

Blog Post Module 2 Final Project:
Andrew Wester
Monday, December 17, 2018

In our Module 2 project, I had a lot of difficulty at first figuring out how best to approach this assignment. We were presented with the Northwind Database from Microsoft, and asked to create 4 scientific hypotheses, with the first being an examination of whether or not discounting products has a statistically significant effect on the quantity of a product in an order. From there, we needed to find at least three other hypotheses to investigate and report on. In going through this project, I found that it was most difficult to scientifically approach each hypothesis, by examining the question, determining the important factors to look at, and then identifying the appropriate methods to statistically examine the results, in order to determine whether there was a statistically significant effect to reject our null hypothesis and support our alternative hypothesis.

As I slowly got further into the project, and moved past investigation and into modeling and analyzing the data, the methods began to come together a bit smoother. When approaching the first hypothesis, as well as every other hypothesis, I set an alpha value of 0.05, indicating a 95% confidence value for my experiment. As I investigated the effects of discount on sales, I ran a Welch's T-Test after noticing a disparity between variance values for my data that was discounted and data that was not, and I found that there was a p-value extremely close to zero, and well below the pre-set 5% confidence value. This low p-value allowed me to reject the null hypothesis that discount had no effect on sales, and that there was, in fact, a statistically significant effect on quantity of items sold with a discount vs without. From there, I used the *Cohen's d* calculation to look at the effect size of the difference, and found the effect size to be 0.286, which is deemed to be a medium effect size. This tells us that while the effect of discounting products will, in fact, help sell more product, it is also not a statistically great enough event that we should revert to this strategy on a regular basis.

For my second Hypothesis, I wanted to examine the top performing employee, and see where their performance differed from other employees, whether other employees may have fewer customers to sell to, or have other responsibilities in their jobs, or some other factor, I was curious to see what effects could be found. As I started digging into this data, I looked at each employee's sales, in terms of the quantity of products sold, the average price of product they sold (were they selling more expensive or cheaper products), and other factors including the number of customers each employee sold to, and what their title was in the organization (VP of Sales might have more responsibilities beyond purely sales, while a newly hired sales representative with fewer customers might be more invested in developing a relationship with their customers before pushing for more sales).

From examining this data, my original hypothesis was that the top selling employee would have higher per-order sales than the average employee, and that this was where the effect of their sales would come in. However, after examining this data, I actually found that the top employee, while having greater total sales than the average employee by a significant margin, was actually selling at a lower per-order value than the average employee. This put my hypothesis down right away, and made me reconsider my hypothesis. I reconsidered the problem, and instead looked at total sales, and saw a statistically significant difference in the total revenue produced by that top employee, compared with the average. I followed this up by also briefly examining the number of clients the top sales representative worked with, and found that they sold to the highest number of customers, at 79 accounts, while the average sales representative sold to only 84 customers. This is a significant gap that gives that top employee a range of customers to increase the yearly sales. This is effective knowledge for the

company, as this employee may simply hustle a lot harder, or perhaps has clients that have been regular purchasers over the years, and simply know how much they need on a yearly basis, so it is easier for the company to hand a large number of clients to a well organized employee and recognize that it will simply take good organization and time management to handle the large amount of orders throughout the course of the year.

My third hypothesis was to examine the effects of regional sales throughout the company, Northwind has lots of territory to cover, and only 9 sales employees listed in the database. I started by trying to categorize the sales into the regions listed as Northern, Southern, Eastern, and Western, but then found that the orders only had the information listed as Northern, Central, and Southern America, Northern, Eastern, Southern, and Western Europe, along with the British Isles and Scandinavia. I decided to examine each of the territories listed as its own entity and look at the information for each. From these groups, I categorized them and looked at the sales information for each region to determine where the majority of sales occurred and whether it was due to the number of orders or the average order price, as well as how much discount was applied regionally. The table is listed below.

	Region	TotalOrders	TotalSales	TotalDiscountSales	AverageOrder	AverageDiscountOrder
1	Western Europe	276	537690	505164	1948.15	1830.31
0	North America	152	318901	295781	2098.03	1945.93
2	South America	145	183902	171856	1268.29	1185.21
4	British Isles	75	117934	108951	1572.45	1452.68
5	Northern Europe	55	94305.9	87156.2	1714.65	1584.66
3	Southern Europe	64	48605.7	45225.7	759.46	706.65
7	Scandinavia	28	25513.6	24545.2	911.2	876.61
8	Central America	28	24073.4	23582.1	859.77	842.22
6	Eastern Europe	7	3531.95	3531.95	504.56	504.56

From this information, I chose to examine the North American information relative to overall values because there were fewer total orders placed than Western Europe, yielding a greater Average Order value. I again ran an independent T-Test, and found another low p-value below the alpha value of 0.05. The effect size this time was 0.225, which is again a medium effect size, and this tells us that while not a huge effect, we can examine this information for future forecasting if the company is interested in boosting overall revenue, by perhaps targeting more customers in North America, or even investing in another employee to focus on the North American region.

The final hypothesis i chose to investigate was whether the top 10% of items with the highest unit price produced different sales values, specifically quantity of items sold, relative to the bottom 10% of items (those that had the lowest unit prices). I figured that total sales for the higher priced items would likely be higher than the lower priced products, since an item priced at \$10 would have to sell more than 35 extra units than a product priced at \$350. However, I was curious to see if there would be any statistical significance between the quantity of items sold that were at the bottom of the pricing ladder compared with those at the top. From investigating this, I found that there was, in fact, little difference between the quantity sold of highest and lowest priced products. There were more units sold of the lowest

priced units than the highest priced units by approximately 12%. This easily confirmed our null hypothesis that there was no difference between highest and lowest priced items in terms of quantity sold. However, upon further examination, I noticed that although there were more units sold of the lowest priced items, the overall revenue produced from the highest priced items was 1,000% higher than the lowest priced items. The highest priced items produced over \$473,000 in revenue, and the lowest priced items produced not even \$40,000 in revenue.

From these above hypotheses, I was able to draw to some conclusions, and with extra time, knowledge, and resources, I feel that I could develop even greater conclusions. However, from what I was able to ascertain from these investigations, there are a few things that this company could take the time to examine or dig deeper into.

1. The first is the effectiveness of the top sales employee, and why they are able to produce so much more revenue at such a greater rate than everyone else. While there will always be situations of an employee hustling harder than others, or another employee perhaps having more responsibilities as a manager or executive that requires extra time away from selling that might impact their final performance metrics. It could still be worth examining the effectiveness that this employee was able to sell so much more than the rest of the employees.
2. Examining the value of adding more representation in North America, and determining why the sales in that region were performing so much better than in Europe and other regions. We saw that Western Europe had a larger number of orders, total sales, and even discounted amount of sales than north America, but on a per-sale basis, North America performed better than Europe did. Perhaps the employees selling in North America do a better job, or the clients are simply purchasing more product on a less frequent basis. It could be something beneficial to examine further to determine whether investing more resources to sales in North America would be beneficial to the company.
3. Finally, spending extra time examining the effects of discounting costs, and looking at the pricing of products would be beneficial to potentially increase sales for the company. Perhaps by examining trends of when sales occurred and what products were sold at what time of the year, the company can better anticipate how much inventory they will need ahead of time. Also, as a food supply company, it is extremely beneficial to the company to recognize the lifespan of the products in order to understand how long a product is able to sit on a shelf before it needs to be sold or thrown away. The food and beverage industry has a difficult time with that aspect of the business as food that goes bad is wasted money on the part of the company.

These would be my recommendations to the Northwind company, and I believe with extra time and resources, all of these recommendations could be examined in more depth, and with greater detail as to provide a more concrete answer to the company and recommendations and strategies for moving forwards into the future.