Investigating the Effect of Diet on Death due to Covid-19

STAT 3080: Project Part 3 (Brittany Nguyen)

```
## [1] "/Users/brittanynguyen/Desktop"
```

Two-Sample t-test: The Meat Food Group

```
###MEAT FOOD GROUP
#Countries with high proportion of deaths due to Covid-19:
samplemeat<-coviddata[which(coviddata$Deaths>=0.02),10]
#histdata<-data.frame(samplemeat)
#ggplot(histdata, aes(x=samplemeat))+geom_histogram(binwidth = 0.8,
#     fill="white",color="black")

#Countries with low proportion of deaths due to Covid-19:
sample2meat<-coviddata[which(coviddata$Deaths<0.02),10]
#histdata2<-data.frame(sample2meat)
#ggplot(histdata2, aes(x=sample2meat))+geom_histogram(binwidth = 0.3,
#    fill="white",color="black")

t.test(samplemeat, sample2meat, mu=0.05, alternative="greater")$p.value

## [1] 0.0003384296

#Reference 6: Professor Martinet's Office Hours
```

Primary Question

The question being investigated in this project is whether the mean consumption of certain food groups by countries containing high percentages of deaths due to Covid-19 is more

than 0.05 (percent) greater than the consumption of the food groups by countries with low proportions of deaths due to Covid-19. Alternatively, "is the difference in the mean consumption of the food groups greater than 0.05%?" If such a difference exists, then conclusions can be made as to the correlation between consumption patterns and Covid death patterns. In non-statistical terms, the question becomes: "Which food group consumption patterns are linked with high proportions of deaths due to Covid-19?" The meat, vegetable, alcohol, fruits, and milk groups were used.

Explanation of Data

The first part of data involves a breakdown of the diet composition of the countries into different food groups (meat, vegetables, alcohol, fruits, milk, etc.) expressed in percentages of the diet (in kilograms) attributed to that food group (Ref.3).
The second portion of the data contains health information relating to the effects of Covid-19 in each country, with variables including deaths, confirmed cases, and recovered (all expressed as percentages of population).(Ref.4 and 5)
This data is appropriate for answering the question because it contains a large set of countries (170), with a good split of countries that have either had large or small percentages of deaths due to Covid-19 (Reference 2). These are the two main samples being compared for each food group (Reference 1).

Justification of Two-Sample t-test

The question was answered utilizing a two-sample t-test for means for each food group tested where each independent sample was obtained by splitting the data into the countries containing high percents of deaths due to Covid and the countries containing low percents of deaths due to Covid.
The two-sample t-test is classified as a parametric, quantitative test which is appropriate for the data, since the sampling distribution of the sample means is known, and because the data for food group consumption is numeric (percentages). This test was also appropriate for the given data because all of the conditions for the test were satisfied. First, using histogram visualizations it was confirmed that the two independent samples used in each food group test were large enough for the Central Limit Theorem to hold (according to normality and shape). Additionally, the t.test() function was used into which both samples were inputted along with mu=0.05,corresponding to whether or not the difference in the means was greater

than 0.05%. Lastly, a t-test specifically was chosen instead of a z-test because the population standard deviation was unknown.

Explanation of the Super Population

The test utilizes the data set as the "sample" data, with the intent of extrapolating the results and conclusions to a "super population" scenario in which a similar pandemic would happen in the future. The answer to the primary question of which food group consumption patterns are linked with high percentages of deaths due to Covid could be applied to the super population. For example, if it was determined that the difference in mean consumption of meat was greater than 0.05%, then higher meat consumption is linked to more Covid-related deaths. This knowledge can be used by experts in the super population to advise the public.

The Results and Conclusions

When the meat food group test was run, the p-value of 0.0003384 and thus the alternative hypothesis was accepted. In non-statistical terms, it was concluded that the mean consumption of meat for the sample of countries containing greater percentages of deaths due to Covid was greater than 0.05% plus that of the sample of countries less negatively impacted. Perhaps
a diet rich in meat is linked to complications and death due to Covid. The magnitude of this p-value was also very small, indicating strong evidence for such a conclusion of the differences in means. When the vegetable food group test was run, the p-value of 0.5157 did not result in a rejection of the null hypothesis. In non-statistical terms, it could not be concluded that the true difference in the mean consumption of vegetables for the two samples of countries was greater than 0.05%. This logically makes sense since vegetables are considered as healthy, and it would be unexpected to see a larger veggie consumption from the group more negatively impacted by the pandemic. Three other food groups, alcohol, fruits, and milk, were tested similarly. All three tests indicated that the countries suffering more from Covid had greater than 0.05% more consumption of that food group than did the countries less impacted. The results are logical since alcohol, fruits, and milk contain sugar and hormones that may be correlated with Covid complications. In conclusion, it can be concluded that diets rich in vegetables are the most consistently correlated with low percentages of deaths due to Covid-19 while diets rich in meat are linked to large percentages of deaths due to Covid-19.

References

1.https://www.kaggle.com/mariaren/covid19-healthy-diet-dataset?select=Food_Supply_Quantity_kg_Data.csv
2.https://www.theguardian.com/world/2020/oct/16/coronavirus-world-map-which-countries-have-the-most-covid-casesand-deaths
3.https://www.fao.org/faostat/en/#home
4.https://www.prb.org/
5.https://coronavirus.jhu.edu/map.html
6. Professor Martinet's Office Hours