**Slide 7**

\_\_Since we want to minimized RSS, we need to find the derivative of RSS based on Beta 0 and Beta 1.

(Introduce proof of Beta coefficient)

\_\_As the result, Beta 1 is equal to covariance divided by variance.

\_\_covariance is a measure of how much the two variables change in the same direction

**Population regression line**

\_\_The property of unbiasedness

\_\_If we estimate beta 0 and beta 1 on the basis of a particular data set, which this estimation is unbiased, this estimation won’t exactly equal to true Beta 0 and Beta 1.

\_\_But if we could average the estimates obtained over a huge number of data sets, then the average of these estimates would be spot on.

\_\_This is the same for sample mean.

\_\_We estimate a sample mean from a data set. This estimate is unbiased. But would no be exactly equal to the population mean

\_\_The difference between a population mean and sample mean is that population mean is the mean over a collection of observation in a population.

\_\_But sample mean is the mean of a portion of the whole collection of observation

\_\_Unfortunately, we are not able to collect data from the entire population, so we need to estimate the population mean from clues and evidence of the sample means.

**How accurate is the sample mean as an estimate of population mean be?**

\_\_To answer this question, we need to calculate the standard error of the sample mean.

\_\_First, what is standard error?

\_\_standard error is the approximate standard deviation of a sample population.

\_\_What is standard deviation?

\_\_standard deviation is the measurement of the variation between each data point relative to the mean.

\_\_further the data points are from the mean, higher the deviation within the data set.

\_\_What does standard error shows?

\_\_Standard error of sample mean tells us the average amount that this estimate mu differs from the actual value of mu.

\_\_Standard error also shows how deviation shrinks with n

\_\_the more observation we have, the smaller the standard error of mu hat.

\_\_In similar vein, below shows the standard errors associated with beta 0 hat and beta 1 hat.

\_\_sigma square is the variance of the error term.

\_\_And this formula requires an assumption of the error term, which for each observation are uncorrelated with common variance.

\_\_**Standard error** of the beta 1 hat represents the average distance that estimated slope differ from the regression line.

\_\_I will find proof of the standard error, but I can’t find it for now.

**Confidence Interval**

\_\_In general sigma square is unknown but can be estimated.

\_\_This estimation is called residual standard error

\_\_A confidence interval is defined as a range of values such that your determined percentage area of probability, will contain the true unknown value of the parameters.

\_\_For example, linear regression with the 95% confidence interval for Beta 1 approximately takes the form of the equation

\_\_The factor 2 might vary slightly based on the number of observations in the linear regression, which is related to the degree of freedom

\_\_I will introduce degree of freedom later in the slides.

**Hypothesis Testing**

\_\_Two type of hypothesis for testing the beta coefficient

\_\_Null Hypothesis, which means if do not reject null hypothesis, then there is no relationship between X and Y

\_\_Alternative Hypothesis: means there is some relation between X and Y.

\_\_In a mathematical way

\_\_Null hypothesis: is to test if Beta 1 is equal to 0, since Y = Beta 0 + error term, which makes X have no association with Y.

\_\_Alternative Hypothesis: Beta 1 is not equal to 0.

\_\_we need to determine whether beta 1 hat is sufficiently far from zero that we an be confident that true beta 1 is non-zero.

\_\_That depends on the standard error of Beta 1 hat.

\_\_If SE() is small, then small values of will provide strong evidence that .

\_\_If SE() is large, then must be large in absolute value in order to reject the null hypothesis

\_\_In practice, we need to compute t-statistic.

**Next Slide**

\_\_The formula measures the number of standard deviations that beta 1 hat is away from 0.

\_\_0 comes from the hypothesis testing

\_\_Now I will introduce t-statistic

**T-statistic**

\_\_T- statistic is a probability distribution that used to estimate population parameter

\_\_We use t-stat if the sample size is small or the population variance is unknown

\_\_also continuing the formula, If there really is no relationship between X and Y, then we expect the formula will have a t-distribution with n – 2 degrees of freedom.

**Degree of Freedom**

\_\_How is degree of freedom important?

\_\_It is important because there are variations of the t-distribution which is determined by degree of freedom

\_\_degree of freedom is the number of independent observations in a set of data

\_\_In linear situation, degree of freedom is the sample size minus the number of coefficient, which is 2.

**T distribution**

\_\_Now back to t distribution

\_\_T distribution has a bell shape, which is similar to normal distribution

\_\_Therefore, just simply compute the probability of observing any value equal to |t| or larger, assuming true beta 1 = 0.

\_\_this probability is called the p-value

\_\_Roughly speaking, in the absence of any real association between the predictor and the response

\_\_a small p-value indicates that it is unlikely to observe such a substantial association between the predictor and the response due to chance

\_\_Therefore,

**Next Slide**

\_\_if the p-value is small enough, we can infer that there is an association between the predictor and the response

\_\_Since we establish that there is a association between the predictor and the response, so we will reject the null hypothesis

\_\_Typically, the p-value cutoff for rejecting the null hypothesis are 5 or 1%

**Example**

\_\_The image shows a table for the result of the advertising data.

\_\_Intercept’s coefficient shows the estimate beta 0

\_\_TV’s coefficient shows the estimate beta 1.

\_\_Next it shows corresponding standard error, t-statistic, and p-value

\_\_You can notice the coefficients are very large relative to their standard errors. This cause the t-statistic are also large.

\_\_This cause the p-value or the probabilities of seeing = 0and = 0 is virtually zero.

\_\_This means the chance of Null hypothesis is true is virtually zero.

\_\_Then, we can conclude Beta 0 and Beta 1 is non-zero