Lab One, Part One

Kevin Lustig, Rebecca Nissan, Anuradha Passan, Giorgio Soggiu

3/1/2022

Contents

1 Part 1: Foundational Exercises

1.1 1.1 Professional Magic

1.1.1 1.1.1 Type I Error of the test

The type I error rate (i.e. false positive) is the probability of rejecting the null hypothesis whe it is correct. The type I error would be the probability of getting 0 or 6, with the assumption that p = 0.5 (null).

This will be the alpha

1.1.2 1.1.2 Power of test given p = 0.75

1.2 1.2 Wrong Test, Right Data - Kevin

In the Likert scale, the meaningful distance between the different scale points is not consistent. That is, assuming the Likert scale for the websites survey includes five points from 1 = "Very Unsatisfied" to 5 = "Very Satisfied," with 3 being "Neutral," we cannot say that a change from 1 to 3 and from 2 to 4 are equivalent quantifiable changes in opinion [1]. In fact, the change in quality of experience necessary for a given respondent to go from Very Unsatisfied with one site to Neutral with the other may be considerably less than the change needed to go from Unsatisfied to Satisfied, though these each consist of a difference of two points. Therefore, though the values produced are numeric, these data violate one of the assumptions for a paired t-test – the use of metric, rather than ordinal, data.

A paired t-test relies on metric data because, like other related tests including the z-test, it is fundamentally a calculation of the difference of means between reference groups. A paired t-test would ask of our survey data: is the mean difference between paired opinion scores different than what we would expect if there were no preference for either website (mean difference within pairs = 0)? Stated otherwise, on average across all respondents, how likely is it that there is really a preference for one site or the other, and how large a preference? However, because of the aforementioned limitation of Likert scale values, we cannot meaningfully parse a mean paired disparity of e.g., +2, because to calculate this requires assuming that non-comparable changes from any one Likert scale point to another are equivalent. The mean of the paired differences is thus meaningless. It is even difficult to trust the directionality of the mean difference across all pairs (respondents like the mobile website more or less than the regular website without regard to how much), as in calculating a mean value purely from the raw Likert scale scores, we may calculate an incorrect value by not correctly taking into account the "weights" of the differences of opinion in, again, Very Unsatsified and Neutral versus Unsatisfied and Satisfied. It's possible to conceive of a scenario in which even the sign of the mean difference is therefore incorrect.

It is this last point on directionality that suggests an alternative approach to this analysis. A non-parametric paired sign test allows us to analyze our ordinal data provided the observations are independent and identically distributed. It does not attempt, like the t-test, to quantify the size of the difference in opinion within pairs, if any. Rather, it treats all positive changes in opinion as equivalent, and does likewise with all negative changes. This alternative test has two main drawbacks. First, it does not have the statistical power of a paired t-test. Second, it loses substantial information present in the original survey responses in the form of the exact values within each paired set of responses. However, in doing so, it allows us to avoid the inaccurate mean calculation of the t-test, and focus on a more accurate analysis of a simpler question: do respondents prefer one website over the other? In looking solely at increases or decreases in opinion score, the paired sign test therefore gives us a reasonable expectation of finding such an effect if one is present in the data.

[1] At least, not with only five scale points; see, e.g.: Huiping Wu and Shing-On Leung, "Can Likert Scales Be Treated as Interval Scales?—a Simulation Study," Journal of Social Service Research 43, no. 4 (June 2017): pp. 527-532, https://doi.org/10.1080/01488376.2017.1329775.

1.3 1.3 Test Assumptions

1.3.1 1.3.1 World Happiness

Scenario: We have two variables: Life.Ladder and Log.GDP.per.Capita, and we want to see whether countries in high GDP per capita are more or less happy than people in countries with low GDP per capita.

Proposed test: Two Sample t-Test

Test Assumptions: 1. Metric variables 2. Random variables are independent and identically distributed (hereby referred to as i.i.d.) 3. Normalcy of random variables

Both of the variables continuous numeric and metric variables, thus satisfying the first condition.They are also independent and identically distributed based on the fact that the respondents were asked to rank the

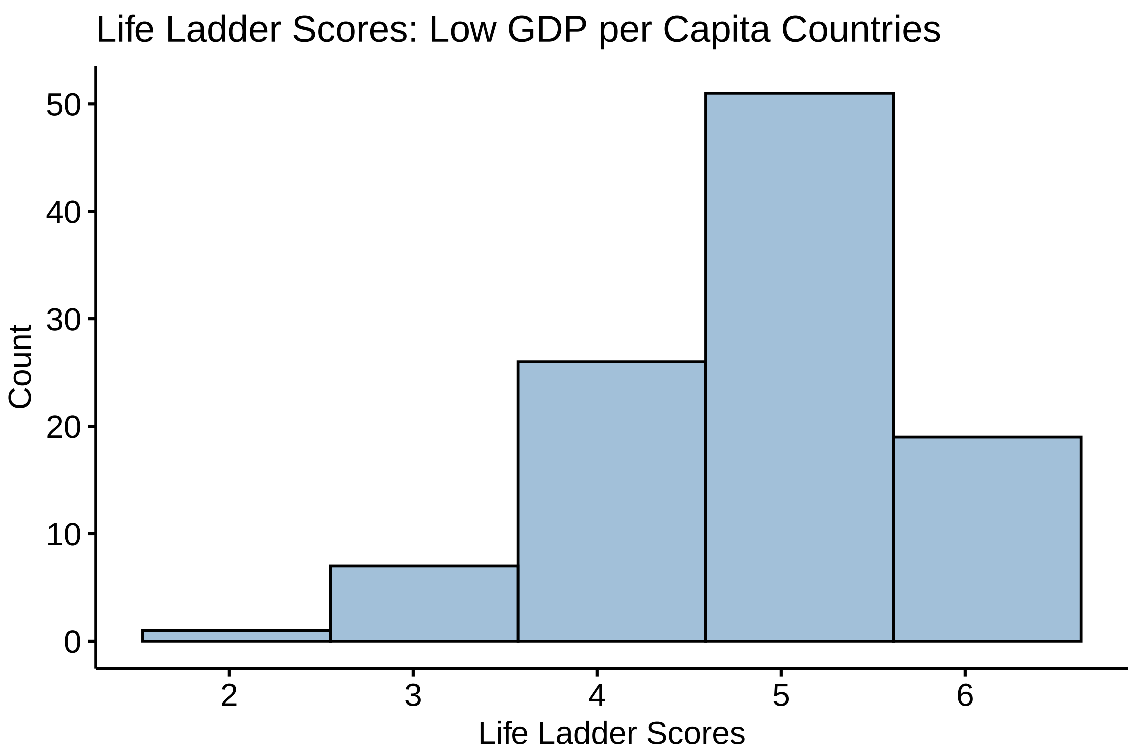
```
  Life.Ladder    Log.GDP.per.capita
 Min.   :2.375   Min.   : 6.966   
 1st Qu.:4.971   1st Qu.: 8.827   
 Median :5.768   Median : 9.669   
 Mean   :5.678   Mean   : 9.584   
 3rd Qu.:6.428   3rd Qu.:10.527   
 Max.   :7.889   Max.   :11.648   
                 NA's   :13       
```

Summary: Life Ladder Score and GDP per Capitas > Sample Mean

```
  Life.Ladder    Log.GDP.per.capita
 Min.   :3.471   Min.   : 9.583   
 1st Qu.:5.917   1st Qu.: 9.993   
 Median :6.291   Median :10.483   
 Mean   :6.349   Mean   :10.415   
 3rd Qu.:7.027   3rd Qu.:10.768   
 Max.   :7.889   Max.   :11.648   
```

Summary: Life Ladder Score and GDP per Capitas < Sample Mean

```
  Life.Ladder    Log.GDP.per.capita
 Min.   :2.375   Min.   :6.966   
 1st Qu.:4.433   1st Qu.:8.100   
 Median :4.992   Median :8.592   
 Mean   :4.919   Mean   :8.609   
 3rd Qu.:5.463   3rd Qu.:9.234   
 Max.   :6.455   Max.   :9.575   
```




```
    Shapiro-Wilk normality test

data:  low_gdp_cap$Life.Ladder
W = 0.97549, p-value = 0.05055
```

p > 0.05 (barely) –> normal-ish

```
shapiro.test(low_gdp_cap$Log.GDP.per.capita)
```

```
    Shapiro-Wilk normality test

data:  low_gdp_cap$Log.GDP.per.capita
W = 0.93951, p-value = 0.0001319
```

```
#https://www.datanovia.com/en/lessons/normality-test-in-r/#check-normality-in-r
```

p < 0.05 –> not normal

```
    Shapiro-Wilk normality test

data:  high_gdp_cap$Life.Ladder
W = 0.97281, p-value = 0.01434
```

p < 0.05 –> not normal

```
shapiro.test(high_gdp_cap$Log.GDP.per.capita)
```

```
    Shapiro-Wilk normality test

data:  high_gdp_cap$Log.GDP.per.capita
W = 0.96977, p-value = 0.00764
```

```
#https://www.datanovia.com/en/lessons/normality-test-in-r/#check-normality-in-r
```

p < 0.05 –> not normal

Conduct the test? NO

1.3.2 1.3.2 Legislators

Scenario: We want to test whether Democratic or Republic senators are older, with two variables party and age (age needs to be calculated from DOB).

Proposed test: Wilxocon Rank Sum Test

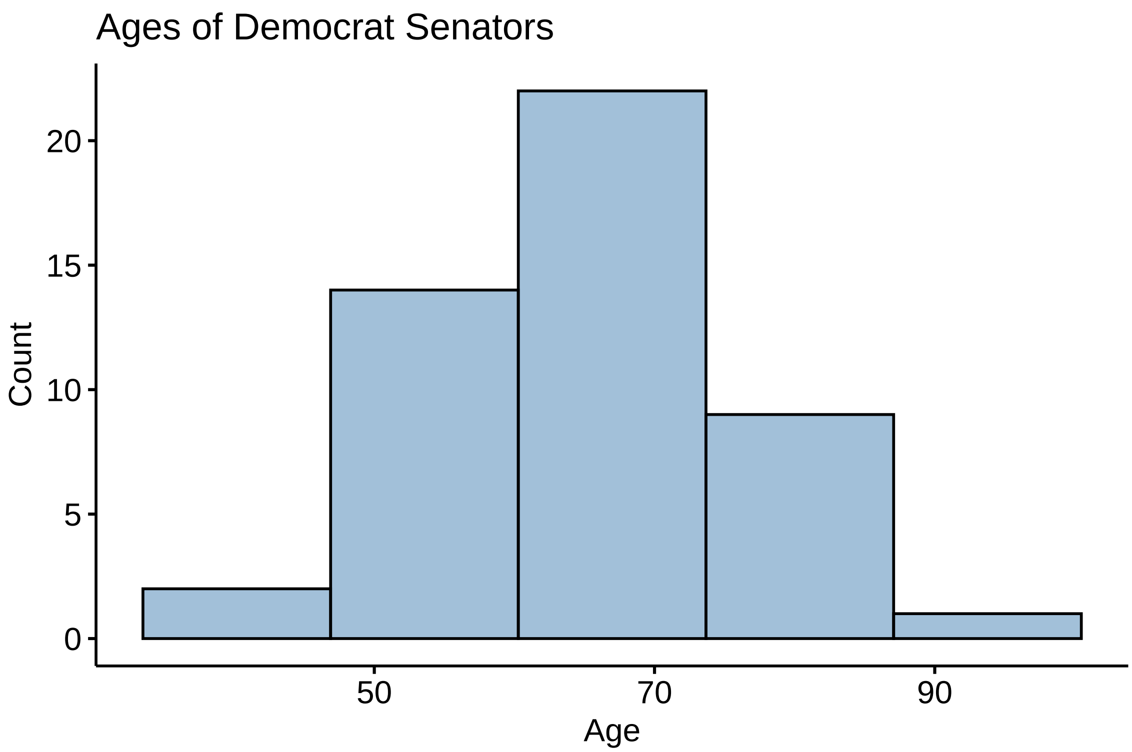
Test Assumptions: 1. Ordinal variables 2. i.i.d. 3. Same shape and spread of the the two variables

Summary: Democrat Senators

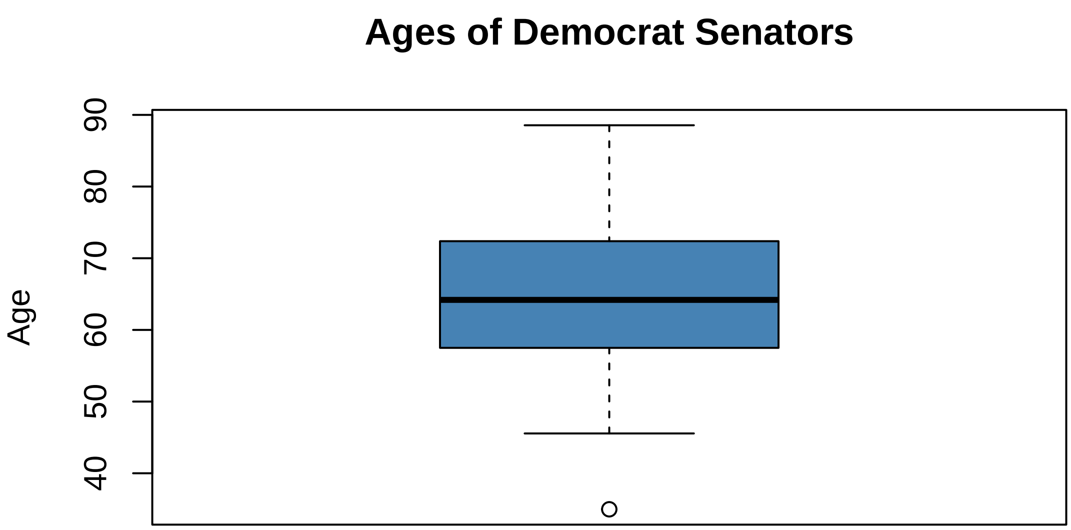
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  34.97   57.66   64.19   64.17   72.28   88.55 
```

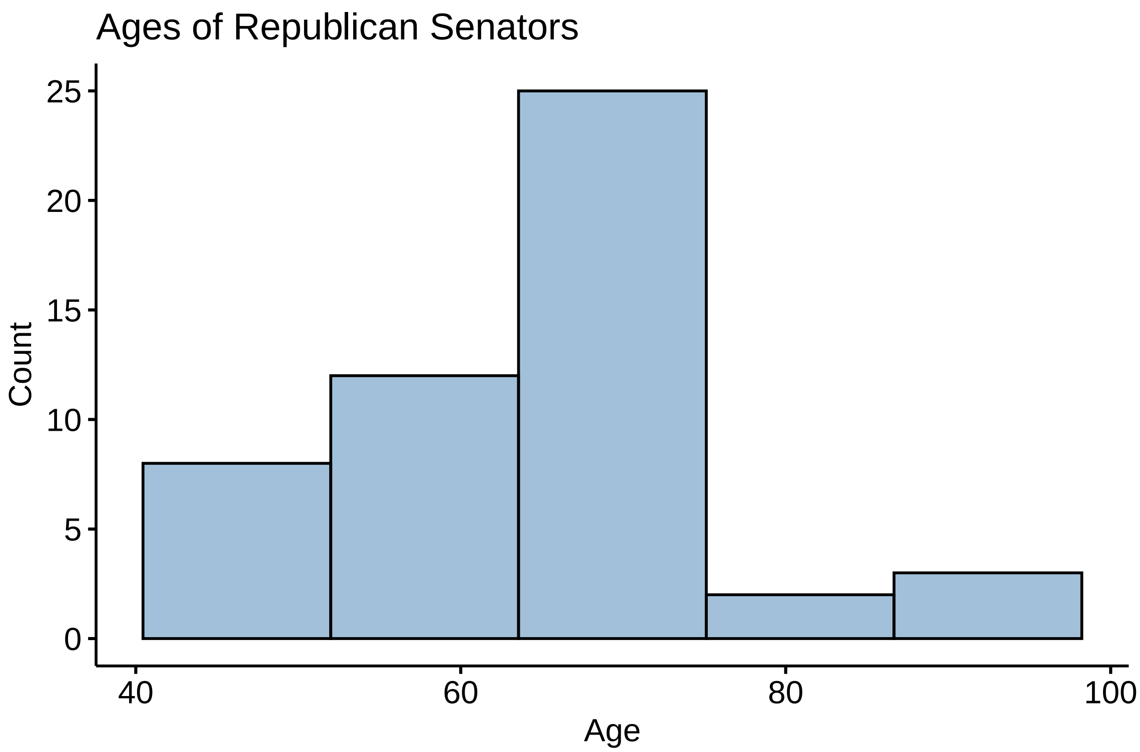
Summary: Republican Senators

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  42.09   59.62   66.34   64.81   69.88   88.31 
```

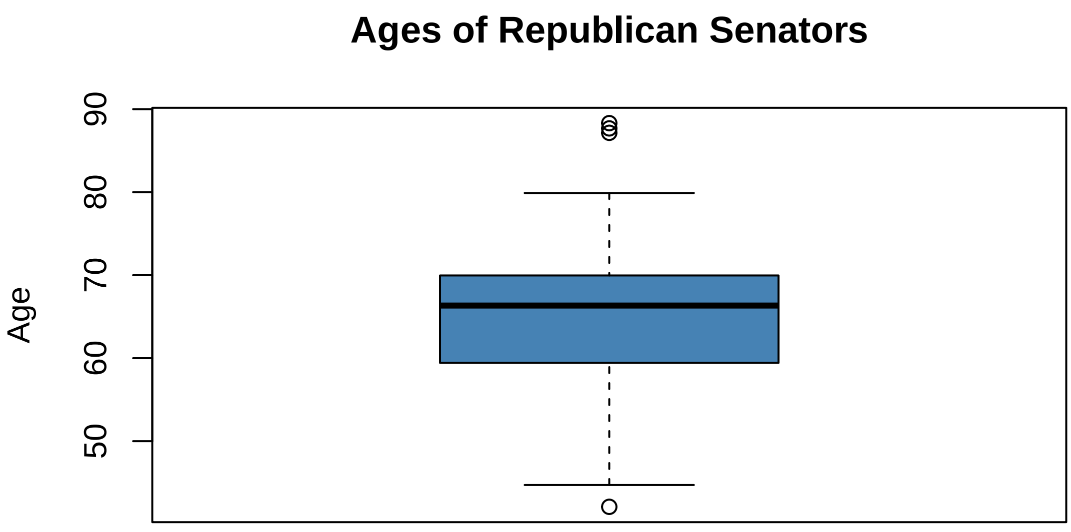



```
boxplot(x=dem_sen$age, data = dem_sen,
        main="Ages of Democrat Senators",
        ylab="Age",
        col = 'steelblue')
```






```
boxplot(x=rep_sen$age, data = rep_sen,
        main="Ages of Republican Senators",
        ylab="Age",
        col = 'steelblue')
```



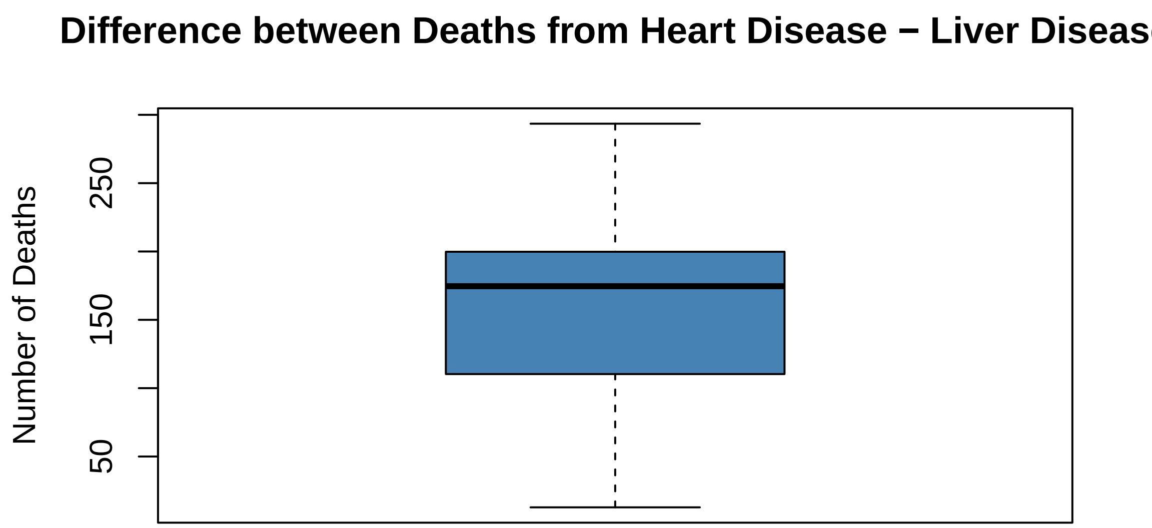
Conduct the test? Maybe yes?

1.3.3 1.3.3 Wine and Health

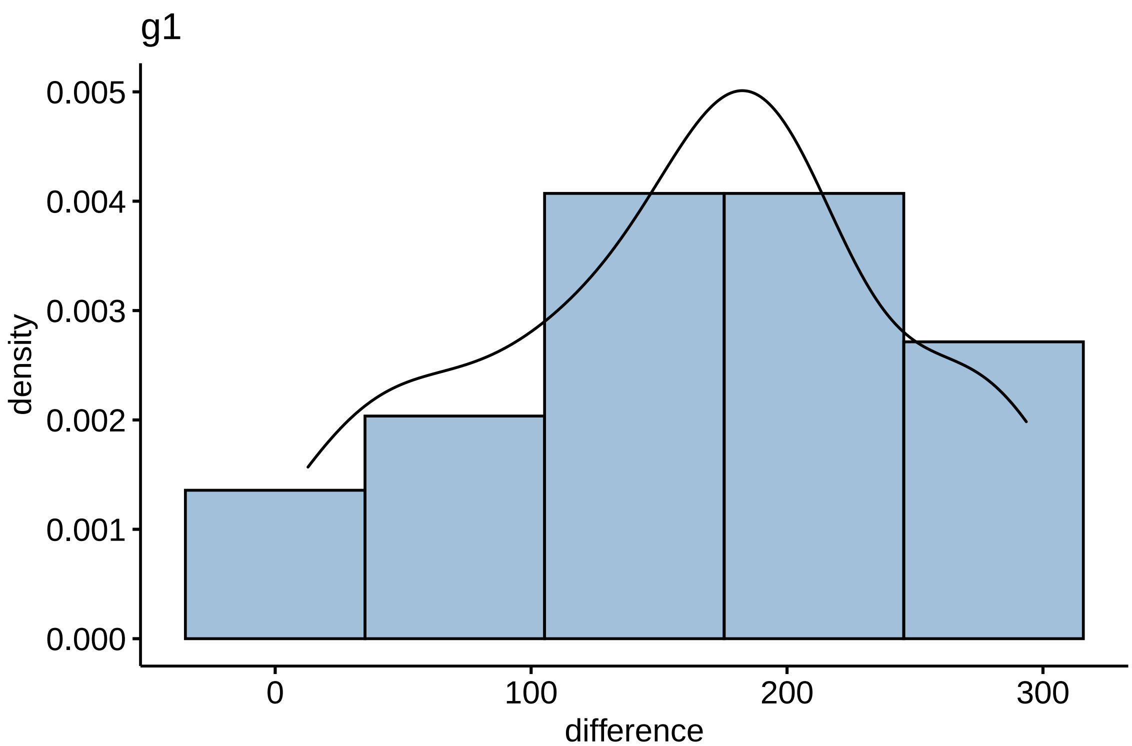
Scenario: We want to test whether these countries have more deaths from heart disease or liver disease.

Proposed test: Wilxocon Signed Rank Test

Test Assumptions: 1. Metric Variables 2. i.i.d. 3. Paired data 4. Difference is symmetric




```
gghistogram(wine_data, main = 'g1', x = 'difference', y = '..density..', fill= 'steelblue', bins = 5, a
```



Conduct the test? Not sure

First box whisker done with advice from https://www.youtube.com/watch?v=Y4-wAT4SNM4&ab_channel =Dr.ToddGrande go to around 6 min

Second chart done with advice from https://www.datanovia.com/en/lessons/wilcoxon-test-in-r/

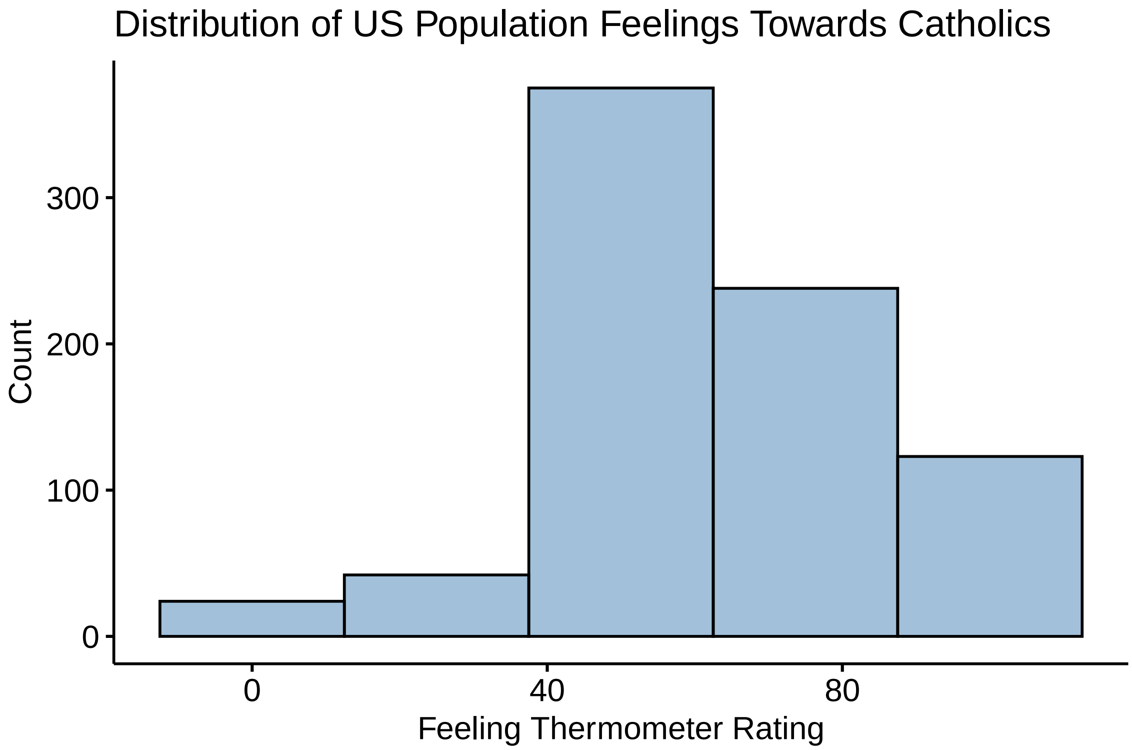
1.3.4 1.3.4 Attitudes Toward the Religion

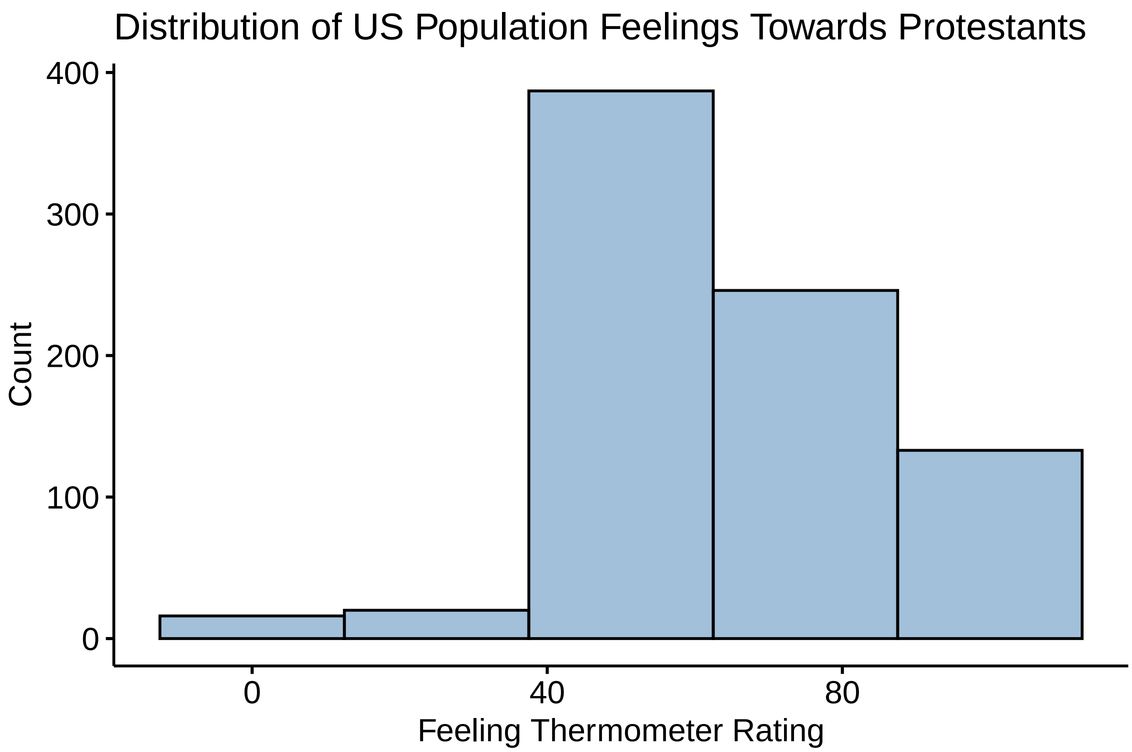
Scenario: We would like to know whether the U.S. population feels more positive towards Protestants or Catholics.

Proposed Test: Paired t-Test

Test Assumptions: 1. Metric Variables 2. i.i.d. 3. Paired 4. Normalcy

```
   cathtemp         prottemp     
 Min.   :  0.00   Min.   :  0.00  
 1st Qu.: 50.00   1st Qu.: 50.00  
 Median : 60.00   Median : 60.00  
 Mean   : 63.16   Mean   : 65.56  
 3rd Qu.: 85.00   3rd Qu.: 85.00  
 Max.   :100.00   Max.   :100.00  
```






```
    Shapiro-Wilk normality test

data:  rel_data$cathtemp
W = 0.93377, p-value < 2.2e-16
```

p-value < 0.05 –> not normal

```
    Shapiro-Wilk normality test

data:  rel_data$prottemp
W = 0.89479, p-value < 2.2e-16
```

p-value < 0.05 –> not normal

Conduct the test? No

Note on Shapiro wilks test The Shapiro-Wilk test is a statistical test used to check if a continuous variable follows a normal distribution. The null hypothesis (H0) states that the variable is normally distributed, and the alternative hypothesis (H1) states that the variable is NOT normally distributed. So after running this test:

If p <= 0.05: then the null hypothesis can be rejected (i.e. the variable is NOT normally distributed). If p > 0.05: then the null hypothesis cannot be rejected (i.e. the variable MAY BE normally distributed).

source: https://quantifyinghealth.com/report-shapiro-wilk-test/