**Q3.1 Assumptions of Two-Sample t-Test:**

The Two-Sample t-Test is a parametric test for unpaired data and has three assumptions.

* **Metric Scale**
* **IID** **data**
* **No major deviations from normality, considering the sample size.**
  + In particular, the t-test is invalid for highly skewed distributions when sample size is larger than 30. It may also be invalid for very highly skewed distributions at higher sample sizes.

**Evaluation of the Two-Sample t-Test Assumptions:**

The World Happiness Report uses data from the Gallup World Poll surveys from which we are asked to evaluate the assumptions of a Two-Sample t-Test with regards to whether people in countries with high GDP per capita are more happy or less happy than people in countries with low GDPs.

The Two-Sample t-Test is a parametric test for unpaired data. Dividing the dataset by countries with GDPs greater than the mean and countries with GDPs less than the mean satisfies the unpaired data intent for this test.

**Metric Scale**: The "Life Ladder" variable, which measures happiness, is ordinal and violates the metric assumption of the Two-Sample t-test. Although it’s evaluated on a scale from 0 to 10, there are not logically equivalent levels of happiness along the ladder. Also, the best possible life one person can imagine will vary greatly as will their interpretation of the values along the ladder.

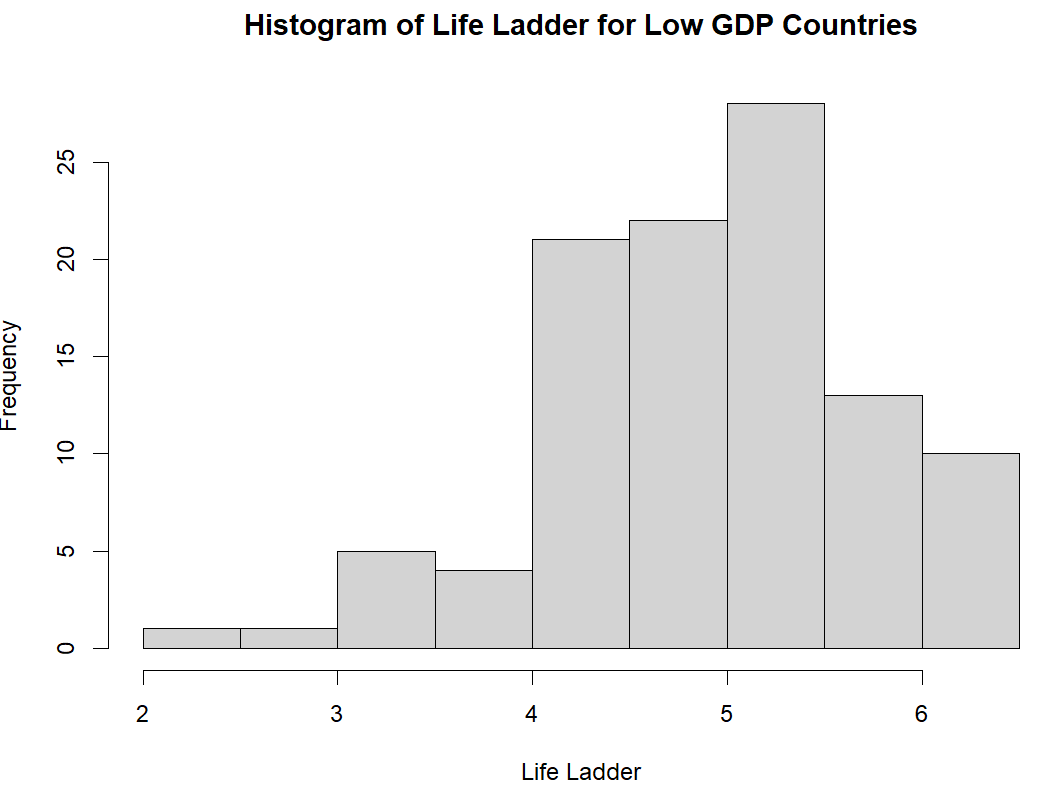
**IID**: The dataset may violate the IID assumption based on possible geographic or temporal clustering.

Although this data is divided by country, GDPs are very connected in a global economy and understanding either the proximity or economic ties between countries may provide insight into the Life Ladder response.

Also, the survey was conducted between 2018 and 2020, so during that time local, regional, or global events may have occurred that impact responses and knowing the location and time of the survey may provide information about the Life Ladder response.

Without knowing more about the means in which the data was collected, at the very least these should be noted as possible limitations to the data.

**Normality**: Normality may be violated based on the skewedness and heavy right tails we see when be evaluate the resulting histograms below.We can also found that the results of the Shapiro-Wilk normality test for each data set resulted in p-values of .011 and .042 for high and low GDP respectively, indicating that both sets are not normally distributed.



**Q3.2 Wilcoxon Rank-Sum Test (Hypothesis of Comparisons) Assumptions**

The Wilcoxon Rank-Sum Test is a nonparametric test for comparing two groups with no pairing.

* **Ordinal Scale**
* **IID data**
  + Each Xi is drawn from the same distribution, each Yi is drawn from the same distribution, and all Xi and Yi are mutually independent.

**Evaluation of Assumptions**

The data from the congress-legislators project contains demographic information about legislators and we are asked to evaluate the assumptions of a Wilcoxon Rank-Sum Test (Hypothesis of Comparison) with regards to whether Democratic or Republican Senators are older.

The Wilcoxon Rank-Sum Test (Hypothesis of Comparisons) is a nonparametric test for comparing to groups with no pairing. Dividing the dataset by Democrat and Republican with the intent of comparing ages meets the intent of comparing two groups with no pairing.

**Ordinal Scale:** Age, derived from the birthday column, is a continuous variable which is beyond ordinal. It can be ranked and is suitable for comparison. Thus, the ordinal scale data assumption is met. However, using metric data in this way will result in a loss of information and likely a loss in overall power. For example, using a metric data in this way you can only apply <,=, and > operations and can no longer utilize simple metric summaries like the mean.

**IID data:** The data may violate IID based on possible geographic clustering. People from various areas, states in this case, tend to vote similarly. For example, knowing that a legislator is from Texas or California may provide insight to the legislator’s age and political affiliation based on tendencies of voters in those states. Without knowing more about the means in which the data was collected, at the very least these should be noted as possible limitations to the data.

Also, dividing the collected data by party affiliation provides the corresponding Xi and Yi distributions. If no senator appears more than once, their ages (and the pairs you form by comparing a Democrat to a Republican) are mutually independent. A simple unique check on identifiers (e.g., bioguide\_id) validates this, and we assume the mutually independent X and Y is not violated acknowledging the above noted geographic clustering limitation.

**Q3.3 Assumptions for a Signed-Rank Test:**

The Signed-Rank Test is a nonparametric test for paired data.

* **Metric scale**: In particular, X and Y are both measured on the same metric scale
* **IID data:** In particular, each pair (Xi,Yi) is drawn from the same distribution independently from all other pairs.
* **Symmetry**: The distribution of the difference X – Y is symmetric around some mean.

**Evaluation of Assumptions**

The wine dataset from the Wooldridge package contains observations and variables related to wine consumption for 21 countries. We are asked to evaluate the assumptions of a Signed-Rank Test with regards to whether countries have more deaths from heart disease or from liver disease.

The Signed-Rank Test is a nonparametric test for paired data. Comparing the number of deaths from heart disease and the number of deaths from liver disease represents unpaired data violates the paired data premise.

**Metric scale:** The two variables being compared are measured as the number of deaths per 100 thousand for the given populations. Therefore, the metric scale assumption is valid.

**IID data:** The data may violate IID based on geographic clustering, specifically with only looking at 21 countries, a significant number of which are European countries. Similar laws, traditions, and tendencies, especially with smaller countries in close proximity, may provide insight into the number of deaths whether linked to wine consumption or not. Without knowing more about the means in which the data was collected, at the very least this should be noted as possible limitations to the data.

**Symmetry**: Symmetry implies that the distribution of differences between the paired sets of data is symmetrical around its central value. Looking at the histogram below, we see that the data is not necessarily appear symmetric, but this may be based on the limited data from 21 countries.

A graph showing different differences

Description automatically generated

**Q3.4** **Assumptions for the Paired t-Test**

The Paired t-Test is a parametric test for paired data.

* **Metric scale:** In particular, the t-test is not valid for variables which only have an ordinal structure.
* **IID data:** In particular, each pair of measurements (Xi,Yi) is drawn from the same distribution, independently of all other pairs.
* **The distribution of the difference between measurements has no major deviation from normality, considering the sample size.**

**Evaluation of Assumptions:**

The dataset is a subset of data from the 2004 General Social Survey (GSS) and consists of measurements of how a respondent feels towards protestants and towards Catholics. We are asked to evaluate the assumptions of a Paired t-Test with regards to whether the US population feels more positive towards Protestants or towards Catholics.

The Paired t-Test is a parametric test for paired data. Since each respondent provides their feelings toward both groups, the resulting data would satisfy the paired premise.

**Metric Scale**: Both the prottemp and cathtemp variables measure the respondent’s feelings towards Protestants and Catholics respectively on the “feeling thermometer” ranging from 0 to 100. The resulting data is ordinal and violates the metric assumption. Although this is evaluated on a scale of 0 – 100, there are not logically equivalent levels of feelings throughout the thermometer’s range and a feeling of 75 for one respondent would likely mean something different to another.

**IID data:** Although GSS is a reputable source of data, the dataset may violate IID based on geographic and regional clustering. This may be found through regional sentiment towards the two groups which may be long term or driven by current events, good or bad, at the local, regional, or global level. This should likely be noted as a limitation.

**The distribution of the difference between measurements has no major deviation from normality, considering the sample size:** Looking at the histogram of the differences below, we can see that there is no major deviation from normality, however, as noted, the data is ordinal so normality does not apply.

A graph of a diagram

Description automatically generated