

## Individual Assignment 1

Descriptive & Inferential Statistics

BUSI650 - Business Analytics

University Canada West

MSc Mario del Pino

## 1- Assignment Objective

The objective of this work is for students to apply the statistical techniques learned on the course to interpret business operations and reach meaningful conclusions that would allow a business to support decision making.

The exercise is based on the Alma-Cafe case study. The case assumes the existence of Boutique Coffee Shop that offers an array of three different coffees from Brazil, Colombia and Peru. The company operates five locations in Metro Vancouver.

### 1- Introduction

You have been assigned to be part of the newly created Business Analytics Team at Alma-Cafe, tasked with improving business performance through the detailed analysis of the data that the company is generating. This task force reports directly to the CEO, as the board of directors has decided to put all digitalization efforts at the center of their corporate strategy with the objective of adapting the company to the new data-driven market.

The company has been operating consistently over the past 3 months, showing a healthy business outcome. One of the most important departments in the company is the Product Procurement Department. This department implemented a pilot project last month and is interested in evaluating the performance of the project to determine if a larger scale project should be put in place.

The Product Procurement Department oversees purchasing all coffee products, with a focus on obtaining the highest quality in products to satisfy the needs of the companies' premium clientele. The most important product they acquire are the coffee beans from South American providers and that are used for making the coffee at each location.

With the intention of improving sales, the Procurement Department decided to partner with a Quality Certification company in Brazil called Bracof. This company claims to have a Quality Assurance process that guarantees a high-quality selection among the many coffee producers that can be found in the country. Early focus groups showed that customers might be willing to increase their daily coffee purchases with this premium coffee. In August, all Brazilian coffee was replaced by the premium mix provided by Bracof. The manager wants to know if the new coffee mix has in fact increased purchases, as there might be the possibility of seeking a similar strategy for imports from Colombia and Peru.

### 2- Provided Dataset

You are provided with a dataset containing a detailed record of all servings for each product line, grouped for each country of origin and location.

<i>Date</i>	<i>product_origin</i>	<i>city</i>	<i>week_day</i>	<i>servings</i>
5/1/2022	Brazil	Downtown	Sunday	1128
5/1/2022	Colombia	Downtown	Sunday	1109
5/1/2022	Peru	Downtown	Sunday	1015
5/1/2022	Brazil	North Vancouver	Sunday	572
5/1/2022	Colombia	North Vancouver	Sunday	741

5/1/2022	Peru	North Vancouver	Sunday	571
5/1/2022	Brazil	Burnaby	Sunday	372
5/1/2022	Colombia	Burnaby	Sunday	573
5/1/2022	Peru	Burnaby	Sunday	257
5/1/2022	Brazil	Surrey	Sunday	364
5/1/2022	Colombia	Surrey	Sunday	505
5/1/2022	Peru	Surrey	Sunday	264
5/1/2022	Brazil	New Westminster	Sunday	217
5/1/2022	Colombia	New Westminster	Sunday	109

The dataset is divided into two tabs. The first one from May 2022 to July 2022 containing 1,380 entries. The second one holds data from August 2022 to September 2022 containing 915 entries.

### 3- Expected results

You are expected to elaborate a Power Point presentation with up to 6 slides containing the following information:

- Descriptive Statistics considering the difference between:
  - o Each Location.
  - o Each Product Origin.

This analysis should include statistical measurements such as median, mean, standard deviation, percentiles 10 and 90, and a histogram to illustrate the data distribution.

This analysis should be done separately for datasets prior and after the pilot project's implementation.

- Inferential Statistics analysis on the effectiveness of the Pilot Project. In the case of the Brazilian Coffee Pilot Project, you need to answer the following questions:
  - o What is the mean and standard deviation for the servings of Brazilian Coffee prior to the implementation of the Pilot? ( $\mu_0$  &  $\sigma_0$ )
  - o What is the mean for the servings of Brazilian Coffee after the implementation of the Pilot? ( $\mu_p$ )
  - o How many data entries are for all Brazilian Coffee servings after the implementation? Keep in mind that the dataset contains entries from multiple locations. ( $n$ )
  - o What is the sampling distribution Mean and Standard Deviation for a sampling size equal to the number of days after the implementation of the pilots? ( $\mu_{\bar{x}}$  &  $\sigma_{\bar{x}}$ )
  - o What is the Z-value for a sampling distribution with an average of  $\mu_p$  given a normal distribution with Mean  $\mu_{\bar{x}}$  and Standard Deviation  $\sigma_{\bar{x}}$ ?
  - o What is the probability of obtaining these values assuming that the Pilot had no effect?
  - o Can we conclude that the Pilot had a positive effect on daily servings?

#### 4- Grading Rubric

Score	50%-60%	60%-70%	70%-80%	80%-90%	90%-100%
Descriptive Statistics (20%)	Mean and Standard deviation are calculated and the histogram is build.	Some metrics are calculated, with all central metrics and standard deviations included allowing to perform inferential statistics.	Most metrics are calculated, and histogram is adequately spaced to allow for clear reading.	All metrics are calculated and stated in a numeric way, with no mistakes or reporting errors.	All metrics are correctly calculated and stated in a numeric way. Relevant metrics are superimposed to the histogram to show relative position.
Inferential Statistics (50%)	Some comparison between datasets was made, with some qualitative description on the difference.	Population statistics are correctly calculated, and an attempt to get the sampling distribution average was made with some minor errors.	Population and sampling distribution statistics are correctly calculated, and an attempt to reach conclusions was made.	Population and sampling distribution statistics are correctly calculated, and calculations allow to reach meaningful conclusions.	Population and sampling distribution statistics are correctly calculated, hypothesis and significance level are correctly stated, and calculations allow to reach meaningful conclusions.
Message Clarity (30%)	Presentation has all relevant information in at least in numeric format.	Presentation has all relevant information in at least in numeric format properly ordered to understand it.	The presentation can uses histograms to show data distribution and information is laid our in an ordered way.	The presentation can convey most of the required information in visually. All plots are properly labeled and multiple pieces of information can be easily found.	The presentation can convey all the required information in a clear and visually appealing way. All plots are properly labeled and multiple pieces of information are contextualized. Recommendations are done to improve future analysis.