Essential Math for Data Analysis Using Excel Online

Module 2, Lab 2: Skew

Learning Objectives
- Find the skew value of a data set.

Description
Learners will examine the coffee consumption data from the last lab and assess the degree to which variables are normally distributed or skewed. They will focus on cups of coffee consumed and temperature. They will use the =SKEW() function in Excel to estimate the exact skewness of the variable.
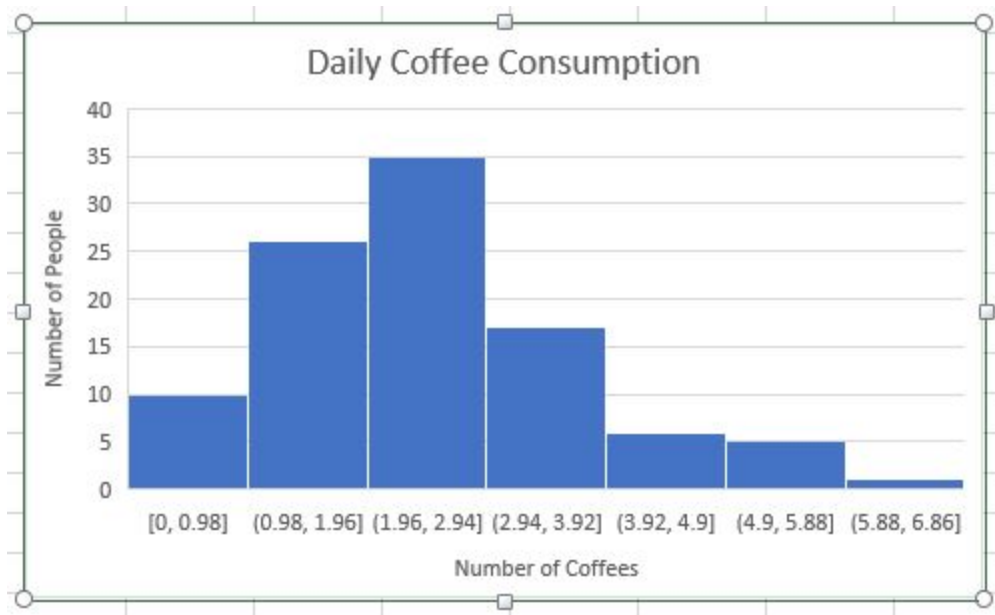
Data set
Mod2Lab.csv

Overview
"Skewness" is all about how asymmetrical the distribution of a set of data is. In this lab, we'll find out exactly how skewed the coffee variables from the last lab are.

What You'll Need
To complete the lab, you will need the online version of Microsoft Excel.

Exercise 1: Coffee Consumption Skew
First, you'll look at the coffee consumption data again. You might remember from the histogram in the previous lab that the coffee consumption variable didn't have that perfect bell shape of a normal distribution. Here's the histogram again.

Daily Coffee Consumption

The "hump" in the data isn't dead-center like it would be in a normal distribution. It's listing a bit. But how skewed is it, exactly?

1. Open up the data set in Excel. There should be 100 different rows, with column headings for various coffee preferences.

2. Use Excel to find the measure of skew in the data. There are several different mathematical formulas for calculating skew, each more horrifying than the last. Luckily, there's an Excel formula we can use: the aptly-named SKEW function. Create a new column on the spreadsheet for the skew.

	A	B	C	D	E	F	G	H
1	id	coffee	preference	black	temp	milk	additions	skew of coffee
2	1	5	Latte	No	182	Yes	Sometimes	
3	2	0	Drip	No	160	Yes	Always	
4	3	1	Latte	No	194	Yes	Never	
5	4	2	Drip	No	169	No	Sometimes	
6	5	1	Espresso	No	168	Yes	Sometimes	
7	6	2	Espresso	No	161	No	Always	
8	7	2	Latte	No	181	Yes	Sometimes	
9	8	3	Drip	No	165	Yes	Sometimes	
10	9	4	Drip	No	177	Yes	Sometimes	
11	10	2	Drip	Yes	177	Yes	Sometimes	
12	11	3	Espresso	No	166	No	Sometimes	

3. In cell H2, enter the SKEW function. The syntax is =SKEW(first cell:last cell). You want the skewness measure of the coffee consumption variable, which is in column B, so your first cell is B2 and your last cell is way down at B101.

	A	B	C	D	E	F	G	H
				f_x =SKEW(B2:B101)				
1	id	coffee	preference	black	temp	milk	additions	skew of coffee
2	1	5	Latte	No	182	Yes	Sometimes	0.689944756
3	2	0	Drip	No	160	Yes	Always	
4	3	1	Latte	No	194	Yes	Never	
5	4	2	Drip	No	169	No	Sometimes	
6	5	1	Espresso	No	168	Yes	Sometimes	
7	6	2	Espresso	No	161	No	Always	
8	7	2	Latte	No	181	Yes	Sometimes	
9	8	3	Drip	No	165	Yes	Sometimes	
10	9	4	Drip	No	177	Yes	Sometimes	
11	10	2	Drip	Yes	177	Yes	Sometimes	

4. The skew measure is about 0.6899, but what does that even mean? A skewness of 0 (zero) would mean the data are normally distributed. A positive value like 0.6899 means the data are *positively skewed*, which means the longer "tail" of the data in the histogram is on the positive side (to the right). That's why the histogram was taller and bulkier on the left side, with a long "tail" on the right side.

    Remember that there is only one 'skew' value to summarize the entire variable. Thus, we only find one skew value.

Exercise 2: Coffee Temperature Skew

This time, you'll find the skewness of the coffee temperature data from the data set. As a reminder, here's what the coffee temperature histogram looked like from the last lab:
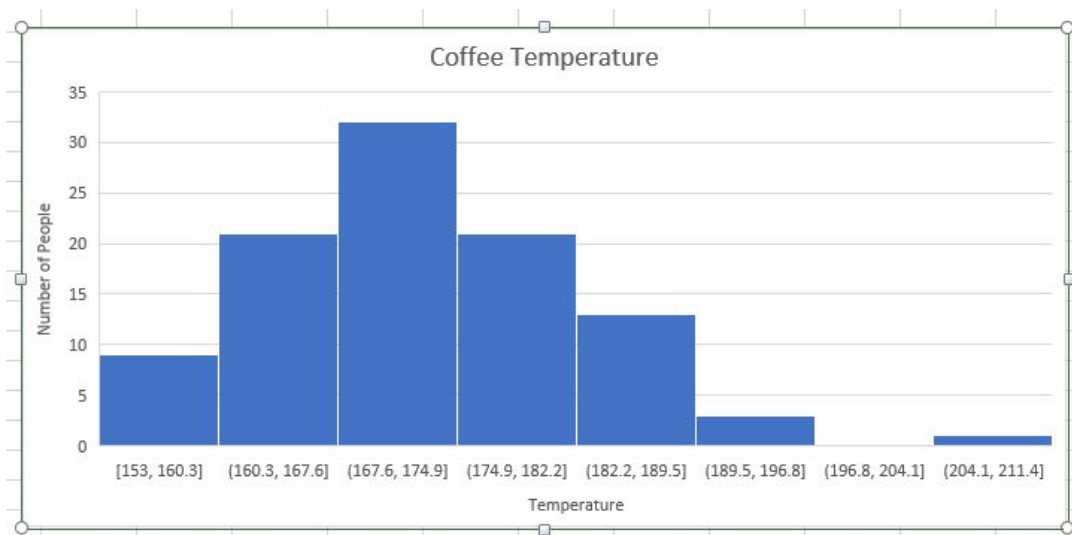
These data also have a little "tail" on the right-hand side, so we can probably expect to see a positive skew value again. Better make sure, though.

1. Use Excel to find the measure of skew in the data. This time, you'll use the values from the "temp" column. Create another new column for the skewness of the temperature.

	skew of temp								
	A	B	C	D	E	F	G	H	I
1	id	coffee	preference	black	temp	milk	additions	skew of coffee	skew of temp
2	1	5	Latte	No	182	Yes	Sometimes	0.689944756	
3	2	0	Drip	No	160	Yes	Always		
4	3	1	Latte	No	194	Yes	Never		
5	4	2	Drip	No	169	No	Sometimes		
6	5	1	Espresso	No	168	Yes	Sometimes		
7	6	2	Espresso	No	161	No	Always		
8	7	2	Latte	No	181	Yes	Sometimes		
9	8	3	Drip	No	165	Yes	Sometimes		
10	9	4	Drip	No	177	Yes	Sometimes		
11	10	2	Drip	Yes	177	Yes	Sometimes		
12	11	3	Espresso	No	166	No	Sometimes		

2. Now use the SKEW function again. This time, use the values from column E.

	=SKEW(E2:E101)								
	A	B	C	D	E	F	G	H	I
1	id	coffee	preference	black	temp	milk	additions	skew of coffee	skew of temp
2	1	5	Latte	No	182	Yes	Sometimes	0.689944756	0.450562957
3	2	0	Drip	No	160	Yes	Always		
4	3	1	Latte	No	194	Yes	Never		
5	4	2	Drip	No	169	No	Sometimes		
6	5	1	Espresso	No	168	Yes	Sometimes		
7	6	2	Espresso	No	161	No	Always		
8	7	2	Latte	No	181	Yes	Sometimes		
9	8	3	Drip	No	165	Yes	Sometimes		
10	9	4	Drip	No	177	Yes	Sometimes		
11	10	2	Drip	Yes	177	Yes	Sometimes		

3. The skew measure this time is about 0.4506. Sure enough, that's still a positive value, which matches the positive "tail" on the histogram.