**Statistical Analysis Presentation**

**based on Health Survey for England 2011**

In this Statistical Analysis Presentation, we are going to talk about some aspects of the population of United Kingdom, including the alcohol issue. The information was provided by Health and Social Care Information Centre ­– usually they are looking at changes in the health and lifestyles of people all over the country. In 2012 the organization had presented the Health Survey for England 2011. The table contains a lot of interesting data for that period of time, and we are glad to use it for today’s practice.

According to the survey of the Office for National Statistics in the UK, which is the executive office of the UK Statistics Authority, in 2016 almost 57% respondents aged 16 and over said they consumed alcohol (which equates to 29 million people in the population). Reading across academic literature we realize, that alcohol affects the brain. Likewise Topiwala and others said, there is no safe level of alcohol consumption for brain health. For example, there is a multiplicative interaction between alcohol and blood pressure and BMI in predicting grey matter volume. Some researchers point that alcohol retail as well as public health authorities are responsible for the situation. Our role is to shed some light on the interrelation within some sample of the population in England, including the issue of alcohol consumption.

We are going performing both descriptive and inferential statistics using RStudio, hence it is free software easy to use even with a basic level of programming, so there is an ideal equation for the presentation.

Let’s started with the descriptive statistics, or “simply statistics”, according to Todd Nick (Nick, 2007). Descriptive statistics often used on a sample to estimate characteristics of a population. The characteristics in statistics are called “variables”, probably due to their ability to vary from individual to individual. First, we describe the dataset. The sample contains 10 thousand six hundred seventeenth observations, in our case it tells how many people are included in the sample in total, also we call it “a sample size”.

73% of respondents among those who gave the answer “yes” or “no” about drinking alcohol gave the positive answer, this level is higher than in survey 2016, which has been demonstrated in the first slide (about 57%). Moreover, keep in mind please that this survey from 2011 has respondents under 16 years old too. Six thousand seven hundred twelve respondents of more than 10 thousand and a half said “yes” for alcohol, it is 63% of the sample size.

54% or respondents are female, 46% are male, you are welcome to find the RStudio code for this and further calculations in the Appendix along with examples of creating pie charts.

The data shows that the highest education level is National Vocational Qualification 4 or 5, Degree or equivalent, 23% of the total sample.

The percentage of divorced and separated people are about 7% and 3% respectively.

Then we dive into specific demographics and health observations to see the mean, median, mode, minimum, maximum, range and standard deviation of household size, body mass index and age. There are some tricks with codes in RStudio, for example with *mode* and *mean*, check the Appendix to get the way we solved that problem.

We can discover a tiny difference in body mass index means between “drinkers” and “abstainers” groups and create a density curve for household size. Household size refers to the number of persons in a private household, a range from 1 to 10 is represented in dataset, so we have got mean of 2.85 and mode of 2, and we can see it clearly in the histogram.

Now we talk about Inferential (or Inductive) Statistics, which based on probability to analyze sample data from a population to improve knowledge about the whole population. The question is which gender drinks more alcohol. Let’s

run a significance test to find out. We should prove that there is statistically significant difference between male and female in alcohol consumption, according to the data from 2011. Gender and drinking are categorical variables, – therefore, we perform Chi-square test to find a relationship between them. (Gonzalez-Chica et al., 2015).

Firstly, we formulate the null hypothesis, that there is no association between gender and alcohol. The alternative hypothesis said that there is an association between gender and alcohol. Beforehand we set the threshold of alpha level on 0.05. That means if p-value will be lower than alpha, we can reject the null hypothesis. After applying a code for Chi-square test we have got the test p-value 2.2e-16, it means in the scientific notation of 0.00000000000000022 – almost zero. It is much lower than alpha, hence we can reject the null hypothesis and conclude that there is an association between gender and drinking alcohol in England.

Indeed, if we go intuitively and draw pie charts, we would distinguish male and female alcohol consumption levels: 84% of male respondents answered positive about drinking, 16% - negative. Among female respondents, who gave the answer about alcohol, the ratio is 74% (“yes”) versus 26% (“no”).

Transit to the next question - which region in England drinks the most alcohol. This time let’s get started with the pie charts. After providing the calculations in RStudio and rounding the results, we are ready to say, that respondents from South West said more frequently about their drinking status: 84%, which is far away from London results – 69%. We are performing a few tests here. For example, one of them is Kruskal-Wallis chi-squared test, which is appropriate for nominal and categorical data if we are working with 3 or more nonparametric variables. All the testes show the p-value lower than alpha threshold with 5% level of significance, so we can reject the null hypothesis and declare that there is an association between region and alcohol consumption.

To present the example of using parametric test, let’s investigate whether there is a statistical difference between men and women in height and weight variables.

The appropriate method is two-sample t-test also known as the independent samples t-test. It is used to test whether the unknown population means of two groups are equal or not (Keselman et al., 2004). The null hypothesis for the test sounds like: two groups have equal means in height. And the same for weight: two groups have equal means in weight. According to the t-scores, degrees of freedom and p-values calculations, we have got p-values are less than 0.05 alpha threshold, so we can conclude these results are statistically significant, and reject the null hypothesis for both height and weight. True difference in means between male group and female group is not equal to 0.

These numbers in the table prove that in the case of height the difference in male and female means is about 10 centimetres, and the difference in male and female weight means is about 10 kilos.

Now we come back to the alcohol issue and trying to find the correlation between whether a person drinks nowadays, income, age and gender. This data in the table shows a very weak positive relationship, close to zero, for all variables in the table ­­– hundredths or thousandths of one percent. However, for the pair Drinking-Gender positive correlation value is about a tenth of a percent, it is still weak. Here some details on the test results. The name of correlation coefficient – Pearson's correlation, t-statistics is equal to 10.782, degrees of freedom 8 thousand 5 hundred and 32, 95% confidence interval is between 0.094 and 0.136, p-value is highly significant, it indicates that the model fits the data well. The rounded positive correlation level between gender and alcohol consumption is 0.12.

The major findings were as follows.

First. There is a little difference in body mass index means between people who answered “yes” about alcohol consumption rather than responders with answers “no”.

Second. In England people drink the most in South West of England, the least in London.

Third. We can conclude that men drink more than women, and our conclusions have been confirmed by other studies.

According to Wilsnack, men are more likely to consume alcohol than women are, it is cross-cultural phenomenon. Other researchers found associations between a wider variety of occupations and the risk of heavy alcohol consumption, that is interesting, it provides the space for further explorations and discussions.

**References:**

Health Survey for England - 2011, Health, social care and lifestyles N (2012) HS Digital. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2011-health-social-care-and-lifestyles [Accessed 25 August 2022]

Windsor-Shellard, B. (2017) Adult drinking habits in Great Britain: 2005 to 2016. The Office for National Statistics. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/bulletins/opinionsandlifestylesurveyadultdrinkinghabitsingreatbritain/2005to2016#:~:text=1.-,Main%20points,million%20people%20in%20the%20population [Accessed 25 August 2022]

Topiwala, A., Ebmeier, K.P., Maullin-Sapey, T. and Nichols, T.E. (2021) No safe level of alcohol consumption for brain health: observational cohort study of 25,378 UK Biobank participants. medRxiv.

Nicholls, J. (2015) Public health and alcohol licensing in the UK: challenges, opportunities, and implications for policy and practice. *Contemporary Drug Problems*, 42(2), pp.87-105.

Nick, T.G. (2007) Descriptive Statistics. In: Ambrosius, W.T. (eds) Topics in Biostatistics. Methods in Molecular Biology™, vol 404. Humana Press. https://doi.org/10.1007/978-1-59745-530-5\_3

Asadoorian, M. O. & Kantarelis, D. (2005) *Essentials of Inferential Statistics*. University Press of America.

Gonzalez-Chica, D. et al. (2015). Test of association: which one is the most appropriate for my study? *An Bras Dermatol.* Jul-Aug; 90(4): 523–528.

Keselman, H.J., Othman, A.R., Wilcox, R.R. & Fradette, K., (2004). The new and improved two-sample t test. *Psychological Science*, 15(1), pp.47-51.

Wilsnack R.W., Wilsnack S.C., Kristjanson A.F., Vogeltanz-Holm N.D. & Gmel G. (2009) Gender and alcohol consumption: patterns from the multinational GENACIS Project. *Addiction*. Sep; 104(9): 1487–1500.

Dempster, S. (2011) I drink, therefore I’m man: gender discourses, alcohol and the construction of British undergraduate masculinities. *Gender and Education*, 23:5, 635-653, DOI: 10.1080/09540253.2010.527824

Thompson, A., & Pirmohamed, M. (2021). Associations between occupation and heavy alcohol consumption in UK adults aged 40–69 years: a cross-sectional study using the UK Biobank. *BMC public health*, 21(1), 1-12.