

# Lab One, Part One

Oscar Garcia, Shuo Wang

June 28th, 2022

## Contents

<b>1</b>	<b>Part 1: Foundational Exercises</b>	<b>1</b>
1.1	Professional Magic . . . . .	1
1.1.1	What is the type I error rate of your test? . . . . .	1
1.1.2	What is the power of your test for the alternate hypothesis that $p = 3/4$ ? . . . .	1
1.2	Wrong Test, Right Data . . . . .	2
1.3	Test Assumptions . . . . .	3
1.3.1	World Happiness . . . . .	3
1.3.2	Legislators . . . . .	4
1.3.3	Wine and health . . . . .	5
1.3.4	Attitudes toward the religious . . . . .	6

# 1 Part 1: Foundational Exercises

## 1.1 Professional Magic

### 1.1.1 What is the type I error rate of your test?

The type I error is  $\alpha = P(\text{reject null} \mid \text{null is true})$ .

Null hypothesis is that  $p = \frac{1}{2}$ . We are told that we reject null if the statistic equals 0 or 6. Therefore,  
 $\alpha = P(\text{reject null} \mid \text{null is true}) = P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 0 \text{ or } 6 \mid p = \frac{1}{2})$   
 $= P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 0 \mid p = \frac{1}{2}) + P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 6 \mid p = \frac{1}{2})$

- First term of the expression:  
 $= P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 0 \mid p = \frac{1}{2}) = P((X_1, Y_1) = (0, 0) \cap (X_2, Y_2) = (0, 0) \cap (X_3, Y_3) = (0, 0) \mid p = \frac{1}{2})$   
We know that  
 $(X_i, Y_i)$  is independent from  $(X_j, Y_j) \forall i, j = 1, 2, 3$ ,  
so we can multiply the individual probabilities. So we can continue to evaluate the above expression:  
 $= P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 0 \mid p = \frac{1}{2}) = P((X_1, Y_1) = (0, 0) \mid p = \frac{1}{2})P((X_2, Y_2) = (0, 0) \mid p = \frac{1}{2})P((X_3, Y_3) = (0, 0) \mid p = \frac{1}{2})$   
We know because we are given the joint distribution that  $P((X_i, Y_i) = (0, 0)) = \frac{p}{2} \forall i = 1, 2, 3$   
Therefore,  
 $P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 0 \mid p = \frac{1}{2}) = (\frac{p}{2})^3 = (\frac{1/2}{2})^3 = \frac{1}{64}$
- Second term of the expression:  $= P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 6 \mid p = \frac{1}{2}) = P((X_1, Y_1) = (1, 1) \cap (X_2, Y_2) = (1, 1) \cap (X_3, Y_3) = (1, 1) \mid p = \frac{1}{2})$   
Using the same independence assumption between different pairs of coin flips,  
 $= P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 6 \mid p = \frac{1}{2}) = P((X_1, Y_1) = (1, 1) \mid p = \frac{1}{2})P((X_2, Y_2) = (1, 1) \mid p = \frac{1}{2})P((X_3, Y_3) = (1, 1) \mid p = \frac{1}{2})$

We know because we are given the joint distribution that  $P((X_i, Y_i) = (1, 1)) = \frac{p}{2} \forall i = 1, 2, 3$

Therefore,  $P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 6 \mid p = \frac{1}{2}) = (\frac{1/2}{2})^3 = \frac{1}{64}$

- Finally, we add first and second expression to calculate  $\alpha$ :  
 $\alpha = \frac{1}{64} + \frac{1}{64} = \frac{1}{32}$

Type I error of the test is  $\frac{1}{32} = 0.031$

### 1.1.2 What is the power of your test for the alternate hypothesis that $p = 3/4$ ?

Power of the test  $= 1 - \beta$

$\beta = P(\text{not reject null} \mid H_a)$ , with  $H_a$  the alternative hypothesis, in this case  $p = \frac{3}{4}$

By the complement rule,

$P(\text{not reject null} \mid H_a) = 1 - P(\text{reject null} \mid H_a)$

and  $P(\text{reject null} \mid H_a) = P(\text{reject null} \mid p = \frac{3}{4})$

$= P(X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 = 0 \text{ or } 6 \mid p = \frac{3}{4})$

This is calculated in similar way as 1.1.1, and we end up in the expression:

$= 2(\frac{p}{2})^3$  with  $p = \frac{3}{4}$ , therefore:

Power  $= 1 - \beta = 1 - (1 - P(\text{reject null} \mid H_a)) = P(\text{reject null} \mid H_a)$

$= 2(\frac{3/4}{2})^3 = \frac{27}{256}$

Power of the test for the alternative hypothesis is  $\frac{27}{256} = 0.105$

## 1.2 Wrong Test, Right Data

The violation of the metric scale assumption if we were to run a paired t-test using data from Likert scales means that the mean and the standard deviation of the t-statistics are inappropriate, and the values calculated with them, like Confidence Intervals, would be misleading.

To remedy this problem while preserving t-test, we need to remember the assumptions for a paired data t-test:

- Metric scale (not ordinal)
- IID data
- No major deviations from normality in the distribution of the difference between measurements

If we want to preserve the t-test, we have to convert the ordinal data from Likert scale into metric data. We can do this by converting it to binary, for instance assigning the lower values 1 to 3, to the value 0 in a binary scale, and the higher values 4 and 5 to a value 1 in the binary scale. Binary scale can be considered to be metric.

If we didn't need to preserve the t-test and wanted to use the full information from the Likert scale data, we would use a nonparametric test such as Sign Test (however, we would lose statistical power).

## 1.3 Test Assumptions

### 1.3.1 World Happiness

Assumptions for the two-sample t-test:

- Metric scale: in particular, the test is not valid for ordinal structure data - This assumption does not hold: the Cantril ladder scale is not metric, but only ordinal. We cannot assume that the distance between a 0 rating for a life and a 1 rating is the same as between say, 8 and 9.

- Grouping variable defined - This assumption holds: the grouping variable is low vs high GDP, lower than the mean or higher than the mean.

- IID data - The samples are not IID. We can argue that there is geographic clustering, because countries that are in the same area close to each other (not always) often have similar living conditions, socioeconomics conditions and similar legislations when compared to other countries far away.

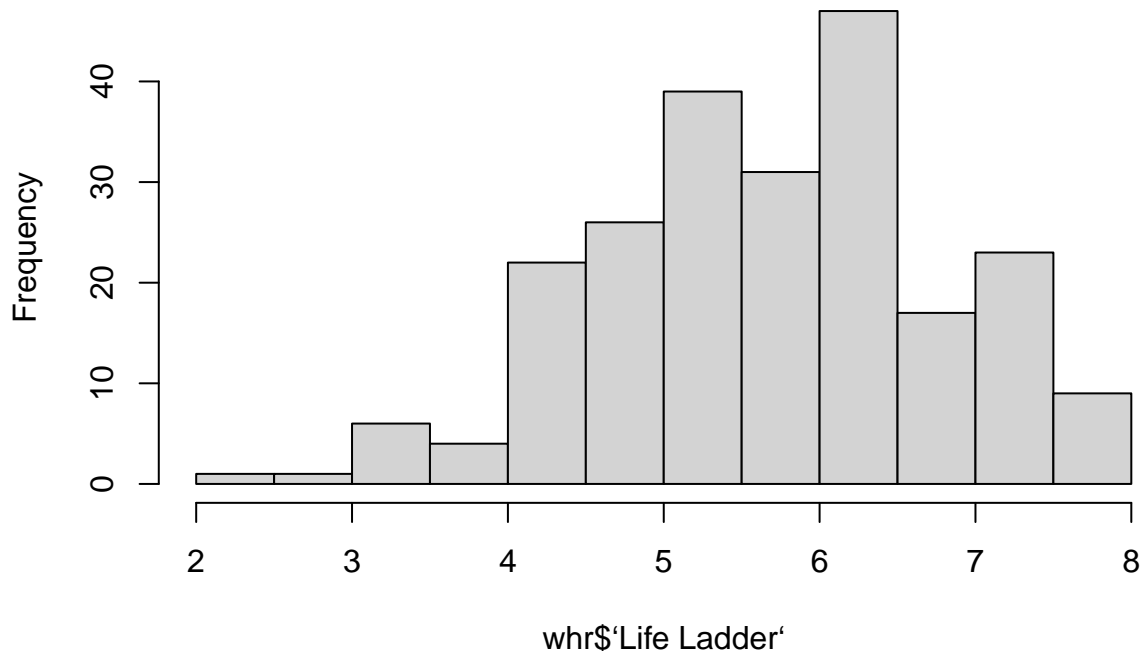
- No major deviations from normality, considering the sample size - In particular, the t-test is invalid for highly skewed distributions when sample size is larger than 30. It may also be invalid for very highly skewed distributions at higher sample sizes.

- In our case, we have 121 countries with higher than average GDP per capita and 105 countries with lower than average GDP per capita. These are large sample sizes, but still we check for skewedness

```
## [1] 121
```

```
## [1] 105
```

**Histogram of whr\$‘Life Ladder‘**



```
## [1] -0.2796529
```

There is very little skew, confirmed by the skewness command (-0.28)

In summary, a two-sample test in this case is not appropriate in this case because the scale of data is not

metric and we don't have IID.

- Equal variance assumption, which was true for the original version of the test, is not required with Welch's version of the test.

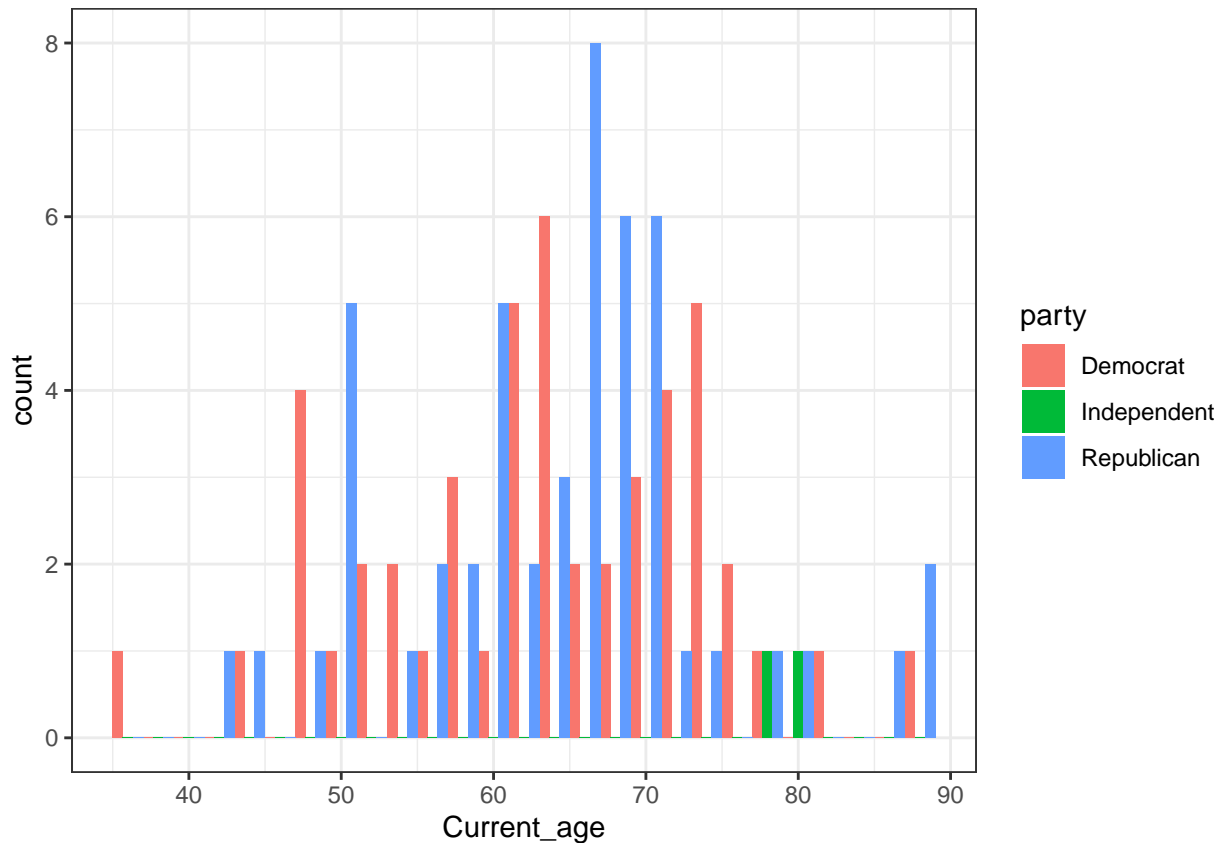
### 1.3.2 Legislators

Assumptions for the Wilcoxon rank-sum test (using the Hypothesis of Comparisons):

- Ordinal scale - This assumption holds; age is a metric variable and therefore, ordinal.
- Grouping variable - We have this, we have Democratic and Republican senators identified in the data
- Similar sample sizes

```
## # A tibble: 3 x 2
##   party      count
##   <chr>      <int>
## 1 Democrat      48
## 2 Independent    2
## 3 Republican    50
```

- We have 48 Democratic senators and 50 Republican senators in the data, so they do have similar sample sizes.
- IID data
  - This assumption means that each Democratic senator's age is drawn from the same distribution, each Republican senator's age is drawn from the same distribution, and all samples are mutually independent.
  - We assume this assumption can hold; we know senators are older in general, but we have no reason to believe that, once we know the age of a senator, that gives us information about the age of another senator, from the same or different party out of the same state or different. Therefore, we assume this assumption also holds.



Because the assumptions for this test hold, we think this test is appropriate for this problem.

### 1.3.3 Wine and health

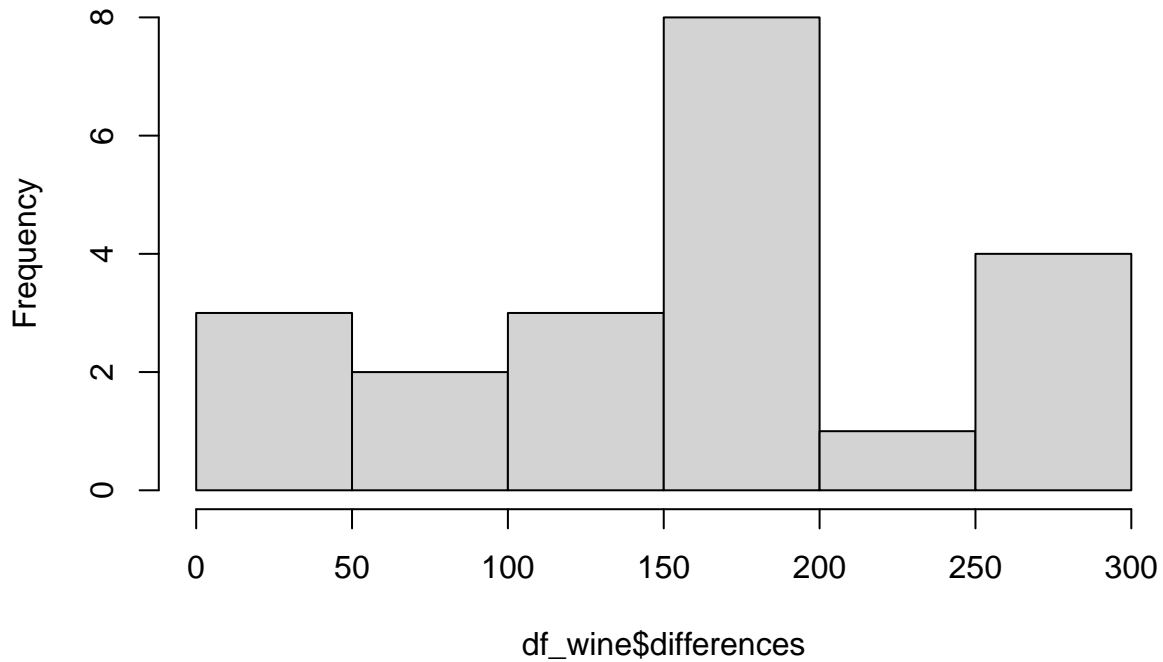
```
##      country alcohol deaths heart liver
## 1 Australia      2.5    785   211  15.3
## 2  Austria      3.9    863   167  45.6
## 3 Belg/Lux      2.9    883   131  20.7
## 4   Canada      2.4    793   191  16.4
## 5  Denmark      2.9    971   220  23.9
## 6   Finland      0.8    970   297  19.0
```

Assumptions for a Wilcoxon signed-rank test:

- There is a pair of random variables for each unit of observation - this assumption holds: we have variables heart and liver for each country
- Dependence within samples - this assumption holds: each sample pair represents one country, and the deaths from heart disease or liver disease due to alcohol consumption have a dependency because both come from the same people
- Metric scale, in particular both X and Y are both measured on the same metric scale - This assumption holds, both heart and liver variables are metric (measured both in deaths per 100,000)
- IID across samples. In particular, each pair is drawn from the same distribution, independently from other pairs. - This assumption does not hold because there is geographic clustering. Countries that are close to each other typically have similar wine consumption, lifestyles and health conditions, so there is geographic dependency between pairs.
- Continuous data (non binary) - this assumption holds, data is numeric (int or dbl)
- The distribution of the difference  $X - Y$  is symmetric around some mean  $\mu$  - As seen below, the shape of differences is approximately symmetric around the mean difference 162.2

```
## [1] 162.2524
```

### Histogram of df\_wine\$differences

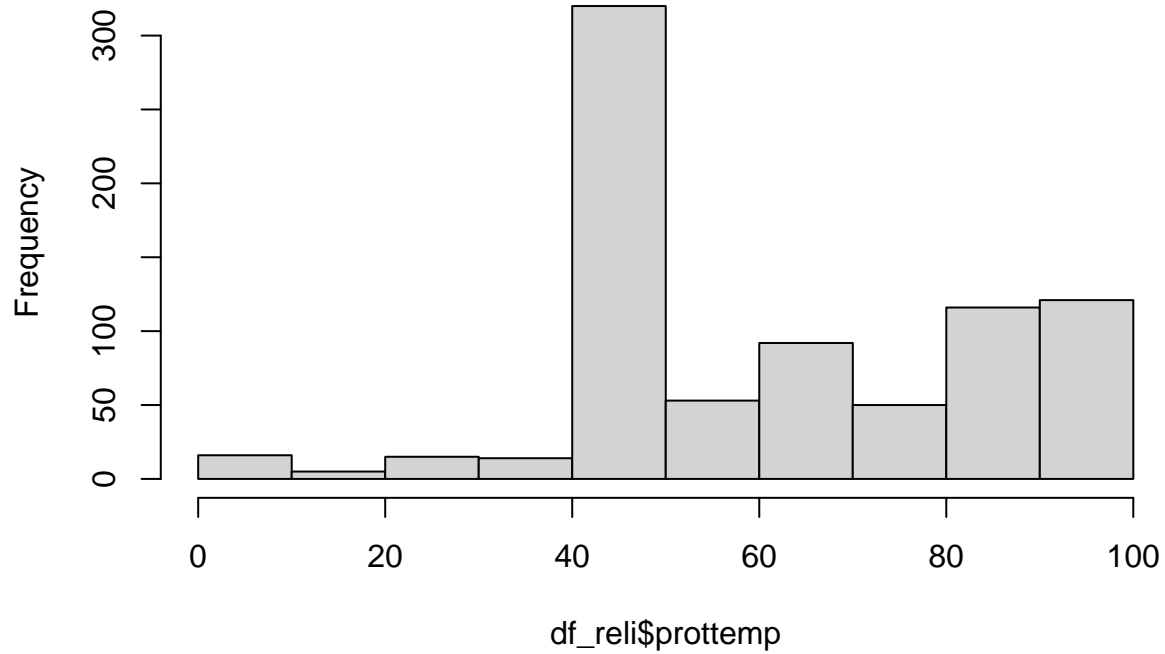


To summarize, this test is not appropriate for this problem because the IID assumption does not hold.

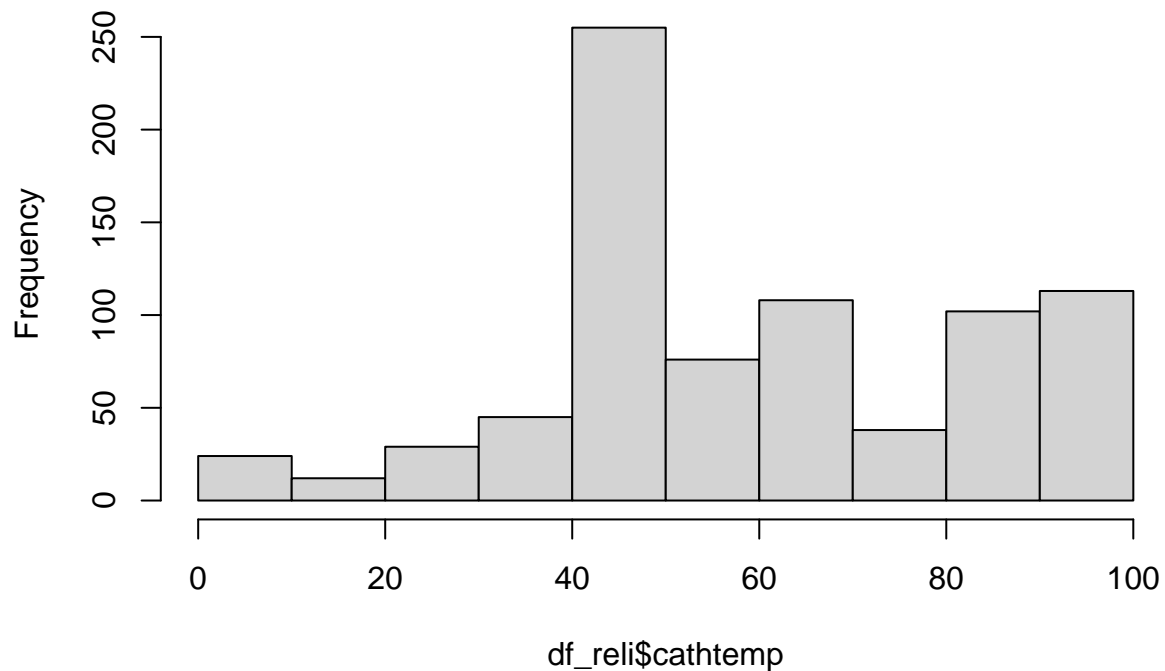
#### 1.3.4 Attitudes toward the religious

```
##   X year id prottemp cathtemp
## 1 1 2004 4      50      50
## 2 2 2004 7       5      85
## 3 3 2004 9      50      50
## 4 4 2004 14     60      60
## 5 5 2004 21    100       0
## 6 6 2004 24      5       5
```

**Histogram of df\_reli\$prottemp**



**Histogram of df\_reli\$cathtemp**



Assumptions for the paired t-test: - Metric scale - This assumption does not hold. The feeling thermometer is not metric, but just ordinal.

- IID data.

- We can safely consider that GSS sampling is IID. There is a finite-sample effect because there is no replacement once a person has been surveyed (so the distribution for the next draw changes), but the change is



likely to be small given the large number of Americans. Overall, GSS data is IID.

- The distribution of the difference between measurements has no major deviations from normality, given the sample size. In particular, for large sample sizes the test is invalid for highly skewed distributions.

- We have a sample of 802 pairs, so we check for high skew. This is a large sample size, and the plots shown above indicate no significant skew, so we can consider this assumption holds.

Overall though, this test is not appropriate because the metric scale assumption does not hold.