

### Q3.1

Assumptions for 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.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.375   4.971   5.768   5.678   6.428   7.889

##      Country.name year Life.Ladder Log.GDP.per.capita Social.support
## 1  Afghanistan 2019      2.375      7.697      0.420
## 2    Albania 2019      4.995      9.544      0.686
## 3    Algeria 2019      4.745      9.337      0.803
## 4  Argentina 2019      6.086     10.000      0.896
## 5   Armenia 2019      5.488      9.522      0.782
## 6  Australia 2019      7.234     10.815      0.943

##      Healthy.life.expectancy.at.birth Freedom.to.make.life.choices Generosity
## 1                      52.4                      0.394      -0.108
## 2                      69.0                      0.777      -0.099
## 3                      66.1                      0.385       0.005
## 4                      69.0                      0.817      -0.211
## 5                      67.2                      0.844      -0.172
## 6                      73.9                      0.918       0.121

##      Perceptions.of.corruption Positive.affect Negative.affect
## 1                      0.924                      0.351      0.502
## 2                      0.914                      0.681      0.274
## 3                      0.741                      0.585      0.215
## 4                      0.830                      0.826      0.319
## 5                      0.583                      0.598      0.430
## 6                      0.430                      0.770      0.202
```

1. Metric scale validation: The data isn't paired so that portion of a conventional t-test would be satisfied. However, the metric scale assumption is not valid here. The Cantril Ladder is ordered from 1 to 10 but the distance between rungs may not be equivalent. These measurements are inherently subjective and we can't perform basic mathematical operations to compare values. Building from this, we can't calculate a mean and therefore can't test for differences in means either.
2. IID data validation: The data is likely not a random sample because countries are not truly independent of each other, due to spatial and temporal clustering. For example, the data includes a variable to measure "freedom of life choices", which may be substantially higher in certain geographic regions (such as the EU). GDP per capita also tends to be correlated in this way and it becomes mathematically difficult to separate potential effects of GDP from location on the Cantril Ladder. For example, it would be difficult to draw statistically sound inferences of the effects of high GDP to areas of the world not represented in that group (such as Africa and South America). The temporal clustering issue is the discrepancy in sampling year, in the range of 2018-2020.
3. No major deviations from normality, considering the sample size: Since the data is not ordinal, not metric, assessment for normality does not apply.

Conclusion: The data is ordinal and not metric and therefore can't be considered normal. The data is also not IID, which makes the t-test an inappropriate choice for this dataset.