_i by including more explanatory variables**, should we expect that the standard error on $\hat{\beta}_1$ will increase or decrease? Why?

By ↓ s^2, we would decrease the standard errors of $\hat{\beta}_1$