© E/Step Software Inc. 2007 All rights reserved

MAD and R² as Measures of Forecast Error

Robert M. Duncan, CFPIM, Training Manager, E/Step Software Inc., (<u>www.EstepSoftware.com</u>) August 1, 2007

In our APICS classes we learned that the easy-to-calculate *mean absolute deviation* (MAD) is an acceptable measure of forecast accuracy. In our statistics classes we learned that the *index of determination* (\mathbb{R}^2 , where \mathbb{R} is the *correlation coefficient*) is a good measure of how well a regression line fits the data. Neither the MAD nor the \mathbb{R}^2 is as useful as the standard deviation for measuring how well a forecast model fits its historical demand, and neither should be used for safety stock decisions. In fact, as demonstrated in this paper, these two measures can even suggest the wrong conclusions.

MAD vs. Standard Deviation

In the graph there are two forecasts. Which is better? According to the MAD calculation, Forecast(1) is better. According to the Standard Deviation calculation, Forecast(2) is better. How can this be? In APICS classes we learned that the Standard Deviation = $1.25 \times MAD$ for normally distributed forecast errors.

The reality is that 1.25 is an approximation, but in this example the ratio of MAD and STDEV vary from 1.70 for the first forecast to 1.10 for the second forecast. The difference is that Forecast(1) has a small deviation of 0, 1, and 2 units, with one exceptionally large deviation of 8 units. Forecast(2) has larger absolute deviations of 1, 2, and 3 but no exceptionally large value.

The MAD calculation is simply the average of the absolute deviations. It gives equal weight to exceptionally large deviations as it does the small deviations. The Standard Deviation calculation weights large deviations much greater than small deviations—by their square.

So if you were to use the MAD and the 1.25 factor to calculate the safety stock for the first forecast, the safety stock would be insufficient and would obtain less service than predicted. Using the MAD for the second forecast would result in inventory and service greater than predicted.

Robert G. Brown invented the MAD in the early days of computing because of the slowness of computing a square root at the time. He points out that now even the cheapest calculator can take a square root. "*MAD is no longer appropriate to the real world of computers. It never was the correct measure of dispersion.*"¹

¹ Brown, Robert G. Materials Management Systems (New York: John Wiley & Sons, 1977) 148

Frequency Distribution

Abs Dev	Freq(1)	Freq(2)
0	2	
1	6	4
2	1	3
3		3
4		
5		
6		
7		
8	1	
MAD	1.60	1.70
STDEV	2.72	1.87
Ratio	1.70	1.10



Which Forecast is Better?

R² vs. Standard Deviation

The demand history for two different SKUs is graphed. Using Excel's Trend *Line* feature, a line is fit to each of 2 SKU's histories using regression. The R^2 value is calculated to determine the index of determination (also called the coefficient of determination) for each SKU.

An R^2 of 1 would be a perfect positive correlation while an R^2 of 0 indicates no correlation at all. Conventional wisdom and even some statistics textbooks incorrectly assume that the greater the coefficient, the better the model: "When the magnitude of the coefficient of determination (R^2) is large, this indicates that the error term for this model is relatively small and the model fits good."²



But this is not always the case. SKU 2 (green) has an R^2 value of 0.6712, a "good" correlation. SKU 1 (red) has an R^2 value of 0.0075, almost no correlation at all. (Note that an R^2 0.6712 is "good" only in comparison to the alternative 0.0075; in general 0.6712—which means only two thirds of the variation in the data is accounted for by the model would be considered poor to marginal.) The R^2 values imply that the forecast model for SKU 2 is superior to the forecast model for SKU 1. But both the standard deviation measurement and our eves disagree.

The standard deviation for SKU 1 is 4.01 units, whereas for SKU 2, it is 22.2 units. Using the standard deviation measure, the amount of safety stock required to cover for forecast error for example 2 needs to be over 5 times greater than for example 1. SKU 1 obviously has a better model.

	SKU 1	SKU 2
R-Sqd	0.0075	0.6712
Std Dev	4.01	22.2

SKU 2

Model

45.6

56.6 67.6

78.7

89.7

100.7

111.7

122.8

133.8

144.8

Demand

30

60

50

120

90

70

150

122

130

130

A few years ago a user requested that FGS add an R^2 calculation. This feature was added to a service pack. If you run the command called RSQUARED it will create the SKU.RSQUARED field. FGS recalculates the value at model fit time. It is interesting to view in SIMULATE, though you will find the results quite disappointing. R^2 is typically a small value and does not always correlate to forecast error. The standard deviation is the best measure and that's why it is used in FGS.

SKU 1

Model

100.1

100.0

99.9

99.8

99.7

99.5

99.4

99.3

99.2

99.1

Data for Examples

MAD Example						<u>R² Example</u>			
							SK		
	Period	Actual	Forecast(1)	Forecast(2)		Period	Demand		
	1	10	9	11		1	105		
	2	11	11	9		2	95		
	3	9	10	10		3	103		
	4	5	3	8		4	97		
	5	5	5	7		5	102		
	6	7	8	6		6	92		
	7	5	13	7		7	102		
	8	7	8	6		8	98		
	9	10	9	7		9	104		
	10	14	13	15		10	98		

² Breyfogle, Forrest W. Implementing Six Sigma (New York: John Wiley & Sons, 1999) 365