

Understanding how Socioeconomic Factors Influence Alcohol Consumption Rates Between the Eastern and Western Hemispheres

Dhanraj Patel - 1003965168

December 17, 2021

Contents

Abstract	2
Keywords	2
Introduction	3
Overview	3
Terminology.	3
Central Research Question	4
Data	4
Data Collection Process	4
Data Summary	4
Numerical Summaries	5
Graphical Summaries	7
Methods	9
Results	10
Conclusions	11
Weaknesses	12
Next Steps	12
Discussion	13
Bibliography	14
Appendix	15
A1: Ethics Statement	15
A2: Materials	15

Abstract

Within this report we will investigate how various socioeconomic factors influences the alcohol consumption rates of countries around the world and compare between the eastern and western hemispheres. We will be using data that was collected by reputable agencies such as the Institute for Health Metrics and Evaluation, the US Census Bureau's International Database, the United Nations Statistics Division, and the World Bank. We used this data in a post stratification method in order to make casual inferences about the connection between socioeconomic factors and alcoholic consumption despite being observational data. By calculating the propensity score using a linear logistic regression we are able to match observations based on their likelihood of being in the treatment groups by using nearest neighbor matching. With a new dataset of only matched observations we are able to investigate the alcohol consumption rates between countries and compare. Using this method we found that between eastern and western hemisphere the drinking rates were relatively similar with the eastern hemisphere countries being a lot more varied compared to the western hemisphere. This is mainly due to the fact that the eastern hemisphere has a lot more countries in it and it contains countries widely varying in their alcoholic consumption rates and socioeconomic conditions. This variation led to the eastern hemisphere having both the highest and lowest rates of alcohol consumption.

Keywords

Propensity Score Matching

Causal Inference

Logistic Linear Regression

Geographical Statistics

Alcohol Consumption Statistics

Global Statistics

Socioeconomic Statistics

Frequentist regression analysis

Introduction

Overview

Ever since the ancient Egyptians alcohol has been part a staple of human culture, as we progressed through the ages alcohol has only gained in popularity being consumed in almost every country in the world. Alcohol has become so widespread that the average person older than 15 years old consumes 6.4 liters per of pure alcohol annually and it has become the second most popular drug used in the world (behind only to caffeine) [2]. While alcohol has its place within social engineering it is important to understand the dangers associated with the abuse alcohol as alcohol abuse contributes to 2.8 million premature deaths per year [2]. Classified as an highly addictive substance a study conducted by the World Health Organization found an association between the use of alcohol and an increased rate of cancer with 1 in 7 of all new cancers in 2020 linked to the use of alcohol [3]. Aside from the health risks associated with alcohol, this study will focus on the effects that Socioeconomic development factors have on countries overall consumption of alcohol and compare the results between countries located in the eastern and western hemispheres. One such factor in this study is the average income levels of countries and the effect that it has on the alcohol consumption among its population. We can see this correlation from a New England Journal of Medicine report where the study found that countries with high income levels had lower rates of adverse health outcomes effecting alcohol drinkers compared to low income countries [4]. In addition to income this study will also focus on socioeconomic factors such as happiness levels, employment rates, urbanization rates and more.

When analyzing the alcohol consumption rates this study will analyze the consumption rates and compare the results between the countries in the eastern and western hemispheres, this is done to identify how the eastern and western culture impacts the results. The study is important to study due to the global impact that the use of alcohol has on the countries all over the world, as previously mentioned that alcohol is an addictive substance that has been linked to many different adverse health effects such as cancer. Through this study we will be able to better understand the impact socioeconomic factors have on the alcohol consumption rates between countries which will help us understand how to reduce the rate of alcohol related fatalities all over the world. Lastly, it is important to note that the study will use alcohol consumption volumes per capita.

Terminology.

Before proceeding with the study it is important to understand some of the terminology that will used in this report. As for the non statistical terminology when discussing the amount of alcohol by volume the amount that will be discussed will be pure alcohol, this refers to the ethanol content by volume. This is mainly done due to the fact that different types of alcoholic drinks tend to have different amount of pure alcohol so to standardized alcohol pure alcohol is used in the study.

In addition it is important in the study to understand what is meant by the western and eastern hemispheres. Countries around the world can be differentiated by being considered in the western or eastern hemisphere due to if they fall on the east or west of the prime meridian which is at longitude 0 (passes through Greenwich London) on a standard world map. There are countries that are in both hemispheres and for the purposes of this study, they are categorized in the hemisphere that most of their country is in.

Lastly as for statistical terminology, it is important to understand what propensity score matching is in this report. Propensity score matching is a statistical method that allows us to make assertions in observational studies that usually are exclusive to experimental studies. Using observational studies we can split up the data based on control and treatment group factors and then create a propensity score that closely matches data observations in both treatment and control groups. Using these observations we can analyze results and make inferences. It is important that the matching done by the propensity score is appropriate, this study will mainly verify the quality of the matching using T-tests and numerical summaries. In addition, in this report the treatment groups are the eastern and western hemispheres and the outcome for study is the alcohol consumption rate.

Central Research Question

Using the data of various socioeconomic and alcohol consumption rates all over the world this study will aim to answer the central research question of how the various socioeconomic factors effect the consumption of alcohol consumption rates between countries in the eastern and western hemisphere.

In this report the preliminary hypothesis is that countries with lower socioeconomic factors will have a higher rate of alcohol consumption rate compared to those with high socioeconomic factors on average. This hypothesis stems from a British Medical Journal (BMJ) report that studied an increase in alcohol consumption in developing countries as a result of alcohol companies turning attention to developing countries as alcohol use in developed countries is decreasing [5]. As for the western and eastern hemispheres it using a map of alcohol consumption rates in all countries in the world [6] it seemed that the consumption rates were relatively equal in both hemisphere so we hypothesis that alcohol consumption rates in both hemispheres will be relatively equal.

Data

Data Collection Process

In this study we will be using a dataset that contains various socioeconomic factors of countries around the world to analyze the effect that they have on alcohol consumption rates. The data can be found in this link <https://www.kaggle.com/sansuthi/alcohol-consumption> [1]. The data was organized by an independent educational non-profit Gapminder (one word) and to create the dataset they used information found collected by sources including the Institute for Health Metrics and Evaluation, the US Census Bureau's International Database, the United Nations Statistics Division, and the World Bank. The data is open and available for everyone to view and use.

Some limitations that could be found in the data collection process is that not all countries around the world can give accurate data that will be used in this study. For example, when it comes to income levels of the population, it is harder to get accurate data from developing countries as they usually don't have a robust system of recording such information developed countries. In addition, another limitation is that using many different sources may create a disconnect the agencies might have different internal standards on the process of collecting information so the metrics recorded by the World Bank may differ from the United Nations Statistics Division creating inconsistencies in the data.

Data Summary

The goal of the study is to understand how various socioeconomic factors effect the alcohol consumption rate in countries around the world. To do this we will use a data that contains the alcohol consumption rate per-capita and various socioeconomic factors of countries around the world. Some countries could not be included and have missing data as the pertinent data could not be collected in those countries. The socioeconomic factors include: income levels, suicide rate per 100 people, employment rate and urbanization rates. Using propensity score matching using this dataset we can find how the aforementioned socioeconomic factors influence alcohol consumption rates between the treatment groups of the western and eastern hemisphere nations. The hemisphere of the countries in the dataset will need to be added manually.

Data Cleaning Process

Before moving onto numerical and graphical summaries its important to understand how the data was cleaned. In this dataset for various reasons a lot of the countries listed were not able to provide all the data points and were missing data as a result. To ensure that missing data would not effect the results of the study, any observation (country) that contained any missing data was removed from the dataset. Moving on there were no duplicated data in the dataset so that was not a factor that needed correcting.

In addition the original dataset did not have a column that stated if each country belonged to the eastern or western hemisphere, since this is vital information for the study a column was added which stated if the countries in the dataset were in the eastern or western hemisphere [7]. For those countries that are located in both hemispheres, in this study they were categorized into the hemisphere where most of their country is in. Lastly in order to aid with the graphical summaries two columns were added that stated if the countries were greater or lesser than the mean income levels and alcohol consumption rates respectively.

Important Variables

Here are the list of important variables that were used in the study.

- country - The name of the country
- alcconsumption - The average pure alcohol consumption of the countries population of those older than 15 per capital in liters.
- suicideper100th - The average number of suicides per one hundred thousand people.
- incomeperperson - Gross Domestic Product per capita in constant \$2000 USD.
- employrate - The percent of population older than 15 who are employed.
- urbanrate - The percent of people who live in urban environments.
- Hemisphere - If the country is in the eastern or western hemisphere
- inc_is_above_avg - If the country is greater or lower than the mean income per person of the countries in the dataset.
- alc_cons_is_above_avg - If the country is greater or lower than the mean alcohol consumption of the countries in the dataset.

Numerical Summaries

With the data cleaned we can move into identifying the numerical summaries of the dataset. Numerical summaries allow to gain insight on features of the dataset with attention of the location and spread of the variables.

Comparing alcohol consumption rates separated by the hemispheres of the countries

Table 1: The trimmed mean is trimmed by 10 percent

Hemisphere	Count	Min	Q1	Median	Q3	Max	IQR	Mean	Trimmean	Var	SD	Range
Eastern	114	0.05	2.1275	5.560	11.0775	23.01	8.950	6.92	6.44	31	5.57	22.96
Western	48	0.51	4.8425	6.475	8.7275	14.92	3.885	6.57	6.49	11	3.30	14.41

When comparing the average consumption of pure alcohol in both hemispheres we first noticed that in this dataset there are many countries who are in the eastern hemisphere compared to the western hemisphere. This is to be expected as there are a lot more countries in the eastern hemisphere compared to the western hemisphere. As for the consumption rates we notice that from the min value to the median the average amount of alcohol drank by the western hemisphere was greater than the countries in the eastern hemisphere; however, from q3 to the max value we see that the eastern hemisphere countries have a higher value. This is mainly due to the fact the eastern hemisphere has a much larger variance, range and SD as they contain the least drinking countries and most drinking countries in the world. Overall the mean and trimmed (10 percent) mean of eastern hemisphere countries was greater than the western hemisphere.

Comparing average Income and employment rates separated by the hemispheres of the countries

Table 2: The trimmed mean is trimmed by 10 percent

Meaure	Min	Q1	Median	Q3	Max	IQR	Mean	Trimmean	Var	SD	Range
Eastern avg Inc	104	592	2202	8602	52302	8010	7342.23	4955.21	118758777	10897.65	52198.0
Western avg Inc	155	1138	3207	7027	37491	5890	7000.40	5202.49	91329551	9556.65	37336.0
Eastern emp rate	35	51	58	65	83	14	58.37	58.03	125	11.18	48.3
Western emp rate	45	56	60	65	82	9	60.76	60.50	63	7.96	36.8

In Table_2 we can see the average income levels and employment rates when comparing the eastern and western hemispheres. When comparing the average income western hemisphere countries have a higher value from the min to the median, but the eastern countries have a higher value from q3 to max. This once again is due to the fact that variance, SD and range of the eastern countries is higher as the eastern hemisphere contains countries with the highest and lowest levels of average income. This disparity can be seen in the fact that the eastern hemisphere countries have a higher mean income average but a lower trimmed (by 10 percent) average. When comparing the percent of the countries that are employed we see that the values are relatively close together with the western hemisphere having higher percent values up to the median and have the same q3 value. The eastern hemisphere has a slightly higher max value with it being just one percent higher. As we have seen with average income the eastern countries hemisphere countries have a higher variance, sd and range; however, unlike average income the western hemisphere has both a higher mean and trimmed mean of employment rates.

Comparing average urbanization and suicide rates (per one hundred thousand) separated by the hemispheres of the countries

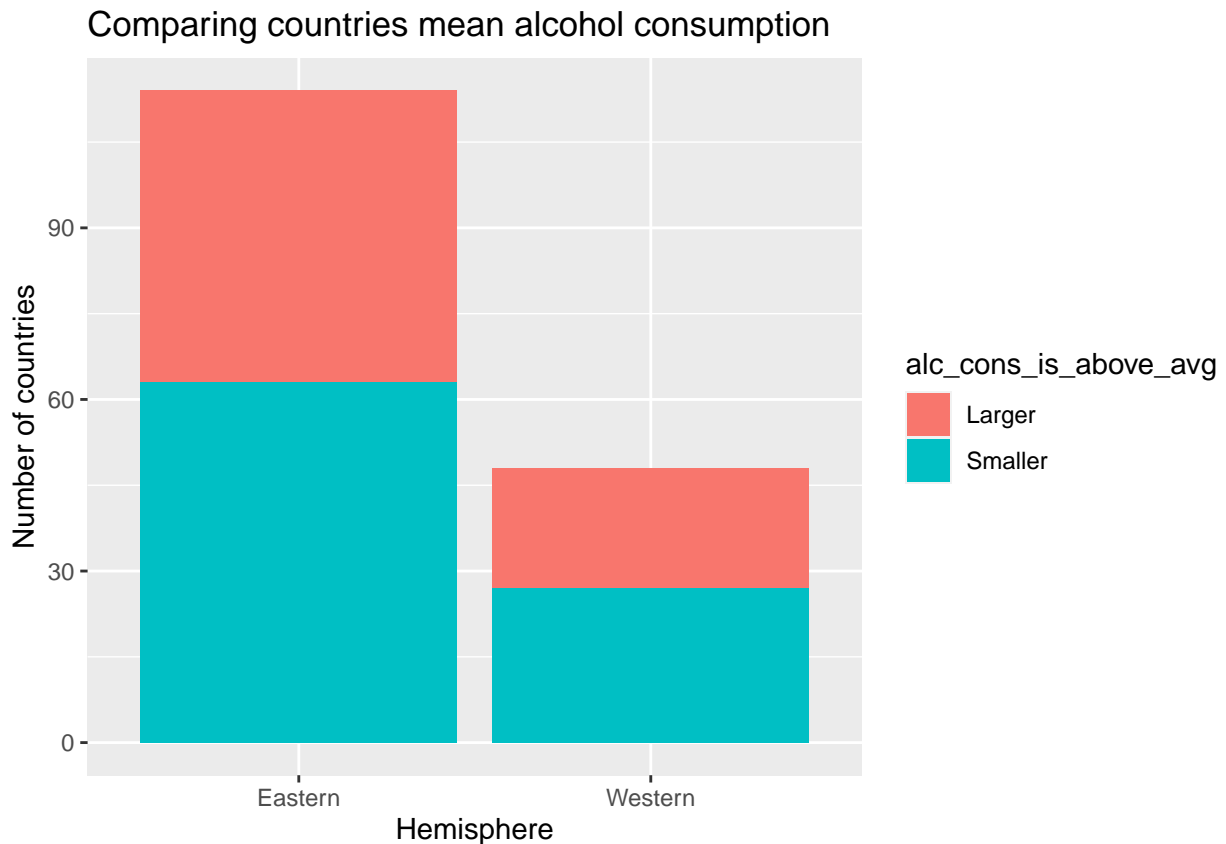
Table 3: The trimmed mean is trimmed by 10 percent

Meaure	Min	Q1	Median	Q3	Max	IQR	Mean	Trimmean	Var	SD	Range
Eastern urbanization	10.0	35	57	70	100	36	53.52	53.51	531	23.05	90.00000
Western urbanization	13.0	48	61	77	93	30	61.28	62.01	419	20.48	80.00000
Eastern suicide rate	0.2	6	10	14	33	8	10.74	10.18	42	6.46	33.14041
Western suicide rate	2.0	5	7	10	36	5	7.99	7.27	31	5.61	34.23362

In Table_3 we can see the numerical summaries for the urbanization and suicide rates of countries, the urbanization rate refers to the percent of a countries population that lives in an urban setting and the suicide rate rates of a country per hundred thousand people. For urbanization rates we see that western countries have a higher urbanization up to q3, but eastern countries have the max urbanization rate. This is due to Singapore which is a city state and thus has 100 percent of its population in an urban environment. In addition, eastern countries have a higher variance, sd and range of urbanization while western countries have a higher mean and trimmed mean values. This is mainly due to the eastern countries containing a lot of developing countries with large amounts of populations who do not live in cities. As for suicide rates the west have the greater min and max values while eastern countries have higher values between the min and max. Despite this, the eastern counties have a higher mean, trimmed mean, variance, IQR, sd and range of values when it comes to suicide rates. From all the tables we have seen the fact that the eastern hemisphere has a lot more countries makes it so they have a larger variance in their data.

Graphical Summaries

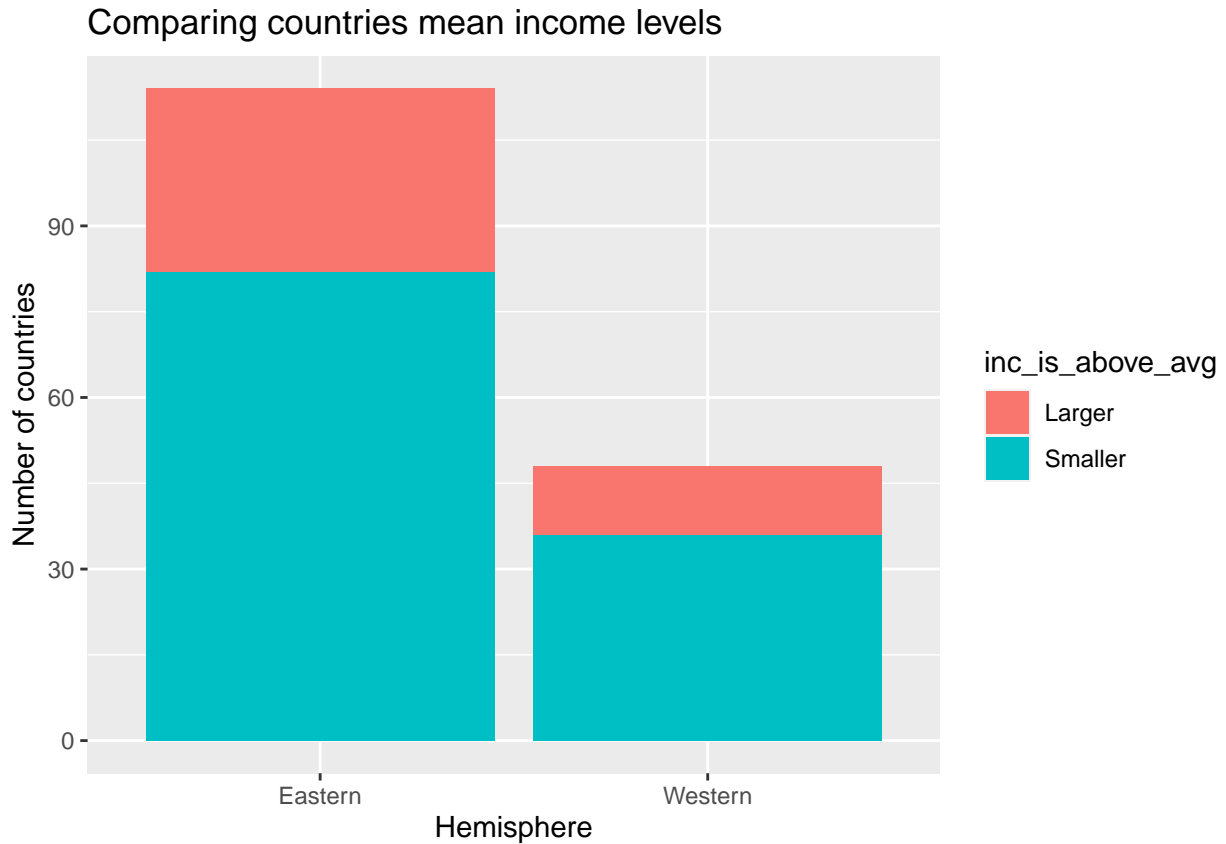
Comparing number of countries below or above the mean alcohol consumption by hemisphere
(Graph_1)



From Graph_1 we are able to see how many countries in the data are in each hemisphere and how many countries in each hemisphere have either larger or smaller average alcohol consumption rate compared to the mean alcohol consumption rate in the dataset. From the graph we see that the number of countries in the eastern hemisphere is much larger than the total number of countries in the western hemisphere. This is to be expected as the eastern hemisphere contains a majority of Asia, Africa and Europe which are continents with a large amount of countries. Also, we see that while both hemispheres have a relatively equal amount of countries with higher and lower average consumption rates compared to the mean that the eastern hemisphere has noticeably higher number of countries with consumption rates smaller than the mean and the western hemisphere has a noticeably higher number of countries with consumption rates higher than the mean.

Comparing number of countries below or above the mean income in each hemisphere

(Graph_2)



From Graph_2 we can see how the how countries in each hemisphere are proportioned based on if they have a greater or smaller average income compared to all counties in the dataset. One thing to keep in mind is that there are a few countries with a large mean average incomes that increase the average and that results in both eastern and western hemispheres having majority of the countries in them below the average income. We see that western countries have a larger disparity of countries above the mean income in contrast to the western hemisphere. In contrast to alcohol consumption rates we see means income levels have a larger amount of disparity, but in bother cases for both hemispheres the majority of countries fell below the mean.

All analysis for this report was programmed using R version 4.0.4.

Methods

Introduction

Before diving into the different methods it is important to reiterate that the main goal of this study is to better understand how various socioeconomic factors influence the alcohol consumption rates when comparing the countries in the eastern and western hemisphere. Usually this study would require a experimental study with random selection between the treatment groups; however, in this case we are unable to do that as the data we are using was collected from a observational study. Thus, in order to make inferences that would otherwise need experimental studies we will be using propensity score matching in conjunction with logistic regression to answer our central research question.

What is propensity score

When it comes to propensity score matching the main idea behind this methodology is to match similar observations in the dataset based on their propensity score. Within the data that was used in this study, the countries in the dataset were not randomly assigned to either hemispheres (treatment groups) and without this randomness component we can not make causality inferences. Instead the data was recorded as it was observed. The idea behind propensity score is if we take our data set and examine what would make observations more likely to be in the treatment group vs non treatment group and then if we match observations based of those likelihoods whether or not they are the treatment group we would have two groups where the observations would be either in the treatment group or not, but they would have the same likelihood of being in the treatment group. The treatment groups in this study are the hemispheres of the countries and by matching these observations as such we would be able to make an inference to the outcome (alcohol consumption) as if this was experimental data.

Steps of propensity score

When utilizing propensity score matching there are steps you generally have to take. The first step is to assign probability to all the observations based of whether or not they would be in the treatment group. In this study we will be using socioeconomic factors (suicide rate, average income, urbanization rate and employment) rate to the likelihood of being in the western hemisphere. We will be using the western hemispheres in this case as in the data set there are many more eastern countries so it makes sense to match towards the western hemisphere. To do this match we will be using a frequentist logistic linear regression model with the socioeconomic factors as the predictors and the hemisphere as the response variable. It is appropriate to use logistic regression as the response is a binary outcome, either in the western hemisphere or not.

After calculating the propensity scores the next step is to match observations based off the propensity scores. In this study we will be using nearest neighbor matching method which matches observations based off propensity scores closest together. After matching the third step is to evaluate the matches and to ensure a balance of the treatment groups. Lastly, with the matches we can then evaluate the outcome of interest which in this study is the alcohol consumption rates in both groups.

Methodologies of logistic regression

As aforementioned to calculate propensity score matching the we will be using frequentist logistic linear regression. This is appropriate as the response variable is a binary outcome The model that we will be using is the following:

$$\begin{aligned} Hemisphere(response) = & \beta_0 + Employment_rate(x_1) + Average_Income(x_2) + ... \\ & ... + Suicide_rate(x_3) + Urbanization_rate(x_4) + \epsilon_i \end{aligned}$$

In the model the response variable will be the hemisphere of the observations and the predictors are the socioeconomic factors. All the predictors are level 2 predictors as this study measures only variables at a national level so no individualistic data was used. Using this method we will be calculating the propensity scores. However, before moving onto the results it is important to understand the assumptions that will be used in the logistic linear regression model. The assumptions are the following:

- Independence: The response (Y values/out comes) values are independent of each other.
- Linearity: There is a linear relationship between the predictors (Numeric x') and log odds of the response (Y).
- Response is binary: That the response variable has binary outcomes.
- Variance: That the variance is a function of a binomial (follows P).

With these methods we are able to create a scatter plot of the alcohol consumption variable in the two groups answer the goal of the study.

Results

In our dataset using logistic regression we calculated the propensity score based on the likelihood of the countries being in the treatment group which in this case is being in the western hemisphere. From there using nearest neighbor matching we match up observations in the dataset based on their propensity scores. Then we evaluated the quality of the matches with T tests and finally with the dataset with the matched pairs we can evaluate the alcohol consumption of both groups to evaluate the goal of the study.

Addressing quality of match

After each observation receives a propensity score we use nearest neighbor matching to create a dataset where with an equal amount of observations from both treatment groups or in the case of this study equal amount of countries from both hemispheres. However, it is important to identify the quality of the matches before moving onto the outcome analysis. To do this we will be using a t test on each variable for both treatment groups. With our T test we will be able to identify how alike the values for each of the variables are. If the P value for the t test is lower or equal then 0.05 our alternative hypothesis will be true meaning that there is a significant difference between the groups. However, if it is greater than 0.05 then our null hypothesis is true meaning there is no significant difference.

(Table_4)

Table 4: t test values for hemispheres

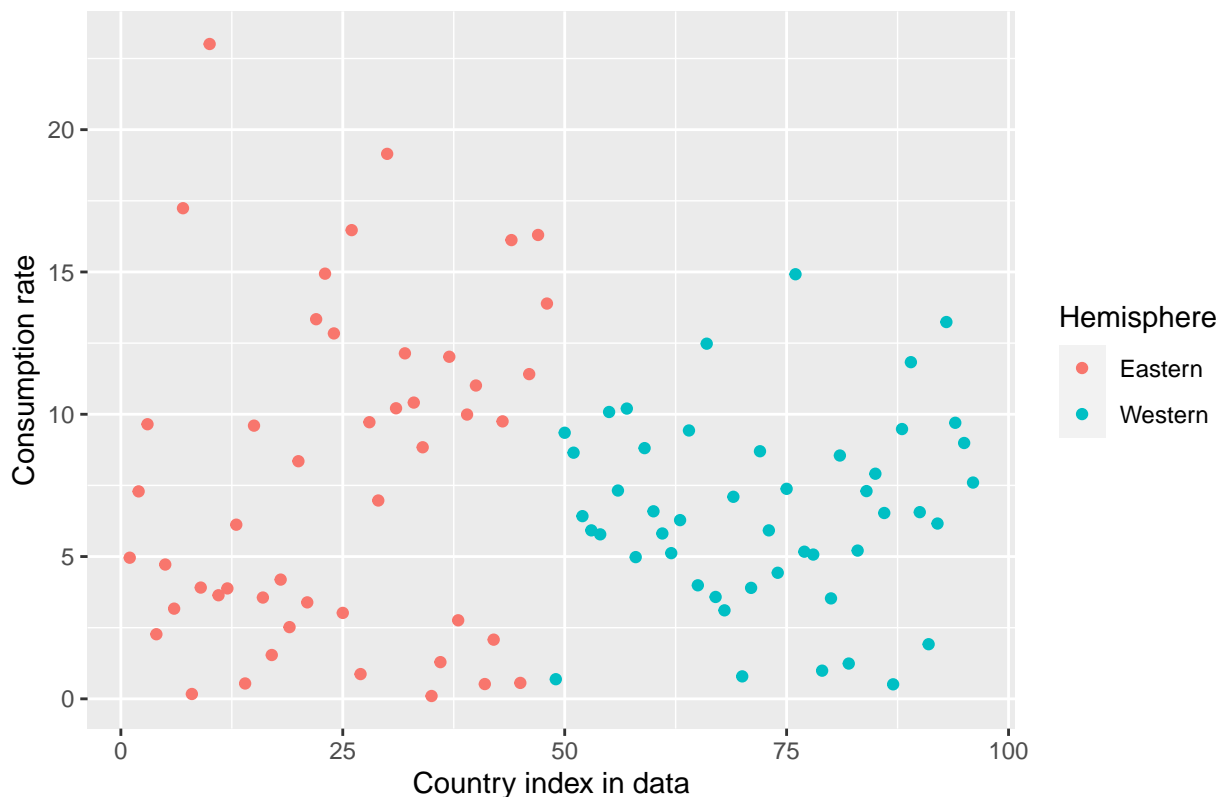
category	T_test_vals
Alc consumption	0.237000
income avg	0.212100
suicide rate	0.002835
emp rate	0.634700
urbanization rate	0.736900

In table_4 we are able to see the values of the T tests. When comparing both hemispheres we see that the p values for all categories besides suicide rates were greater than 0.05. This means that for these categories the null hypothesis is correct and thus there is no significant difference between the values for each hemisphere. For the suicide rate we see that the value is much lower than 0.05 meaning there is a significant difference between the values in the 2 categories. Overall when we compare the values of the outcome with the hemispheres it is important to keep in mind how alike these values are for each category.

Comparing alcohol consumption for both hemispheres

(graph_3)

Comparing Hemisphere alcohol consumption



From graph_3 we are able to see the results of the study which outlines the alcoholic consumption rates for both hemispheres. From what we can see both hemispheres have relatively equal amount of alcohol consumption between them with the two hemispheres being relatively equal. One hemisphere does not have a dramatically higher or lower consumption rate compared to the other hemisphere. However, it is important to note that although relatively equal the Eastern hemisphere has countries that have the highest and lowest alcohol consumption amounts while the western hemisphere countries sit in between. The eastern hemisphere countries also seem to be a lot more varied compared to the western hemisphere where the values are closer together.

From our data the results do make intuitive sense. It makes sense that the eastern hemisphere being a lot more varied as they have a lot more countries in them and those countries are very culturally different. For example the eastern hemisphere contains many European countries famous for their alcohol and countries where alcohol is completely banned. The west has a lot less countries in comparison So due to these reasons we can conclude the results do make sense.

Conclusions

With alcohol being such a popular drink all over the world the goal of this study was to examine how the average consumption rate of alcohol was effected by various socioeconomic factors of different countries. The socioeconomic factors included in this study were suicide rate, employment rate, average income and urbanization rate. From these results we compared the results between countries in the eastern and western hemisphere to see if there is any noticeable difference. The initial hypothesis was that the two hemispheres would have a relatively equal consumption rates and that developing countries with low socioeconomic factors would have higher alcohol consumption rates.

From our data section we preformed numerical and graphical analysis of the variables in the dataset. From

the numerical analysis we compared the numerical summaries of the alcohol consumption rates, the income and employment together rates and the suicide rates and urbanization rates together. The key results from this sections were that the in most cases up to the median the western hemisphere had the higher values in all the categories (except suicide rates), but from the median to the max value the eastern countries had the higher values. This is mainly due to the diverse countries in the eastern hemisphere and the greater number of countries. Then from the graphical summaries we compared the average income and alcohol consumption of both hemispheres. The key results were that for alcohol consumption both hemispheres had relatively even split between countries who drank above and below the average. With the average income we noticed that both hemispheres had a majority of countries with lower than the mean income of all the countries. This is because a relatively small number of countries have a high mean average, while there a lot of countries with low mean incomes.

Then in the methods section we discussed the processes that went into post stratification and the justifications and characteristics of the logistic linear equation in our model. and then finally in our results we were able to answer the study central question. Before moving onto the alcohol consumption rates it was important to evaluate the quality of the post stratification matches. From the T tests we found that in the data set most of the values were similar between the hemispheres except for suicide rate which significantly dissimilar with a very low p value (< 0.05). Then from the results we noticed saw that both hemispheres had similar levels of alcohol consumption rates, but that eastern hemisphere was more varied and had the highest and lowest levels of consumption.

From the results if we take a step back and focus on the big picture we see that in the world there are a lot of factors that influence a countries consumption of alcohol. When using the socioeconomic factors including employment rate, suicide rate, employment and urbanization we found that generally the overall trend of consumption between the two hemispheres are very similar. The eastern hemisphere is more varied in their consumption and the western hemisphere is more homogeneous. A large part of this is cultural difference, economic factors and the fact that the east has many more countries.

Weaknesses

When looking at the results its important to understand the limitations of the results. The limitations can be divided into two categories.

The first section of limitations has to do with the data, when comparing rates of alcohol consumption it is very difficult to source the data that is both accurate and does not have any biases. The first issue I noticed is that its hard to get data on many countries due to many reasons such as the people in the country living very remote lives. In this study if a country contained any missing data they were removed from the dataset which impacted results as those countries were not included in the data.

In addition there are many limitations with using post stratification that impacted results. The first issue is that when matching, the eastern countries that were not matched with a western country were removed from the dataset. This resulted in a large lose of data which impacted results. Secondly the matching algorithm that was used was nearest neighbor matching which has issues. The nearest neighbor matching method has issues where when matching once matched the algorithm does not recheck the matched values for further unmatched observations. This means that it is possible that observations ended up with matches that were not the best fit as the best fit was already matched. This impacted the final dataset used for the outcomes.

Next Steps

If one is to do this study again some of the next steps are the following. One idea is the include more socioeconomic factors than than the four that were used in this study. There are a lot of socioeconomic that effect alcohol use besides the four used in this study such as depression rates, addiction rates and personal freedoms. By including more socioeconomic factors we will get a more detailed and complete idea of what influences that consumption of alcohol.

In addition, another suggestion is to use a more thorough matching algorithm to match observations based off propensity score. The issue with the nearest neighbor algorithm is that it does not recheck observations that were already matched, despite the fact that a matched observation could match even better with an unmatched observations. A more through algorithm would increase the quality and accuracy in the matches.

Finally in this study a lot of countries had missing data causing them to be removed entirely from the dataset. If possible the next study should attempt to find the missing data if possible to get a more accurate idea of both hemispheres.

Discussion

From this study we were able to analyze the effect various socioeconomic had on alcohol. From our results we found that both the eastern and western hemispheres had a generally equal amount of consumption, while the east was more varied and had the min and max values. In contrast, the western hemisphere was more homogeneous with the values being closer together. The main reasons for this is that the socioeconomic factors in the eastern countries is also a lot more varied than the west. In addition a large part of the variance has to do with the cultural aspects of the eastern countries. The countries in the eastern hemisphere contain countries with cultures that hold alcohol in very high regard like European countries and countries who out right ban alcohol like Saudi Arabia. In addition the eastern countries vary greatly in socio economic conditions containing both the poorest and richest countries. As culture is linked heavily to socioeconomic conditions the variation in one led to another. In contrast, while still varied the western countries were not as varied and this led to less variation in the data, a large part of this has to do with the fact that there are many more eastern countries compared to western countries. From this study we were able to better understand the different socioeconomic condition of countries around the world and the impact it had on the alcohol consumption rates around the world.

Bibliography

1)

Sandhya S. (2021, May) *Alcohol Consumption Suicide Rate, Income, Employment & Urban Rate*. kaggle. <https://www.kaggle.com/sansuthi/alcohol-consumption>. (Last Accessed: Dec 17, 2021)

2)

Ritchie H, Roser M. (2019, Nov) *Alcohol Consumption*. OurWorldInData. <https://ourworldindata.org/alcohol-consumption>. (Last Accessed: Dec 17, 2021)

3)

(2021, July 21) *New WHO study links moderate alcohol use with higher cancer risk*. Centre for Addiction and Mental Health (CAMH). <https://www.camh.ca/en/camh-news-and-stories/new-who-study-links-moderate-alcohol-use-with-higher-cancer-risk> (Last Accessed: Dec 17, 2021)

4)

Smyth A et al. (2015, sept 16) *Effects of Alcohol Consumption Differ Between Rich and Poor Countries*. NEMJ Journal Watch <https://www.jwatch.org/na39085/2015/10/06/effects-alcohol-consumption-differ-between-rich-and-poor>. (Last Accessed: Dec 17, 2021)

5)

Walls H, Cook S, Matzopoulos R, et. (2020, April 16) *Advancing alcohol research in low-income and middle-income countries: a global alcohol environment framework*. BMJ Global Health 2020;5:e001958. <https://gh.bmj.com/content/5/4/e001958.info>. (Last Accessed: Dec 17, 2021)

6)

(2021) *Alcoholism by Country 2021*. worldpopulationreview. <https://worldpopulationreview.com/country-rankings/alcoholism-by-country>. (Last Accessed: Dec 17, 2021)

7)

(2021) *Western Hemisphere Countries 2021*. worldpopulationreview. <https://worldpopulationreview.com/country-rankings/western-hemisphere-countries>. (Last Accessed: Dec 17, 2021)

Appendix

A1: Ethics Statement

Many ethical considerations were made in this study. One such ethical consideration is that the study was meant to as reproducible and transparent as possible. The data used in this study is available and open for anyone to use freely. There is no barrier to attaining this data except access to the internet. In addition to make the study transparent and reproducible as possible all major decisions and outcome were reported in the study accurately and in great detail. Some of these decisions include the steps and reasoning to use post stratification. In addition, the logistic model and its results were also outlined in great detail. This study was designed to be as reproducible as possible. Lastly, the study is transparent with all the steps that it took and the limitations of its methodologies and results.

The second main ethical consideration was that the data was selected as to not misrepresent the countries in the dataset and to limit the amount of bias in the data. When conducting a study on factors that effect countries it is important to source your data adequately as to not be impacted by false data, selection bias and skewed data. In order to endure that this study was not effect by that the source of the data came from reputable sources that can be trusted to give data as accurate and non biased as possible. The institutions that collected the data including the Institute for Health Metrics and Evaluation, the US Census Bureau’s International Database, the United Nations Statistics Division, and the World Bank. All of these sources can be trusted to give ethical data.

A2: Materials

The data that was used for this study was found using an online data sharing website called kaggle. The data can be found at this link [1].

Table 5: Data used in study

country	alconsumption	incomeperperson	suicideper100th	employrate	urbanrate	Hemisphere
Afghanistan	0.03	NA	6.684385	55.7	24.04	Eastern
Albania	7.29	1914.997	7.699330	51.4	46.72	Eastern
Algeria	0.69	2231.993	4.848770	50.5	65.22	Western
Andorra	10.17	21943.340	5.362179	NA	88.92	Eastern
Angola	5.57	1381.004	14.554677	75.7	56.70	Eastern
Antigua and Barbuda	8.17	11894.464	2.161843	NA	30.46	Western