Module 6 - Orange Juice Sales and Prices

Background

- In this module, you will be looking at sales and price data for orange juice in grocery stores. You have data from 83 stores on three brands (Tropicana, Minute Maid, and the store brand, Dominick's) over about two years. Using this data, you will build a model to estimate the quantity that will sell at a specified price point and assess what price point will maximize profit.
- The Excel workbook for this module consists of 1 tab.
    1. The "Filtered" tab contains all the data, but it is filtered to show only data for Tropicana at Store 5. It also includes some additional calculations. Note that if you change the filters different rows of the data are displayed and hidden. If you are unfamiliar with working with filters, first view the screencast *Working with Filters*.

Part 0: Linear Models

Linear equations are very simple equations, but they can be quite powerful at modeling real-world phenomenon. Once a model is created, it can be used to make predictions.

1) To begin this Module, view the screencast *Linear Models*.

Part 1: Price vs. Quantity Scatter Plots and Correlations

1) Tropicana at Store 5

Here is a scatter plot for the price and quantity pairs for **Tropicana at Store 5** for the 116 weeks of data available**.** A (modified) version of this scatter plot shows on the **Filtered** tab**.**

What is the **correlation** between **price and quantity** for **Tropicana at Store 5**?
(If you need a review on correlation, view the screencast *Correlation Coefficient*.)

2) Minute Maid at Store 5

   a. Create a similar scatter plot that shows the price and quantity pairs for **Minute Maid at Store 5**.

   b. What is the **correlation** between **price and quantity** for **Minute Maid at Store 5**?

Part 2: Estimating the Price and Quantity Relationship for Tropicana at Store 5

Above we looked at the scatter plot of price and quantity for Tropicana in Store 5. In the chart below, you see a line that passes through the cloud of points. About half the points are above the line and about half of them are below it.

The line above intersects the horizontal axis at ($3.4375, 0). In other words, on that line, for a price of $3.4375, the quantity estimated by the line is 0.

1) Looking at the **graph above** and using the data in the spreadsheet, find the equation of this line. (Note that the initial line appearing in the Excel workbook **may not** be the same as the line shown above.)

Hint: you can find the equation of a line if you know two points on the line. One of the points on this line is ($3.4375, 0). What is another one? If you move your curser over the data points in the graph on your spreadsheet, you can see the exact values.

2) There is a graph with a scatter plot of price versus quantity combined with the graph of an equation of quantity as a function of price.

   a. If you type in 110,000 for the Intercept (b) and -32,000 for the Slope (m) under the linear model, what is the value for the quantity estimated for a price of $3.00?

   b. Adjust the intercept (b) and slope (m) values in the spreadsheet until you have what looks to you from your graph like a "best" fit. What are your intercept (b) and slope (m) values?
   b =
   m =

   c. Notice that there is a column for "Estimated Quantity" and one for "Squared Difference."
      - The *Estimated Quantity* is the y-value of the line (y=mx+b) where the m and the b are in cells on the spreadsheet, and the x is the price value for that row.
      - To find the *Squared Difference*, subtract the Quantity Sold (actual value from the data) from the Estimated Quantity (predicted by the line). Then, square that difference.

      Looking at the graph, how would you describe what Squared Difference represents?

   d. There is also a cell on that tab that shows Total Squared Difference (i.e., the sum of the values in the Squared Difference column). The lower that total is, the better the fit of the line to the data. Find the Total Squared Difference when the intercept is 100,000 and the slope is -30,000?

   e. Find the Total Squared Difference when the intercept is 100,000 and the slope is -40,000?

   f. Find the values of the intercept and slope that make the Total Squared Difference as low as possible. (Hint: Try using the Solver tool in Excel. This is an add-in and you may need to add it.) If you are unfamiliar with the Solver Tool, first view the screencast ***SOLVER***.

b=
m=

Part 3: Estimating the Price and Quantity Relationship for Minute Maid at Store 5

1) Answer the following questions about Minute Maid at Store 5:
    a. What was the price of Minute Maid in Store 5 in week 53?

    b. What was the quantity sold of Minute Maid in Store 5 in week 68?

2) Determine the equation of the linear model that passes through the (price, quantity) pair for Minute Maid at Store 5 in Weeks 58 and 60. What is the Total Squared Difference for all weeks for Minute Maid at Store 5 with this specific slope and intercept?

3) Find the values for the intercept (b) and the slope (m) that make the Total Squared Difference as low as possible.

    b=
    m=

4) The steps above are the building blocks for running a simple regression, which is procedure for finding the best fit, y, as a linear function of x by minimizing the sum of squared differences between the points and the line. There are (at least) two other ways to find that best fit line in Excel. See if you can get them to replicate your optimal m and b.

    a. Look for an option to add a Trendline to your scatter plot. There should also be a formatting option for the Trendline to display to equation of the line, if it doesn't display by default.

    b. The SLOPE() and INTERCEPT() functions can by used to determine the slope and intercept of the line of best fit. Excel also contains a function for a linear regression, called LINEST() ("linear estimate"). In this case, you want to regress Quantity Sold on Price. (You will almost certainly have to look online for help on using the LINEST function.)

Part 4: Price and Profit

You will calculate the store's actual profit (Actual profit = (price – cost) * actual quantity sold) for each product for each week.

1) For **Tropicana at Store 5,** assume that Tropicana costs the store $0.90:
   a) Create a scatter plot of price (on the horizontal axis) and actual profit (on the vertical axis) for all weeks.

   b) What was the price during the week with the highest profit?

2) For **Minute Maid at Store 5,** assume that Minute Maid costs the store $0.80:

   a) Create a scatter plot of price (on the horizontal axis) and actual profit (on the vertical axis) for all weeks.

   b) What was the price during the week with the highest profit?

3) For **Tropicana at Store 5**,

   a) Using the cost for Tropicana listed above, calculate estimated profit for price levels of $1.00, $1.50, $2.00, …, $4.00.

      Estimated profit = (price – cost) * estimated quantity.

      For estimated quantity, use the quantity predicted by the best-fit line from the end of Part 2. These come from the m and b values that minimized Total Squared Difference. What is the estimated profit for a price of $3.00?

   b) Write a function (by hand) for the weekly estimated profit f(x) where the price is x dollars.

   c) Create a graph that has both the data for actual profit in a scatter plot and the curve for estimated profit.

d) Looking at the curve, what was the price associated with the highest estimated profit? (Note that this not a precise method, but the visual tool of the graph is a very quick method for approximating the optimal price point for the model.)

4) For **Minute Maid at Store 5**,

a) Using the cost for minute maid listed above, calculate the estimated profit for price levels of $1.00, $1.50, $2.00, …, $4.00.

Estimated profit = (price – cost) * estimated quantity.

For estimated quantity, use the quantity predicted by the best-fit line from the end of Part 3. What is the estimated profit for a price of $3.00?

b) Write a function (by hand) for the weekly estimated profit f(x) where the price is x dollars.

c) Create a graph that has both the data for actual profit in a scatter plot and the curve for estimated profit.

d) Looking at the curve, what was the price associated with the highest estimated profit?

Selected Answers

Part 1

1) -0.6716
2b) For Minute Maid at Store 5, correlation is negative: -0.5403.

Part 2

1) Using the point ($2.39, 32449) which also falls on the line, an equation is y = -30978 x + 106485

2a) At $3.00, the estimated quantity is 14,000.

2d) Total Squared Difference = 19,251,134,076 for intercept = 100,000 and slope = -30,000

2e) Total Squared Difference = 141,308,445,076 for intercept = 100,000 and slope = -40,000

2f) The Total Squared Difference is minimized when the intercept is approximately 66,441 and the slope is approximately -17,951.

Part 3

1a) The price of Minute Maid in Week 53 at Store 5 is $2.19

1b) The quantity of Minute Maid sold in Week 68 at Store 5 is 10,049.

2) The intercept is approximately 88,104.31 and the slope is approximately -29,897.14. The Total Square Difference for this linear model is approximately 29,452,839,689 .

3) The Total Squared Difference is minimized when the intercept is approximately 74,801 and the slope is approximately -25,191 (with Total Square Difference approximately 28,280,498,834).

Part 4

1b) The price during the week with the highest profit is $2.49. (Profit is $146,332; Week is 143.)

2b) The price during the week with the highest profit is $1.99. (Profit is $118,963; Week is 128.)

3a) At $3.00 the estimated profit $26,438.

3b) Using the approximate values from Part 2 Question2f, the function is
f(x)=(x-0.90)(-17951x+66441)

3d) With the given data points, the high point of the curve is 2.50. If we consider a smaller step size of 0.25, then the high point of the curve is at about $2.25.

4a) At $3.00 the estimated profit is approximately -$1,702.

4b) Using the approximate values from Part 3, f(x)=(x-0.80)(-25191x+74801)

4d) The high point of the curve is at about $2.00.

The creation of this work "Mathematical Analysis in Business" was supported by Open CU Boulder 2020-2021, a grant funded by the Colorado Department of Higher Education with additional support from the CU Office of the President, CU Office of Academic Affairs, CU Boulder Office of the Provost, and CU Boulder University Libraries.