Essential Math for Data Analysis Using Excel Online

Module 5, Lab 2: Forecasting

Learning Objectives

- Visualize time-series data in Excel.

Description

Learners will be given a time-series data set (output from a company and time points) and will visualize the increasing output over time. Users who have Excel 2016 can also use the built-in time-series forecasting tool; everyone can visualize and extrapolate from the historical trend.

Data set

Mod5Lab2.csv

Overview

One of the most powerful tools available to data analysts in Excel is the built-in time-series forecasting feature (though it's only available in the Windows desktop version). In this lab, you'll examine a company's time-series data set showing output values at regular daily intervals, and then predict how those output values will look 100 days *after* the historical trend in the given data set.

What You'll Need

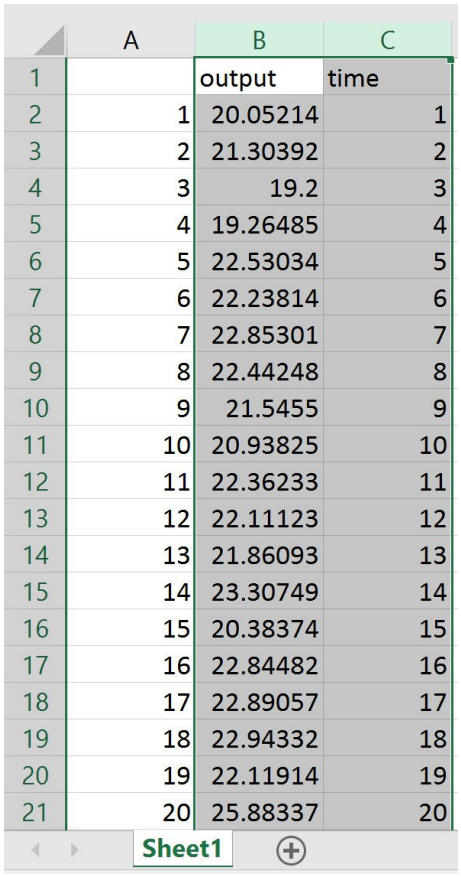
To complete the lab, you will need access to either Excel Online or the Windows version of Excel 2016. The way you'll participate in this lab will depend on the version you have.

- Note: The forecasting tools presented in this lab are *only* available in the desktop version of Excel 2016 for Windows. They aren't available in Excel Online or Excel for Mac. You can still participate in the lab if you don't have access to the Windows version of Excel, though — see below.
- **If you only have access to Excel Online**, then this will be a hands-off lab: You'll read through the exercise, see how we came up with the answer, and make a note of how forecasting works in the desktop version of Excel 2016 for Windows. You will be able to answer the "Lab Check" question from this document.
- **If you have access to Excel 2016 for Windows**, then this will be a hands-on lab: You'll walk through each step and perform the calculations like you would in a regular lab.

Exercise 1: Tomorrow's Forecast

1. Open the data set in Excel. There should be 100 different data points, with columns for ID, output, and the time point associated with each output (in days).

2. Select everything in column B and column C by holding Shift and clicking on the B and C at the top of both columns. (Again, if you only have access to Excel Online, you can just follow along by reading these steps.)

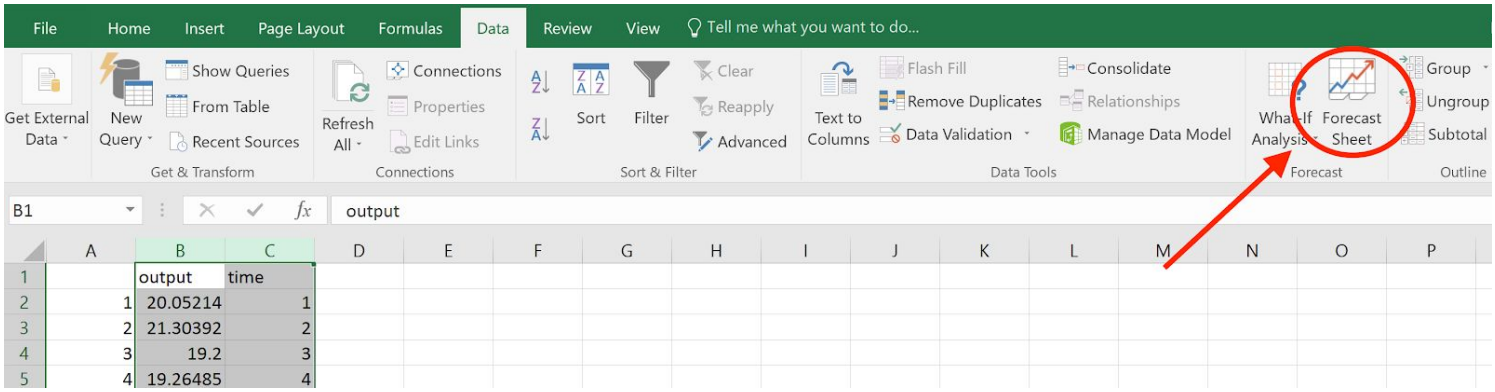


	A	B	C
1		output	time
2	1	20.05214	1
3	2	21.30392	2
4	3	19.2	3
5	4	19.26485	4
6	5	22.53034	5
7	6	22.23814	6
8	7	22.85301	7
9	8	22.44248	8
10	9	21.5455	9
11	10	20.93825	10
12	11	22.36233	11
13	12	22.11123	12
14	13	21.86093	13
15	14	23.30749	14
16	15	20.38374	15
17	16	22.84482	16
18	17	22.89057	17
19	18	22.94332	18
20	19	22.11914	19
21	20	25.88337	20

Sheet1

Make sure that *all* the data in both columns are selected (i.e. all 100 data points, not just the first 20 shown in the screenshot.)

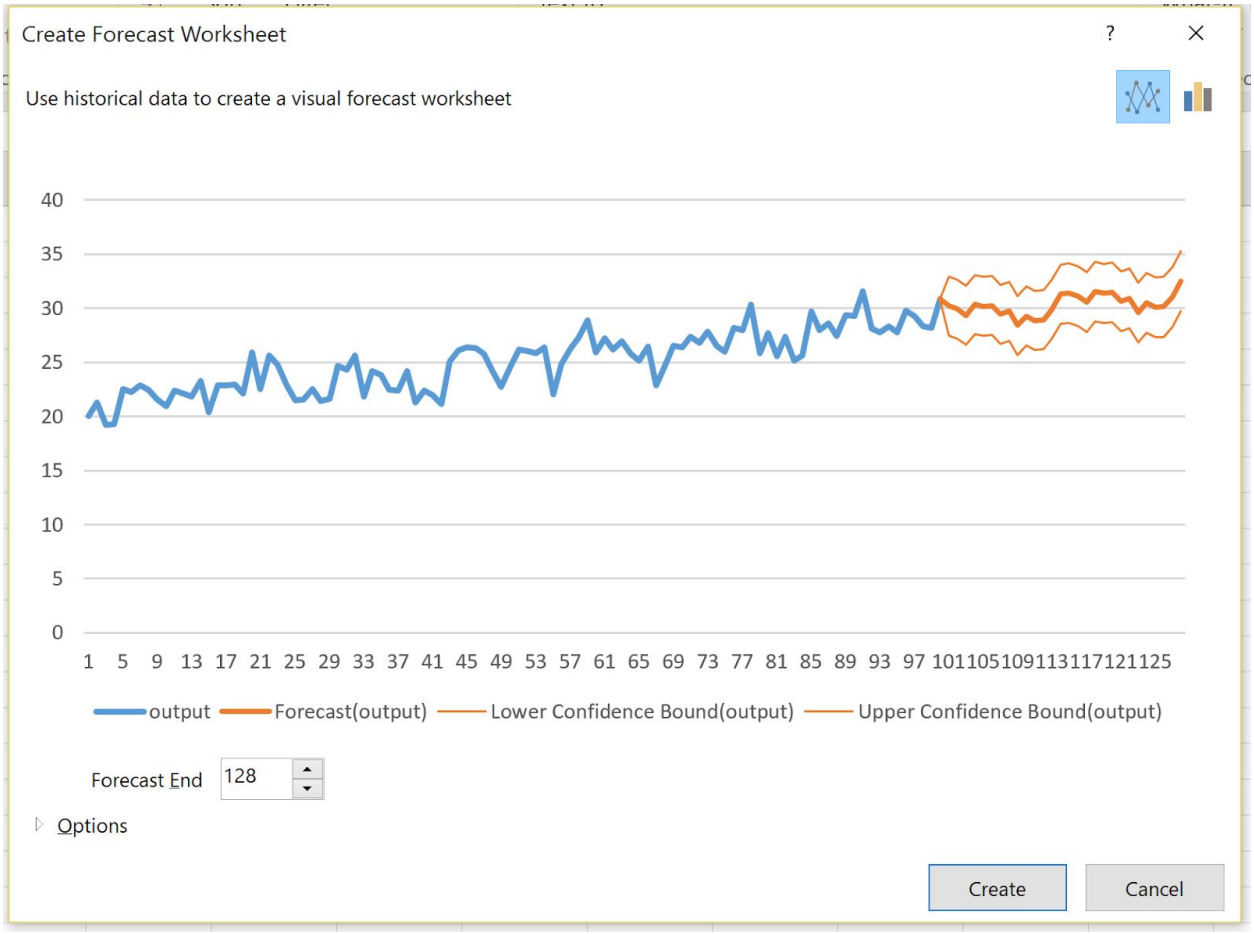
3. With both columns of data selected, go to the ribbon and click the Data tab, then select the Forecast Sheet option on the right-hand side. It looks like this:



Note: The math behind time-series forecasting is pretty advanced, so we won't go into the details of what the algorithm is doing mathematically. What you *do* need to know is that Excel 2016 for Windows has this powerful black-box tool that will do the work for you. It's an extremely useful tool for data analysis, so even if you're just following along with Excel Online, make a note of this "Forecast Sheet" tool for future reference.

For those of you with access to the Windows version, click the Forecast Sheet button.

4. The Create Forecast Worksheet will pop up.

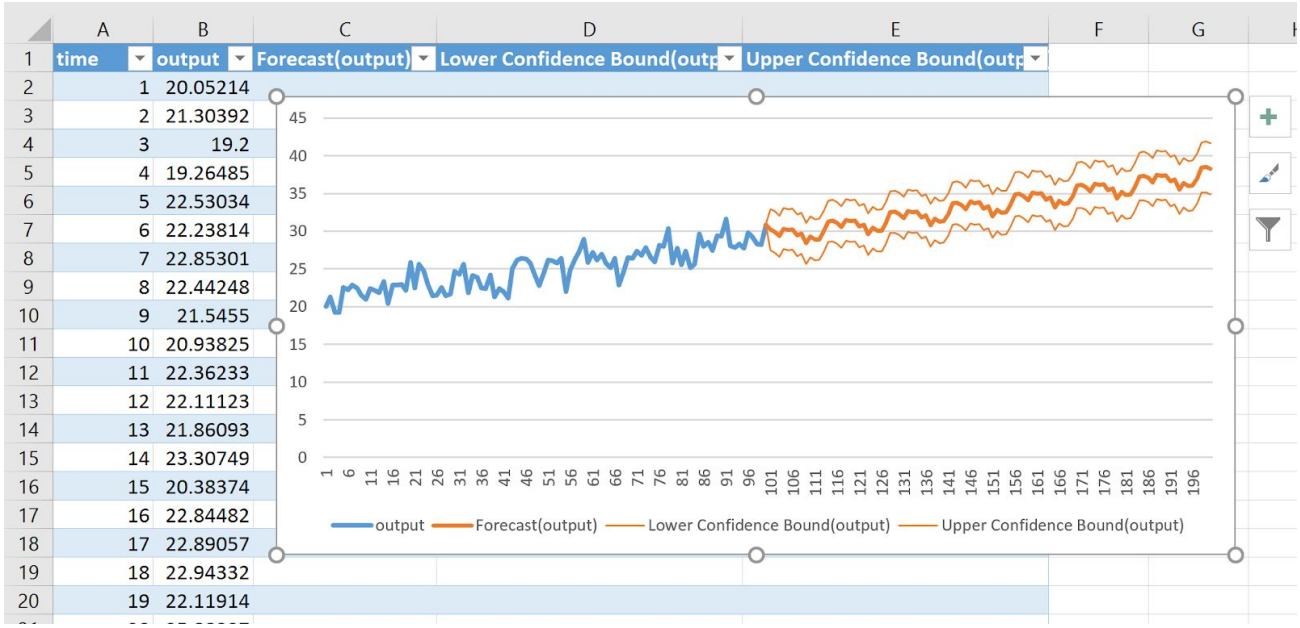


The blue graph shows the actual historical data set, while the orange shows the forecasted future values of the output. Notice that there's a band of uncertainty between the Lower Confidence Bound and the Upper Confidence Bound.

The Forecast End box at the bottom allows you to increase the future values up to 200 days. The default is currently set to forecast out to 128 days (though your own program's default might be slightly different).

Go ahead and click into the Forecast End box and type in 200, then click Create at the bottom.

5. Once you hit Create, Excel will create a new worksheet with the forecast graph. It'll also copy the data into an Excel table, which will allow you to adjust or add values on the fly.



time	output	Forecast(output)	Lower Confidence Bound(outp	Upper Confidence Bound(outp
1	20.05214			
2	21.30392			
3	19.2			
4	19.26485			
5	22.53034			
6	22.23814			
7	22.85301			
8	22.44248			
9	21.5455			
10	20.93825			
11	22.36233			
12	22.11123			
13	21.86093			
14	23.30749			
15	20.38374			
16	22.84482			
17	22.89057			
18	22.94332			
19	22.11914			

Now you have a handy visualization of your data's predicted values in the future, all the way out to Day 200. The forecast graph even includes dips and spikes to show the predicted seasonal trends and random noise.