

Transcript

Note: This transcript was automatically prepared using Adobe Premiere Pro and corrected only for transcription typos.

Hello, this is Matthew Bullen, and this is my final presentation for Numerical Analysis April 2025. Let's get started. First, the scope of the project was to evaluate the statistical significance of a variety of demographic factors drawn from the 2011 Health Survey for England. In particular, analysis of demographic factors that were relevant to alcohol consumption levels by adults.

For a quick executive summary of the findings, I found that men were more likely to report consuming alcohol for women, but a large percentage of both genders reported consuming alcohol regularly. That percentage was well over 70% for each. Despite high alcohol consumption rates for both genders, a review of the current literature found evidence of disparities by gender in terms of how alcohol consumption is studied and by which investigative framework, that being medical, legal, sociological, public policy, or otherwise.

And, similarly, public health policies in general that were focused on alcohol consumption exhibited a distinct lack of awareness of gender differences and modalities of consumption. In other words, public health policies trended to assuming that gender does not matter while the literature, and the findings from this project, suggest that it does matter.

So to jump into the stages of the project, the first stage was to do some preliminary data cleaning, which I did by importing the source file, taken directly from the UK survey results, which I then converted to CSV format. I trimmed the CSV file columns to only those I knew I would need to use in the project, and then did a sweep of all row values to replace missing or otherwise marked as unusable values with a standardized placeholder. That placeholder is "NA" in R Studio. I then imported the data into R Studio as a new data frame. And on this slide, we see a clip of the first 25 values of the culled data frame.

The first stage of project was to run some basic descriptive statistics. For example, based on the total sample size of 10,617 participants, what percent consumed alcohol? What was the highest educational level? The second step was to find some basic values such as the mean, median, mode, mineral maximum range, and standard deviation for three columns of data, which were household size, BMI and age at last birthday. I think the significant thing to note about the results was that BMI did not have a mode. It looks like body mass index across the sample was unique for each person, which is reasonable and aligns with common sense.

And on this slide, we see the results for gender. And it is clear that men, at

approximately 84%, reported consuming more alcohol than women at approximately 74.5%. It's also worth noting that these figures are based on only those values for men and women that made sense. For example, children were excluded. Participants who did not answer or, for whatever other reason, were unable to answer, were excluded.

After the baseline percentages were found, I ran a Chi-square test with the null hypothesis being that there is no significant association between gender and alcohol consumption rates, and the alternative hypothesis being that there is a significant association. And as you can see from the results, the p-value was nearly zero, definitely well under an 0.05 confidence level, which means that the null hypothesis had to be rejected and we had to accept the alternative hypothesis, which led to the conclusion that men were more likely than women to report consuming alcohol.

The second inferential exercise was to repeat the same process for region. For the sake of brevity, I'll skip going over null versus alternative hypotheses going forward, since they're fairly clear. They all repeat the same steps of the first exercise. As we can see from the results, results range between approximately 75% to just shy of 84%, which aligns with the raw percentages based on gender. There was one outlier, for London. The London metropolitan area reported lower alcohol consumption rates, less than 70%.

I think the only other detail to note is that while residents of the South West region, at 83.9%, were the most likely to report consuming alcohol, that figure was in line with several other regions. The South West was simply the region that had the highest numerical percentage.

The next stage of inferential statistics in the project was to investigate if there were any differences in height between the genders. To do that, I first needed to perform a normality test to see if the values in the height column followed the normal distribution. To do this, I used a Shapiro-Wilk test, which came back with an extremely small p-value well under 0.05, meaning that height values did not fall on a normal distribution.

It's important to note, though, that the Shapiro-Wilk normality test can only use the first 5000 row values, not every value, so I thought it would be prudent to run a second test to confirm the results that was capable of evaluating all height values in the data frame, which was the Anderson-Darling test. And as before, I found it came back with an extremely small p-value well under 0.05, and an R-squared value significantly larger than one, which meant that the null hypothesis had to be rejected, and I was able to confirm height values did not fall on a normal distribution.

The presence or absence of normal distribution was important because the next step was to perform an evaluative test, which depended on whether or not the independent variable, gender, excuse me, height, was normally distributed, since height was a continuous numerical float / decimal value.

And from the results of the Mann-Whitney-Wilcoxon test, once again, I found an extremely low p-value approaching zero, which led to the conclusion that there is a significant variance in height between the two actors.

Moving on, the next step in the project was to repeat the same process for weight to determine if there was any detectable association between weight and gender, the exact same process as for height. I first did a Shapiro-Wilk normal distribution test, which returned an extremely small p-value, which led to the conclusion that weight values did not follow normal distribution.

I then repeated, or confirmed, those results with an Anderson-Darling normality test, which also came back with an extremely small p-value, which led to the conclusion that weight values did not fall on a normal distribution. As before with height, I performed a Mann-Whitney-Wilcoxon test, which came back with an extremely low p-value, which led to the conclusion that there is a significant variance in weight between the genders.

The final step for the inferential statistics section of the project was to find correlation values between several variables or several data columns, which were current alcohol consumption, total household income, age at last birthday, and gender. It's worth noting that alcohol consumption and gender are categorical binary values. Yes or no? Male or female? Age at last birthday was recorded as a discrete integer value in terms of age being recorded as a full year, no month or day values.

But household income, in contrast, was a continuous value, and since it was a continuous value, it might not have been normally distributed, and that presence or absence of a normal distribution in turn helped me determine which type of correlative test to use. After applying the Shapiro-Wilk and Anderson-Darling normality tests as before, both of which returned very small p-values, I had to conclude that in household income was not normally distributed, and because it was not normally distributed, the Spearman correlation test was appropriate, as opposed to other varieties of correlation tests that are capable of working, excuse me, incapable of working with anything other than normally distributed data.

And as we see from the findings, there were a number of positive and negative correlations. Each of them was fairly weak. All of them lower than 0.10, excuse me, all of them lower than 0.11, with the highest correlation between gender and alcohol consumption at 0.0174. I think the significance of these results is that while there are correlations present, they are fairly weak. None of them are larger than 0.25 or 0.5. None of them approach positive or negative one.

The final stage of the project was to review the literature for current findings on the association between gender alcohol consumption, which I did via a Google Scholar search. I found in the current literature that men are more likely to report consuming any amount of alcohol and larger quantities of alcohol in general, and they're also more likely to be recorded as experiencing health, legal, and other harms from excessive alcohol consumption.

However, a recent survey of surveys, a meta survey, found that in aggregate, the majority of studies focused on women who consume alcohol, but not from the same perspective. Rather than a medico-legal perspective, in most cases, most studies were from a sociological or cultural perspective. One striking example was a study that looked into women's alcohol consumption in England through the lens of wine-mom culture, which is a culture where alcohol consumption is viewed as normal or

even desirable as a means to cope with the stresses of parenting and to socialize with other mothers or parents.

But aside from specific studies, the most interesting finding that I gained from current literature was that public health policy, in broad strokes, tends to ignore gender differences. Or another way to say it would be that it tends to assume that gender does not matter when considering how to develop public policies around alcohol consumption, which I interpret as inadequate in light of both the current literature suggesting that gender differences do matter, and in light of the findings from, these fiscal policies in the project.

So, in conclusion, my recommendations would be that further research focusing on alcohol consumption by gender through a medical lens would be especially worthwhile for women because they're typically not viewed through that lens, or at least not as often, as well as vice versa. Further research into male alcohol consumption modalities should be undertaken in any form, but it would be helpful to have more sociological and culturally focused studies.

For example, a study of male alcohol consumption at large sporting events would seem like a fairly fruitful area for further research. And, finally, public health policy development should focus on gender aware policies that appropriately, with the emphasis on appropriately, account for and respond to gender differences in alcohol consumption so that potentially beneficial public policies are not overlooked or ignored because of a false assumption that gender does not matter to alcohol consumption rates.

That's my presentation. Thank you for your time.