Notes on Assignment II (s discussed in class on Thursday 1/6/2022)

Present the following analysis in a brief report (max 3 pages):

1. Find the country with the greater consumption of antibacterial, based on worksheet 8,

and report its descriptive statistics (mean, standard deviation, median).

Does your decision agree with the conclusion from Assignment I.

(20 points= 14 + 6 (justification))

For this point you should find the country with the highest consistent consumption in antibacterials. Use data from worksheet 8 to calculate descriptive statistics for each European country. Specifically you should look at the mean, median & sd for columns B:F and K:L. The country with the highest mean and the lowest sd, should be the one selected.

FOR THE REMAINING POINTS YOU SHOULD COMPARE DATA FOR THE COUNTRY ASSIGNED TO YOU AND THE COUNTRY WITH THE GREATEST ANTIBACTERIAL CONSUMPTIONS FOUND IN 1.

2. Use a parametric test to compare the 2 countries (the one found in 1. and the country assigned to you) in terms of tetracylines consumption for systemic use (ATC group J01), for the community- sheet 1. Discuss your results reporting the hypotheses tested, statistic value, p-value and final decisions.(30 points)

Now (and for points 2-4), we are back at worksheet 1 (ATC group J01). Compare the two countries by using a parametric test. It makes sense in this case to use an independent parametric test because we can assume that the national health systems of the two countries are taking independent decisions, possibly different, also affecting the question of interest in our analysis (i.e. antibacterial consumptions). The open issue is whether to use a z-test or a t-test to compare the population means of antibacterial consumptions for the two countries. Form the hypotheses as non-directional.

3. Use a non-parametric test to compare the 2 countries in terms of tetracylines consumption for systemic use (ATC group J01), for the community- sheet 1. Discuss your results reporting hypotheses tested, statistic value, p-value and final decisions.(30 points)

It is independent populations as discussed above, thus the only non-parametric statistic option is the Wilcoxon-Mann-Whitney test.

4. Justify the decision to use a non-parametric test. Do you think there are reasons to use a parametric test instead? (20 points)

Decide based on the normality of the data and the sample size, since the choice of a parametric rather than a non-parametric test depends on whether data can be assumed to follow a normal distribution. Use the right tools to prove it, that is qq-plot or Shapiro-Wilks test.

Your report to should include the following sections:

1. Summary: a short paragraph explaining the data sets and your findings.

2. Sections dedicated to each of the questions above, showing specific results and graphs if required.

3. Please include your RStudio code in the Appendix.