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?

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 $X = (0, 0, 0, 0, 0, 0, 0, 0, 1)$

   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 $Var(X)$?

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
1	2	65
2	3	70
3	5	75
4	6	78
5	8	85

   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 $\dfrac{\sum_i \hat{u}_i}{n-k}$ where n is the number of observations and k are the number of regression parameters. What is σ^2 for this model?

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

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

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?