Lab 1 - Part 1: Founcational Exercises

Applied Practice

Question 1: Professional Magic

Our aunt claims to have magic coins always flipped at the same time. We want to test this because we believe the coins are independent. We also assume: $P(X_i = 0) = P(X_i = 1) = P(Y_i = 0) = P(Y_i = 1) = 1/2$, which means $P(X_i = Y_i) = p$.

Our aunt claims that $p > \frac{1}{2}$

To validate this claim, we design a test. We will flip the coins 3 times and defint our test statistic to be the sum $X_1 + Y_1 + X_2 + Y_2 + X_3 + Y_3$. Our null hypothesis is that $p = \frac{1}{2}$. We will reject H_0 if our test statistic is 0 or 6.

(a) What is the type 1 error rate of this test?

This is the probability we reject H_0 when we should not. We can calculate this by looking at the probability of getting a test statistic of 0 or 6 givent he coins are fair.

Type 1 error = $(\frac{1}{2})^6 * 2 = .03125$

With this test, we have a 3.125% chance of committing a type 1 error.

(b) What is the power of the test for the H_a that $p = \frac{3}{4}$?

We assume $H_0 = \frac{1}{2}$ and $H_a = \frac{3}{4}$ with 6 coin flips.

power $= 2 \cdot \frac{1^6+3^6}{4^6} = 2 \cdot \frac{729}{4096} = 0.35595$

Question 2: Wrong Test, Right Data

If we run a paired t-test on a set of survey results, because the data are on a likert scale, we violate the metric scale assumption of the t-test. This creates a number of problems for us.

1. The responses are equally spaced apart. Neutral to agree is the same step as agree to strongly agree. We are assuming the same internal between each step on the likert scale.
2. The larger implication is that we can average answers. If one responds that they strongly agree, and one responds neutral, we are assuming that on average people agree when this does not accurately reflect the intent of the respondents, mischaracterizing our customers opinions of our websites.

Test Assumptions

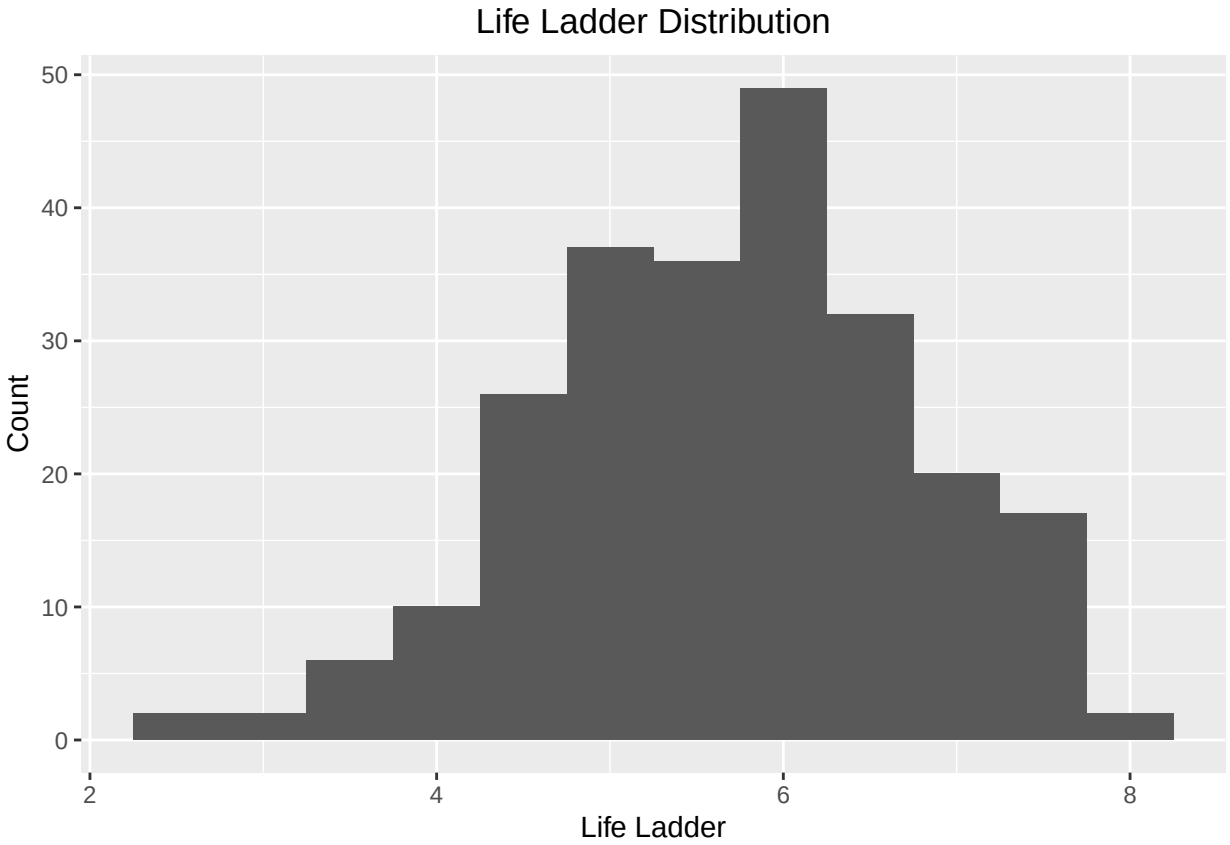
1. Assumptions for a two-sampled t-test to evaluate whether countries with high GDP per capita

(higher than the mean) are more or less happy than people in countries with low GDP (lower than the mean)

We will use both the Life Ladder varaible defining happiness and the log of GDP per capita variable to determine if this test meets the assumptions.

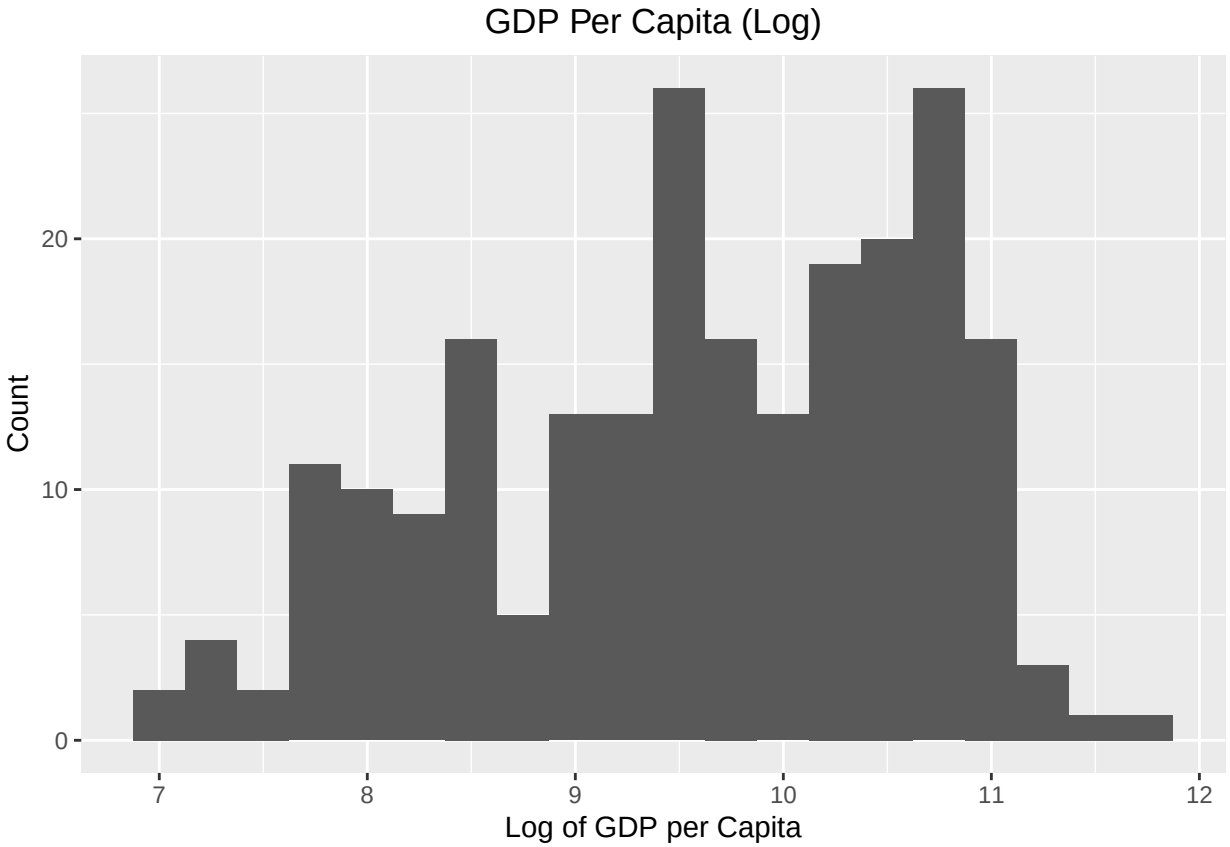
- Assumption 1: The data is metric. In particular, the variables are not oridnal.

  - The first assumption is not met for a two-sampled t-test because the Life Ladder variable that measures respondents' happiness is an oridinal variable. Respondents were asked to rank on a scale of 1 (lowest) to 10 (highest) how happy they were. This value is not metric because the difference in happiness between a ranking of 9 and 10 may not equal the difference in happiness between a ranking of 6 and 7. This means that the measuring system is not uniform and there for not measurable or metric.
- Assumption 2: Data is Independent and Identically Distributed
  - The data is a subset of data from the annual Gallup World Poll. Per Gallup's website the survey is "the most comprehensive and farthest-reaching survey of the world. [The] survey is administered face to face or by telephone in more than 160 countries and areas and in over 140 languages." Additionally, random-route procedures to select sampled households are used, and whether or not telephone interviews are conducted is dependent on if at least 80% of a country has telephone coverage. If not, Gallup uses an area frame design for face-to-face interviewing in randomly selected households. Given this information, we have determined that the second assumption of Independent and Identically Distributed data is met because there should be no dependent variables collected and all variables should have an identical distribution given the collection methods. Most should have access to telephones, if those interviews are being conducted, otherwise the survey method is changed to ensure all households can be reached, and the households are selected randomly with the oldest family member being interviewed in each household.
- Assumption 3: Data has no major deviations from normality given the sample size. In particular, with sample sizes larger than 30 the data is not heavily skewed.
  - To determine if assumption 3 is met, let's look at a graph showing the distribution of the Life Ladder variable that indicates happiness and the GDP per capita.




```
## [1] 239
```

- The above shows a relatively normal distribution for the Life Ladder variable with a sample size of 239.



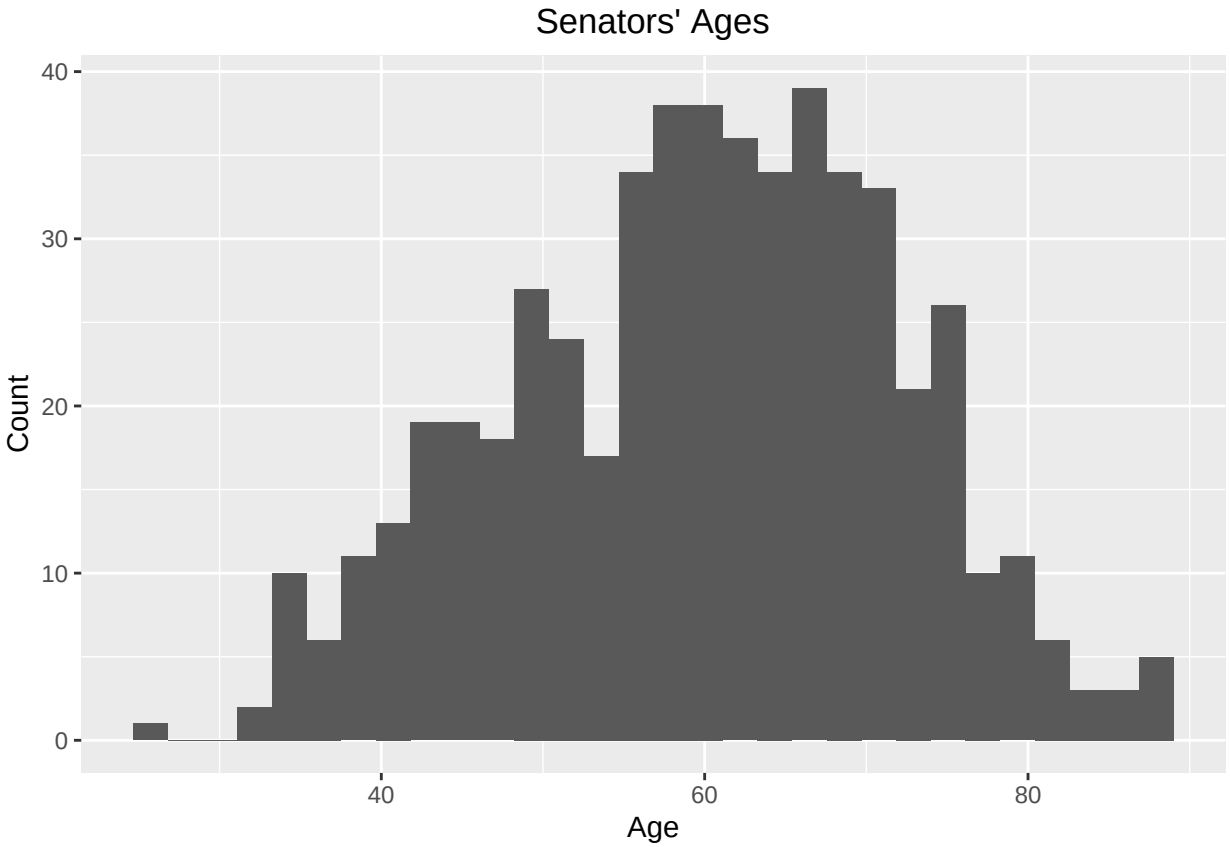
- The GDP per capita does not seem to be normally distributed and therefore assumption 3 is not met.

In conclusion, the ordinal Life Ladder data breaks assumption 1 for a two-tailed t-test, that the data is metric, and the distribution of the GDP per captia is not normally distributed, breaking assumption 3 of normally distributed data.

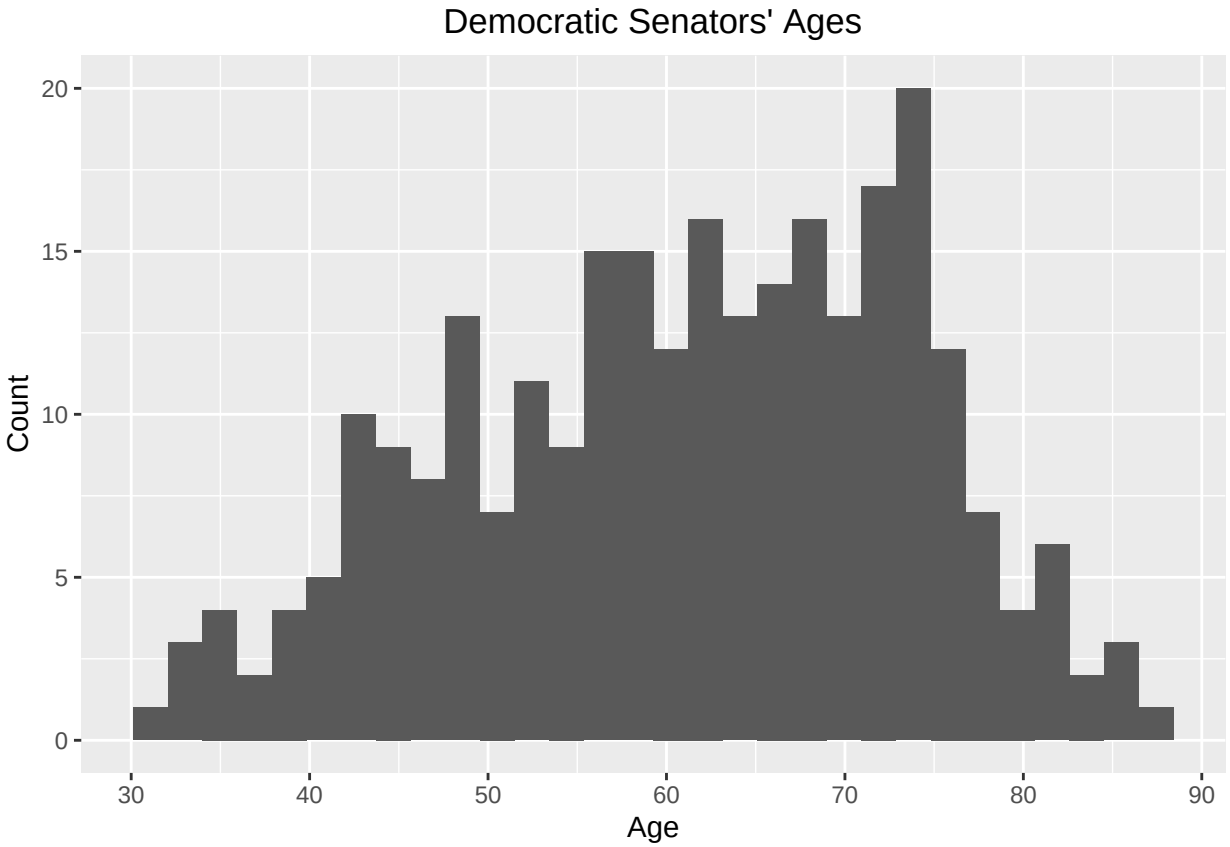
2. Assumptions, and evaluation of assumptions, for a Wilcoxon rank-sum test

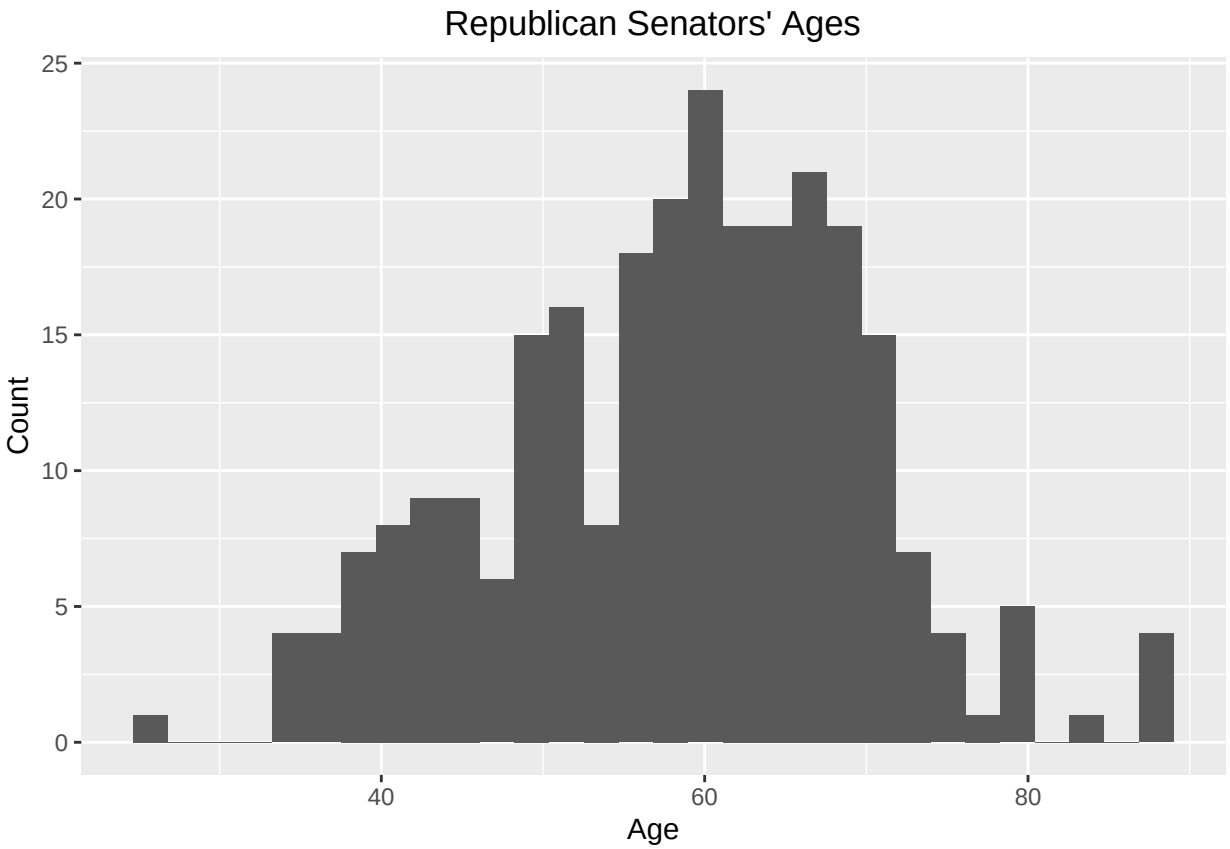
Are Democratic or Republican senators are older. We will use the hypothesis of means version. X variables will be the ages of Democratic senators and Y variables will be the ages of Republican senators.

- Assumption 1: The data is metric.
  - Ages are metric because they are measurable in intervals. The measuring system of birth dates to determine age is uniform, meaningful, and measurable.
- Assumption 2: Data is Independent and Identically Distributed. Each X_i is drawn from the same distribution, each Y_i is drawn from the same distribution, and all X_i and Y_i are mutually independent.



- The data is independent because each senator's age doesn't depend on their party affiliation. The data is identically distributed with a relatively uniform distribution.
- Assumption 3: The distributions of the age of republicans and age of democrats have the same shape, but are just translated from each other (there is/could be a delta).


Democratic Senators' Ages
Count
20
15
10
5
0
30
40
50
60
70
80
90
Age

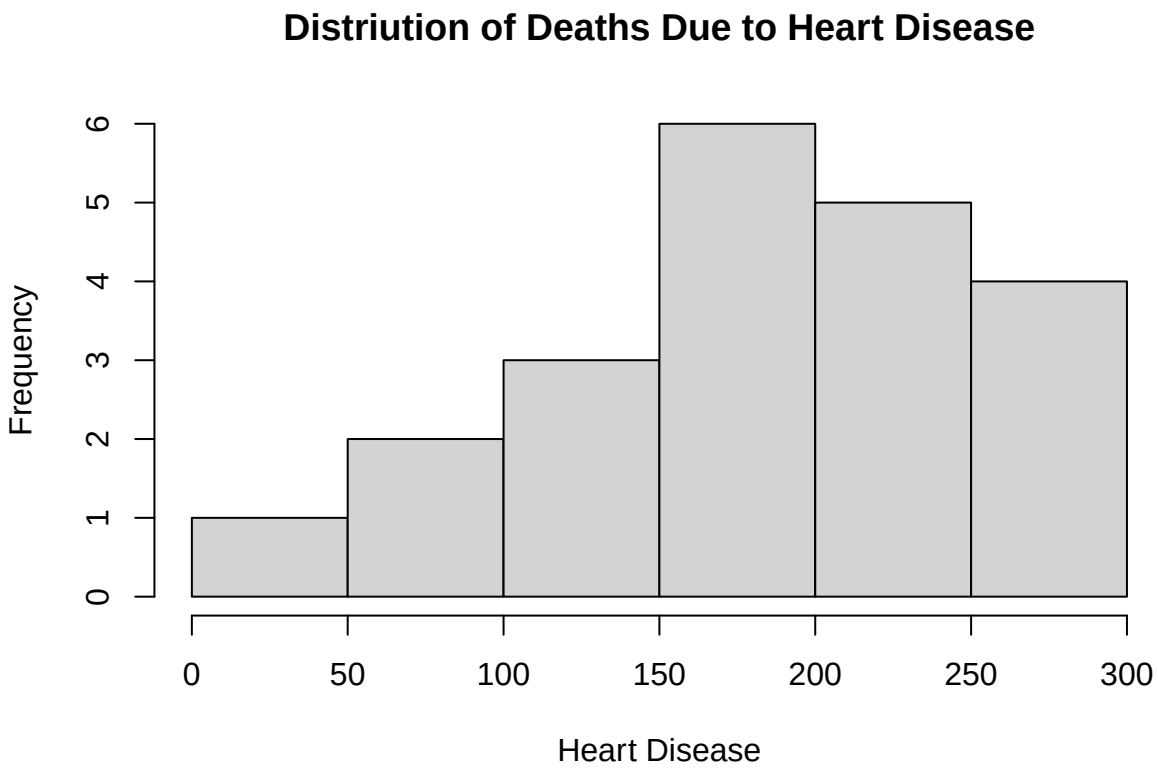


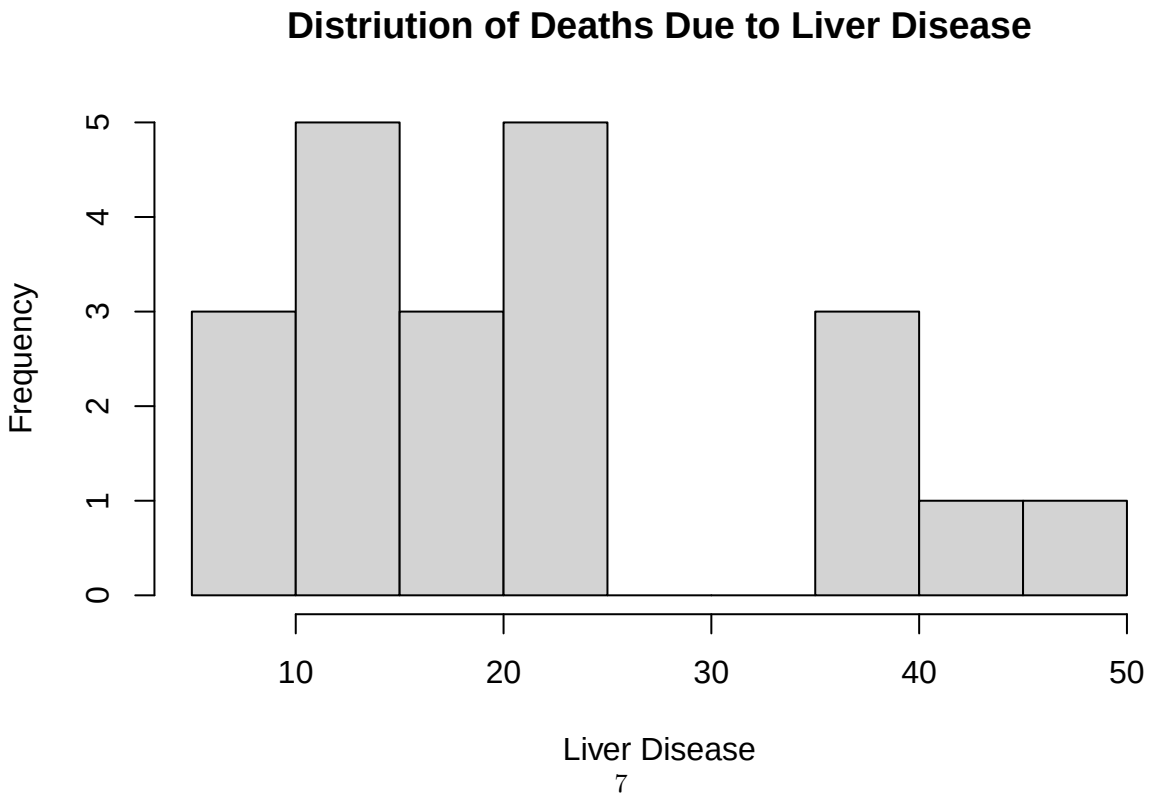
- Both distributions are highly similar but could potentially have different means, meaning the distributions are identical but potentially translated from each other.

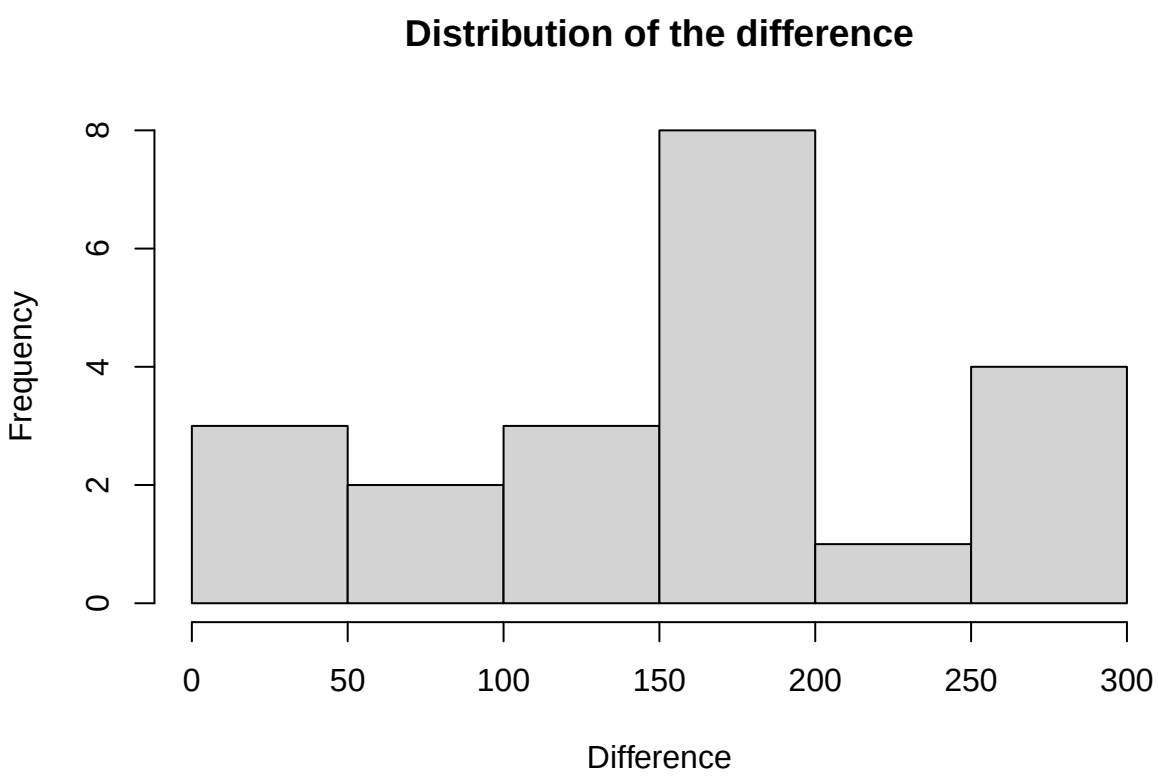
Given the fact that all 3 assumptions are met, the Wilcoxon rank-sum test would be a good test to measure whether Democratic or Republican senators are older.

3. Assumptions for signed test (Wilcoxon Rank Sign Hypothesis of Means Test)

Do more die of heart or liver disease?



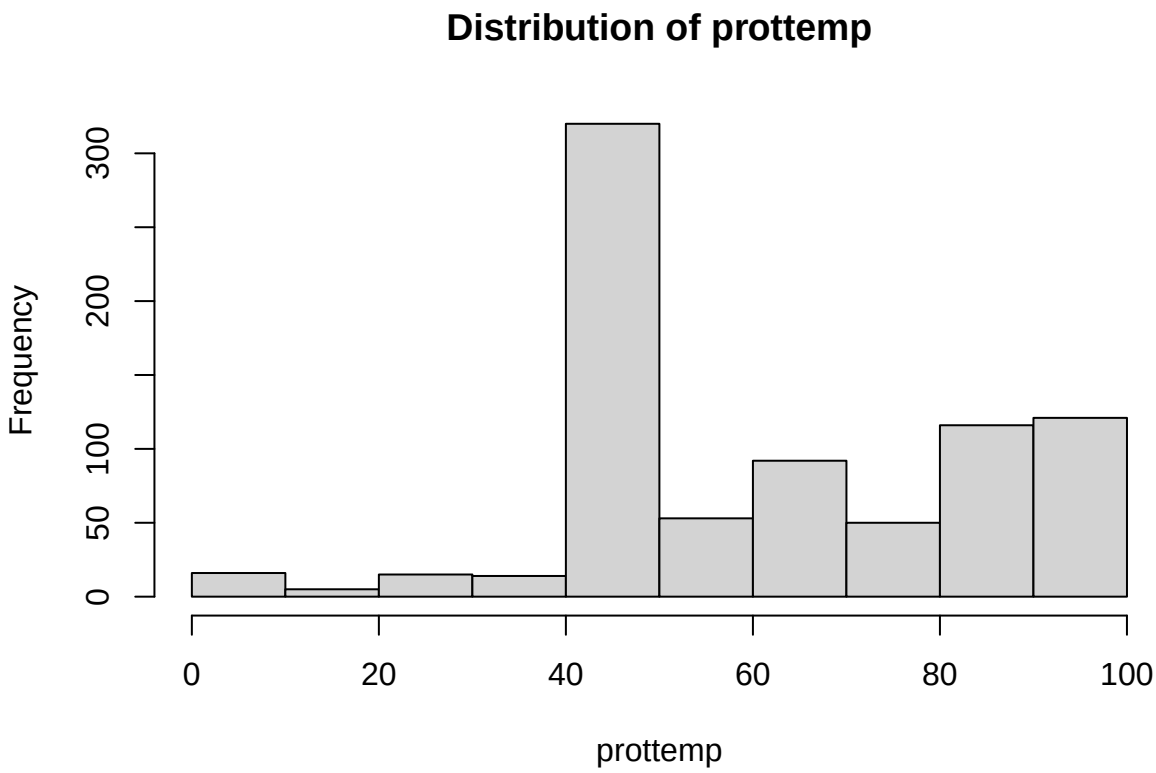


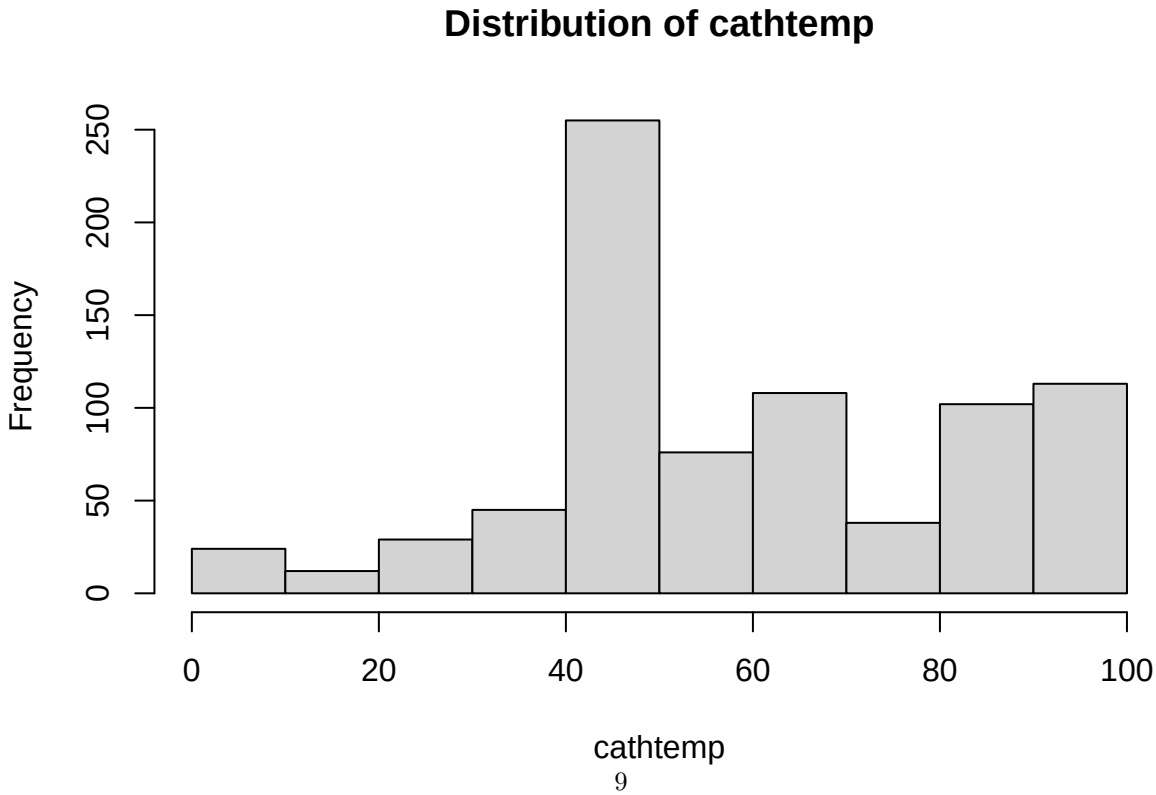


- Non-parametric
  - The data does not have a normal distribution and the sample size is small (n = 21)
- Unpaired Data (no dependency between population)
  - There is no dependency between the populations because the group of people who die from liver disease is completely separate from the group of people that die from heart disease.
- I.I.D.
  - We are assuming data is independent because the population is from random individuals in different countries. The data of people who died from liver disease is distributed similarly but not identically to the data of people who died from heart disease. Because the sample size is so low it is difficult to have the same distribution in both populations. The assumption of IID is not met.
- Metric scale
  - The data is metric because it is a count of the amount of people who die from liver and heart disease. The data is measurable with a uniform difference between each increment.
- Symmetric Distribution
  - The data is NOT symmetrical as shown by the distribution of the difference

4. Assumptions for paired t-test

Do more like Protestants or Catholics





```
## [1] 802
```

- Parametric- Relatively normal distribution
  - The t-test is invalid for highly skewed distributions with sample sized greater than 30
  - The data is not distributed normally and the sample size is over 800 rows of data, therefore the assumption is not met.
- Paired
  - The data is paired because the same sample of people was surveyed and their opinions were recorded on both Protestants and Catholics.
- I.I.D.
  - The data is identically distributed (see graphs above). The data is potentially not independent because someone liking Protestants and Catholics could be dependent on how much they like religious people in general. If someone dislikes all religious people, that would affect both their admiration of Protestants and Catholics. Therefore, the criteria is not met.
- Metric Scale
  - The data is ordinal because survey respondents rated how much they liked Protestants and Catholics on a likert scale from 0 (not at all) to 100 (extremely like). Therefore the criterion is not met.