As discussed lightly in the introduction, standard errors can be used to compute the **confidence intervals**. A 95% confidence interval is defined as a range of values such that with 95% probability, the range will contain the true unknown value of the parameter. The range is defined in terms of lower and upper limits computed from the sample of data.
A 95% confidence interval has the following property: if we take repeated samples and construct the confidence interval for each sample, 95% of the intervals will contain the true unknown value of the parameter. For linear regression, the 95% confidence interval for β_1 approximately takes the form:

$$\hat{\beta}_1 \pm 2 \cdot \text{SE}(\hat{\beta}_1)$$

There is approximately 95% chance that the interval

$$\left[\hat{\beta}_1 - 2 \cdot \text{SE}(\hat{\beta}_1), \hat{\beta}_1 + 2 \cdot \text{SE}(\hat{\beta}_1)\right]$$

Standard erros can also be used to perform **Hypothesis tests** on the coefficients. The most common hypothesis test involves testing the **null hypothesis** of

$$H_0 : \text{There is no relationship between } X \text{ and } Y$$

verus the **alternative hypothesis**

$$H_a : \text{There is some relationship between } X \text{ and } Y$$

Mathematically, our test is

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

since if $\beta_1 = 0$ then the model reduces to $Y = \beta_0 + \varepsilon$ and X is not associated with Y. To test the null hypothesis, we need to deternube whether β_1, our estimate for β_1, is sufficiently far from zero that we can be confident that β_1 is non-zero.
How far is enough? depends on the accuracy of $\hat{\beta}_1$ or depends on $\text{SE}(\hat{\beta}_1)$

1. For a small $\text{SE}(\hat{\beta}_1)$, it may provide strong evidence that $\beta_1 \neq 0$ and hence we know that there is a relationship between X and Y
2. For a large $\text{SE}(\hat{\beta}_1)$, $\hat{\beta}_1$ must be large in absolute value in order for us to reject the null hypothesis. We compute a **t-statistic** given by

$$t = \frac{\hat{\beta}_1 - 0}{\text{SE}(\hat{\beta}_1)}$$

t-statistic measure the number of standard deviations that $\hat{\beta}_1$ is away from 0. If there really is a no relationship between X and Y, then it is expect that their is a t-distribution with $n - 2$ degrees of freedom.

p-value is the probability of observing any number equal to $|t|$ or larger in absolute value, assuming $\beta_1 = 0$. This probability is called the p-value.

1. small p-value indicates that there is an association between the predictor and the response. We *reject the null hypothesis* and we can declare that a relationship exists between X and Y if the p-value is small enough.
2. large t-statistics are also large for large coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ are very large relative to their standard errors, si tge t-statistics are also large; the probability of seeing such values if H_0 is true are virtually zero. Hence $\beta_0 \neq 0$ and $\beta_1 \neq 0$

Assessing the Accuracy of the Model

After rejecting the null hypothesis in favour of alternative hypothesis. The next task is to quantify the *the extent to which the model fits the data.*
The quality of a linear regression fit is typically assessed using two related quantities:

1. Residual Standard Error (RSE)
2. the R^2 statistic

Residual Standard Error

RSE is an estimate of the standard deviation of ε. It is the average amount that the response will deviate from the true regression line.

$$\text{RSE} = \sqrt{\frac{1}{n-2}\text{RSS}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$$

$$\text{RSS} = \sum_{i=1}^{n}(y_i - \hat{y_i})^2$$

RSE is considered a measure of the *lack of fit* of the model to the data.

1. A small RSE indicates that the model fits the data very well or in other words $\hat{y_i}$ is very close to y_i for one or more observations
2. If $\hat{y_i}$ is very far from y_i for one or more observations, then the RSE may be quite large indicating that the model doesn't fit the data well.

R^2 Statistic

R^2 statistic provides an alternative measure of fit. It is the proportion of variance and takes a value between 0 and 1 and is independent of the scale of Y. It is a measureof the linear relationship between X and Y.

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

1. $\text{TSS} = \sum (y_i - \overline{y})^2$ is the total sum of squares and measures the total variance in the response Y. It can be considered as the amount of variability inherent in the response before the regression is performed.
2. RSS measures the amount of variability that is left unexplained after performing the regression.
3. TSS − RSS measures the amount of variability in the response that is explained by performing the regression.
4. R^2 measures the proportion of the variability in the response (Y) that is explained using (X)
   (a) R^2 statistic that is close to 1 indicates that a large proportion of the variability in the response that is explained by the regression
   (b) R^2 statistic that is close to 0 indicates that the regression does not explain much of the variability in the response that is explained by the regression

R^2 statistic has an interpretational advantage over the RSE since it is a proportion. The definition of *good* R^2 value depends upon the area of application, like in physics we're most likely to get get an extremely close value to 1 while it is a significant model if we get a value close to 0.1 in the fields of biology, psychology and marketing.
$r = \text{Cor}(X, Y)$ is also a measure of the linear relationship between X and Y.

$$\text{Cor}(X, Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

In a simple regression setting, R^2 statistic and the squared correlation r^2 are identical