Title: Alcohol Consumption and Health Expenditure vs. Life Expectancy in the World

Background and Introduction:

As more guidelines come out depicting how one should prioritize their health, there has been an unfortunate increase in Alcohol Consumption. However, some countries have still decreased their overall health expenditure (Health Service Data). Since, alcohol consumption and health expenditure are values that can have a significant impact on the life expectancy of individuals within these countries, it is important to visualize and manage the trends seen within the data in order to provide guidance to increase average life span. I am utilizing data from the World Health Organization (WHO) on Life Expectancy to conduct my analysis. This particular data set addresses the previous oversight of not considering the impact of immunization and the Human Development Index when comparing life expectancy with other variables. The purpose of this analysis is to discern key relationships between alcohol consumption, health expenditure, and life expectancy as well as examining the difference of this effect within developed vs. developing countries.

The data set comprises several columns, including country, year, development status, life expectancy (age), adult mortality rates, infant deaths, alcohol consumption, percentage expenditure on health, hepatitis B immunization, and measles cases. Data is collected by WHO's member states that compile health data mainly derived from population-based sources (household surveys, civil registration systems) and institution-based sources (health facilities) (Data Collection Tools). By considering these various variables collectively, we can gain insights into the areas within countries that require funding or encouragement to enhance life expectancy. Additionally, by examining variables such as alcohol consumption and health expenditure, we can unravel connections between specific behaviors/variables and life expectancy.

Study Design and Description:

I obtained this data set on life expectancy from the World Health Organization (WHO) via Kaggle.com. User compiled the data set: KumarraJarshi and included information from the Global Health Observatory data repository. This repository tracks the health status of all countries concerning immunization, mortality, economic, social, and other health-related factors.

The WHO database contains data from 193 countries from 2000 to 2015. However, my project will specifically focus on 2010 to maintain data independence. Unfortunately, the exact number of individuals included in the study is not indicated. It is important to note that this study is observational.

The data collection process involved the World Health Organization conducting an observational study. Multiple sources were used from each country to gather data on life expectancy, resulting in a dataset covering 193 countries over 15 years. This dataset includes 20 predicting variables.

Scientific Questions:

Overall Goal: To assess and analyze the multiple predictors for life expectancy in 2010, focusing on alcohol consumption and health expenditure.

Question 1: How do changes in alcohol consumption affect life expectancy in developed vs. underdeveloped countries?

Question 2: How do changes in health expenditure affect life expectancy in developed vs. underdeveloped countries?

Statistical Questions:

Question 1: Is the effect of increasing alcohol consumption, along with adjusting for developmental status, associated with an increase in the average life-span in the world?

Question 2: Is the effect of increasing health expenditure, along with adjusting for developmental status, associated with an increase in average life span in the world?

Variables:

Variable	Description	Classification	Unit	Meaning	Role
Country	Country Name	Categorical	N/A	Demographic/Location	Identifier
Year	Year of Study	Quantitative	Years	Demographic/Time	Identifier
Alcohol Consumption	Amount of Alcohol Drank	Quantitative	Average Liters	Baseline Risk Factor	Predictor of Interest
Developmental Status	Whether the country is still developing or not	Binary	Yes/No	Baseline economic status of country	Predictor of Interest
Life Expectancy	The amount of years someone lives in this country	Quantitative	Years	Measure of exposure effect alcohol and other variables	Outcome
Adult Mortality Rate	Probability of dying between 15 and 60 years per 1000 population	Quantitative	Count per 1000 population	Measure of mortality rate from effects of other variables	N/A
Health Expenditure	Amount of Money spent on healthcare	Quantitative	% GDP per capita	Measure of health expenditure effect on life expectancy	Predictor of Interest

Methods:

Our study was fundamentally designed to scrutinize the probable indicators of life expectancy for 2010, with a distinct focus on alcohol intake and health expenditure. We implemented an observational study approach to explore the potential associations among various health and economic factors. The dataset, gathered by the World Health Organization (WHO) and publicly available via Kaggle, encapsulates an exhaustive compilation of data for 193 countries from 2000 to 2015.

The scope of this research was constrained to the year 2010 to ensure data independence. The parameters chosen for our exploration incorporated Life Expectancy, Adult Mortality, Alcohol Consumption, Percentage Expenditure on Health, Country, Year, and Status (Developing versus Developed).

Initial data manipulation entailed processes of data purging and transformation. The original dataset was trimmed to include only the 2010 data and the chosen parameters. Absent data points were dealt with based on their nature and abundance. The corresponding observations were discarded if a parameter had a few missing data points. However, appropriate data imputation methods were used for parameters with many missing values.

We utilized various visualization techniques, like scatterplots and boxplots, to discern the relationship between life expectancy and other parameters partitioned by the 'Status' of the country. Scatterplots facilitated the comparison of 'Life expectancy' and 'Percentage expenditure' based on 'Status'. Boxplots depicting 'Alcohol' by 'Status' provided a better understanding of the patterns of alcohol consumption and health expenditure in developed and developing countries.

To address our research questions, we deployed multiple linear regression analysis. We formulated two models - one examining the influence of variations in alcohol consumption on life expectancy and the other inspecting the impact of changes in health expenditure on life expectancy. Both models took into account the country's developmental Status. These models aimed to elucidate the associations between life expectancy and the selected predictors and to discern further how these associations differ in context to developed and developing countries.

The condensation of the regression model yielded a valuable understanding of the model's coefficients, serving as symbols of the strength and direction of the relationships. Additionally, the analysis integrated critical statistical parameters like p-values. These elements highlight the model's suitability while affirming the coefficients' statistical significance.

While the research brings forth potential predictors influencing life expectancy, it is paramount to accentuate the study's observational nature. This methodological approach, albeit informative, inherently restrains us from affirming any causative relationships. Moreover, the analysis is elementary and does not regard all the premises of linear regression or potential confounding variables. Future studies could contemplate these factors for a more exhaustive understanding of life expectancy predictors.

Summary Statistics:

```
  Country               Year           Status    Life_expectancy
Length:121         Min.   :2010   Developed : 18   Min.   :36.3
Class :character   1st Qu.:2010   Developing:103   1st Qu.:66.3
Mode  :character   Median :2010                    Median :72.8
                   Mean   :2010                    Mean   :70.6
                   3rd Qu.:2010                    3rd Qu.:75.1
                   Max.   :2010                    Max.   :88.0
Adult_Mortality    Alcohol        percentage_expenditure  Hepatitis_B
Min.   :  2.0   Min.   : 0.010   Min.   :   0.00        Min.   :56.00
1st Qu.: 73.0   1st Qu.: 1.330   1st Qu.:  17.48        1st Qu.:83.00
Median :138.0   Median : 4.190   Median :  99.08        Median :94.00
Mean   :146.6   Mean   : 4.772   Mean   : 260.66        Mean   :89.75
3rd Qu.:196.0   3rd Qu.: 7.910   3rd Qu.: 396.63        3rd Qu.:97.00
Max.   :682.0   Max.   :14.440   Max.   :1423.80        Max.   :99.00
   Measles
Min.   :  0.00
1st Qu.:  0.00
Median :  5.00
Mean   : 97.96
3rd Qu.:103.00
Max.   :853.00
```

The dataset comprises data from 121 countries, with 18 being developed and 103 developing.
The dataset only contains information from the year 2010.
Life expectancy ranges from 36.3 to 88 years, with a median of 72.8 years, suggesting a skewed distribution towards higher life expectancy.
The adult mortality rate varies significantly from as low as 2 to as high as 682, with a median of 138.
The alcohol consumption data shows a wide range, from 0.01 to 14.44, with a median of 4.19, hinting that alcohol consumption varies significantly among countries.
The percentage expenditure on health ranges from 0 to 1423.8. The extensive range might suggest some countries spend much more on health than others.

Inferential Analyses:

Model 1:

It uses Alcohol consumption and the Status of the country (Developing or Developed) as predictors. From the model summary, we can see that the intercept is 77.0276. This is the expected value of life expectancy when all other predictors are zero. In this context, it would represent the life expectancy in a developed country (as the base category) with zero alcohol consumption. The coefficient for alcohol is 0.1602 but is not statistically significant ($p = 0.47$), meaning there is not enough evidence to conclude that changes in alcohol consumption significantly predict changes in life expectancy. This result should be interpreted cautiously, as it doesn't mean that alcohol consumption has no impact on life expectancy, but rather that within the confines of this model and data, we couldn't find a statistically significant relationship. The coefficient for StatusDeveloping is -8.4454, which is statistically significant ($p < 0.001$). This suggests that all else being equal (i.e., for the same level of alcohol consumption), life expectancy is approximately 8.4454 years lower in developing countries compared to developed countries.

Model 2:

It uses percentage expenditure on health and the Status of the country (Developing or Developed) as predictors. The intercept in this model is 76.023044, which represents the expected life expectancy in a developed country with zero health expenditure. The coefficient for percentage expenditure is 0.004998, which is statistically significant ($p = 0.01$). This suggests that for each unit increase in health expenditure, we expect life expectancy to increase by about 0.004998 years, holding the country's status constant. The coefficient for StatusDeveloping is -7.897322 and is statistically significant ($p < 0.001$). This suggests that all else being equal (i.e., for the same level of health expenditure), life expectancy is approximately 7.897322 years lower in developing countries compared to developed countries. Comparing the two models, both models indicate that a country's development status significantly predicts life expectancy. Also, neither of these models explains a large proportion of the variance in life expectancy, suggesting other significant predictors might not be included in these models.

Assumptions:

The assumptions for multiple linear regression analysis were considered for these models. These assumptions include linearity, independence, and normality of residuals.

Linearity: The linearity assumption implies a linear relationship between the response and predictor variables. If this assumption is violated, the forecasts may be inaccurate, and the statistical tests associated with the regression may be unreliable. Scatterplots visually checked linearity for all predictors against the response (life expectancy). Any non-linearity might be addressed by transforming variables or using non-linear models. Met.
Independence: The observations are assumed to be independent of each other. This assumption was likely met, as the data was collected for different countries and should be independent. Met.
Normality: This assumption can be checked using a QQ plot. If the residuals depart significantly from the straight line in the QQ-plot, the normality assumption might not be met, and transformations might be considered. Met.

Discussion:

The analytical results yielded several essential revelations. Primarily, a country's development status surfaced as a substantial determinant of life expectancy, as both models suggested lower life expectancy in developing countries compared to developed ones, all other factors being equal. This finding is harmonious with established expectations and prior research.

Additionally, the analysis revealed that health expenditure significantly influences life expectancy, reinforcing the understanding that more significant health investments likely result in superior health outcomes, thereby amplifying life expectancy. However, the influence of alcohol consumption on life expectancy did not display statistical significance within this study's scope, warranting careful interpretation.

A significant limitation to acknowledge is that these models only account for approximately small percent of the variance in life expectancy, respectively. This implies the existence of other influencing elements that these models do not consider. Moreover, the study's

observational framework restricts our capacity to identify explicit causal associations. Missing data/altered from certain developing/developed countries can add a degree of uncertainty to the models. Lastly, erroneous observational measurements by the methods of the World Health Organization can skew the data towards certain trends.

Considering the cross-sectional characteristic of our data, it presents an opportunity to explore the temporal dynamics of these attributes and assess their influences on the evolution of life expectancy. Any imputation for missing data also adds a degree of uncertainty.

From a broader scientific perspective, these findings underscore the significance of health expenditure and a country's development status in determining life expectancy. Future research should aim to pinpoint other crucial predictors and investigate the reasons behind the observed disparities in life expectancy between developing and developed countries. This information could inform health policy and interventions to enhance life expectancy.

Appendix of Distributions and Regressions:

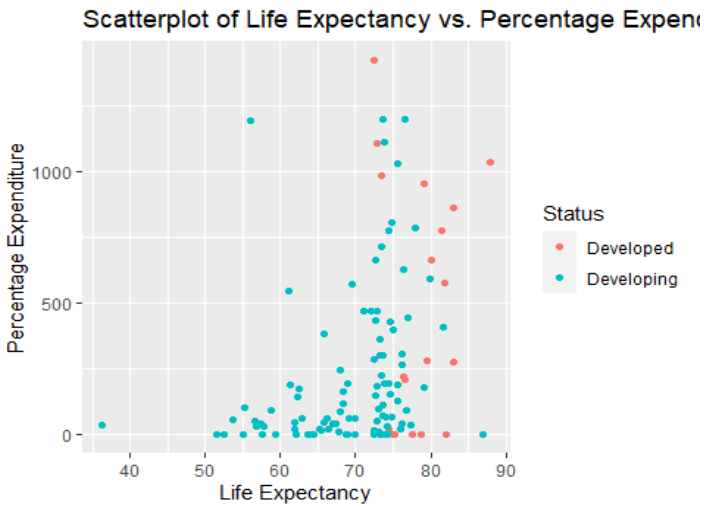

Figure 1

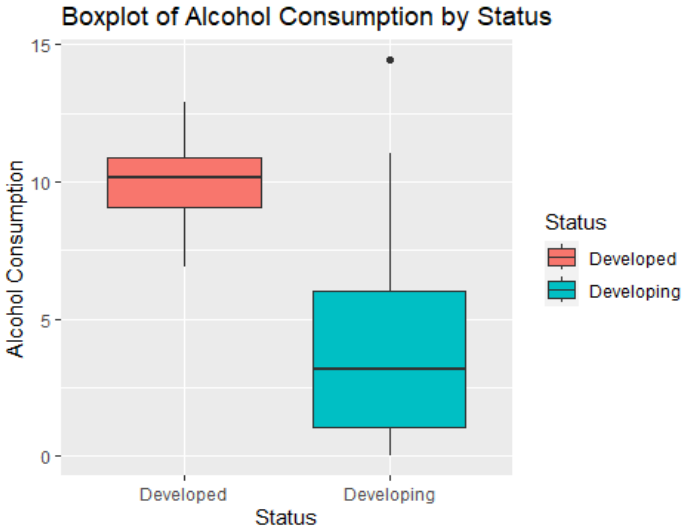

Figure 2:

```
Call:
lm(formula = Life_expectancy ~ Alcohol + Status, data = LED_2010_imputed)

Residuals:
    Min      1Q  Median      3Q     Max
-33.205  -3.551   2.104   4.770  16.939

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      77.0276     2.8078  27.433  < 2e-16 ***
Alcohol           0.1602     0.2205   0.727 0.468884
StatusDeveloping -8.4454     2.3182  -3.643 0.000402 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.35 on 118 degrees of freedom
Multiple R-squared:  0.1792,    Adjusted R-squared:  0.1653
F-statistic: 12.89 on 2 and 118 DF,  p-value: 8.681e-06
```

Figure 3:

```
Call:
lm(formula = Life_expectancy ~ percentage_expenditure + Status,
    data = LED_2010_imputed)

Residuals:
    Min      1Q  Median      3Q     Max
-32.007  -2.813   1.477   4.908  18.874

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            76.023044   1.995527  38.097  < 2e-16 ***
percentage_expenditure  0.004998   0.002023   2.471   0.0149 *
StatusDeveloping       -7.897322   1.937368  -4.076 8.33e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.183 on 118 degrees of freedom
Multiple R-squared:  0.2161,    Adjusted R-squared:  0.2028
F-statistic: 16.27 on 2 and 118 DF,  p-value: 5.76e-07
```

Figure 4:

References:

"Data Collection Tools - Who." *World Health Organization*, www.who.int/data/data-collection-tools. Accessed 2 June 2023.

"Health Service Data - Who." *World Health Organization*, www.who.int/data/data-collection-tools/health-service-data. Accessed 2 June 2023.

"Modules." *World Health Organization*, www.who.int/data/data-collection-tools/health-service-data/toolkit-for-routine-health-information-system-data/modules. Accessed 2 June 2023.