



Café Data

Concetta A. DePaolo

David F. Robinson

Indiana State University

Journal of Statistics Education Volume 19, Number 1 (2011),

www.amstat.org/publications/jse/v19n1/depaolo.pdf

Copyright © 2011 by Concetta A. DePaolo and David F. Robinson all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Time series; Forecasting; Trend; Decomposition; Regression; Correlation.

Abstract

In this paper we present time series data collected from a café run by business students at a Midwestern public university. The data were collected over a ten-week period during the spring semester of 2010. These data can be used in introductory courses to illustrate basic concepts of time series and forecasting, including trend, seasonality, and the use of time series decomposition. Since the data relate to a student-run enterprise, we believe that statistics students, especially those in business disciplines, will find the application interesting and engaging. In addition to exercises in which students can perform statistical analyses, we also provide several examples in which the data and results are related to the business context, thereby showing the relevance and importance of data-driven business decisions.

1. Introduction

We, like many statistics instructors, prefer to use real and relevant data to interest and motivate our students. This approach has been advocated by many, including the American Statistical Association's GAISE (Guidelines for Assessment and Instruction in Statistics Education) College Report ([2005](#)), which recommended that instructors use real data in teaching. Furthermore, other authors have advocated the connection of data to its context ([Chance, 2002](#)) and the use of data that relates to student interests ([Dargahi-Noubary & Gowney, 1998](#)). In this paper, we describe a dataset that we have compiled that may be especially interesting to business students. It involves sales data from a café run by undergraduate business students to serve clientele in their College of Business.

There are many resources for statistics instructors wishing to use “real” data in the classroom. There are several online data repositories, including the JSE Data Archive (http://www.amstat.org/publications/jse/jse_data_archive.htm), CauseWeb’s dataset archive (<http://www.causeweb.org/>), the Data and Story Library (<http://www.stat.cmu.edu/DASL/>), and the Time Series Data Library (<http://robjhyndman.com/TSDL/>), which as its name implies, has a large number of time series data sets. There are many published articles with real data sets, such as those listed by [Singer and Willett \(1990\)](#). There are also a number of student-run businesses (for example, Flyer Enterprises at the University of Dayton) and some are actually student-run statistical consulting companies who analyze other businesses’ data (see, for example, [Boomer, Rogness, & Jersky, 2007](#); [Gunaratna, Johnson & Stevens, 2007](#)). We believe that this dataset is interesting because it involves data generated from a student-run business, which appears to be an unusual form of “real” data that should appeal to students.

In the remainder of this paper, we discuss the background, development and startup of the student-run café, as well as its current operations. We discuss how the data were collected, describe the dataset, and offer some interesting pedagogical uses. Throughout the discussion, we attempt to highlight the relationship of the data to the business context.

1.1. Background

The College of Business is located in one of the buildings of a two-building complex; the College of Education was originally located in the second building. The campus foodservice vendor operated a submarine sandwich shop in the lobby of the College of Business until 2009, when the College of Education moved to a renovated building nearby that had a new deli-café in its lobby. With the loss of half of the customers in the College of Business/College of Education complex, the campus vendor decided to close the “sub” shop, leaving the College of Business with only vending machines for drinks and snacks. Loss of the sub shop left the lobby of the College of Business with a run-down appearance and a less vibrant atmosphere. Surveys revealed that most people wanted food and drinks to be available in the College of Business but the small numbers of remaining customers made a traditional for-profit café unlikely to survive.

An analysis by a team of ten students in a senior-level business strategy course found that the expenses of running a foodservice in the College of Business were simply too high and the potential sales were too low for a viable for-profit café. The team predicted, based on survey data from students, faculty and staff, that a café could not sell enough product to break-even between expenses and income. If the College of Business was going to have a foodservice, a traditional approach would not work. This realization forced the team to create an innovative café plan.

1.2. The Café Startup

A key point in the mission of the university is to provide experiential and service learning opportunities in which students can apply what they have learned in the classroom to real-world situations. A faculty member in the college saw the opportunity for students to make a contribution to their college and engage in a “hands-on” learning experience by starting a new business. There was also great potential for future students as a café could be used as a learning laboratory for projects from multiple business disciplines. During the fall of 2009, an associate

dean suggested creating a student-run foodservice and encouraged the faculty member to pursue it. The Dean of the college supported the venture and allowed it to go ahead but issued an important challenge to the students, to get the café up and running in just six weeks!

The student team next needed to obtain some investment capital to improve the equipment and the look of the café. A chance comment from one of the team members to the President of the university led to a request from the President, who liked the idea of a student-run business. The team created a short, formal business plan presentation and vetted it with their professor and the Associate Dean before meeting with the President to request \$13,000 for start-up funding. The President challenged the team by offering them a \$6500 no-interest loan payable in three years.

To keep expenses down the team worked closely with the foodservice vendor and the maintenance staff of the university. The student team worked with the campus foodservice vendor to taste the items, create a menu, negotiate a supply agreement, design procedures to order and pickup food, and create an internal accounting and inventory system. The café opened two months later and operated briefly for three days in December before closing for the winter break. The test-run provided proof of concept that a café could be operated by students in the College of Business. The state health inspector licensed the café and it passed an inspection shortly thereafter.

In January of 2010, the café re-opened with a new student team running the operations (see [Figure 1](#) below). The new team was charged with developing the café's strategy, codifying its operations and demonstrating if it was financially and practically viable. That team completed the redecoration of the café space, refined ordering quantities and procedures to minimize waste, and started marketing promotions to boost sales. The students kept café prices as low as possible to provide value to the customers, opting for a mix of moderately priced items compared to other campus food vendors. This pricing strategy made the reduction of waste items crucial to the survival of the café because each unsold item meant a loss of the cost of the item that had to be made up by sales of other items.

Figure 1: The café, Executive Express, during its grand opening in January, 2010. (Photo by Tony Campbell, printed with permission from University Photography Services.)



The key to being economically viable was the decision to staff the café with volunteers from the café consulting team and College of Business student-professional organizations (e.g. from the accounting, marketing, and insurance areas). To encourage the organizations to supply volunteers, the organizations received 10% of gross sales on the days they staffed the café.

By using volunteers, the café team had to adjust café operations to cope with having many different students working in the café. To facilitate using volunteer labor, the students developed a training program and provided supervised on-the-job training. The food handling required was simplified by having the foodservice vendor prepare and package the food. Workers would not need to be “Food Safe Certified” saving training time and certification expenses while it minimized contamination issues.

The café teams kept costs low enough and sales high enough to be able to repay the loan from the President in the Spring of 2010 after just one semester of operations (over two years ahead of schedule). Continuing operation of the café is planned for the fall semester of 2010 and hereafter.

1.3. Current Café Operations and Inventory

The café serves sandwiches, bagels, muffins, cookies, sodas, juices and coffee. The food products are purchased pre-wrapped and labeled for sale from the campus foodservice operator. The café began regular operations during the second week of the spring semester of 2010. Team members and volunteers staff the café from 8am to 2pm Monday through Friday.

The data in this case come from the order sheets that are used to order and receive the prepared food and to record the sales of sodas and coffee. Every afternoon a café team member orders food for pick-up at a campus location for the next day at 10:30 am. The quantity of food items on hand are listed in one column and the number of each item received that day is recorded in

another column; then the number remaining is listed and the total number sold is determined by subtracting the remaining number from the total on hand after the order is received. Any items that have expired cannot be sold and have to be given away to team members or volunteers or thrown away. These waste items are deducted from the total available for sale before the total sold is calculated. During the first week of regular operations in Spring 2010, the café had a great deal of waste but the team quickly adjusted to an ordering pattern to fit the level of sales that the café was experiencing.

2. Data Collection

Data for the café dataset come from three sources.

- 1) Total dollar sales are from the daily sales register tapes from the cash register.
- 2) The number of items sold for each food item is taken from the daily inventory reports filled out by the student workers each day. The daily reports are filled out manually and then summarized into a spreadsheet by the inventory specialist (also a student team member). For coffee and soda, the beginning inventory of cups less the ending inventory (per a daily physical count) determined the number sold. That is, for coffee and soda:

$$\text{Number Sold} = \text{Beginning Inventory of Cups} - \text{Ending Inventory of Cups}$$

Each day, at the beginning of the first shift (at 8 am), a student takes a physical count of all of the food items in the café, as well as the coffee and soda cups. When the new food items arrive for the day (usually 10:30 am) they are counted and recorded by a team member on the inventory sheet. At the end of the day (2 pm) all food items are counted by a team member; any item that expired that day and was not sold is separated out and recorded as waste. The number sold for any item is calculated for each food product, coffee, or soda, by adding the beginning inventory plus the number of items received and subtracting ending inventory and waste items, equaling the physical count that was recorded as ending inventory. That is, for food items:

$$\text{Number Sold} = \text{Beginning Inventory} + \text{Items Received} - \text{Waste} - \text{Ending Inventory}$$

Although this inventory process was somewhat time consuming, it provides good physical control with counts rarely being off due to error. Tight control and a strong sense of mission among the team and volunteers have thus far led to little or no theft. In spite of the controls, there have been occasional discrepancies in the counts or when calculated sales were compared to actual cash register receipts. Errors were possible in the physical counts and in the recording of products received each day but we found that these errors were infrequent and typically isolated to a single item on any single day.

- 3) The daily temperature data were found on an internet-based historical weather site called Weather Underground (2010, <http://www.wunderground.com/history/>). These data were collected because weather was thought to affect sales. The theory formed by staff was

that if it was cold, people were reluctant to leave the building and therefore ate at the café, but if it was warm they went out more and café sales dropped.

3. Description of the Data

The raw data are contained in the file [cafedata.dat](#). This is a tab-delimited data file with 22 variables collected over 48 days. The variables include:

- Time in days ($t = 1, 2, 3, \dots, 48$);
- Date, Day of Week Code, and Day of Week;
- Number of bread sandwiches, wraps, muffins, cookies, fruit cups, chips, juices, sodas, and coffees sold per day;
- Number of bread sandwiches, wraps, muffins, cookies, fruit cups, and total number of items that expired and were thrown away each day;
- Total number of coffees and sodas sold per day;
- Reported daily sales in dollars (excluding taxes and donations);
- Maximum daily temperature reported in degrees Fahrenheit for the city.

(An Excel version of the data file is contained in [cafedata.xls](#).)

Data was collected Monday through Friday from January 19, 2010 to April 1, 2010, with the exception of Friday, March 5. On that day, the Friday before the university's spring break, the café was closed due to an anticipated lack of customers. More detailed information about the data set appears in the [Appendix](#) and the associated documentation file [cafedata_documentation.txt](#).

4. Pedagogical Uses

In this section we provide examples of how the café data can be used to illustrate some basic statistical concepts and techniques. Examples are listed by statistical technique and teaching notes for instructors are included. We begin with basic analyses and conclude with more advanced analyses, at each step illustrating how deficiencies in one method can be addressed by employing the next method. In each section, we include exercises that tie the statistical analyses to the business context.

4.1. Missing Data Values

There is one day (day 34, March 5, 2010) for which sales data is missing because the café was closed. This Friday was the day before Spring break and the café team decided to close due to lack of both staff and anticipated customers. All other sales data are present.

If time series analyses are done with statistical software such as Minitab, these missing values should not cause any issues. For example, Minitab recognizes the value as missing and performs the analyses without the value. On the other hand, if analyses are to be done using hand-entered formulas in a spreadsheet such as Excel, the missing values need to be addressed prior to analysis since formulas cannot refer to empty cells. Also, Excel's built-in functions for time

series, trend, and regression within the Analysis Toolpak do not have the capability to skip or impute missing values.

For beginning students who will be doing calculations themselves or with a spreadsheet, an instructor might wish to cleanse the data before disseminating to students. Among more advanced students, an instructor may wish to initiate a discussion as to how the missing values might effectively be estimated. This discussion might refer to estimation of the missing value through averaging other values, for example, from the Friday before and the Friday after this date. Of course, students may come up with other viable options as well.

4.2. Issues with Using Descriptive Statistics for Time Series Data

When beginning to discuss time series data, we like to illustrate that descriptive statistics can be misleading and can hide a lot of information if time is not considered. Below is a series of simple exercises, along with the associated Minitab output (Figures 2a and 2b), to illustrate that point.

1. Compare and contrast the descriptive statistics for Sodas and Coffees sold per day (ignoring day of the week).
2. How different are the number of Sodas sold on different days of the week? What about the number of Coffees sold by day of the week? What factors could cause such differences? How can this information be used by the café's managers?
3. Graph the time series of both Sodas and Coffees and then compare the trends over time. What do you see?
4. Briefly discuss how the descriptive statistics for these time series could be misleading and could mask important information from café managers and staff.

Figure 2a: Minitab Time Series Plot

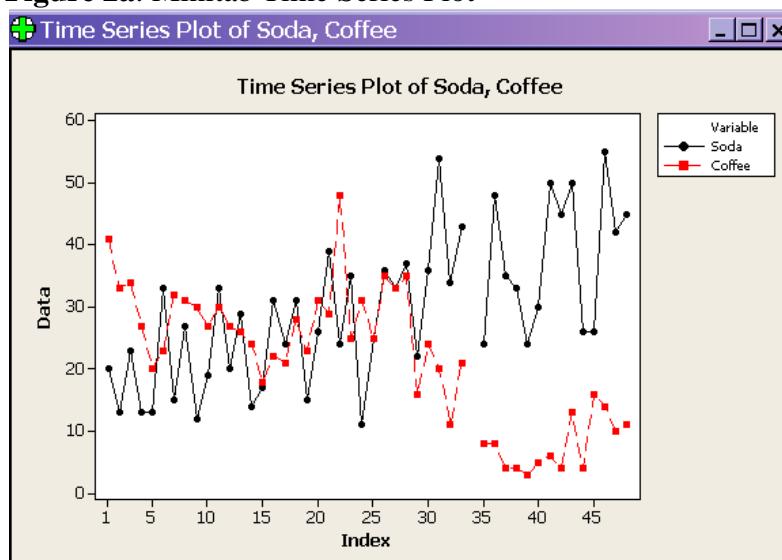


Figure 2b: Minitab Descriptive Statistics Output

Descriptive Statistics: Sodas, Coffees												
Variable N N* Mean SE Mean StDev Minimum Q1 Median Q3 Maximum												
Sodas	47	1	29.57	1.73	11.86	11.00	20.00	29.00	36.00	55.00		
Coffees	47	1	21.51	1.62	11.08	3.00	11.00	23.00	30.00	48.00		
Descriptive Statistics: Sodas, Coffees												
DayofWeek												
Variable	Week	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	
Sodas	Fri	8	1	17.13	2.09	5.91	11.00	12.25	14.50	23.50	26.00	
	Mon	9	0	24.00	2.31	6.93	13.00	18.00	25.00	28.00	36.00	
	Thu	10	0	35.30	2.70	8.54	23.00	28.50	34.00	43.50	50.00	
	Tue	10	0	39.90	3.62	11.45	20.00	32.50	37.50	51.00	55.00	
	Wed	10	0	28.50	3.47	10.97	13.00	18.75	28.50	36.75	45.00	
Coffees	Fri	8	1	19.75	3.91	11.06	3.00	7.00	23.50	29.25	31.00	
	Mon	9	0	19.33	2.88	8.63	5.00	12.00	20.00	26.00	31.00	
	Thu	10	0	22.80	3.29	10.41	4.00	12.50	25.50	31.75	35.00	
	Tue	10	0	22.80	3.59	11.34	6.00	12.50	22.50	31.25	41.00	
	Wed	10	0	22.30	4.66	14.74	4.00	8.50	24.00	33.00	48.00	

Important points to note are that, if time is not considered, the mean number of coffees sold is lower than the mean number of sodas sold, but the variation of the two is very similar. However, when the data are examined by day of the week, we find that coffees sold are a lot more consistent between the days, averaging about 19 to 23, while sodas seem to vary more by day, averaging from about 17 on Fridays to 40 on Tuesdays. These differences are probably due to the fact that there are more classes on Tuesdays and Thursdays than there are on the other days, especially Friday when very few classes take place. This could have implications for the café's managers who need to schedule employees according to how busy they believe the café will be on each day. Lastly, the time series shows that while the number of sodas appears to increase over the course of the semester, coffee sales decrease; thus, the two products that appear so similar with regard to descriptive statistics are actually very different over time. If the café's staff had not analyzed the trend in the data, they might conclude that they need similar quantities of both coffee and sodas, thus eventually having too much of one and not enough of another to satisfy demand.

4.3. Trend Analyses

After a plot with obvious trend is constructed for Coffee and Soda sales, it is reasonable to segue into a trend analysis. We like this data for encouraging students to think about factors that influence trends in sales, both overall and in certain products, and therefore the business aspects of these data. For example, in this case data collection occurred over the course of a spring semester, from January to April, during which time the temperatures varied from 20 to 80 degrees Fahrenheit. We find a relationship between temperature and sales of soda and coffee. In addition, trend exercises can be used to illustrate the need for analysis of seasonality in time series data. Below are some simple correlation and regression exercises, along with the resulting Minitab output (Figures 3a, 3b, and 3c), that can be used to illustrate these ideas related to trend and the business context.

1. Estimate the daily trend in number of sodas sold and in number of coffees sold using regression. What do you notice about the slopes? Are the trends/slopes significant? To what do you attribute the rather low r-squared values?
2. Compare the two models you constructed in #1 above.
3. Consider the time period (January to April) over which these data were collected. What factors could be affecting coffee and soda sales?
4. Examine the correlations between the variables soda, coffee and maximum daily temperature. Are there significant correlations?
5. If sodas and coffees are combined (Total Soda and Coffee variable), is there a trend in the number of drinks sold per day? From a business perspective, use these and earlier results to try to explain what is happening with your coffee and soda sales over the course of the semester.
6. Is there a significant trend in overall café Sales? Comment on the r-squared value and the desirability of using only trend to predict overall sales figures.

Figure 3a: Minitab Trend Output for Soda analysis and Coffee analysis**Regression Analysis: Sodas versus t**

The regression equation is
 Sodas = 16.0 + 0.559 t

47 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	15.995	2.632	6.08	0.000
t	0.55889	0.09398	5.95	0.000

S = 8.97511 R-Sq = 44.0% R-Sq(adj) = 42.8%

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	2848.6	2848.6	35.36	0.000
Residual Error	45	3624.9	80.6		
Total	46	6473.5			

Regression Analysis: Coffees versus t

The regression equation is
 Coffees = 35.4 - 0.570 t

47 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	35.352	2.267	15.60	0.000
t	-0.56965	0.08094	-7.04	0.000

S = 7.72917 R-Sq = 52.4% R-Sq(adj) = 51.3%

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	2959.4	2959.4	49.54	0.000
Residual Error	45	2688.3	59.7		
Total	46	5647.7			

Figure 3b: Minitab Trend Output for Total of Soda and Coffees & for Total Café Sales**Regression Analysis: Total Soda and Coffee versus t**

The regression equation is
 Total Soda and Coffee = 51.3 - 0.011 t

47 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	51.347	3.724	13.79	0.000
t	-0.0108	0.1330	-0.08	0.936

S = 12.6987 R-Sq = 0.0% R-Sq(adj) = 0.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.1	1.1	0.01	0.936
Residual Error	45	7256.6	161.3		
Total	46	7257.7			

Regression Analysis: Sales versus t

The regression equation is
 Sales = 154 - 0.228 t

47 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	153.77	13.45	11.44	0.000
t	-0.2284	0.4801	-0.48	0.637

S = 45.8494 R-Sq = 0.5% R-Sq(adj) = 0.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	476	476	0.23	0.637
Residual Error	45	94597	2102		
Total	46	95073			

Figure 3c: Minitab Correlations Output**Correlations: Sodas, Coffees, Max Daily Temperature (F)**

	Sodas	Coffees
Coffees	-0.402 0.005	
Max Daily Temper	0.530 0.000	-0.741 0.000
Cell Contents: Pearson correlation P-Value		

When these analyses are done, we see that coffee sales decreased as temperatures increased, but soda sales increased, as shown by the following regression equations (see [Figure 3a](#)):

Equation 1: Sodas = 16.0 + 0.559 t
Equation 2: Coffees = 35.4 - 0.570 t

Important points about these analyses include that while soda sales have a positive significant trend ($b_1 = +0.559$, $T = 5.95$, $p < 0.001$), coffee sales have a significant negative trend of almost identical size ($b_1 = -0.570$, $T = -7.04$, $p < 0.001$). The trend analyses for the variable that combines coffee and soda sales and the overall sales figure give the following results (see [Figure 3b](#)):

Equation 3: Total Soda and Coffee = 51.3 - 0.011 t
Equation 4: Sales = 154 - 0.228 t

Interestingly, when coffee and soda sales are combined, there is no significant trend ($b_1 = -0.01$, $T = -0.08$, $p = 0.936$), nor is there a significant trend in the overall sales figures ($b_1 = -0.23$, $T = -0.48$, $p = 0.637$). These results suggest that the decreases in coffee sales are almost entirely offset by the increases in soda sales and that the net effect on overall sales is negligible. The unifying factor here is the temperature. As the weather continues to get warmer, coffee sales decrease and soda sales increase. This can be verified by observing in [Figure 3c](#) the significant correlations between coffees sold and temperature ($r = -0.741$, $p < 0.001$) and sodas sold and temperature ($r = 0.530$, $p < 0.001$). From a business perspective, we might surmise that we are keeping a fairly consistent customer base, but that these customers are switching from hot to cold drinks as the weather becomes warmer. Finally, it is clear from the near-zero r -squared value that the use of only time as a predictor provides extremely poor explanation of variation in sales.

Helpful Hint: Depending on how much experience students have with time series at the point when these data are used, students may need some guidance when addressing the question in #2 above to realize that the seasonality is what causes the relatively low r -squared values (44-52%). Similarly, with regard to the questions in #6, while students will recognize the inability of the time variable to predict sales from the near-zero r -squared value, they may need to examine a time series plot or otherwise be prompted to think about what is causing the variations in sales.

Alternative Application: The variables sodas and coffees are included because they have significant trends. Other variables with significant negative trends (Bread Sandwich Waste, Wrap Waste, and Total Items Wasted) can also be analyzed. These negative trend items might be used to relate to the business context of how the café's managers became better at ordering perishable items over the course of the term, therefore leading to decreased waste. There are also other variables in the dataset that do not show significant trends (Bread sandwiches, Wraps, Muffins, Cookies, Fruit Cups, Chips, Juices, and Sales). Another possible application would be to ask students to examine the business context to conjecture why trends are not present for these items.

4.4. Multiple Regression Analysis with Seasonal Dummy Variables

After a trend-only analysis of the overall Sales receipts variable proves inadequate and students recognize the presence of seasonality within the week and that temperature is a predictor of sales of at least one item, a possible next step in forecasting sales is the use of multiple regression with predictors that include time, days of the week, and possibly temperature. In this approach, we use four dummy variables representing Tuesday, Wednesday, Thursday and Friday. We used Monday as the base case, so it is represented by zeros for all four day-of-the-week dummy variables. Then, each Tuesday is given a value of “1” for the Tuesday dummy variable and a value of “0” for each of the other dummies. Similarly, Wednesdays are indicated by a “1” for the Wednesday dummy variable and a “0” for the other dummies, etc.

Some suggested multiple regression exercises and the associated Minitab results (Figures [4a](#) and [4b](#)) and calculations are as follows:

1. Use a multiple regression model with time and days of the week to predict sales. Remember that you will need to substitute several dummy variables (how many?) instead of using the DayofWeek variable as it currently appears in the data set. How good is this model?
2. Use your model to predict Sales for $t = 49$ through 53.
3. How much more variation in Sales can be explained if the Max Daily Temperature variable is included?
4. If you were a café manager, which model would you prefer? How would you use this information to better manage the business?

Figure 4a: Minitab Multiple Regression Output for Time and Day-of-the-Week

Regression Analysis: Sales versus t, Tuesday, ...

```

The regression equation is
Sales = 129 - 0.279 t + 70.8 Tuesday + 22.7 Wednesday + 51.1 Thursday - 29.9 Friday

47 cases used, 1 cases contain missing values

Predictor      Coef    SE Coef      T      P
Constant      129.33   12.75  10.14  0.000
t            -0.2794   0.3154  -0.89  0.381
Tuesday       70.83    13.81   5.13  0.000
Wednesday     22.70    13.81   1.64  0.108
Thursday      51.10    13.81   3.70  0.001
Friday        -29.91   14.62  -2.05  0.047

S = 30.0496    R-Sq = 61.1%    R-Sq(adj) = 56.3%

Analysis of Variance
Source        DF      SS      MS      F      P
Regression     5  58051  11610  12.86  0.000
Residual Error 41  37022    903
Total          46  95073

```

Figure 4b: Minitab Multiple Regression Output for Time, Temperature, and Day-of-the-Week**Regression Analysis: Sales versus t, Tuesday, ...**

The regression equation is
 Sales = 144 + 0.269 t + 72.9 Tuesday + 25.3 Wednesday + 54.7 Thursday
 - 29.1 Friday - 0.706 Max Daily Temperature (F)

47 cases used, 1 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	143.62	15.83	9.07	0.000
t	0.2686	0.4828	0.56	0.581
Tuesday	72.95	13.69	5.33	0.000
Wednesday	25.25	13.72	1.84	0.073
Thursday	54.66	13.82	3.96	0.000
Friday	-29.06	14.42	-2.02	0.051
Max Daily Temperature (F)	-0.7055	0.4755	-1.48	0.146

S = 29.6187 R-Sq = 63.1% R-Sq(adj) = 57.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	59982.4	9997.1	11.40	0.000
Residual Error	40	35090.8	877.3		
Total	46	95073.2			

The multiple regression equations to predict sales based on time (t) and seasonality (day of the week) can be summarized as follows:

Equation 5:

$$\text{Sales} = 129 - 0.279 t + 70.8 \text{ Tuesday} + 22.7 \text{ Wednesday} + 51.1 \text{ Thursday} - 29.9 \text{ Friday}$$

In this equation, the coefficient of the time variable, t, is -0.279, which indicates a slight downward trend in overall sales. The coefficients of the day-of-the-week dummy variables represent the average amount that each weekday differs from the base day. For example, the coefficient of Tuesday is 70.8, indicating that, on average, Tuesday sales are about \$70.80 more than Monday sales. Using only time and the four day of the week dummy variables explains about 61% of variation in sales, suggesting that this model is fairly good.

The predictions for the next week can be calculated as follows:

t = 49 (Friday),	Sales = 129 - 0.279(49) - 29.9	= \$85.43
t = 50 (Monday),	Sales = 129 - 0.279(50)	= \$115.05
t = 51 (Tuesday),	Sales = 129 - 0.279(51) + 70.8	= \$185.57
t = 52 (Wednesday),	Sales = 129 - 0.279(52) + 22.7	= \$137.19
t = 53 (Thursday),	Sales = 129 - 0.279(53) + 51.1	= \$165.31

The model shown in [Figure 4b](#) is the same except that the temperature variable is added as a predictor. For that model, the regression equation is as follows:

Equation 6:

$$\text{Sales} = 144 + 0.269 t + 72.9 \text{ Tuesday} + 25.3 \text{ Wednesday} + 54.7 \text{ Thursday} \\ - 29.1 \text{ Friday} - 0.706 \text{ Max Daily Temperature (F)}$$

As the negative coefficient of temperature indicates, sales tend to decrease as the temperature increases, a finding that supports anecdotal evidence from café managers and workers, who feel that business is very good on cold days when people seem less willing to leave the building in search of food and drinks. However, upon examining the regression output, we find that the temperature variable is not statistically significant ($T = -1.48$, $p = 0.146$); thus, this analysis shows that, while temperature is significant in explaining coffees and sodas sold, it does not appear to have a significant impact on overall sales.

Potential Pitfalls: Students often have trouble deciding the correct number of dummy variables to include in the model. Also, if they choose a different day of the week for the reference category, their output will obviously vary from what is shown here. For example, if Friday were used as the base day and the dummy variables used represented Monday through Thursday, then the regression coefficients for time (t) and temperature will remain unchanged, but the coefficients of the day-of-the-week dummy variables will all increase by 29.06, thus adjusting the coefficients to use Friday as the “0” case rather than Monday.

Helpful Hint: At this point, it is helpful to discuss with students the limitations of using regression for such an analysis. First, it is dangerous to predict for values of independent variables outside the observed range of values. This, of course, would apply to forecasts for future time periods ($t > 48$) and might also apply to temperatures as well. We often use a contrived example (though not necessarily using these data) to illustrate this point. Also problematic is the assumption of regression that the errors are independent. It is fairly easy to convince students with an example of a cyclical time series that errors are often correlated with one another. After this discussion, an introduction to time series analysis and the forecasting of data with both trend and seasonality is natural.

4.5. Time Series Decomposition/Seasonally Adjusted Forecasts

Once students have observed that trend alone may not explain a large amount of variation in the time series and that using dummy variables in a multiple regression model has its limitations, then time series decomposition is a logical next step. Decomposition is a method for analyzing multiplicative time series and calculating seasonally adjusted forecasts. Observations in a multiplicative time series are assumed to be the product of four components: trend (long term growth or decline), seasonality (periodic increases or decreases occurring within a year), cyclical behavior (periodic increases or decreases occurring over multiple years, such as those associated with economic conditions), and irregularity (the random or error component). This method is commonly covered in introductory business statistics textbooks (e.g. [Anderson, Sweeney and Williams, 2011](#); [Groebner, Shannon, Fry and Smith, 2011](#); [Bowerman, O'Connell and Murphree, 2009](#); [Aczel and Sounderpandian, 2009](#)).

Decomposition (as it is called in Minitab) involves “decomposing” a multiplicative time series into its four components as follows. First, the seasonal components or indices are found by what is called the Ratio-to-Moving-Average Method. These indices are multipliers that represent how much larger or smaller observations in a given period (e.g. day, month, quarter) are compared to

the baseline measure. For example, a seasonal index of 1.4 would represent a period that is 40% higher than the baseline measurement. Then, the time series is de-seasonalized by dividing each observation by the seasonal index for that period, thereby removing what seasonality is present. Next, a linear trend component is calculated for the de-seasonalized data and a trend-only forecast is computed for future periods using the regression equation. Lastly, the trend-only (or unadjusted) forecasts are adjusted by multiplying each unadjusted forecast by the appropriate seasonal index. While this process is a bit complex when done in a spreadsheet, it is done automatically within some statistical software packages. (For example, within Minitab, see Stat > Time Series > Decomposition.)

Emphasizing the business applications, one may wish to forecast sales for the coming week(s) using both trend and seasonality representing the days of the week, as the following exercises and Minitab output ([Figure 5](#)) illustrate.

1. Develop a time series decomposition model to forecast Sales using trend and seasonality for days of the week.
2. What is the trend component? Are total sales increasing or decreasing over time?
3. Is there seasonality within a week? If so, indicate the seasonal indices for each day (MTWRF). Which days appear to have higher or lower sales? (Note that data, $t = 1$, starts on a Tuesday so this should be indicated in the model that the first observation corresponds to the 2nd period during the week.)
4. Calculate the seasonally adjusted forecasts for Sales for one week into the future. How accurate do you think these forecasts might be? How might this information be used by the café's managers?
5. Compare your results to the forecasts you generated using multiple regression with day-of-the-week dummy variables.

Figure 5: Time Series Decomposition Minitab Output

Time Series Decomposition for Sales	
Multiplicative Model	
Data Sales	
Length	48
NMissing	1
Fitted Trend Equation	
Yt = 158.08 - 0.465231*t	
Seasonal Indices	
Period	Index
1	0.84971
2	1.26086
3	0.99401
4	1.31980
5	0.57562
Accuracy Measures	
MAPE	16.786
MAD	22.743
MSD	922.034
Forecasts	
Period	Forecast
49	77.870
50	114.555
51	169.397
52	133.083
53	176.088

The Minitab outputs the linear trend equation that indicates a slight downward trend in sales over the course of the term:

$$\text{Equation 7: } Yt = 158.08 - 0.465231*t$$

Next, the output gives the seasonal indices, numbered “1” through the number of periods in the data. For these data, which are measured daily for five days per week, there are five seasonal indices: 1 = Monday, 2 = Tuesday, 3 = Wednesday, 4 = Thursday, 5 = Friday. In the output shown in [Figure 5](#), it was specified under the Minitab “Options” that the first observation occurred on a Tuesday, which corresponds to the second “period” during the 5-day week. If this adjustment is not made, then the seasonal indices will correspond to 1 = Tuesdays (the day corresponding to $t = 1$). The seasonal indices indicate that there is quite a bit of seasonality within a week, with Fridays having about half of the sales of Wednesdays, and Tuesdays and Thursdays being about 25-30% higher than Wednesdays.

Accuracy measures given by the output include MAPE (Mean Absolute Percent Error), MAD (Mean Absolute Deviation) and MSD (Mean Square Deviation). In this analysis, MAPE suggests that these forecasts have a mean absolute percent error of about 17%, which is somewhat accurate. MAD suggests that the average error is about \$23. Managers might use this information

to help with staffing or forecasting, for example, when they might be able to pay back the café's startup loan.

The final portion of the output shows the forecasts for the next five days. Each forecast (F_t) is calculated by multiplying the forecasted trend component (T_t) by the appropriate seasonal index (S_t). For example, for $t = 49$, which corresponds to a Friday (period 5), we have:

$$\begin{aligned} F49 &= T49 * S49 \\ F49 &= (158.08 - 0.465231*49) * S5 \\ F49 &= 135.28 * 0.57562 = 77.870 \end{aligned}$$

The analysis also predicts sales of between \$115 to about \$176 for the following four days (Monday through Thursday). Note that the forecasts generated by this model are very close to those generated by the multiple regression model with day-of-the-week dummy variables.

Alternate Application: Though we have used sales here, students may be asked to compute seasonally adjusted forecasts for any of the individual items. The business application can be motivated by the need to have accurate order quantities, especially on the perishable items such as fruit cups and sandwiches.

5. Conclusions

In this paper, we present data collected from a café run by undergraduate business students. The data allow us to illustrate several statistical concepts related to time series and forecasting. We hope that other instructors, especially those teaching business statistics, will find this data relevant and interesting for students. We believe that the inclusion of data collected by a student-run business will engage other students more heavily with its analysis, and will seem appealing to undergraduate students. We also hope that the business context we provide is seen as accessible to students in introductory courses, and that the accessibility serves to motivate students and illustrate the importance of making data-driven business decisions.

Appendix: Key to Variables in Café Data File

For the file cafedata.dat (saved as tab delimited text)

Variable	Description	Label
1	Time (in days)	t
2	Date	Date
3	Day of Week Code (1 = Mon, 2 = Tue, 3 = Wed, 4 = Thur, 5 = Fri)	Day Code
4	Day of Week	Day of Week
5	Number of bread sandwiches sold	Bread Sand Sold
6	Number of bread sandwiches expired and thrown away	Bread Sand Waste
7	Number of wraps & croissant sandwiches sold	Wraps Sold
8	Number of wrap & croissant sandwiches expired and thrown away	Wraps Waste
9	Number of muffins sold	Muffins Sold
10	Number of muffins expired and thrown away	Muffins Waste
11	Number of cookies sold	Cookies Sold
12	Number of cookies expired and thrown away	Cookies Waste
13	Number of fruit cups sold	Fruit Cup Sold
14	Number of fruit cups expired and thrown away	Fruit Cup Waste
15	Number of bags of chips sold	Chips
16	Number of bottles of juice sold	Juices
17	Number of cups of soda sold	Sodas
18	Number of cups of coffee sold	Coffees
19	Total number of cups of soda plus coffee sold	Total Soda and Coffee
20	Reported daily sales, not including tax and donations	Sales
21	Maximum daily temperature reported in the city	Max Daily Temperature (F)
22	Total number of items expired and thrown away	Total Items Wasted

Day 34 (03/05/10) has missing data for all sale items because the café was closed.

Acknowledgement

The authors wish to thank Dr. Connie McLaren for her helpful suggestions in the preparation of this manuscript.

References

Aczel, A.D. & Sounderpandian, J. (2009). *Complete Business Statistics* (7th ed.). New York: McGraw-Hill/Irwin.

Anderson, D.R., Sweeney, D.J. & Williams, T.A. (2011). *Statistics for Business and Economics*, (11th ed.). Mason, OH: South-Western.

Boomer, K.B., Rogness, N. & Jersky, B. (2007). Statistical Consulting Courses for Undergraduates: Fortune or Folly? *Journal of Statistics Education*, 15(3), <http://www.amstat.org/publications/jse/v15n3/boomer.html>

Bowerman, B.L., O'Connell, R.T. & Murphree, E.S. (2009). *Business Statistics in Practice* (5th ed.). New York: McGraw-Hill/Irwin.

Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education* , 10 (3), www.amstat.org/publications/jse/v10n3/chance.html.

Dargahi-Noubary, G., & Gowney, J. A. (1998). Risk-A Motivating Theme for an Introductory Statistics Course. *The American Statistician* , 52 (1), 44-48.

GAISE. (2005). Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report. Retrieved May 14, 2010, from The American Statistical Association (ASA): <http://www.amstat.org/education/gaise/GAISECollege.htm>.

Groebner, D.F., Shannon, P.W., Fry, P.C. & Smith, K.D. (2011). *Business Statistics: A Decision-Making Approach* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.

Gunaratna, N.S., Johnson, C.A. & Stevens, J.R. (2007). Service-Learning for Graduate Students through a Student-run Consulting Program. *Journal of Statistics Education*, 15(2), <http://www.amstat.org/publications/jse/v15n2/gunaratna.html>

Singer, J.D. & Willett, J.B. (1990). Improving the Teaching of Applied Statistics: Putting the Data Back Into Data Analysis. *The American Statistician*, 44(3), 223-230.

Weather Underground (2010). History & Data Archive, Retrieved June 10, 2010 from <http://www.wunderground.com/history/>.

Concetta A. DePaolo
Associate Professor of Operations Management & Analysis
Scott College of Business
Indiana State University
Terre Haute, IN 47809
cdepaolo@indstate.edu
Phone: 812-237-2283
Fax: 812-237-8133

David F. Robinson
Assistant Professor of Management
Scott College of Business
Indiana State University
Terre Haute, IN 47809
drobinson15@indstate.edu
Phone: 812-237-8829
Fax: 812-237-8133

[Volume 19 \(2011\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Resources](#) | [Editorial Board](#) |
[Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Guidelines for Readers/Data
Users](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)