3.3.2: The Simple Linear Regression Model

In the scatterplot example shown above, we saw linear correlation between the two dependent variables. We are now going to create a statistical model relating these two variables, but let's start by reviewing a **mathematical linear model** from algebra:

$Y = \beta_0 + \beta_1 X$

Y: Dependent Variable

X: Independent Variable

β_0: Y - intercept

β_1:Slope

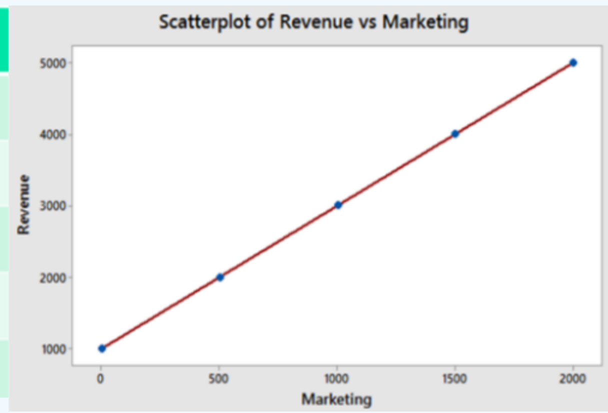
Example

You have a small business producing custom t-shirts. Without marketing, your business has revenue (sales) of $1000 per week. Every dollar you spend marketing will increase revenue by 2 dollars. Let variable X represent the amount spent on marketing and let variable Y represent revenue per week. Write a **mathematical model** that relates X to Y.

Solution

In this example, we are saying that weekly revenue (Y) depends on marketing expense (X). $1000 of weekly revenue represents the vertical intercept, and $2 of weekly revenue per $1 marketing represents the slope, or rate of change of the model. We can choose some value of X and determine Y and then plot the points on a scatterplot to see this linear relationship.

X=marketing	Y=revenue
$0	$1000
$500	$2000
$1000	$3000
$1500	$4000
$2000	$5000

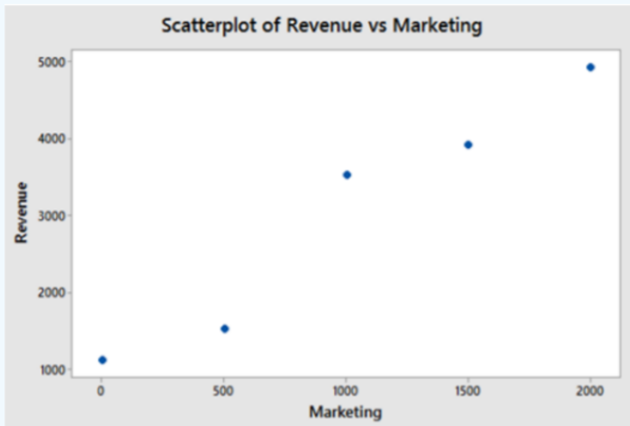


We can then write out the mathematical linear model as an equation:

Linear Model	Example
$Y = \beta_0 + \beta_1 X$	$Y = 1000 + 2X$
Y: Dependent Variable	Y: Revenue
X: Independent Variable	X: Marketing
β_0: Y – intercept	β_0: $1000
β_1: Slope	β_1: $2 per $1 marketing

We all learned about these linear models in Algebra classes, but the real world doesn't generally give such perfect results. In particular, we can choose what to spend on marketing, but the actual revenue will have more uncertainty. For example, the true revenue may look more like this:

X=Marketing	Expected Revenue	Y=Actual Revenue	ε=Residual Error
$0	$1000	$1100	+$100
$500	$2000	$1500	-$500
$1000	$3000	$3500	+$500
$1500	$4000	$3900	-$100
$2000	$5000	$4900	-$100



The difference between the actual revenue and the expected revenue is called the **residual error**, ε If we assume that the residual error (represented by ε) is a random variable that follows a normal distribution with $\mu=0$ and σ a constant for all values of X, we have now created a **statistical model** called a **simple linear regression model**.