Category of Items Which Grosses the Highest Amount of Sales

Ashley V. Hester

Grand Canyon University, USA

AHester8@my.gcu.edu Ashley.vhester@gmail.com

Abstract

The study focused on category of items which grosses the highest amount of sales at my husband's family's small business- Crown Liquor. I want to determine this by selecting data from 30 different sales orders. My second question is related to the ATM fee located on the sales orders. I want to consider the probability that a customer will most likely pay cash or credit (use their debit/ATM card). Also, at what total sale amount is the customer more likely to use their credit or debit card? Based on the small sample, or part of the population of interest, I will use inferential statistics to estimate a property of a population based on a sample (Lind et al., 2018). When I can evaluate the 30 sales orders as a sample of the entire population, I can assume that the same probabilities occur in both settings. The likelihood of a customer using their debit/ATM card or credit card will help my husband's family decide on the fee amount. Is the $0.50USD ATM fee too small, too much, or is it not necessary? The answer will depend on the number of customers who use the credit/debit ATM service, how frequently, and for how much. The purpose of this statistical study is aimed at reducing the costs. By applying statistical inference approaches to data, we can aim to reduce the costs (Mansour & Aboshady, 2022).

Keywords*: inferential statistics, probability, observations of frequency*

Introduction

According to Lindt et al., (2018), "When an object or individual is observed and recorded as a nonnumeric characteristic, it is a qualitative variable or an attribute." By organizing the data on the sales orders, I can use the qualitative data (sales items) and place them into categories (type of sales items). By making observations of qualitative variables that I measured and recorded as labels or names; I can create a nominal level of measurement. Data recorded at the nominal level of measurement is represented as labels or names, contains no particular order, and is only classified and counted (Lind et al., 2018). By placing the qualitative variables into certain classifications, I can use my qualitative data (sales items) and turn it into something that I could count or quantify. The rest of my variables are quantitative; total amount (plus CRV, and 7.75% tax[1] when applicable), ATM fee, and sales order total.

I can further develop my analysis based on observations of frequency. Whichever items are bought most frequently would have the most entrants. I was able to place each item on the 30 sales orders into a separate category: 1-grocery, 2-beer, 3-liquor, 4-lotto, 5-tobacco, 6-wine, 7-discount. However, I want to add another element to analyze my data a step further. Not only do I want to know what category the most sales have, grocery vs. non-grocery items, but I would like to determine by how much, in US dollars. That way I can know which category is producing the most profits in sales amounts and quickly determine the cost of those items as well. For example, the cost of grocery items, beer, and wine is 30% less than the total on the sales order. The mark-up on liquor is 35%, tobacco is 15%, and lotto is 5%. In conclusion, by separating the

[1] *According to CA state tax laws, " In General. Tax does not apply to sales of food products for human consumption except as provided in Regulations 1503, 1574, and 1603. (Grocers, in particular, should note that tax applies to sales of "hot prepared food products" as provided in Regulation 1603(e).) Food products of California Constitution, Article XIII, Section 34 (California Department of Tax and Fee Administration, 2022)."

items into 7 distinct categories, I can quickly reference the costs vs. profits and determine how each sales category is performing.

Creating a Data Set

The next step was to create two frequency tables based on two separate questions. The frequency observations and statistical analysis of data collected on 30 sales orders could help my husband's family understand the frequency statistical characteristics more deeply (Deng et al., 2019). The questions that I chose in this step were, question 1- "Which category has the largest number of top-dollar sales items, mid-range of sales items, and low-range of sales items?" and question 2- "What is the probability that a customer will pay using their ATM/debit card?" To answer question 1, I categorized the sale items found on each receipt into the 7 categories (grocery-1, beer-2, liquor-3, lotto-4, tobacco-5, wine-6, discount-7) and found the category totals and individually placed them in their correct range of sale amount. The average itemized receipt item was $4.52USD, the median itemized receipt item was $2.85USD, the mode was $1.99USD, the minimum value was $1.99USD, the maximum value was $10.44USD, there were 5 outliers in the data, or anything above $10.44USD, and a standard deviation of $4.64USD.

According to the steps in constructing a frequency distribution, we must decide on the number of classes, determine the class interval, set the individual class limits, and tally the raw data into classes and determine the frequency in each class (Lind et al., 2018). The class intervals for the itemized sale items will be categorized at $4.00USD, evenly rounding from the standard deviation at $4.64USD. Starting at $-3.00USD that would compute at $-3.00USD up to $3.70USD with a value/frequency distribution of 45, $3.70USD up to $7.70USD with a value/frequency distribution of 12, $7.70USD up to $11.70USD with a value/frequency distribution of 7, $11.70USD up to $15.70USD with a value/frequency distribution of 0,

$15.70USD up to $19.70USD with a value/frequency distribution of 2, $19.70USD up to $23.70USD with a value/frequency distribution of 1, and $23.70USD up to $27.70USD with a value/frequency distribution of 1. When the mean is greater than the median, the trend is now positively skewed, with a long right tail. This can be interpreted as meaning there are more outliers on the upper/right side of the data.

The low-range category was determined to be $0.30USD up to $8.00USD, the mid-range category came in at $8.00USD up to $16.00USD, and the high-range category came in at $16.00USD up to $24.00USD. Grocery items had the highest number (60 frequency count out of 77 low-range itemized sales) of low-range sales. Tobacco items had the highest number (5 frequency count out of 9 mid-range itemized sales) of mid-range sales. Finally, liquor items had the highest number (2 frequency count out of 2 high-range sales) of high-range sales. There was a total of 88 categorized itemized sale items.

To answer question 2- “What is the probability that a customer will pay using their ATM/debit card?” I assessed the 30 sales orders and categorized them by ATM fee. If there was an ATM fee it was added as $0.50USD in the category, and if there was no ATM fee there was a $0.00USD added to the category. The class intervals will be categorized in USD, starting at $0.24, evenly rounding from the standard deviation at $0.217. Starting at $0.00 that would set the class intervals at (in USD): $0.00 up to $0.24 with a value/frequency distribution of 23, $0.24 up to $0.48 with a value/frequency distribution of 0, and $0.48 up to $0.72 with a value/frequency distribution of 7. Overall, the probability that a customer will pay using their ATM/debit card is 7 out of 30, or 0.2333, or 23.33%. Keep in mind, this exact percentage of probability may differ under some environmental factors and become reduced to other tools of uncertainty (Munir et al., 2021).

When creating a frequency distribution table, I also used the same 3 categories- low-range ($0.30USD up to $8.00USD), mid-range ($16.00USD up to $24.00USD), high-range ($24.00USD up to $32.00USD). This way I could further analyze the ATM charges to the independent variable- receipt total (plus CRV, and 7.75% tax[1] when applicable). Does the receipt total amount have a contributing factor to the amount of ATM/debit card usage? In the low-range category there were 2 ATM/debit card fees, in the mid-range category there were 3 ATM/debit card fees, and in the high-range category there were 0 ATM/debit card fees. Overall, the probability that a customer will pay using their ATM/debit card is 7 out of 30, or 0.2333, or 23% and the chances of when the customer will use their ATM/debit card does not really vary based on total sale amount. However, there is a slight increase in the middle of the data, and a peak at the mid-high range, $16.00-$24.00USD.

Descriptive Statistics

For the next analysis, I chose the two variables; total quantity/number of items per transaction and receipt total (plus CRV, and 7.75% tax [1] when applicable). The mean, median, and mode for total quantity/number of items per transaction was 2.9, 2, and 2, respectively. The mean, median, and mode for receipt total (plus CRV, and 7.75% tax[1] when applicable) was $11.35 USD, $10.38 USD, and $6.23 USD, respectively. The median is the middle value of a set of data arranged from the minimum to the maximum the median. That is, 50% of the observations are larger than the median and 50% are smaller (Lind et al., (2018)). The mode is the value that occurs most often in a set of data (Lind et al., 2018).

According to Lind et al., (2018), "The variance and the standard deviation use all the values in a data set and are based on deviations from the arithmetic mean." According to the definition of variance by Lind et al., (2018), "The arithmetic mean of the squared deviations

from the mean." The sample variance for total quantity/number of items per transactions was 4.92069. This is the average of the variations from the mean, or the amount in which all quantity totals per transaction vary, on average, from the sample mean, or average point in the 30 observations (mean= 2.9) (Lind et al., 2018). The sample variance for receipt total in $USD (plus CRV, and 7.75% tax[1] when applicable) was $41.1663. This is the average of the variations from the mean, or the amount in which all sales transactions vary from their sample mean, on average (average transaction = $11.35USD, average variance from $11.35USD = $41.17USD). This is where, even though $11.35USD is your average transaction, the majority of your transactions are in fact going to lie within $41.17USD from $11.35 USD.

The total quantity/number of items per transaction's standard deviation equals 2.21826. According to Lind et al., (2018), "a small standard deviation for a set of values indicates that these values are located close to the mean. Conversely, a large standard deviation reveals that the observations are widely scattered about the mean." According to Chebyshev's Therorem, "For any set of observations (sample or population), the proportion of the values that lie within k standard deviations of the mean is at least 1 – 1/k2, where k is any value greater than 1 (Lind et al., 2018). " Chebyshev's Theorem allows us to conclude that at least three out of every four, or 75%, of the values must lie between the mean plus two standard deviations and the mean minus two standard deviations (Lind et al., 2018). Further, at least eight of nine values, or 88.9%, will lie between plus three standard deviations and minus three standard deviations of the mean (Lind et al., 2018). And lastly, at least 24 of 25 values, or 96%, will lie between plus and minus five standard deviations of the mean (Lind et al., 2018). With that being said, although the average sales order contains 2.9 items, 75% of transactions will range between 7.3 items (2.9 + 2.2*2) to -1.5 or, 0 (2.9 - 2.2*2), transactions. 88.9% of transactions will range from 3 standard deviations

of the mean: (2.9 + 2.2*3) = 9.5 to (2.9 - 2.2*3) = -3.7 or 0 transaction items. Finally, 96% of transactions will range from 5 standard deviations of the mean: (2.9 +2.2*5) = 13.9 to (2.9 - 2.2*5) = -8.1 or 0 transactions.

The total receipt total's (plus CRV, and 7.75% tax[1] when applicable) standard deviation equals $6.41609USD. The average sale at Crown Liquor is $11.35USD, with 75% of transactions ranging between $24.18USD (11.34 + 6.42*2) to $-1.14USD (11.34 - 6.24*2). 88.9% of transactions will range from 3 standard deviations of the mean; (11.35 + 6.42*3) =$30.61USD to (11.35 - 6.42*3) =$-7.91USD or $0USD. Finally, 96% of transactions will range from 5 standard deviations of the mean; (11.35 +6.42*5) = $43.46USD to (11.35- 6.42*5) = $-20.75USD or $0USD.

According to Lind et al., (2018), "A value between zero and one, inclusive, describing the relative possibility (chance or likelihood) an event will occur." According to classical probability, based on the assumption that the outcomes of an experiment are equally likely, the probability of an event happening is computed by dividing the number of favorable outcomes by the number of possible outcomes: probability of an event = number of favorable outcomes/total number of possible outcomes. For itemized sale items, the class intervals will be categorized at 2.5, starting at 1.0 that would set the class intervals at: 1.0 up to 3.5 with a value of 22 and a probability of 73.33%, 3.5 up to 6 with a value of 5 and a probability of 16.67%, 6 up to 8.5 with a value of 2 and a probability of 6.67%, and finally 8.5 up to 11 with a frequency count of 1 and a probability of 3.33%. For total sale amount transaction, the class intervals will be categorized into 4 classes in USD, by the class interval of at $7.20 starting at $2.00 up to $9.20 with a frequency of 14 and a probability of 46.67%, $9.20 up to $16.40 with a frequency of 9 and a

probability of 30.00%, $16.40 up to $23.60 with a frequency of 5 and a probability of 16.67%, and $23.60 up to $30.80 with a frequency of 2 and a probability of 6.67%.

Confidence Intervals

The population mean of receipt total is unknown, but the best estimate is $11.35USD, based on the sample mean. Hence, our best estimate of the unknown population value is the corresponding sample statistic. Thus, the sample mean of $11.35USD is a point estimate of the unknown population mean (Lind et al., 2018). The findings are significant enough to conclude that at 95% level of confidence. Only 5% of the total receipts would not contain the population mean. That is not too bad, considering the margin of error is $4.79. The findings are significant enough to conclude that at 90% level of confidence, that 90% of the population receipt totals will lie between $9.36USD and $13.34USD. 10% of the total receipts would not contain the population mean. That is not too bad, considering the margin of error is $3.98USD. The findings are significant enough to conclude that at the highest level of confidence, 98%, that 98% of the population receipt totals will lie between $8.96USD & $14.24USD. Only 2% of the total receipts would not contain the population mean. That is not too bad, considering the margin of error is $5.28USD.

Conclusion

To answer question 1- “Which category of items grosses the highest number of sales at Crown Liquor’ The answer was grocery items. Grocery items had the highest number of sales, with a 60 frequency count out. Together on 30 receipts, there was a total of 88 categorized itemized sales items. With 60 out of 88 itemized items, grocery items also came in at highest number of low-range sales ($-0.30USD-$8.00USD). The second category of items grossing the highest number of sales was beer. Beer was the second highest number, with a total number of

10 frequencies count out of 88 total itemized sales. The third category of items grossing the highest number of sales was tobacco. Tobacco products had the third highest ranking category out of 7, with a total of 5 frequency count out of 88 total itemized sales. Tobacco products also had the highest number of mid-range sales ($8.00USD-$16.00USD) with a total number of 5 frequency count out of 9.

Question 1 can be further determined by; the receipt total amount (plus CRV, and 7.75% tax[1] when applicable). The average receipt total (plus CRV, and 7.75% tax[1] when applicable), in USD is $11.35, the receipt total amount (plus CRV, and 7.75% tax[1] when applicable) median in USD is $10.38, the receipt total amount (plus CRV, and 7.75% tax[1] when applicable) mode in USD is $6.23, the receipt total amount (plus CRV, and 7.75% tax[1] when applicable) standard deviation in USD is $6.42. The largest receipt total amount in USD (plus CRV, and 7.75% tax[1] when applicable) was $26.35 and the smallest receipt total amount in USD (plus CRV, and 7.75% tax[1] when applicable) was $2.00.

To answer question 2- “What is the probability that a customer will pay using their ATM/debit card?” I assessed the 30 sales orders and categorized them by ATM fee. If there was an ATM fee it was added as $0.50USD in the category, and if there was no ATM fee there was a $0.00USD added to the category. Overall, the probability that a customer will pay using their ATM/debit card is 7 out of 30, or 0.2333, or 23%. Furthermore, the chances of when the customer will use their ATM/debit card does not really vary based on total sale amount. However, there is a slight increase in the middle of the data, and a peak at the mid-high range, $16.00-$24.00USD.

References

California Department of Tax and Fee Administration. (2022). Laws, regulations and Annotations. Sales and Use Tax Regulations - Article 8. Retrieved June 1, 2022, from https://www.cdtfa.ca.gov/lawguides/vol1/sutr/sales-and-use-tax-regulations-art8-all.html

Deng X., Li, H., Yu, W., Weikang, W., & Liu, Y. (2019). Frequency Observations and Statistic Analysis of Worldwide Main Power Grids Using FNET/GridEye. *2019 IEEE Power & Energy Society General Meeting (PESGM), Power & Energy Society General Meeting (PESGM), 2019 IEEE*, 1-5. Retrieved June 29, 2022 from https://lopes.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.8973560&site=eds-live&scope=site&custid=s8333196&groupid=main&profile=eds1

Lind, D. A., Marchal, W. G., & Wathen, S. A. ((2018)). Statistical techniques in business and economics (17th ed.). New York, NY: McGraw Hill Education, ISBN-13: 9781259666360.

Mansour, M.M.M., & Aboshady, M.S.A. (2022). Assessing the Performance of Insulating Fluids via Point of Statistical Inference View. *TWMS Journal of Applied & Engineering Mathematics*, *12* (2), 469-480. Retrieved June 29, 2022 from https://lopes.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=aci&AN=156307949&site=edslive&scope=site&custid=s8333196&groupid=main&profile=eds1

Munir, M., Mahmood, T., & Hussain, A. (2021). Algorithm for T-spherical fuzzy MADM based on associated immediate probability interactive geometric aggregation operators.

Artificial Intelligence Review, *54*(8), 6033-6061. Retried June 29, 2022 from https://lopes.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2021-38002-001&site=eds-live&scope=site&custid=s8333196&groupid=main&profile=eds1.