Lab 1: Question 3.1

World Happiness

Bailey Kuehl, Erik Sambrailo, Jenna Sparks, Roshni Tajnekar

March 7, 2023

We would like to compare the happiness of people in countries with high GDP-per-capita (referred to as GDP going forward) to those from countries with low GDP-per-capita. For this we define two random variables H (happiness of people from high GDP countries) and L (happiness of people from low GDP countries), where both $h, l \in \mathbb{N} \cap [0:10]$. A *two sample t-test* is proposed for this analysis. Because the population variances are unknown, and we assume that they are not equal between the groups, we are assuming a *Welch's t-Test* is desired. We will now evaluate the standard assumptions of this test.

Assumption 1: Metric Scale

Both H & L need to be measured on a metric scale. The Cantril Ladder scale proposed is an ordinal scale, not a metric scale. Measurements of happiness are subjective, and the difference in happiness between any two points on the scale likely vary. Therefore, the mean happiness of the population cannot be derived by taking the mean Cantril Ladder score, which is a foundational requirement for a t-test. **We would be violating this assumption if we proceeded with an ordinal scale.**

Assumption 2: Independence

All $(H_1, H_2, ...H_n)$ and $(L_1, L_2, ...L_n)$ need to be I.I.D. There could be geographical & socio-economic clustering issues with this assumption given that the samples drawn from neighboring countries could have similar cultures where their beliefs around happiness and wealth affect their scoring. A happiness score of "5" for one group may look very different than it would for another. This would violate the assumption of independence. Additionally, time-dependent data is a concern. As we are looking at data from 2018-2020, we could be comparing scores from one country in 2018 to another country in 2020. Happiness scores can vary greatly based on time and it's relation to major events, such as COVID-19 in 2020. **Thus, there are concerns with the IID assumption for this data set.**

Assumption 3: Normality

The distribution of both H & L need to both be normal. Considering both H and C are drawn from an ordinal scale, for which normality cannot be derived, **normality of the variables cannot be assumed.**