Fine-Tuning Your Forecasting for Cycles and Patterns

By Maggie Klenke

In basic courses on forecasting, you learned to apply the forecasting techniques of time series analysis. These include predicting the trend which could be a rate of growth over time or perhaps a rate of shrinkage in contact volume. The other primary component of time series analysis is the seasonality pattern – the peaks and valleys of the workload from month to month throughout the year. These were also broken down from monthly forecasts to weekly, daily and finally to intervals such as half-hours.

In many organizations, there are also business drivers to be considered. These include one-time events such as a competitive announcement or system down time. The impact of these events might be confined to the hours of a single day. However, most of the events impacting the contact center are repeatable events and often spread over several days or even weeks. Some of these can be predicted for impact based on historical patterns but it is not easy to know when it might happen again. These include such things as weather impacts or marketing campaigns. Others are repeated at regular intervals and both impact on workload and timing are predictable. These more predictable patterns are associated with such things as bill mailings that happen at the same time each month.

Any type of repeatable pattern that can be predicted should be so that the center can staff appropriately. Those patterns that spread over more than one day are typically referred to as cycles.

Cycle Characteristics

There are three primary components that need to be identified in order to forecast the impact of a cycle. The first is the total volume of work that is associated with the business driver. This volume may spread out over several days or weeks in a variety of patterns. The second component is the distribution pattern. The third element is identification of when the workload change will start and when it will no longer be a factor. Each of these elements requires its own process.

Predicting the volume of work associated with a cycle starts with identifying the business driver that causes it. When that is known, then an analysis of the history of the activity that drives the contacts can be associated with the number of contacts that resulted. For example, if a company sends out a catalog that prompts customers to call to place orders, there is generally a correlation between the number of catalogs that are mailed and the volume of additional calls that results. A marketing email that targets a defined number of potential customers may drive a volume of calls that can be attached to it. A storm that impacts a defined number of citizens may drive calls to the insurance claims center in a predictable way.

In the example below, there are two historical periods when marketing sent out a quantity of emails with a special offer to potential customers. The calls were tracked so that there is a clear association of those calls that are a direct result of the emails and those numbers are given on the right.

Emails Sent	Call Volume
123,000	6,150
135,000	6,823

The implication of this history is that if we know how many people are targeted to receive the emails, we have a good chance of predicting the additional calls that will be driven to the contact center.

Regression Analysis

This is where regression analysis goes to work. It is the relationship that is important. In this case it is the ratio of calls to items mailed. This process can be used for almost any kind of mailing or event.

When a regression analysis is done, there are two sets of data to be identified. One is called the independent variable or the *cause*. In this case it is the mailings, or the number of emails sent. The other is called the dependent variable and it is the *effect*. In this case, it is the calls that are a direct result of the mailing.

For many cycles, the impact can be experienced over several days or weeks. It is important to know not only how many extra calls will be generated, but how they will spread out over those days. It is common to find that the starting day of week can be as important a factor as the fact that it is the third day in the cycle, so identifying both the pattern that occurs across several days and how it is influenced by the day of week of the starting point will often be important. Other cycles are more dependent upon the date of month, regardless of what day of the week it falls upon. For example, bills might be mailed on the 5th of the month.

One of the hardest challenges in cycled forecasting is to separate the normal contact volume from the incremental so that the influence can be clearly seen. This is best done at the data collection stage as the contacts are happening so that you capture the incremental workload in a different file from the normal, but that is not always possible. Some centers use different toll-free telephone numbers for each campaign, or the associates may code the contacts to the appropriate driver as they are handled. This can also help to identify if the average handle time for these contacts is appreciably different from regular contacts.

Below is the historical data for the email campaigns and resulting calls for this example. It is important to understand that the call volume noted here is the incremental call volume, not the total including all the normal calls. There clearly seems to be a relationship here between the number of emails and the calls. However, knowing that relationship with precision will improve the accuracy of the forecast.

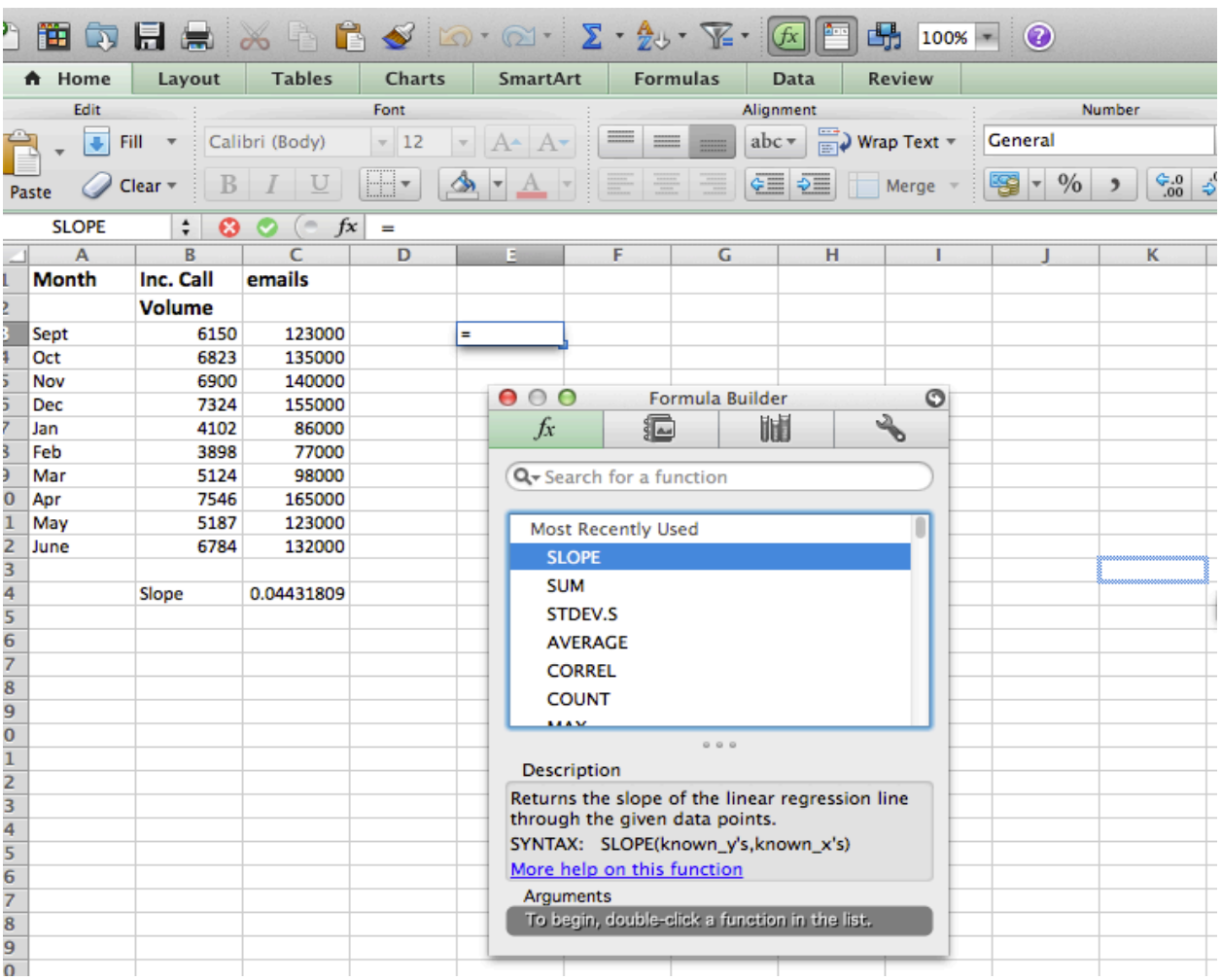
Month	**Incremental Call Volume**	**Emails**
September	6150	123,000
October	6823	135,000
November	6900	140,000
December	7324	155,000
January	4102	86,000
February	3898	77,000
March	5124	98,000
April	7546	165,000
May	5187	123,000
June	6784	132,000

When doing the regression analysis in Excel, follow these steps.

1. Enter the data for the number of emails or other drivers in one column.

2. Enter the data for the number of calls or other results received in another column.
3. Choose any open cell for the answer and click on the down arrow next to the AutoSum character.
4. Choose “more functions” and then the function “slope” from the choices given.
5. The system will display two boxes. In the first one labeled X array, highlight the column for the calls (the dependent variable).
6. In the second box labeled Y array, highlight the column for the emails (the independent variable).
7. Press enter to see the result.

In this example, the result is 0.044318093 (rounded off to 0.0443). That means that we should receive 4.4 calls for every 100 emails sent. The picture below may make this a bit easier to understand.



In the example below, we apply this analysis to a forecast. The regression analysis has been done and found that for every 100 emails sent, there should be 0.0443 calls.

If the normal forecast based on history suggests that the center will get 3322 calls during the upcoming period when the impact of the mailing will be felt (perhaps over a 5 to 6 day period). Marketing has informed the forecaster that 90,000 emails will be sent. This allows the forecaster to calculate the total call volume for that upcoming period.

$$3322 + (0.0443 \text{ X } 90{,}000) = 7309 \text{ calls}$$

Add the 3322 normal calls to the incremental calls from the email campaign. Since 90,000 emails were sent and the ratio of calls is .0.0443 per email, multiply those together and get 3987. Adding 3987 and 3322 results in a prediction of 7309 total calls for the period when the email campaign call volume will be felt. This is the total impact of the entire campaign but of course it will not happen all on one day.

Once the total impact of the business driver is identified, the next step is to analyze historical data from similar events and try to determine how long the impact will affect the center. Look for the beginning of the impact (which may not be the same day the driver occurs). For example, if the driver is postal mailings, it takes a couple of days for the items to arrive in the customer mailboxes. However, if the driver is the collection department making outbound calls to notify customers that their service will be discontinued if they don't call to make a payment in the next 24 hours, the calls are likely to start very quickly.

Once the impact can be seen in the historical data and the start point is identified in terms of how long after the cause, then the data should be analyzed to determine when the impact is no longer being felt and contact volume has returned to normal. With both start and stop identified, the total length and volume distribution will be known. Of course, not all cycles (even driven by the same type of event) will be the same length, so analyzing several cycles of historical data can be helpful in computing averages.

There are a few things to consider in applying regression analysis:

1. The sample data may not be sufficient to form an accurate basis for regression. As for most statistical analyses, the more data the better.

2. Other business changes may have had an impact. The example uses a linear regression analysis based on a direct relationship between one cause and the effect. Where there are two or more factors influencing the result, a variation called multiple regression analysis may be needed. This is applied in pricing a house for sale based on such things as square footage, lot size, age, exterior makeup, etc. for example.

3. The range of data may not include the size of variable planned, making the new assumption unreliable. For example, the new mailing may be 10 times larger than any in the past which makes the calling ratio more unpredictable.

Correlation Analysis

In the regression analysis, three primary components were identified that are the characteristics of a cycle and the volume of work was predicted to match the driver that causes it. Now the analysis moves to figuring out how the incremental volume will distribute across the days of the cycle including the timing of the start and end of the cycle.

Remember that regression is the tool used to identify how many incremental contacts the center will receive. Correlation coefficient is the tool for identifying the pattern of distribution over the days of that cycle.

As with most forecasting, history is the source of the data for analysis. In this case, the idea is to find similar incidents and analyze those patterns and look for a good match that can be used for the forecast.

The process using correlation coefficients is another math function that can easily be applied using a spreadsheet tool like Microsoft Excel. This function is also found in the drop down choices next to the AutoSum feature.

In Excel, the function is called CORREL and the formula asks you to highlight two arrays. An array is either a column or row of continuous numbers and the two arrays must have the same number of cells for the formula to work. The output will be a number that will be 1.0 or less and that number tells you how closely one set of numbers matches the other in terms of the pattern. A result of 1.0 means there is a perfect match.

Analyzing weather patterns is a great example of how correlation coefficients can be used. Consider the challenge of an insurance claims call center whose customers call whenever a weather pattern threatens or hits them. As expected, the behavior of callers will be quite different when the weather issue is a tornado than when it is a hurricane or flood.

In the tornado, there is little warning and within an hour, the insured knows if the house is gone or not. While he may not know all the damage that has been experienced, he knows he needs to call the insurance adjuster to get the claim going so he will have some place to live tonight and a rental car to replace the one upside down in the neighbor's front room. Therefore, the insurance company's call arrival pattern is generally unaffected prior to the event, and has a significant surge immediately after it, dwindling over the next few days as people get past dealing with their grief and get to the insurance questions.

A hurricane or flood can produce quite a different pattern. There is usually warning over several days prior to the actual event during which the insurance company might get calls from customers wanting to know if they are covered and how they can adjust coverage as needed. Then the hurricane or flood actual hits. It might be days before the customers can get back into their homes or neighborhoods to truly assess the damage and call for an adjuster and file a claim. Therefore, the pattern of calls will start prior to the actual event and probably have a lull for a few days around the event itself and then pick up again over a period of weeks thereafter.

To predict the impact of one weather pattern versus another, look for periods that have high correlation coefficients over the several weeks prior to and after the actual weather event. If you can find several of these and they are similar to each other, then you will have a pretty good predictor of the pattern of the upcoming event.

Correlation coefficient analysis is used in the upcoming example so you can see how it applies and how to use Excel to help with the math.

Billing Cycle Example

The example that follows is based on billing cycles as this is a common occurrence in many contact centers. This process can be applied in a similar way to almost any kind of driver that creates additional calls starting at some predictable point in time. Any kind of mailing, a calling campaign, even such things as the day Social Security checks arrive at the bank would fit this model.

In the example, invoices are mailed to customers on the 5th of each month. It could be any day of the week except Sunday when the center is closed, and the post office does not operate. There are noticeable spikes in the call volume after the mail is received by the customers and it generally lasts for 5 to 6 days. The normal call volume has been separated from the incremental in the historical records.

Whether the center works a six-day or seven-day work week, the fact remains that the 5th is definitely a "moving target". While it could have been any day of the week, mailings sent out on Fridays or Saturdays are likely to accumulate so that the calling begins on the following Monday because of the mail delivery delay over the weekend.

Over time and as data is accumulated, look for definite cycles to emerge in the cumulative numbers. One way of quantifying whether a cycle exists is by running a correlation of the individual series of data vs. the cumulative data series. Unless consistently high correlation coefficients are found, there is no definitive cycle.

Remember to take out the normal call volume and be sure to discard any data that is skewed or has anomalies.

Below is the set of assumptions used in this billing cycle example. The center is open for calls 6 days per week. There are 4 months of history to look at as a starting point. If a good match of pattern is found in some of this data, the history can be expanded to improve our prediction accuracy. The analysis will focus on those 5 to 6 days that receive the calls and that starts when the mail is delivered, not when it is mailed.

It is important to understand that the correlation coefficient analysis is only comparing one set of data to another set to determine if the pattern is a good match or not. The actual numbers in one set can be dramatically higher than the other and it will not matter so if there has been major growth or shrinkage between sample data points, it is OK.

On the next 4 tables, the data for each of the 4 months of history is provided. It is fortunate in this case because there are two months that start on a Saturday and two that start on a Tuesday. That wouldn't typically be the case for January through April, but it is important to find pairs of months with the same start day of week for this type of analysis. It is also important to note that it really doesn't matter if these months are just prior to the one being forecast or even a continuous series. They just need to be good examples of the pattern for this business driver with pairs starting on the same day of week.

The assumption has been made that it takes two days from the date of mailing for the bills to be received by the customers. This is a key assumption that will be needed for any analysis like this. It is reasonable to assume emails or calls are received on the same day made, even if they sit in the customer inbox for a few hours before opening. But postal mail takes a while to arrive.

January Data

Date	Day	Actual CV	Normal CV	Incremental CV
1-5	Sat	2703	2700	n/a
1-6	Sun	0	0	n/a
1-7	Mon	6875	5289	1586
1-8	Tue	5754	4875	879
1-9	Wed	5398	4750	648
1-10	Thu	4802	4523	279
1-11	Fri	4200	3898	302
1-12	Sat	3109	2700	409

Above is the January data found in the history file. Notice the 5th is a Saturday and while the center is open, customers have not yet gotten the bills so the call volume is pretty normal. The example center is closed on Sunday.

By Monday, many of the customers will have received their bills and the calling will start. Notice the incremental call volume of 1586. As the cycle (and week) progresses, the call

volume continues to drop with a little surge on Saturday, perhaps driven by customers who work during the week and wait to open bills until the weekend.

February Data

Date	Day	Actual CV	Normal CV	Incremental CV
2-5	Tue	4887	4875	n/a
2-6	Wed	4734	4750	n/a
2-7	Thu	5904	4523	1381
2-8	Fri	5387	3898	1489
2-9	Sat	3156	2700	456
2-10	Sun	0	0	0
2-11	Mon	5869	5289	580
2-12	Tue	5076	4875	201

Above is the February sample data. Once again, no change to the 5th Tuesday or even Wednesday as the bills haven't arrived in the customer mailboxes yet. But by Thursday, there is incremental calling and even more on Friday. Saturday is relatively light but there is still residual calling on Monday and Tuesday of the following week.

Comparing the incremental call volume pattern for February to January, it is clear the day of week arrival pattern and even the amounts on each of the 5 to 6 days affected are quite different.

March Data

Date	Day	Actual CV	Normal CV	Incremental CV
3-5	**Sat**	**2703**	**2700**	**n/a**
3-6	**Sun**	**0**	**0**	**n/a**

3-7	**Mon**	**6843**	**5289**	**1554**
3-8	**Tue**	**5590**	**4875**	**715**
3-9	**Wed**	**5402**	**4750**	**652**
3-10	**Thu**	**4760**	**4523**	**237**
3-11	**Fri**	**4208**	**3898**	**310**
3-12	**Sat**	**3098**	**2700**	**398**

The next example of data for this business driver is March and notice that it has the same 5th of the month starting on Saturday that was seen in the January data. Notice that the calling pattern is quite similar as well with the biggest volume on Monday, gradually decreasing across the week, with a little surge on Saturday.

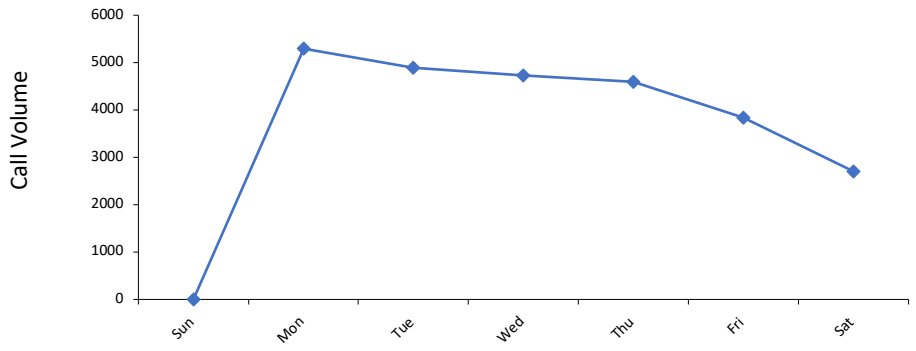
It is beginning to look like the day of week is a definite factor in the distribution pattern.

April Data

Date	Day	Actual CV	Normal CV	Incremental CV
4-5	**Tue**	**4887**	**4875**	**n/a**
4-6	**Wed**	**4734**	**4750**	**n/a**
4-7	**Thu**	**5943**	**4523**	**1420**
4-8	**Fri**	**5217**	**3898**	**1319**
4-9	**Sat**	**3509**	**2700**	**809**
4-10	**Sun**	**0**	**0**	**0**
4-11	**Mon**	**5780**	**5289**	**491**
4-12	**Tue**	**5209**	**4875**	**334**

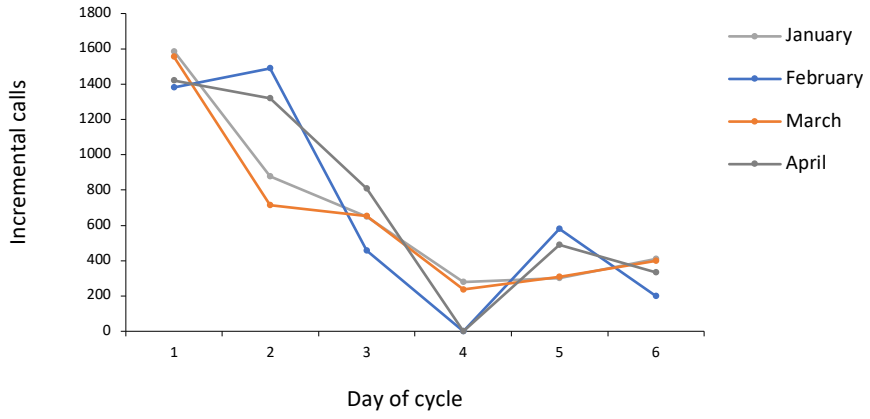
The April start day matches February and we see a similar pattern here. It is not an exact match as the first day of this month's cycle has higher volume than the second while February had a higher second day. The correlation coefficient analysis will reveal if it is close enough to consider it worth using.

Just to make it easier to visualize, here are graphs of the patterns. The first one is the normal call volume and it starts out at the peak on Monday and declines steadily across the week.



These are the 4 months of just the incremental call volumes. The blue line for February and the grey line for April are fairly close to each other, and that the green for January and red for March are close but the two sets are not very close to each other.

We can confirm the patterns with the math.



Here are the instructions for computing the correlation coefficients using a Microsoft Excel function. The analysis centers on the incremental call volume and the day of week. First, enter all the data as shown in the following chart so that the columns of numbers are set up beside one another for the two months to be analyzed.

Choose an open cell for the result and click on the down arrow next to the AutoSum and choose "Correl". Highlight the data for one month for Array 1 and the other for Array 2 and click OK. Remember that a perfect match is 1.0 and anything less than perfect is less than 1.0. It does not matter if one set of numbers is significantly higher than the other. The process is only looking for a match of the distribution pattern.

1-5	Sat	0	2-5	Tue	0
1-6	Sun	0	2-6	Wed	0
1-7	Mon	1586	2-7	Thu	1381
1-8	Tue	879	2-8	Fri	1489
1-9	Wed	648	2-9	Sat	456
1-10	Thu	279	2-10	Sun	0
1-11	Fri	302	2-11	Mon	580
1-12	Sat	409	2-12	Tue	201

Here are the results of the comparisons of each set of months in our 4-month analysis. It is clear that the correlations are in the high 90s for January/March and for February/April and all the others are much lower. A correlation coefficient below .8 generally means a poor match.

- Each set of months was analyzed
  - Jan – Feb = .811069
  - Jan – Mar = .991872
  - Jan – Apr = .880543
  - Feb – Mar = .752364
  - Feb – Apr = .955501
  - Mar – Apr = .842633

Looking at the two sets of data that gave the best match, they are the ones that have the 5th on the same day of week. Therefore, this analysis has proven the hypothesis that the day of week is a critical factor in determining the pattern of our call arrivals.

The process has proven that day of week is important, so the next step is to dig into the history and find as many examples of 5th starting on each day of the week as possible. It really doesn't matter if the call volume overall has changed over time, it is only the pattern that matters.

Distribution Pattern Forecast

The next step will be to determine the percent of the incremental cycle workload that happens on each day based on the average of the historical periods we have available.

With the total amount of incremental calling determined from the regression analysis and this percent of the total that will appear on each day, the process will identify the amount of calls that will be received on each day of the cycle. Adding this to the normal call volume will produce a good combined forecast.

The example that follows applies some numbers to make the process a bit easier to follow. Here is our incremental call data from the January and March analysis. It would be far better if there were 6 or 8 months of data to use but use what you have that matches the selection criteria. The first step is to average the call volumes for each of the days of the week separately. On Monday, for example, the average of 1588 and 1554 is 1571.

Compute the average for each day of the week and compute the average total for the week as well. (If this were a 10-day or 30-day cycle, it would be done for each day in the entire cycle, not just for the week.)

The next step is to determine what percent of the total cycle call volume will fall on each day. Monday's 1571 is divided by the week's total of 3986 to reveal that 39.4% of the incremental call volume of the cycle will fall on that day.

Day	**Jan CV**	**Mar CV**	**Avg CV**	**Day %**
Mon	1588	1554	1571	39.4
Tue	879	715	797	20.0
Wed	648	652	650	16.3

Thu	279	237	258	6.5
Fri	302	310	306	7.7
Sat	409	398	403	10.1
Total	4107	3866	3986	100

Remember that even though this example has use only 2 sets of data, more data typically will make the results more accurate.

Computing the Final Forecast

In the first column of the example below, the normal call volume percentages for each day of the week are provided and the normal week is forecast to have 3322 calls. Applying the assumed daily percentage to each day of the week computes the normal call volume for each of the days in the second column.

In the next step, take into account the total number of additional calls that are expected based on the regression analysis. In this case problem we have assumed 3987 additional calls based on the regression analysis performed earlier.

Day	**Normal CV %**	**Normal CV**	**Added CV Daily %**	**Added CV**	**Total Forecast**
Mon	24.5	814	39.4	1571	2385
Tues	19.4	644	20.0	797	1441
Wed	18.9	628	16.3	650	1278
Thu	18.0	598	6.5	259	857
Fri	16.4	545	7.7	307	852
Sat	2.8	93	10.1	403	496

Total	100	3322	100	3987	7309

The correlation coefficient analysis revealed that 39.4% of the total incremental calls are expected to arrive on Monday, so applying .394 X 3987 indicates that the added call volume will be 1571 calls that day. Adding the normal call volume of 814 to the added 1571 from the bill mailing results in a total Monday forecast of 2385 calls as shown in the last column.

Calculate the added call volume expected on the remaining days in the week and compute the total forecast for each day and the week.

The incremental or added call volume from the bill mailing is provided for each day of the week. By adding that number to the normal call volume for that day, the total forecast for the day is computed. The total new forecast for the week is 7309 calls.

The forecaster will now be able to use this new forecast to determine the volumes by interval of the day so that schedules can be created to meet the service goal set by the center. Of course, the distribution pattern of the incremental calls may be different by half-hour than the normal calls and an intraday pattern analysis may improve the accuracy of the forecast even further.

Summary

Planning for repeatable cycles is an important part of achieving accuracy in planning and staffing for the center. Regression analysis is a useful tool for determining the relationship between the drivers and results, such as bills mailed and calls received. Correlation coefficients help us to find a pattern that is a good match to the one we are forecasting so that the distribution of the incremental work can be identified by which day (or interval) it is likely to arrive. With a total impact and the distribution pattern identified, the incremental work can be effectively predicted and added to the normal workload. The result is a forecast that will provide the basis of a staffing and scheduling plan that should produce more consistent achievement of service goals.

The good news is that you don’t have to be a math wizard to use these tools. If you understand how they work and apply to the problem at hand, just rely on the spreadsheet tools to do the heavy lifting. No formula needed. Of course, if you are a math wizard, so much the better.