Reprinted from IFPress (2007) 10(1)2,3

Discovering Time-Dependent Trends

Richard E Lampe and Daniel M. Ennis

Background: Product changes in manufacturing and storage are often time-dependent. Sometimes these changes are linear, sometimes cyclical and at other times chaotic. It is often not obvious that a time-dependent change is in fact occurring in some variable of interest. In this report we discuss methods for studying whether a trend exists and, if so, what the nature of that trend is. Data used to study this problem, observations collected sequentially over time, are referred to as a time series. Time series techniques can prove to be important in bringing to light underlying trends in data that may not be evident through other techniques[1].

Scenario: Food and beverage color is an important component of product acceptability and value. Various techniques have been developed for the objective color measurement of citrus juices. These include colorimetric methods and spectrophotometric techniques. The relevance of these instrumental measurements to human perception is often of interest in establishing standard methods for judging citrus quality. For instance Anthony, Ennis and Cook[2] studied the relationship between visual rankings of lemon juice and colorimeter color scores.

Your company markets a range of citrus juice products that include lemon juice. You have recently been made aware of consumer complaints about variation in the color of the lemon juice that your company produces. The majority of these complaints express the opinion that sometimes the juice appears to be too dark. You know that during lemon juice pasteurization and concentration procedures, the resulting lemon juice concentrate can be darkened to an unacceptable brown-yellow color under certain temperature and time-dependent circumstances. In order to study potential changes in the color of your company's lemon juice, you have obtained color data taken twice each day over an eight-week period from your company's production facilities. The data consists of a total of 112 color readings of lemon juice taken 12 hours apart with the two daily readings corresponding to the day shift's production and the night shift's production, respectively.

Moving Averages and Autocorrelation: The simplest and most basic smoothing technique is a moving average. A moving average of time series data is an example of a filter, a linear mapping of one time series to a new time series. Each point in the new time series has a value equal to the average of k successive points in the original time series for some value of k, called the smoothing factor. For example when $k = 3$, if the original time series is given by the data $x_1, x_2, x_3, \ldots, x_n$, the new time series would be $y_1, y_2, y_3, \ldots, y_{n-2}$, where $y_i = (x_i + x_{i+1} + x_{i+2})/3$.

Autocorrelation is a measure of the correlation between observations separated by fixed units of time called *lag*. For example, suppose that there are n time-based observations given by $x_1, x_2, x_3, \ldots, x_n$. When the lag = 1, the autocorrelation is the correlation between the two sets $x_1, x_2, x_3, \ldots, x_{n-1}$ and $x_2, x_3, x_4, \ldots, x_n$. This is a measure of the relationship between the value of each point in time on the value of its successor. Similarly, when the lag = 2, the autocorrelation is the correlation between the two sets $x_1, x_2, x_3, \ldots, x_{n-2}$ and $x_3, x_4, x_5, \ldots, x_n$. This is a measure of the relationship between the value of each point in time on the value of data points 2 time units removed from it.

Line Plot of the Lemon Juice Color Data: As a first step in assessment it is helpful to look at a line plot or a scatter plot to visually identify any apparent trends in the data. The line plot of lemon juice color values over the 112 chronological observations appears below in Figure 1. There is no apparent trend in the data. A trend towards an increase in darkness, for instance, would be evident through relatively higher color values appearing on the right side of the plot and relatively lower values appearing on the left side. A review of the plot does not seem to suggest very much in the way of such a trend. In fact, a linear regression fit of the data yields a line with slope not significantly different from 0 ($p = 0.99$.) There is a good deal of variation present in the color data. To get a clearer picture of potential trends, it is desirable to remove some of the random variation or noise from the data. Smoothing techniques are one way to reduce or to cancel the effects of random variation, and they can help to reveal underlying trends in the data.

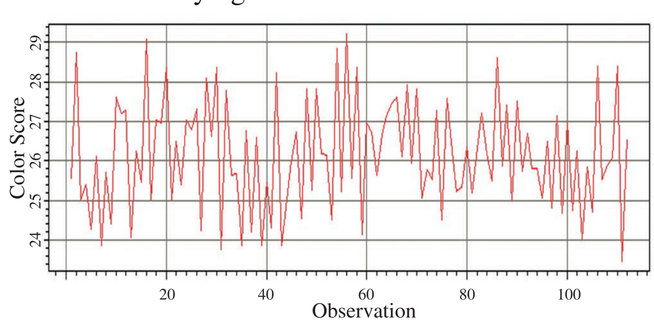

Figure 1. Lemon juice color values plotted against chronological order of observations. Observations were taken twice a day, twelve hours apart. On each day, the first observation corresponds to day shift production and the second to night shift production.

Time Series Analysis: For the purpose of investigating the lemon juice data, you opt to use moving averages that calculate the average over a week's worth of data.

In other words, $k = 14$ since there were two observations per day (day shift and night shift.) The line plot of moving averages, shown in Figure 2, exhibits a smoother graph of the data though it still shows no consistent trend. To further investigate potential trends you next opt to use the autocorrelation, an exploratory time series tool, on the original data. Exploratory analysis of the lemon juice data was performed using autocorrelation on the complete period of the data, observations 1 through 112.

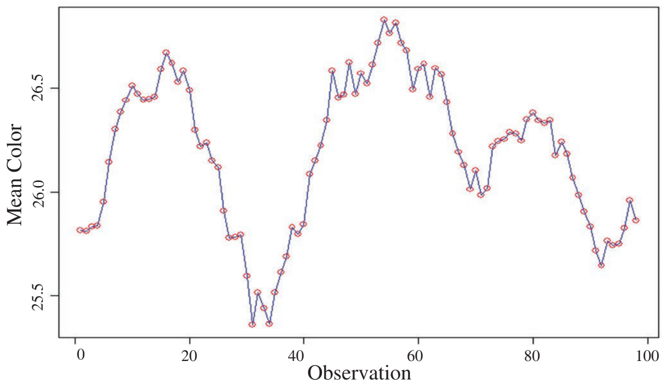

Figure 2. Fourteen-point moving averages of lemon juice color values plotted against chronological order of observations.

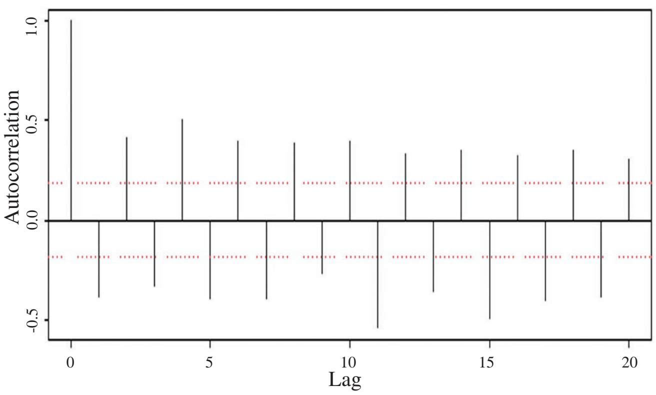

Figure 3. Autocorrelations of lemon juice color values plotted for day 1 through 56.

A plot of these autocorrelations is shown in Figure 3. The heights of the vertical lines shown are the values of the correlation for the given lag on the horizontal axis. The horizontal dotted lines indicate the 95% upper confidence level (UCL) and the 95% lower confidence level (LCL). If the data were completely random, 95% of the autocorrelation values should fall within the UCL and LCL. We see significant positive correlations corresponding to even-valued lag times and significant negative correlations corresponding to odd-valued lag times. This suggests a cycle occurring between the shifts and confirms that there is significant evidence for a trend in the data. The fact that the odd correlations are negative indicates a potential difference between the day shift color and night shift color.

Because significant negative correlations corresponding to odd valued lag times are indicative of a potential issue between shifts, you separate the day shift and night shift data and investigate these separately. Scatter plots of the two shift samples appear random with no obvious trends. You next run a 7-day moving average for each shift, and plot them both chronologically on the same axis. This line plot appears in Figure 4 and provides visual evidence of the periodic nature of the data that was suggested by the autocorrelation. From this you learn that the color values for the night shift samples have a higher mean than the average of the day shift samples. The night-shift mean differs significantly from the day-shift mean ($p < 0.001$), confirming the source of the darkened lemon juice.

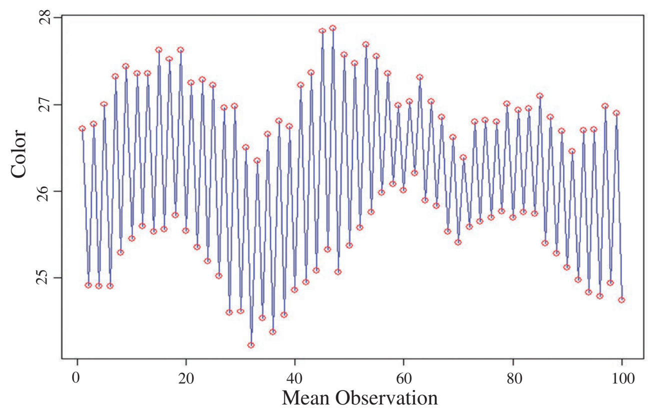

Figure 4. Line plot of 7-day moving averages for day and night shift color, plotted chronologically on the same axis.

As a result of discovering the day/night cycle in the color data, the investigation turns to a mechanical check of pasteurization controls. One of the night shift production thermostats is found to be faulty causing the lemon juice to be heated to temperatures higher than the pasteurization specification, causing some partial oxidation of the citrus juice and subsequent darkening.

Conclusion: Exploratory time series tools are useful in research when product performance and/or quality is time-dependent. Basic descriptive tools such as smoothing processes like the moving average, as well as exploratory tools like autocorrelation can be instrumental in helping to locate and identify trends otherwise hidden in data.

References

1. Chatfield, C. (1996). *The analysis of time series: an introduction*, 5th ed. Chapman & Hall.
2. Anthony, K., Ennis, D., and Cook, P. (1984). Lemon juice color evaluation: sensory and instrument studies. *Journal of Food Science*, **49**(6), 1435-1437.