1. Suppose we have a dataset that shows the number of bedrooms and the selling price for 20 houses in a particular neighborhood
   Bedrooms = [1,1,1,2,2,2,2,3,3,3,3,3,3,3,4,4,4,5,5,6]
   Prices = [120,133,139,185,148,160,192,205,244,213,236,280,275,273,312,311,304,415,396,488]
2. Using the Bedrooms data points:
   - Extract a random sample of size 10
   - Compute the mean and standard deviation of the sample
   - Compute the the mean of the population
   - Computer the confidence interval of 95% using the sample
   - Check if the population mean lies between the upper and lower bounds of the interval

$$CI = \bar{x} \pm z \cdot \frac{s}{\sqrt{n}}$$

Standard deviation

Mean value

Sample size

Lower/Upper limit

z-value for the confidence level

3. Using both of the Bedrooms and the Prices:
   - Compute the covariance and the correlation between the two variables
   - Build a regression model and estimate the regression parameters (slope and intercept).
   - Predict the house price for a house with 7 rooms
   - Estimate the upper bound and the lower bounds of the prediction interval of a house with 7 rooms
     - Alpha = 0.1
     - N = 20
     - Df = N-2

$$\boxed{\hat{y}} \pm t_{(1-\alpha/2,n-2)} \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}\left(1 + \frac{1}{n} + \frac{(\boxed{x} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2}\right)}$$

4. Using as sample of houses that have only 3 bedrooms
5. Try to find the lower price that can be payed for a house with 3 rooms using Tolerance interval
6. Try to find the highest price that can be payed for a house with 3 rooms using Tolerance interval

$$\bar{x} \pm \ k_1 s$$
$$\bar{x} \pm \ k_2 s$$

Some Extra stuff:

Try to use these Formulas as well and compare the results

We use the following formula to calculate a **confidence interval**:

$$\hat{y}_0 \pm t_{\alpha/2,n-2} * S_{yx}\sqrt{(x_0 - \bar{x})^2/SS_x + 1/n}$$

We use the following formula to calculate a **prediction interval**:

$$\hat{y}_0 \pm t_{\alpha/2,n-2} * S_{yx}\sqrt{(x_0 - \bar{x})^2/SS_x + 1/n + 1}$$

where:

- **$\hat{y}_0$**: Estimated mean value of response variable
- **$t_{\alpha/2,n-2}$**: t-critical value with n-2 degrees of freedom
- **S_{yx}**: Standard error of response variable
- **x_0**: specific value of predictor variable
- **$\bar{x}$**: mean value of predictor variable
- **SS_x**: Sum of squares for predictor variable
- **n**: Total sample size