EMC 445 EMS SYSTEMS MANAGEMENT
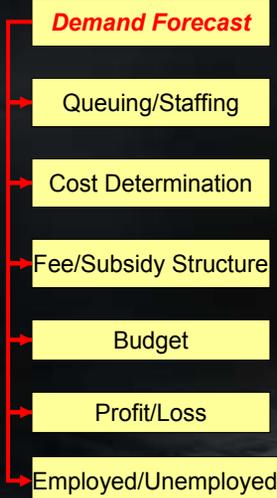
Call Demand Forecasting I:
Linear Regression Primer

Where Does Call Demand Forecasting Fit?

Demand Forecast

Queuing/Staffing

Cost Determination

Fee/Subsidy Structure

Budget

Profit/Loss

Employed/Unemployed

Review of Regression

- Use of one or more variables to predict another variable.
- The variables used to predict are the "predictor", "explanatory", or "independent" variables. Denoted as "X".
- The variable to be predicted is called the "dependent" variable. Denoted as "Y".
- The general form of the equation is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ...\beta_i x_i + e$$

where,

y is the independent variable (the one you wish to predict)
each x represents an independent variable
β_0 is the y-intercept term, or the value of y when all x's are zero
β's are the regression coefficients (to be calculated by the software)
e is the error term, or prediction error

# Applying Regression to Call Demand Forecasting

- In the setting of predicting EMS call demand, monthly call volume will be the dependent (Y) variable (i.e., the one we wish to predict), and time will be the independent variable.

Year	Month	Time Series (X)	Call Volume (Y)
2000	January	1	253
	February	2	254
	March	3	243
	April	4	260
	May	5	268
	June	6	250
	July	7	270
	August	8	281
	September	9	259
	October	10	291

Forecasting Call Volume
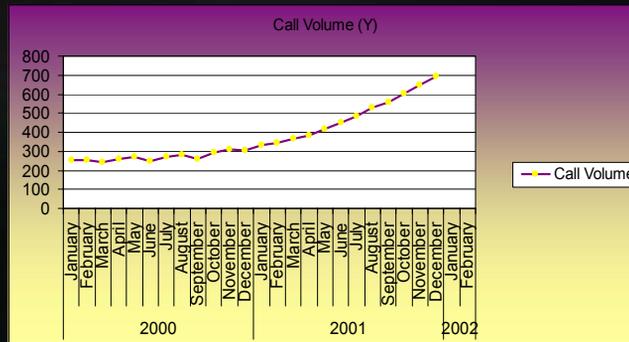
Call Volume (Y)

Based on our historical data, we want to forecast EMS call demand for the months of January and February, 2002 (series numbers 25 and 26).

Forecasting Using Excel

- Select a column in Excel for the dependent variable (i.e., call volume) and enter the data.
- Select a column for the independent variable(s) (time series) and enter the data.
- Click on "TOOLS" then click on "DATA ANALYSIS" then scroll down and highlight "REGRESSION" and click "OK".
- Input the "Y Range" which is the dependent variable (call volume). Do this by highlighting the column of the spreadsheet with the call volumes.
- Input the "X Range" which is the independent variable (Month number). Do this by highlighting the column of the spreadsheet with the month numbers.
- Click the button for "Output Range" then highlight the upper left cell of the area of the spreadsheet where you want the output to be pasted.
- Then click OK.
- The output block will appear in the spreadsheet.

Forecasting Using Excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.924754679
R Square	0.855171217
Adjusted R Square	0.84858809
Standard Error	54.13072463
Observations	24

ANOVA

	df	SS	MS	F	Significance F
Regression	1	380634.8557	380634.9	129.9035063	0.000000000106
Residual	22	64462.97768	2930.135		
Total	23	445097.8333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	149.5036232	22.80800712	6.554874	0.000001360808	102.2026607	196.804586	102.2026607	196.804586
X Variable 1	18.19304348	1.596227986	11.39752	0.000000000106	14.8826657	21.5034213	14.8826657	21.5034213

- Adjusted R square indicates how well the equation "fits" the data. The range is 0 to 1, where 1 represents a perfect fit. The higher the R^2 the more accurate the regression equation.
- Significance F indicates how likely the equation really represents the data and not coincidental. Reported as a decimal, the lower the number the better. It should be less than 0.05.
- X Variable 1 is the coefficient for the first independent variable. This coefficient is multiplied by the value of each independent variable. There may be many X variables.
- Intercept is the point where the regression "line" intersects the Y axis. Also referred to as the "constant term."
- P value is similar to the significance F above. It indicates how well individual independent variables predict the dependent variable, and not done so by chance. "Good" values are less than 0.05. Any independent variables with p values greater than 0.05 should be eliminated and the regression recomputed.

Forecasting Using Excel

- The equation for predicting call demand for January, 2002 (month #25) is:
- Call volume = $\beta_0 + (\beta_1 x_1) + (\beta_2 x_2) + (\beta_3 x_3) + \dots$

  Where,
    - β_0 = intercept term
    - β_1 = coefficient for first independent variable
    - x_1 = value of first independent variable
    - β_2 = coefficient for second independent variable
    - x_2 = value of second independent variable

- In this example, we only have the intercept term and one variable which is month number.
- Call volume = 149.50 + (18.19*25) = 604 calls
- For February, 2002 (month #26 in our example), the call volume is 149.50 + (18.19*26) = 622 calls.

Forecasting Using Excel

Comparison of Forecast and Empirical Call Volume
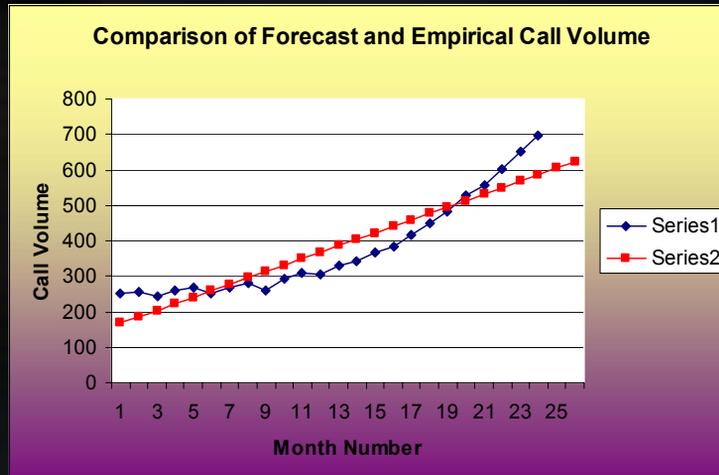
Using "Dummy" Variables

- So far, we have only used one independent variable-the sequential number of the month. It is possible to transform this variable (and all variables in general) to another variable. Looking at the graph of the call volume, it is apparent that there is a curve in the call volume trend. If we graph our equation and compare it against the raw data (see above), it is apparent that our equation is "linear" and does not "fit" our data well, despite the high R^2 value. In fact, the further we go into the future, the less accurate our forecast becomes. What we need to do is "capture: the curve in the trend. To do this, we will use dummy variables.

- To set up a dummy variable, go to the empty column adjacent to the column of independent variables. For each observation (i.e., month) in the independent column, create a dummy variable by transforming the independent variable with exponentiation. For example, we can square each term in the independent variable column: Month 1 is still 1, month 2 becomes 4, month 3 becomes 9, etc. Then, when we specify the independent variables in the regression, we will specify both columns of independent variables (i.e., the month and the dummy variable).

- In this example, we created a dummy variable by squaring the independent variable. You can also cube, raise to a power of 4, or even use exponents less than one. In general, the higher the value of the exponent, the steeper the curve in the fitted equation.

- After we run the regression with the two independent variables, we must redo the equation. Our equation will now have an intercept and two coefficients for each independent variable.

- This is where art and science meet in forecasting. You will have to search for the best value of the exponent by varying the value of the exponent, run the regression, graph the results, and evaluate your forecast both statistically and graphically

Using Dummy Variables

SUMMARY OUTPUT WITH DUMMY VARIABLE

Regression Statistics	
Multiple R	0.997213835
R Square	0.994435433
Adjusted R Square	0.993905474
Standard Error	10.86009921
Observations	24

ANOVA

	df	SS	MS	F	Significance F
Regression	2	442621.0565	221310.5	1876.439168	0.000000000000
Residual	21	2476.77685	117.9418		
Total	23	445097.8333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	278.3013834	7.245876606	38.40824	0.000000000000	263.2327555	293.370011	263.2327555	293.370011
X Variable 1	-11.52951657	1.33546625	-8.63333	0.000000023769	-14.30677115	-8.752262	-14.3067712	-8.752262
X Variable 2	1.188902402	0.051859997	22.92523	0.000000000000	1.081053615	1.29675119	1.081053615	1.29675119

Setting up the Spreadsheet

- To set up our equation in the spreadsheet, our equation with the additional variable becomes:

- **Call volume = 278.30 + (-11.52*month number) + (1.18*month number^2)**

- Thus, for January, 2002 (month #25), the forecast is:

  Call volume = 278.30 + (-11.52*25) + (1.18*625) = 733 calls

Verifying Results Graphically

Comparison of Empirical Data and Forecast with Dummy Variable