Lab\_1 Problem 1.3: Test Assumptions

QUESTION: For the four following questions, your task is to evaluate the assumptions for the given test. It is not enough to say that an assumption is met or not met; instead, present your evidence in the form of background knowledge, visualizations, and numerical summaries. If you produce a histogram as part of your evaluation, be sure to consider what the most appropriate bin width is. The test that we ask you to evaluate may or may not be the most appropriate test for the scenario. Because the goal of this task is to evaluate whether the data satisfies the assumptions necessary for the test to provide meaningful results, you do not need perform the test (you may perform the test, but we will not be marking for the test results).

# 1.3.1 World Happiness

QUESTION: The file datasets/Happiness\_WHR.csv is subsetted from the World Happiness Report, a yearly publication that uses data from the Gallup World Poll surveys. The variable life ladder is a measure of happiness, described in the FAQ as follows: This is called the Cantril ladder: it asks respondents to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0. They are then asked to rate their own current lives on that 0 to 10 scale. The rankings are from nationally representative samples,for the years 2018-2020. You would like to know whether people in countries with high GDP per capita (higher than the mean) are more happy or less happy than people in countries with low GDP (lower than the mean).

# Solution

## Test Assumptions for a Two Sample t-Test

1. Metric scale (In particular, the t-test is not valid for variables which only have an ordinal structure)
2. IID data
3. 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)

### Metric Scale

Upon evaluating the dataset happinees\_WHR.csv, although the evaluation was on a scale of 1-10, the Cantril Ladder is an example of a Likert Scale meaning that we have a non-metric scale. In other words the values used in the dataset rely upon opinion (qualitative measurement) rather than a quantitative measurement. Therefore a two sample t-test fails on this assumption. View the query below to get an understanding of the happinees\_WHR.csv dataset.

## Country name year Life Ladder Log GDP per capita  
## Length:239 Min. :2019 Min. :2.375 Min. : 6.966   
## Class :character 1st Qu.:2019 1st Qu.:4.971 1st Qu.: 8.827   
## Mode :character Median :2019 Median :5.768 Median : 9.669   
## Mean :2019 Mean :5.678 Mean : 9.584   
## 3rd Qu.:2020 3rd Qu.:6.428 3rd Qu.:10.527   
## Max. :2020 Max. :7.889 Max. :11.648   
## NA's :13   
## Social support Healthy life expectancy at birth Freedom to make life choices  
## Min. :0.4200 Min. :48.70 Min. :0.3850   
## 1st Qu.:0.7590 1st Qu.:62.00 1st Qu.:0.7360   
## Median :0.8480 Median :67.20 Median :0.8220   
## Mean :0.8256 Mean :65.84 Mean :0.8038   
## 3rd Qu.:0.9175 3rd Qu.:70.45 3rd Qu.:0.8910   
## Max. :0.9830 Max. :77.10 Max. :0.9700   
## NA's :8 NA's :2   
## Generosity Perceptions of corruption Positive affect Negative affect   
## Min. :-0.28900 Min. :0.0700 Min. :0.3220 Min. :0.0830   
## 1st Qu.:-0.11900 1st Qu.:0.6470 1st Qu.:0.6450 1st Qu.:0.2250   
## Median :-0.04500 Median :0.7780 Median :0.7330 Median :0.2830   
## Mean :-0.01524 Mean :0.7168 Mean :0.7149 Mean :0.2891   
## 3rd Qu.: 0.07400 3rd Qu.:0.8480 3rd Qu.:0.7900 3rd Qu.:0.3440   
## Max. : 0.56100 Max. :0.9630 Max. :0.8910 Max. :0.5320   
## NA's :14 NA's :14 NA's :2 NA's :2

### IID Data

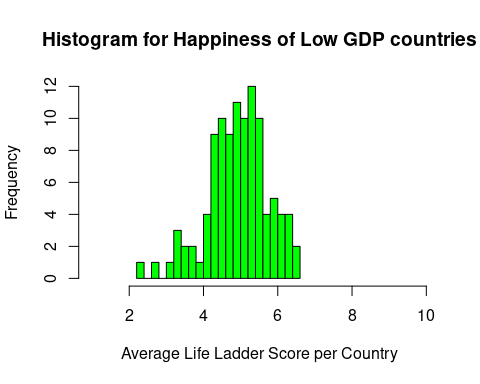
Based on background knowledge, given that each country is fairly independent of one another. There are no apparent violates of independents, some examples being clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. The data also appears to be identically distributed as will be shown in the plots below. This assumption for the data is meet, but as stated previously, a t-test would not be a wise test to use for a non-metric statistical analysis.

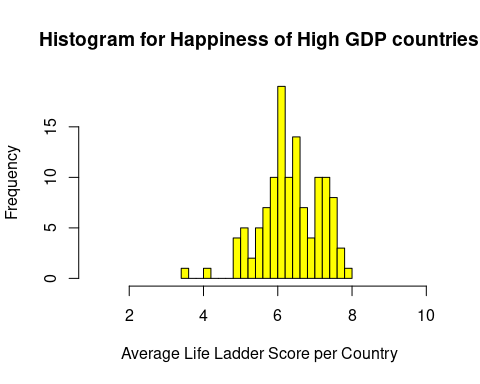
### No Major Deviations from Normality

This assumption would not be met if the distribution was highly skewed for distributions when the sample size is larger than 30. While our data is slightly left skewed, it does not meet the criteria for highly skewed data and therefore this assumption is meet. Please see the histograms below

The steps taken here to create the histograms were to first separate the data based off of high and low GDP as done by splitting across the mean GDP of the given 239 countries.

Next, two histograms were made and show the left skewness of the distribution for high and low GDP countries.





Lastly, a simple unpaired t-test was run to see what the results would have been and given the p-value for this example, it would have been likely that the null hypothesis would have been rejected.

##   
## Welch Two Sample t-test  
##   
## data: h\_GDP$`Life Ladder` and l\_GDP$`Life Ladder`  
## t = 13.251, df = 218.47, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.218086 1.643727  
## sample estimates:  
## mean of x mean of y   
## 6.355488 4.924581

# 1.3.2: Legislators

QUESTION: The file datasets/legislators-current.csv is taken from the congress-legislators project on Github. You would like to test whether Democratic or Republican senators are older. List all assumptions for a Wilcoxon rank-sum test (using the Hypothesis of Comparisons). Then evaluate each assumption, presenting evidence based on your background knowledge, visualizations, and numerical summaries.

# Solution

## Test Assumptions for a Wilcoxon Rank-Sum Test (Hypothesis of Comparisons)

1. Ordinal Scale
2. IID Data

Null Hypothesis: The probability that a draw from ranks higher than a draw from is the same as the probability that a draw from ranks higher than a draw from .

### Ordinal Scale

Based on background knowledge, ordinal scale measures data of categorical nature where ordered categories and the distances between the categories are not known. Since we are measuring age given by the birthday of each congressman in this analysis, and age would be considered a metric variable, this test would fail on this assumption and not be a viable option for statistical analysis. View a summary of the data below.

## last\_name first\_name middle\_name suffix   
## Length:538 Length:538 Length:538 Length:538   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## nickname full\_name birthday gender   
## Length:538 Length:538 Min. :1933-06-09 Length:538   
## Class :character Class :character 1st Qu.:1953-03-30 Class :character   
## Mode :character Mode :character Median :1961-03-05 Mode :character   
## Mean :1961-11-29   
## 3rd Qu.:1970-08-14   
## Max. :1995-08-01   
##   
## type state district senate\_class   
## Length:538 Length:538 Min. : 0.000 Min. :1.00   
## Class :character Class :character 1st Qu.: 3.000 1st Qu.:1.00   
## Mode :character Mode :character Median : 6.000 Median :2.00   
## Mean : 9.984 Mean :2.01   
## 3rd Qu.:13.000 3rd Qu.:3.00   
## Max. :53.000 Max. :3.00   
## NA's :100 NA's :438   
## party url address phone   
## Length:538 Length:538 Length:538 Length:538   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## contact\_form rss\_url twitter facebook   
## Length:538 Length:538 Length:538 Length:538   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## youtube youtube\_id bioguide\_id thomas\_id   
## Length:538 Length:538 Length:538 Length:538   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## opensecrets\_id lis\_id fec\_ids cspan\_id   
## Length:538 Length:538 Length:538 Min. : 260   
## Class :character Class :character Class :character 1st Qu.: 45591   
## Mode :character Mode :character Mode :character Median : 79718   
## Mean : 543374   
## 3rd Qu.:1003305   
## Max. :9275683   
## NA's :156   
## govtrack\_id votesmart\_id ballotpedia\_id washington\_post\_id  
## Min. :300018 Min. : 119 Length:538 Mode:logical   
## 1st Qu.:412199 1st Qu.: 22411 Class :character NA's:538   
## Median :412570 Median : 52964 Mode :character   
## Mean :412042 Mean : 75411   
## 3rd Qu.:412772 3rd Qu.:133024   
## Max. :456862 Max. :188334   
## NA's :62   
## icpsr\_id wikipedia\_id   
## Min. :14066 Length:538   
## 1st Qu.:21106 Class :character   
## Median :21564 Mode :character   
## Mean :24264   
## 3rd Qu.:21972   
## Max. :94659   
## NA's :77

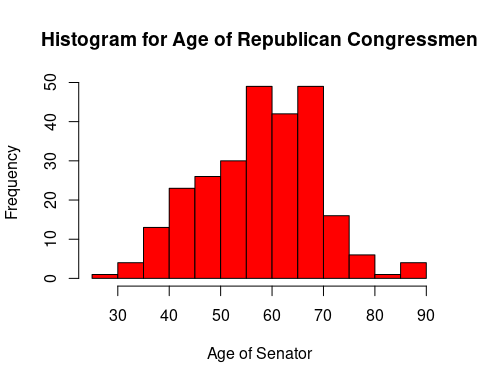
### IID Data

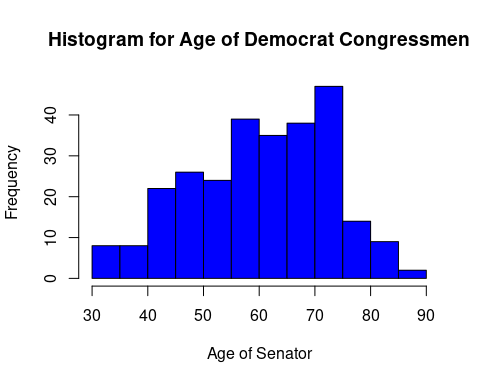
For this assumption, each has to be drawn from the same distribution, each has to be drawn from the same distribution, and all and are mutually independent. There are no apparent violates of independents, some examples being clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. The data also appears to be identically distributed as will be shown in the plots below. This assumption for the data is meet, but as stated previously, the Hypothesis of Comparison version of the Wilcoxon Rank-Sum would not be a wise test to use for a metric statistical analysis.

For the analysis of this dataset, a simple extraction of the each congressmen’s birthyear subtracted from the current year yields their age.

Next, subsetting the data by Republicans and Democrats in rep and dem respectively.

And finally, plotting the resulting of ages for Republicans and Democrats.





And Finally below, I have the Wilcoxon Rank-Sum Test. I’m not sure how to make sure that I’m evaluting the Hypothesis of Comparison version, nevertheless, this version would not be the correct hypothesis test to run because of the failed assumption.

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: rep$age and dem$age  
## W = 31022, p-value = 0.006447  
## alternative hypothesis: true location shift is not equal to 0

# 1.3.3 Wine and Health

QUESTION: The dataset wine can be accessed by installing the wooldridge package. install.packages(“wooldridge”) library(wooldridge) ?wine wine It contains observations of variables related to wine consumption for 21 countries. You would like to use this data to test whether countries have more deaths from heart disease or from liver disease. List all assumptions for a signed-rank test. Then evaluate each assumption, presenting evidence based on your background knowledge, visualizations, and numerical summaries.

# Solution

## Test Assumptions for Wilcoxon Signed-Rank Test

1. Metric Scale
2. IID Data
3. The distribution of the difference is symmetric around the same mean.

### Metric Sacle

Based on background knowledge, the X and Y measured here have to be measured on the same scale since we are using a paired test. For this question, both the liver and heart deaths are on the same scale of 100,000 deaths. This will be shown in the dataset below.

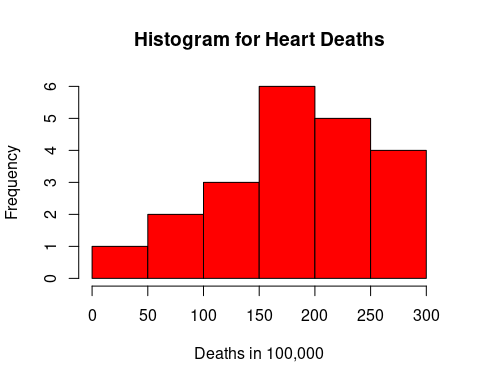
## country alcohol deaths heart liver  
## 1 Australia 2.5 785 211 15.3  
## 2 Austria 3.9 863 167 45.6  
## 3 Belg/Lux 2.9 883 131 20.7  
## 4 Canada 2.4 793 191 16.4  
## 5 Denmark 2.9 971 220 23.9  
## 6 Finland 0.8 970 297 19.0  
## 7 France 9.1 751 71 37.9  
## 8 Iceland 0.8 743 211 11.2  
## 9 Ireland 0.7 1000 300 6.5  
## 10 Israel 0.6 834 183 13.7  
## 11 Italy 7.9 775 107 42.2  
## 12 Japan 1.5 680 36 23.2  
## 13 Netherlands 1.8 773 167 9.2  
## 14 New Zealand 1.9 916 266 7.7  
## 15 Norway 0.8 806 227 12.2  
## 16 Spain 6.5 724 86 36.4  
## 17 Sweden 1.6 743 207 11.2  
## 18 Switzerland 5.8 693 115 20.3  
## 19 UK 1.3 941 285 10.3  
## 20 US 1.2 926 199 22.1  
## 21 West Germany 2.7 861 172 36.7

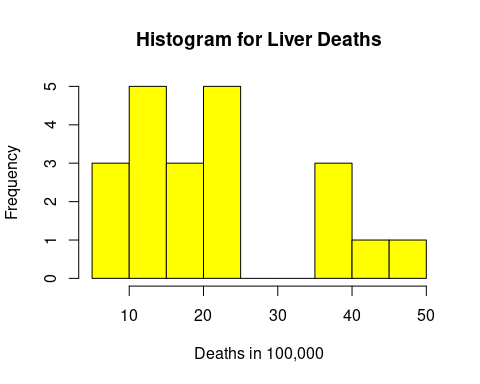
### IID Data

Using background knowledge, each pair has to be drawn from the same distribution independently of all other pairs. Even though this is a small dataset, it appears to be independent due to each observation being a different country and the data also appears to be identically distributed as will be shown in the plots below. In this case, There are no apparent violates of independents, some examples being clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. This assumption for the data is meet and thus far, the Wilcoxon Ranked-Summed Test appears to be a viable test.

For the data visualization, two histograms of heart deaths and liver deaths were made in the plots below.

## [1] 162.2524

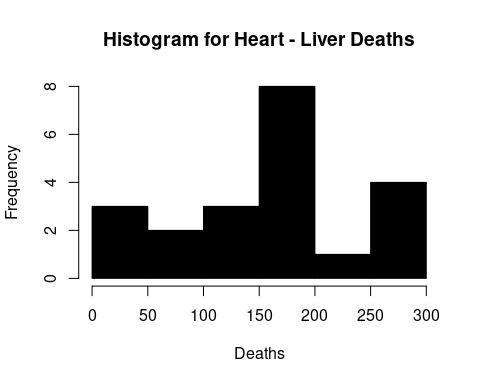




### The Distribution of the difference is the Same Around Some Mean

Using the data visualization below, we have taken the difference of Heart and Liver deaths over the 21 different observations and take the difference of their means (162) as the point for the data to be symmetrical around. The data below shows that even with this small dataset it appears that the data meets this assumption being symmetrical around the mean of 162.

In conclusion, all of the assumptions have been met for this test and it appears to be a viable test to use for this dataset.



The test was ran below to see what the results might be.

##   
## Wilcoxon signed rank exact test  
##   
## data: wine$heart and wine$liver  
## V = 231, p-value = 9.537e-07  
## alternative hypothesis: true location shift is not equal to 0

# 1.3.4 Attitudes Towards Religion

QUESTION:The file datasets/GSS\_religion is a subset of data from the 2004 General Social Survey (GSS). The variables prottemp and cathtemp are measurements of how a respondent feels towards protestants and towards Catholics, respectively. The GSS questions are phrased as follows: I’d like to get your feelings toward groups that are in the news these days. I will use something we call the feeling thermometer, and here is how it works: I’ll read the names of a group and I’d like you to rate that group using the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the group. Ratings between 0 degrees and 50 degrees mean that you don’t feel favorable toward the group and that you don’t care too much for that group. If we come to a group whose name you Don’t recognize, you don’t need to rate that group. Just tell me and we’ll move on to the next one. If you do recognize the name, but you don’t feel particularly warm or cold toward the group, you would rate the group at the 50 degree mark. How would you rate this group using the thermometer? You would like to test whether the US population feels more positive towards Protestants or towards Catholics. List all assumptions for a paired t-test. Then evaluate each assumption, presenting evidence based on your background knowledge, visualizations, and numerical summaries.

# Solution

## Test Assumptions Paired t-Test:

1. Metric Scale
2. IID Data
3. No major deviations from normality, considering the sample size.

Null Hypothesis: The expectation of equals the expectation of .

### Metric Scale

Using background information, the t-test is not valid for variables which only have an ordinal structure. In this case, the feeling thermometer used to determines ones feelings towards Catholics or Protestants would fall under the Likert Scale and also be considered of ordinal structure. Therefore, the test would fail under this assumption. A summary of the given data is shown below as supporting evidence.

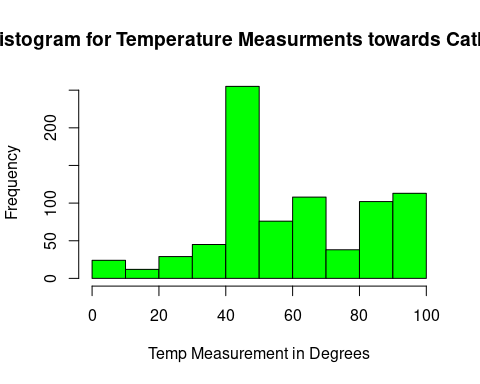
## ...1 year id prottemp   
## Min. : 1.0 Min. :2004 Min. : 4.0 Min. : 0.00   
## 1st Qu.:201.2 1st Qu.:2004 1st Qu.: 728.8 1st Qu.: 50.00   
## Median :401.5 Median :2004 Median :1373.5 Median : 60.00   
## Mean :401.5 Mean :2004 Mean :1381.9 Mean : 65.56   
## 3rd Qu.:601.8 3rd Qu.:2004 3rd Qu.:2053.5 3rd Qu.: 85.00   
## Max. :802.0 Max. :2004 Max. :2808.0 Max. :100.00   
## cathtemp   
## Min. : 0.00   
## 1st Qu.: 50.00   
## Median : 60.00   
## Mean : 63.16   
## 3rd Qu.: 85.00   
## Max. :100.00

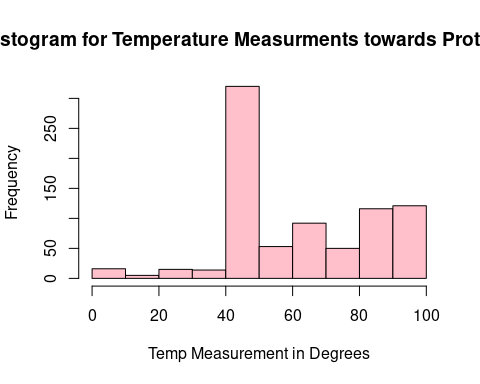
### IID Data

Making use of background knowledge, each pair of measurements is drawn from the same distribution, independently of all other pairs. This data appears to have paring, independence and identical distribution and does not violate. There are no apparent violates of independents, some examples being clustering of data, in geographical regions, school cohorts, or families, strategical interaction, like competition among sellers or imitation of a species, or autocorrelation were one time period may affect the next. The data also appears to be identically distributed as will be shown in the plots below. This assumption for the data is meet, but as stated previously, the Paired t-Test would not be a wise test to use for an ordinal scale statistical analysis.

### No Major Deviations from Noramlity, Considering the Sample Size

Using background knowledge 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. This does not appear to be the case as shown in the graphs below. This assumption for the data is meet, but as stated previously, the Paired t-Test would not be a wise test to use for an ordinal scale statistical analysis.





A Paired T-Test was run out of curiosity for the results.

##   
## Paired t-test  
##   
## data: GSS$prottemp and GSS$cathtemp  
## t = 2.9249, df = 801, p-value = 0.003543  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.7902444 4.0152419  
## sample estimates:  
## mean of the differences   
## 2.402743

In conclusion, through our analysis of the different scenarios where hypothesis tests could be run, the World Happiness statistical test would not be best evaluated with a Two-Sample T-Test, the Legislators problem would not be best evaluated with a hypothesis of comparisons version of the Wilcoxon Rank-Sum Test, the Wine and Health problem does fit all the assumptions for a Wilcoxon Signed-Rank Test and finally the Attitudes towards Religions problem does not fit the assumptions for a Paired T-Test.