Forecasting

**Forecast errors**

The forecast performance of various weather systems is affected by large variations between original and forecasted values. More errors arise from the errors in initial conditions as compared to errors in model formation. Various sources have to be examined before any clear conclusion can be drawn about the sources of forecasting errors. There are many ways to separate the analysis and model errors including running one model with different initial conditions or running different models with same initial conditions (F, E.Klinker, P.Courtier, & Hollingsworth, 1996).

The simplest way of calculating forecast error is to subtract the actual values from the forecast values. A positive forecast shows that there is a higher value forecasted as compared to the actual value and a negative forecast is the situation when some value has been over estimated. Generally, there has been an over estimation in the data given for the assignment. Only two times out of 11 there is an under estimation of values.

|  |  |  |  |
| --- | --- | --- | --- |
| **YEAR** | **ACTUAL (MILES FROM JOHNSTOWN)** | **24-HOUR FORECAST (MILES FROM JOHNSTOWN)** | **Forecast Error** |
| 1 | 4 | 6 | 2 |
| 2 | 5 | 5 | 0 |
| 3 | 30 | 40 | 10 |
| 4 | 2 | 10 | 8 |
| 5 | 12 | 13 | 1 |
| 6 | 7 | 5 | -2 |
| 7 | 11 | 11 | 0 |
| 8 | 21 | 25 | 4 |
| 9 | 12 | 8 | -4 |
| 10 | 12 | 15 | 3 |
| 11 | 6 | 9 | -3 |
|  |  |  |  |

It can be seen from the last column of the above table that most forecasts have resulted in a higher value as compared to the actual values.

**Mean Absolute Deviation**

This is used as a measure of variability in the data presented. Firstly the mean of data has to be taken by adding all values and dividing the sum of values by their number. Next step is to subtract mean from all values and take the absolute of these values. It is necessary to take the absolute of these values because the sum of deviations taken from mean is zero. These differences are then summed up and divided by the number of values to obtain the final result.

|  |  |  |
| --- | --- | --- |
| **YEAR** | **ACTUAL (MILES FROM JOHNSTOWN)** | **Difference from mean** |
| 1 | 4 | │4-10│ = 6 |
| 2 | 5 | │4-10│ = 6 |
| 3 | 30 | │30-10│ = 20 |
| 4 | 2 | │2-10│ = 8 |
| 5 | 12 | │12-10│ = 2 |
| 6 | 7 | │7-10│ = 3 |
| 7 | 11 | │11-10│ = 1 |
| 8 | 21 | │21-10│ = 11 |
| 9 | 12 | │12-10│ = 2 |
| 10 | 12 | │12-10│ = 2 |
| 11 | 6 | │6-10│ = 4 |
|  |  | Sum = 65 |

The mean absolute deviation is 65/11 which is approximately 6. This is simply the mean average of the absolute differences between actual reading and the overall mean showing the amount by which mean differs from the original values. This method gives proportional place to each entry in the result and is easier to be taken up by new researchers. Some other measures of variation have been considered too complicated for the new researchers to understand (Gorard, 2014). Another issue regarding the use of other measures of dispersion is that they do not take into account the relative distance of values from the mean.

**Naïve Model**

A naïve model is based upon the estimates from the past year or time periods. In the present study, a naïve model could have been used to estimate temperature for any future time period. Naïve models were first suggested by the economists in 1940s as benchmarks of forecasting accuracy. A simple naïve model will state that the temperature based on the previous year value. The naïve model can be traced back to the caveman who used to say that tomorrow’s weather will be same like today. The documented proof for this method can be seen from Braddock Hickman in 1942. (McLaughlin, 1983). The following table will show the forecasts based on previous years’ values.

|  |  |  |
| --- | --- | --- |
| YEAR | ACTUAL (MILES FROM JOHNSTOWN) | Naïve Forecasts |
| 1 | 4 | - |
| 2 | 5 | 4 |
| 3 | 30 | 5 |
| 4 | 2 | 30 |
| 5 | 12 | 2 |
| 6 | 7 | 12 |
| 7 | 11 | 7 |
| 8 | 21 | 11 |
| 9 | 12 | 21 |
| 10 | 12 | 12 |
| 11 | 6 | 12 |

There is a considerable difference between the original forecasts made by Monica and the naïve forecasts which shows that there can be some errors made by her. The differences will be judged by using the mean squared error which is a measure of dispersion for forecasting errors, taking the average of squared individual errors.

|  |  |  |
| --- | --- | --- |
| **ACTUAL (MILES FROM JOHNSTOWN)** | **24-HOUR FORECAST (MILES FROM JOHNSTOWN)** | **MSE** |
| 4 | 6 | (4-6)^2 |
| 5 | 5 | (5-5)^2 |
| 30 | 40 | (30-40)^2 |
| 2 | 10 | (2-10)^2 |
| 12 | 13 | (12-13)^2 |
| 7 | 5 | (7-5)^2 |
| 11 | 11 | (11-11)^2 |
| 21 | 25 | (21-25)^2 |
| 12 | 8 | (12-8)^2 |
| 12 | 15 | (12-15)^2 |
| 6 | 9 | (6-9)^2 |

|  |  |  |
| --- | --- | --- |
| **ACTUAL (MILES FROM JOHNSTOWN)** | **Naïve Forecasts** | **MSE** |
| 4 | - | (4-0)^2 |
| 5 | 4 | (5-4)^2 |
| 30 | 5 | (30-5)^2 |
| 2 | 30 | (2-30)^2 |
| 12 | 2 | (12-2)^2 |
| 7 | 12 | (7-12)^2 |
| 11 | 7 | (11-7)^2 |
| 21 | 11 | (21-11)^2 |
| 12 | 21 | (12-21)^2 |
| 12 | 12 | 0 |
| 6 | 12 | (6-12)^2 |

The above tables show the mean squared error for both the actual forecasts taken by Monica and naïve forecasts based on the values of previous years. When we calculate the values for two data sets, we see that MSE for naïve forecasts is 162.18 and for Monica’s forecasts, they are 20.27. Thus we will prefer the forecasts put in place by Monica.

**Mean Absolute Percent Error**

The mean absolute percentage error is one of the most widely used measures for forecast accuracy. It is the average of absolute percentage errors and is scale independent and is easy to interpret. One major drawback with this method is that it produces undefined or indefinite values when there is a zero involved in the calculations. If the values are very small, this method may produce very large percentage errors. An approach used in this case is to exclude the outliers or values counting to less than 1 (Kim & KJim, 2016).

**Forecast Bias**

This problem refers to having either very high or very low forecasts as compared to the actual values leading to a forecasting error. Many reasons may result in this situation. The optimism bias is shown mainly by sales and other relevant forecasting teams. This is because there is a bonus system put in place for the sales people and they want to show higher sales. Sandbagging is the reverse of optimism bias where executives tend to lower the forecasts so that no subordinate is able to win any bonus. An anecdote bias is a situation where something bad had been experienced by the employees of a company and any future estimates will be made considering the past event or happening. The recent data bias is true for all processes which are undertaken by humans. Our minds remember the most recent happenings more than the past ones which means that people will tend to over react to the most recent happenings. The forecast can be affected by some silly things which are shown to the participants before they make their estimates. In the data provided by Monica, there is optimism bias reflected.

Developing a regression model over a period of time will better suit Monica and she will be able to make better forecasts using it. This will also help her to know the changes that will take place over a period of time and also forecast independently for a particular point in time.

# **References**

F, R., E.Klinker, P.Courtier, & Hollingsworth, A. (1996). Sensitivity of forecast errors to initial conditions . *Quarterly Journal of the Royal Meteorological Society*, 121-150.

Gorard, S. (2014). Introducing the mean absolute deviation 'effect' size. *International Journal of Research & Method in Education*, 105-114.

Kim, S., & KJim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 669-679.

McLaughlin, R. L. (1983). forecasting Models: Sophisticated of Naive? *Journal of Forescasting*, 274.