**Data Analytics Capstone Topic Approval Form**

**Student Name: Taylor Sullivan**

**Student ID: 000825214**

**Capstone Project Name:** **The** **Predictive Power of Food Consumption**

**Project Topic**: **A Predictive Model of Food Consumption versus Health Outcomes in Developing Nations**

**Research Question: Does the consumption rate of various food products influence health outcomes?**

**Hypotheses**: **H0 : Food consumption has no impact on health outcomes in developing nations, H1: in nations with lower consumption of high-fat and high-caloric foods, there are more favorable health outcomes, H2: nations with greater consumption of high-fat and high-caloric foods result in better health outcomes.**

**Context:**

Nutrition is the foundation of disease prevention. In studies on noncommunicable disease (NCD), poor diet has been linked to increased heart disease, stroke, diabetes, and cancer. These health outcomes disproportionately affect low-income societies that do not have the same access to fresh produce and often cannot afford healthier options. According to the World Health Organization (2011), “Poverty is closely linked with NCDs, and the rapid rise in the magnitude of these health problems is therefore predicted to impede poverty reduction initiatives in low-income countries and communities” (p. 33). Developing nations in particular are therefore expected to face a nearly insurmountable challenge when attempting to eradicate these diseases and lift their citizens out of poverty. As federal governments attempt to intervene, they should consider food consumption’s impact on these health outcomes. This study looks at the implications of food consumption rates in developing nations to develop a predictive model for health outcomes. Parametric statistical analysis will be used to develop the model.

**Data:** A 2010 food consumption data set, 2010 health outcomes data set, and nutrition facts data set are used for this study.

<https://fdc.nal.usda.gov/download-datasets.html>

<http://datatopics.worldbank.org/consumption/product/Rice>

<https://apps.who.int/gho/data/view.main.NCDDEATHCAUSESNUMBERv?lang=en>

This is public information presented to the public at-large by non-governmental organizations.

**Data Gathering:**

Description of the Data: The World Bank offers data on the 2010 consumption rate of different foods (binned by like items) for each developing nation, subset by income levels and quantified in millions of U.S. dollars. The dataset consists of the following variables: Area (Rural, Urban, National), Consumption Segment (High, Middle, Low Income), Country, Measure Names (US$, $PPP, Local Currency), Product, and Measure Values for the amount consumed measured in cost. There are five, independent, qualitative, categorical variables and one, dependent, quantitative, continuous variable. The World Health Organization dataset contains the health outcomes for each country subset by gender. There are six variables in the dataset, three continuous, one date, and two categorical. This includes the following variables: country, year, causes (malignant neoplasms, diabetes mellitus, cardiovascular diseases, chronic obstructive pulmonary disease), both sexes (count of cases resulting in death), male, and female. The next dataset used in the study is from FoodData Central and will necessitate the merge of four different datasets involving food names and nutritional values. This merged dataset will consist of one character value, food name, and three continuous variables, protein, carbohydrate, and fat value. This dataset will show the nutritional value of each food item. The final dataset, gathered from JohnSnowLabs on DataHub, contains the population rates for each country in 2010. This is a key addition to allow for the calculation of per capita rates, and demonstrate the real correlation between food consumption rates and higher rates of noncommunicable diseases.

**Data Analytics Tools and Techniques**:

Parametric statistical methods will be used in this study. The analysis will be performed in SAS. SAS graphics, R graphics, and Tableau dashboards will be provided for the presentation of findings.

**Justification of Tools/Techniques:** SAS University Edition will be used for the cleaning, merging, and analysis. SAS is an integrated software suite for advanced analytics, business intelligence, data management, and predictive analytics. SAS software can be used through both a graphical interface and the SAS programming language, or Base SAS. SAS software can access data in almost any format, including SAS tables, Microsoft Excel tables, and database files. This is ideal for the varied formats of the datasets in this study. SAS also facilitates the managing and manipulation of the datasets, such as merging, subsetting, and creating new columns with ease. SAS Software has the statistical power to conduct correlations, logistic regression, model selection, and Bayesian hierarchical models, which will be necessary in this study. Finally SAS provides the variety of reporting formats, including HTML, PDF, and RTF, that will be fundamental to the data analytics report (SAS Studio: Help Center, 2016).

**Project Outcomes**: An impact analysis of nutritious food consumption rates on health outcomes. A statistical model that can be used to predict health outcomes in order to direct the flow of resources.

**Projected Project End Date**: **06/30/2020**

**Reference Sources**:

SAS Studio: Help Center. (2016, June 21). Retrieved May 08, 2020, from https://support.sas.com/software/products/sas-studio/faq/SAS\_whatis.htm

Alwan, A., Dr. (Ed.). (2011). *Global status report on noncommunicable diseases 2010* (p. 33, Rep.). Italy: World Health Organization. Retrieved May 8, 2020, from https://apps.who.int/iris/bitstream/handle/10665/44579/9789240686458\_eng.pdf;jsessionid=B563EF98D7EC35CD759F3FC2408B9F85?sequence=1

Data Hub. (2018). In Population Figures by Country. John Snow Labs. Retrieved July 10, 2020, from https://datahub.io/JohnSnowLabs/population-figures-by-country#readme

U.S. Department of Agriculture. (2020, April). Download FoodData Central Data. Retrieved May 08, 2020, from https://fdc.nal.usda.gov/download-datasets.html

World Bank Group. (2010). Global Consumption Database. Retrieved May 08, 2020, from http://datatopics.worldbank.org/consumption/product/Rice

World Health Organization. (2018, April 5). NCD Deaths by Cause and Sex - Data by Country. Retrieved May 08, 2020, from https://apps.who.int/gho/data/view.main.NCDDEATHCAUSESNUMBERv?lang=en

**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 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.

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).

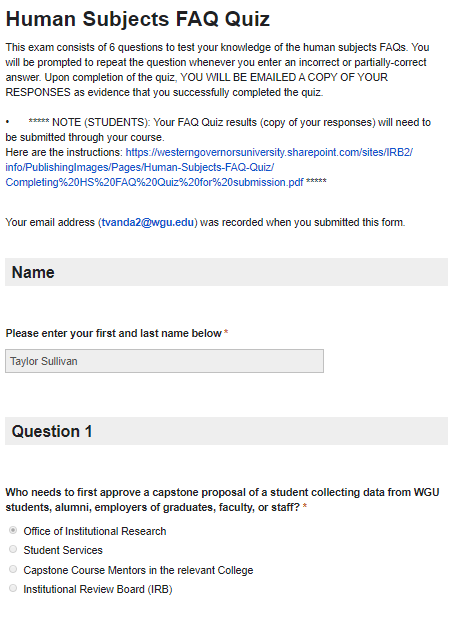
Course Mentor’s Approval Status: Select one

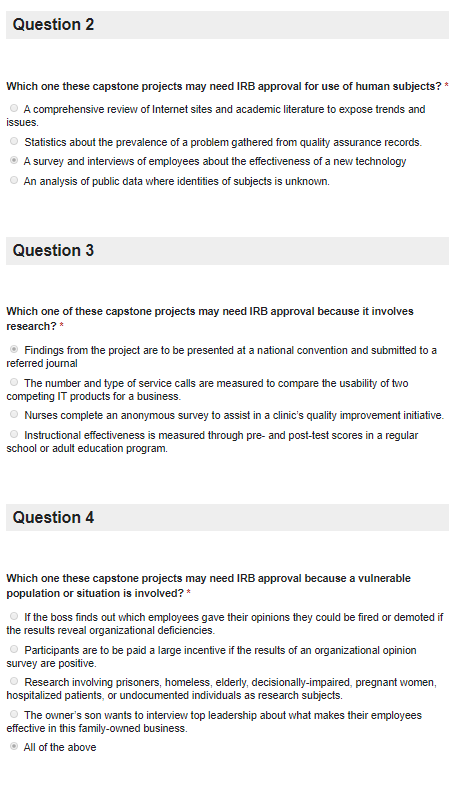
Date: Click here to enter a date.

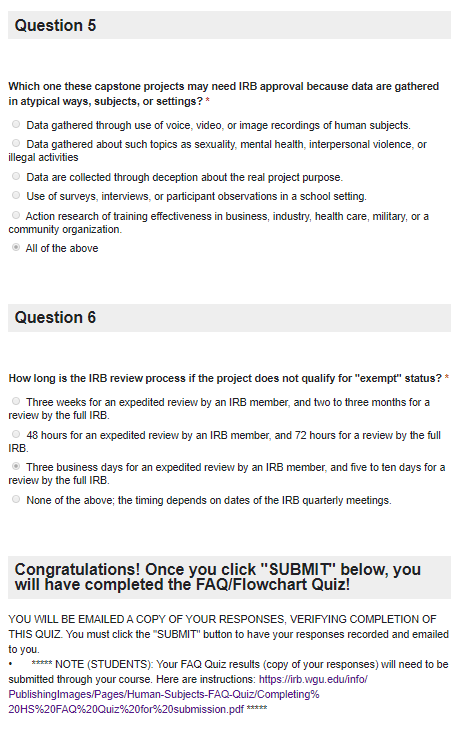
Reviewed by: Click here to enter text.

Comments: Click here to enter text.

**APPENDIX A:** Email showing completed “Human Subjects FAQ Quiz”

****

****

****

**APPENDIX B:** IRB approval (not applicable)