Data Analytics Capstone Topic Approval Form

Student Name: Tony Dunsworth

Student ID: 000329564

Capstone Project Name: Non-parametric statistical analysis of 911 PSAP response data

Project Topic: Kruskal-Wallis statistical analysis on 911 PSAP response research data.

X This project does not involve human subjects research and is exempt from WGU IRB review.

Research Question: Is there a statistically significant difference in 911 call processing times among the study group?

Hypothesis: H_0: There is no statistically significant difference concerning call processing times in the compared datasets.
H_1: There is a statistically significant difference concerning call processing times in the compared datasets.

Context: The contribution of this study to the field of Data Analytics and to the City of Alexandria's Department of Emergency and Customer Communications (DECC) is to statistically analyze the impact of SARS-COV2 (COVID19) on the operations of the Public Safety Answering Point (PSAP) for the City of Alexandria, Virginia. Similar to the study by Marazzo et al analyzing the effects of SARS-COV2 on the PSAP serving the city of Milan, Italy (Marazzo, et al 2020), this study will utilize Kruskal-Walls analysis to compare response times from 2019 to 2020 to see the impact of the pandemic and the operational response to ensure operational continuity for the community. By conducting this analysis, the results will contribute to better emergency operational planning and adding to industry wide best practices.

Data: The data needed for this analysis comes from the CentralSquare Inform Computer Aided Dispatch system database; specifically, the Response_Master_Incident table. (CentralSquare, 2020) The full dataset for 2019 and 2020 consists of 90,273 rows. Per the data dictionary (CentralSquare, 2020), the Response_Master_Incident table consists of 146 columns. Of those, the analysis will only use a small subset of these columns. The variables used are broken down as follows:

https://github.com/trdunsworth/WGU_Capstone

Field	Type
Response_Date	DateTme
Year	Numeric (Discrete)
Month	Categorical
WeekNo	Numeric (Discrete)
DOW	Categorical
Hour	Numeric (Discrete)
Shift	Categorical
Priority_Number	Numeric (Discrete)
Problem	Categorical
Agency	Categorical
MethodOfCallRcvd	Categorical
Fixed_Time_PhonePickup	DateTime
Fixed_Time_CallEnteredQueue	DateTime
Time_First_Unit_Assigned	DateTime
Fixed_Time_CallTakingComplete	DateTime

T2Q	Numeric (Continuous)
T2Disp	Numeric (Continuous)
ProcTime	Numeric (Continuous)

. The data is owned by the Department of Emergency and Customer Communications of the City of Alexandria, Virginia. Permission for use and analysis was given by Renee Gordon MBA, SHRM-CP, CPE, ENP Director of the department (Gordon, 2020). Signed permission for use is attached to the approval form. Limitations: The data harvested goes back only two years calendar years 2019 and 2020 and is limited to the completeness of the entries into the table. Delimitations: The dataset will be delimited by selecting only a subset of the available columns. All potential identifying information, such as address, or identification of employees have been removed. Also, all rows which are requests for mutual aid from surrounding jurisdictions have been removed since there is no call taker involvement and during 2020, steps were taken to automate the dispatch process for these calls; allowing units to be assigned to these calls without dispatcher intervention. Rows are limited to ensure those listed were handled by DECC employees and that the service calls were dispatched to units for response. Finally, calls which weren't answered by DECC staff and processed are also removed from the resultant dataset. The retained rows pertain to the date and time of call taker and dispatcher service call contact points, the agency dispatched to the call, the nature of the call, how the service call entered the system, and the priority level given to the call.

Data Gathering: The Treatment of the Data: The data has been gathered from the Response_Master_Incident table using a T-SQL query. In the first iteration of the query, there were a significant number of empty strings and NULL values in the MethodOfCallRcvd column. The rows were retained in order to perform data analysis using this column. The NULL and empty values were transformed in the T-SQL query and transformed into the string "Not Reported". Per Reddy (2019), this method of transforming data during gathering can be used to address the missing and NULL values. The quality of the resultant data set is relatively high and contains Qualitative and Quantitative variables. Additionally, there are several computed and derived columns created through the T-SQL query in order to allow for additional analysis based on temporal markers, shifts. The computed columns are the elapsed times for the various temporal markers in the service calls.

Data Analytic Tools and Techniques: The Design of the Study: 1. A QQ plot and Shapiro-Wilk will be run to determine the normality of the data. Per Tanya Hoskins (2012) of the Mayo Clinic, if the assumptions of normality is not validated, then the nonparametric Kruskal-Wallis test will be appropriate for the continuous numeric variables. 2. The Kruskal-Wallis tests to compare means between three or more groups will be run on the computed columns to create the means comparisons (Hoskins, 2012). Univariate, bivariate, and multivariate analyses will be run on the remaining variables in order to glean additional insights into the data.

Justification of Tools/Techniques: R will be used for normality tests and the Kruskal-Wallis tests for the continuous numeric variables. Per Dancho (2012), R is well-built for all of the different analytical methods, is business friendly, and has excellent tools for professional quality output of the results of the analyses. Per Professor Norm Matloff, "R is written *by* statisticians, *for* statisticians," giving it an advantage over Python for detailed statistical analysis (Matloff, 2019)

Project Outcomes: The anticipated outcomes are insights into how the SARS-COV2 (Covid19) virus impacted 911 Dispatch operations throughout 2020. Support for the alternative hypothesis can be found in Marrazzo et al (2020) that this analysis can show how the measures taken by the Department of Emergency and Customer Communications impacted the times to receive and dispatch emergency calls for service in the PSAP Operations Center.

Projected Project End Date: 6/30/2021

Institutional Review Board Quiz and Approval

Have you read and understood the "Human Subjects FAQ" page and completed the "Human Subjects FAQ Quiz" at the WGU Institutional Review Board (IRB) website?
(https://irb.wgu.edu/info/Pages/Home.aspx)

⊠ Yes, I have read and understood the "Human Subjects FAQ" and have provided email proof of my completed quiz in appendix A. (https://irb.wgu.edu/info/Pages/Human-Subjects-FAQ-Quiz.aspx)

☐ No, I have not completed the Human Subjects FAQ quiz.

Assess whether your capstone proposal complies with WGU's IRB standards for exemption status. Explain why you believe the proposed project complies with the standards for exemption status. If it does not, make arrangements with a course mentor and the IRB for approval.

⊠ The research complies with WGU's IRB exemption status because:

- Research involving the collection or study of freely available de-identified underlined existing data
- Research that does not employ methodology on human subjects.

☐ The research requires approval from WGU's IRB because:

☐ Yes, I would like to schedule a conference to discuss my project.

Projected Project End Date: 2/28/2021

Marrazzo, F, Spina, S, Pepe, PE, et al. Rapid reorganization of the Milan metropolitan public safety answering point operations during the initial phase of the COVID-19 outbreak in Italy. *JACEP Open*. 2020; 1: 1240– 1249. https://doi.org/10.1002/emp2.12245

Hoskin, Tonya. "Parametric and Nonparametric: Demystifying the Terms." Mayo Clinic, 2012. https://www.mayo.edu/research/documents/parametric-and-nonparametric-demystifying-the-terms/doc-20408960

Dancho, M. (2020, December 17). *6 Reasons To Learn R For Business [2021]*. Business Science. https://www.business-science.io/business/2020/12/17/six-reasons-to-use-R-for-business-2021.html

CentralSquare, Inc. (2020). *Computer-Aided Dispatch | CAD Dispatch Software | CentralSquare*. https://www.centralsquare.com/public-safety/cad

Gordon, R. M. (2020, October 8). *Capstone Data Approval*. https://github.com/trdunsworth/WGU_Capstone/blob/main/Dunsworth_Capstone_Data_Approval.pdf

Matloff, N. (2019, October 14). *R-vs.-Python-for-Data-Science*. GitHub

https://github.com/matloff/R-vs.-Python-for-Data-Science

Reddy, M. A. (2020, February 15). *A Comprehensive guide on handling Missing Values - #ByCodeGarage*. Medium.

Course Instructor Signature/Date:

To be filled out by a course mentor:

☒ The research is exempt from an IRB Review.

☐ An IRB approval is in place (provide proof in appendix B).

Course Mentor's Approval Status: Approved

Date: 3/8/2021

Reviewed by:

Comments: Click here to enter text.