**Basic information about linear regression**

\_\_Linear Regression is a simple approach for supervised learning.

\_\_It is a useful tool for predicting a quantitative response.

\_\_Widely used statistical learning method even it is somewhat dull compared to modern statistical learning method.

\_\_However, it serves more as a good jumping-off point for newer approaches.

\_\_In later chapter, many statistical learning methods can be seen as generalization or extension of linear regression.

**How to use Linear Regression?**

\_\_Using the advertising data from Chapter 2 as an example.

\_\_200 market observation on sales for a particular product as a function of advertising budgets for TV, radio, and newspaper.

\_\_The figure is for illustration, later, I will be focusing on sales vs TV.

\_\_The goal is to suggest a market plan that will result in high product sales on the basis of this data.

**Important Questions we seek to address**

\_\_*Is there a relationship between advertising budget and sales?*

\_\_Determine whether the data provide evidence of an association between advertising budget and sales.

\_\_*How strong is the relationship between advertising budget and sales?*

\_\_Assuming there is a relationship between advertising budget and sales.

\_\_How accurate can we predict the sales by given a certain advertising budget.

\_\_*Which media contribute to sales?*

\_\_There are three media---TV, radio, and newspaper.

\_\_Do all three contribute to the sales? Or just one or two contributed?

\_\_Must find a way to separate out the individual effects of each medium when we have spent money on all three media.

\_\_*How accurately can we estimate the effect of each medium on sales?*

\_\_For every dollar spent on advertising in a particular medium, by what amount will sales increase?

\_\_How accurate ca we predict this amount of increase?

\_\_*How accurately can we predict future sales?*

\_\_For any given level of TV, radio, or newspaper advertising, what is our prediction for sales, and what is the accuracy of this prediction?

\_\_*Is the relationship linear?*

\_\_If there is approximately a straight-line relationship between advertising budget in the various media and sales, then linear regression is an appropriate tool.

\_\_*Is there synergy among the advertising media?*

\_\_For example: spending $50,000 on TV advertising and $50,000 on radio results in more sales than spending $100,00 to either TV or radio individually.

\_\_In marketing, this is known as synergy effect.

\_\_In statistics, it is called interaction effect.

\_\_Will specific answer those question in Section 3.4

**3.1 Simple Linear regression**

\_\_Simple linear regression is a very straightforward approach for predicting a quantitative response Y on the basis of a single predictor variable X

\_\_It assumes that there is a linear relationship between X and Y

\_\_Regressing Y on X

\_\_In the advertising example, X may represent TV and Y may represent sales.

\_\_then this is regressing sales on TV.

\_\_In the equation, there (beta 0) and (Beta 1) are unknown constants

\_\_(Beta 0) represents the intercept

\_\_The expected value of Y when X = 0.

\_\_(Beta 1) represents the slope

\_\_The average increase in Y associated with a one-unit increase in X.

\_\_Simple Linear regression has two parameters, which are (Beta 0) and (Beta 1).

**Estimating the parameters**

\_\_In the advertising data, it consists the TV advertising budget and the product sales in 200 different market.

\_\_the goal is to use the training data to obtain Beta 0 and beta 1.

\_\_so that the linear model fits the data and the resulting line is as close as possible to the 200 data points

\_\_the common approach involves minimizing the least squares criterion.

\_\_Let be the prediction for Y based on the ith value of X.

\_\_ith residual is the difference between the observed value and the estimated observed value.

\_\_difference between error and residual is that error is comparing observed value to the true unknown value, but residual is comparing the observed value to the estimate value predicted by the linear model.

\_\_Defines residual sum of squares (RSS).

\_\_It is a technique used to measure the amount of variance in a data set that is not explained by the regression model

\_\_variance is the amount the function would change if changes happened to the data set.

\_\_The least squares approach chooses the parameters to minimize the RSS.

**Next slide**

\_\_Beta 1 shows the mathematic equation for calculating the slope.

\_\_And Beta 0 uses the result from estimating Beta 1 and inserted it into the linear model to find out Beta 0.

\_\_bar y and bar x are the sample means.

Figure

\_\_The figure uses the 200 market observations of the TV advertising budget vs the sales on a product.

\_\_The linear fit is found by minimizing the sum of squared errors.

\_\_Each grey line segment represents an error, and the fit makes a compromise by averaging their squares

\_\_In this case, the linear fit capture the relationship between TV advertising budget and sales.

\_\_However, linear model is a little bit deficient in the left of the plot.

**Population regression line**

\_\_Assumes that the true relationship between X and Y takes the form of Y = f(x) + epsilon for some unknown function f and epsilon is a mean-zero random error term.

\_\_If function f is approximated by a linear function, then the equation is Y = Beta 0 + Beta 1 times X + epsilon.

\_\_The error term serves as a catch-all for what we miss with this simple model.

\_\_true relationship is probably not linear

\_\_There might be other variable that cause variation in Y.

\_\_Typically assumes that error term is independent of X.

\_\_Example: The graph shows 100 random X with corresponding Y. which are the black circles

\_\_The red line represents the unknown true relationship. For illustration, the linear relationship of the red line is 2+3X. It is also known as population regression line.

\_\_The blue line is the least squares line. The line is the estimation of f(x) based on the observed data.

\_\_We only have one data set, what does it means that two different lines describe the relationship between predictor and the response.

\_\_This is the concept of these two lines is a statistical approach of using information from a sample to estimate characteristics of a large population.

\_\_ Suppose we want to find the population mean (mu) of some random variable Y.

\_\_mu is unknown, but we have access to the n observations from Y.

\_\_A reasonable guess is that estimated mu is equal to the sample mean.

\_\_However, sample mean and population mean is different, but sample mean can provide a good estimate of population mean.

\_\_In a similar way, the unknown parameters Beta 0 and Beta 1 define the population regression line.

\_\_We need to estimate theses unknown parameter using the estimated Beta 0 and Beta 1.

\_\_Suppose that we only have estimated one sample mean and try to estimate the population mean.

\_\_This estimate is unbiased, because we expect the sample mean equal to population mean.

\_\_But only using one data set could result in overestimate or underestimate the population mean.

\_\_But if we could average out a huge number of the sample means from a huge number of the data sets.

\_\_then the average would be super close to population mean.

\_\_This the concept taught before called the Weak law of Large Numbers.

\_\_If n is very large, then the sample mean is very close to the population mean.

\_\_This also applies to estimating parameters. Average of a large number of estimates would result in very close or exactly equal to population regression line.

\_\_In the right, the population regression line is shown in red, and least squares line in blue.

\_\_It also generated 10 more different data set and plotted 10 corresponding least squares lines.

\_\_The different sets generated from the same true model result in slightly different least squares line.

\_\_But population line does not change.

\_\_According to the book, the average of the least square lines would be pretty close to the population regression line.

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

\_\_How accurate is the sample mean as an estimate of population mean?

\_\_We can answer this by computing the standard error of the sample mean.

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

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

\_\_The data points are further from the mean, means higher deviation within the data set.

\_\_Standard error also shows how the deviation shrinks when n is getting bigger.

\_\_In a similar way, we wonder how close (Beta 0 and Beta 1) hat are to the true values Beta 0 and Beta 1.

\_\_We need to compute standard error associated with (Beta 0 and Beta 1) hat.

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

\_\_For this formula to work, we need to assume the error term for each observation are uncorrelated with common variance sigma squares.

**Confidence interval**

\_\_in general, sigma squares is unknown, but can be estimated as residual standard error.

\_\_example, 95% confidence interval for :

\_\_The factor 2 may vary slightly based on the number of observations in the linear regression

\_\_Later introduce in the chapter.