

Assignment 6. Your Project or Mine

© 2006 Samuel L. Baker

If you have data from your work place or for a course project that you would like to analyze, you may do so for this week. Contact me so we can discuss how to apply this course's ideas to your data.

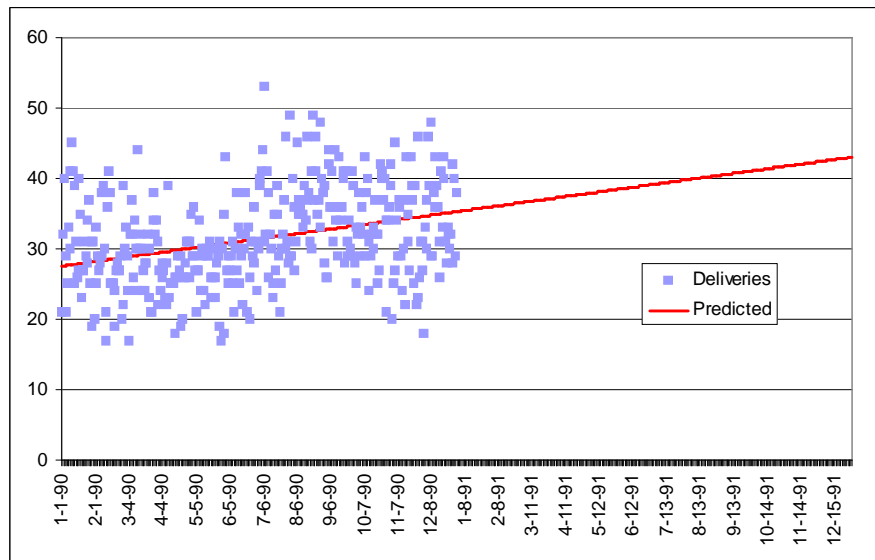
If you do not have a project for your own, I have one for you. We'll go back in time, so we can use real data.

Your job is to project Medicaid-paid deliveries of babies for 1991, based on the 1990 daily data you have gotten from the Medicaid Management Information System. The data are in the file ASSIGN6.XLS. You can get this file by clicking the link on the syllabus for Data for Assignment 6.

Another staffer fitted a simple linear equation to these data, using Time as the independent variable. The diagram to the right shows what he or she did.

The staffer first calculated the least squares simple regression line through the data for 1990. That line showed a pretty substantial increase in births per day over the year, on average. The staffer extended that line through all the days of 1991.

The staffer then added up the projected 1991 daily deliveries. The resulting prediction for the year 1991 was 14,300 Medicaid-paid deliveries. This is a substantial increase from the 11,477 deliveries in 1990. If your agency uses that prediction, it'll have to budget for about a \$9 million additional expenditure. The State Budget and Control Board will require your agency to cut services somewhere else to compensate.



You suspect that much of the increase during 1990 was actually due to a one-time increase in the Medicaid eligibility income ceiling, to 185% of the poverty level, that took effect July 1, 1990. You do not expect the eligibility ceiling to increase again in 1991, because 185% of poverty is the most that Federal law allows.

Your assignment is to make your own projection for 1991. Create a model that takes into account the mid-year eligibility extension. Also, take advantage of the daily data to find out if certain days of the week tend to have more or fewer deliveries than others. This will give you some insight into how much discretion doctors have in managing Medicaid deliveries. (January 1, 1990, was a Monday.)

Your deliverable for this assignment is a one-page memo and a graph.

The one-page memo will present and explain your prediction for 1991. Pretend you are writing for a boss

who is smart but has not had a course in multiple regression. You must therefore emphasize concepts and not details of method.

Your memo will also do some analysis of the daily fluctuations in Medicaid-paid births.

Details for doing the assignment

The data file has a column for the date and a column for the number of deliveries. You have to fill in the time trend variable, an eligibility shift variable, and the day-of-the-week dummy variables.

You may want to open the Answer Checker now, to check what you are doing as you go along.
<http://hspm.sph.sc.edu/Courses/J716/a06/Births.html>

Filling in the Data. Here is a method to speed up filling in the time and dummy variables.

Download (from the syllabus) and open the assign6.xls file. Follow the instructions below to make the first nine rows look like this:

	A	B	C	D	E	F	G	H	I	J	K
1		Deliv	Time	Eligib	Mon	Tue	Wed	Thu	Fri	Sat	Sun
2	1-1-90	21	1	0	1	0	0	0	0	0	0
3	1-2-90	32	2	0	0	1	0	0	0	0	0
4	1-3-90	40	3	0	0	0	1	0	0	0	0
5	1-4-90	21	4	0	0	0	0	1	0	0	0
6	1-5-90	29	5	0	0	0	0	0	1	0	0
7	1-6-90	25	6	0	0	0	0	0	0	1	0
8	1-7-90	33	7	0	0	0	0	0	0	0	1
9	1-8-90	41	=C8+1	=D8	=E2	=F2	=G2	=H2	=I2	=J2	=K2

(When you type those 9th row formulas, you will see numbers rather than the formulas.)

First fill in the cells in row 1. These are the variables names.

Then type 1 through 7 in C2 to C8. This is the start of the time variable.

Put a 0 in D2. Copy and paste that cell to the D2 to K8 block. That will give you a rectangular array of 0's.

Type 1's in the cells on the diagonal (E2, F3, G4, H5, I6, J7, K8).

Type these formulas in the 9th row:

C9 is =C8+1, the cell above it plus 1, because time increases by 1 each day.

D9 is =D8, so its value is equal to that of the cell above it.

Cells E9 through K9 are set to equal the cells seven rows above, because the day-of-the-week dummies repeat on a seven-day cycle.

Row 9, cells C9 through K9, should show 8 0 1 0 0 0 0 0 0. If you would like to see the formulas as in the picture above, go to Excel's menu and select Tools | Options | View and check the box to show the

formulas. Before you continue with the assignment, change this option it back so the numbers show.

Now we can fill in the rest of the rows of the spreadsheet.

Highlight C9:K9 and use the copy command to copy this to the clipboard. Then highlight the cells in column C from C9 all the way down to C366, at the bottom of the data. Paste from the clipboard. The big block C2:K366 will all fill in.

The Eligibility Shift dummy needs to be fixed. Right now it's all 0's. Move the cell selector to D183. That's in the row for 7-1-90.

Type 1 in cell D183. When you press or move the cell selector off of that cell, Excel will recalculate the D column, changing everything below D183 to 1's. Neat!

Save the file to your storage device. Then select all the data for copying and pasting to LS.

Use LS to do your regressions, with Deliveries as your dependent variable. When choosing your independent variables, first do the regression with all the variables in except for one of the day dummies. I suggest that you choose Monday to leave out of the regression. However, you can leave out any day you want. The prediction will be the same regardless. Inspect the results from that regression. Then run a second regression that includes only the independent variables that had significant coefficients in the first regression.

Interpretation

Suppose you get the following results. (These are made up for illustration purposes. Your results will be different. Different significant variables and different numbers.)

Dependent Variable: Deliv
Observations: 365

Variable	Coefficient
Time	.01234567
Eligibil	4.4343434
Tue	-3.2121212
Thu	1.2323232
Intercept	45.454545

The Time coefficient means that Medicaid-paid births per day rose at an average rate of .0123 births per day, separate from the effect of the eligibility shift.

The Eligibility Shift coefficient means that the eligibility change caused a 4.434 increase in the number of Medicaid-paid births per day. (It's OK to say "cause" here because of the situation.)

The daily dummies are interpreted by comparing with the days left out of the equation. Tuesdays averaged 3.21 fewer births than the average of Mondays, Wednesdays, Fridays, Saturdays, and Sundays. Thursdays averaged 1.232 more births than the average of Mondays, Wednesdays, Fridays, Saturdays, and Sundays.

Predictions. Please don't do 365 predictions in LS, one for each day of 1991. Instead, put the coefficients from your LS results into a spreadsheet formula. Specifically:

Go back to your spreadsheet. Move the cell selector to C366.

Highlight across the bottom row, cells C366 through K366.

Copy (such as with **Ctrl**+c) these cells to the clipboard.

Now highlight the C column below where our data are, from C367 all the way down to 731. We want to add to add another whole year of rows of independent variables, one row for each of 365 days. (731=366+365.)

Paste (such as with **Ctrl**+v) from the clipboard, and you have filled in all the independent what you need to make predictions for every day in 1991.

Go to cell B367.

Where to type in the regression equation formula.

Enter your regression equation formula in that cell. For example, based on the regression results I made up above, I would type this in B367:

$$= .01234567 * C367 + 4.4343434 - 3.2121212 * F367 + 1.2323232 * H367 + 45.454545.$$

	A	B	C	D
359	12-24-90	33	358	1
360	12-25-90	30	359	1
361	12-26-90	32	360	1
362	12-27-90	28	361	1
363	12-28-90	42	362	1
364	12-29-90	40	363	1
365	12-30-90	29	364	1
366	12-31-90	38	365	1
367			366	1
368			367	1

Your formula will be different. It will have different coefficients and different column letters than F and H.

You have probably noticed that, in the formula above, I just added Eligibility Shift dummy's coefficient, rather than putting in $+ 4.4343434 * D367$. This is OK because D367 has a 1 in it, as do all of the D column cells from D367 down to D731. I would have been just multiplying by 1.

When the B367 formula is complete, cell B367 will have a prediction, based on your regression results, for January 1, 1991.

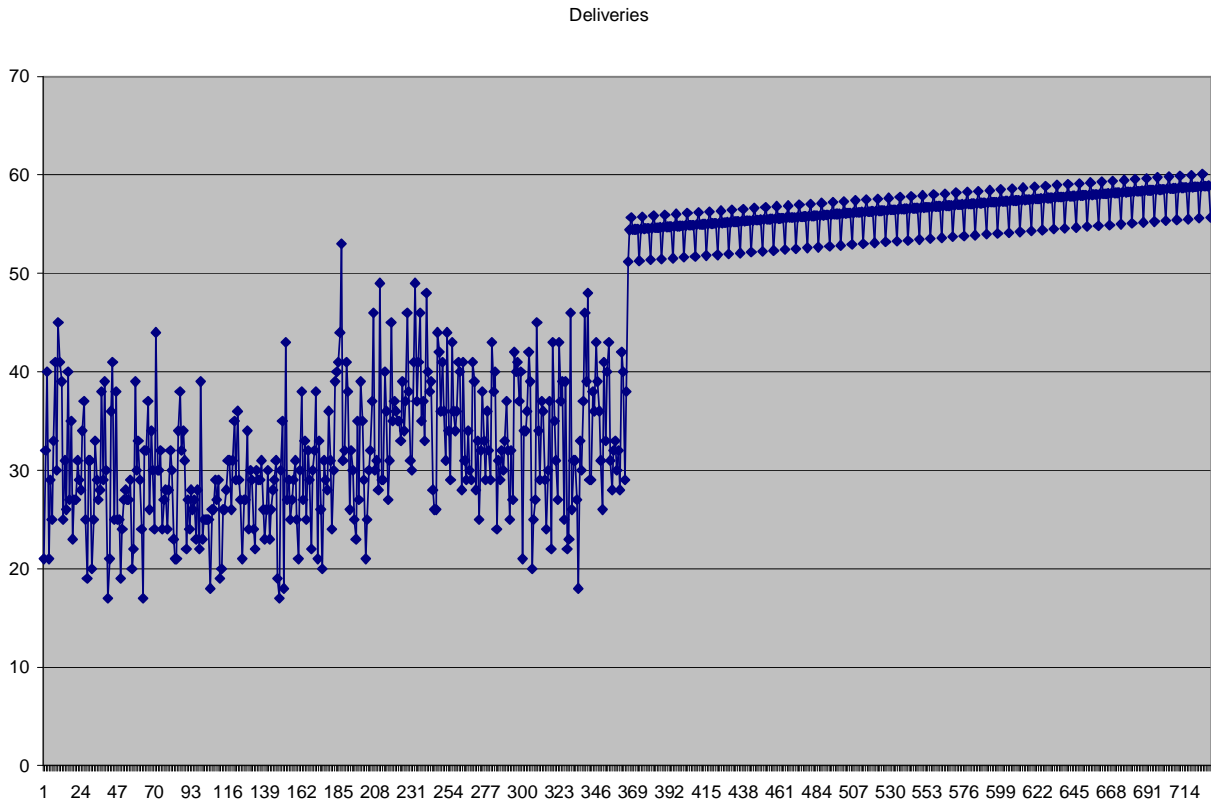
Copy B367, the cell with the prediction for January 1, 1991, and then paste to cells B368 down through B731. That will give you a prediction for every day of 1991.

Move the cell selector to an empty part of your spreadsheet. Type in $=sum(B367:B731)$ to get the predicted total for 1991.

Do a line chart, with the entire column from B2 through B731 as your 1st series. That will show you your observed values for 1990, followed by your 1991 prediction.

In your chart, what appear to be bars going across are your predictions for the different days of the week. The example formula I show on the preceding page would generate a chart with one bar for Tuesdays,

another bar for Thursdays, and a third bar for the other days, like this:



Your graph should look something like this, except that the right half should look like a continuation of the trend of the left half. It should not be cockeyed like the right half of the above graph is.

In this graph, for observations 366-730, there are three apparent bars. The top bar is the Thursdays. The lower bar is the Tuesdays. The middle bar is all the other days.

Now to writing the memo. Start your memo with a sentence reporting your prediction of the total number of Medicaid-paid births for the year 1991. Then explain how you got that prediction. Explain that your method, unlike the simple regression method, separates the effect of the one-time eligibility change from the underlying time trend. Then say much difference the eligibility change made. Also say what time trend you found once the eligibility change was taken into account.

In a separate paragraph, explain what you found about fluctuations by the day of the week. Say which days have more or fewer births than which others, and by how much. Only mention differences that are statistically significant. Speculate in one or two sentences about a possible reason for what you found.

Submit the memo and the graph. Those two pages are all you need to give me.