Research Questions and Background Dive

We know how energy directly impacts the environment, analyzing renewable energy by countries and continents and what the difference is between renewable from 2015 to 2020. This is for critical endeavor for several reasons, primarily driven by the global need to transition towards sustainable and low-carbon energy sources.
There are several reasons we decided to take this topic except for environmental impacts, there are many countries that have implemented policies and incentives to promote renewable energy adoption. Analyzing this data allows for an evaluation of the effectiveness of these policies so that's the first thing that we are going to answer statistically, and the second would be what is the trend of all percentages of all renewable sources in the year 2015 to 2020 i.e. 5 years. If we check the news and do a little research these were the years when people started evaluating the impact of our activities on the environment and ecology. Basically, people started getting aware of consequences more during these years as NASA recorded ozone depletion.

Critique the data.

The study was made to see how countries have adapted to renewable sources of energy and what the difference between the percentage of all renewables is from 2015 → 2020. The data appears to be observational, as it presents information from various countries without any manipulation. It is not from a controlled experiment. Basically **Is there a significance difference between renewables and countries in the year 2015 → 2020**

The Response variable or the dependent variable is likely to be different measures of renewable energy. We are going to use 'all renewables 2020' and all 'renewables 2015' for our linear regression. Explanatory variables or our independent variables could be 'continents'. The dataset has countries as their control variable as countries would influence the outcome of the continents.

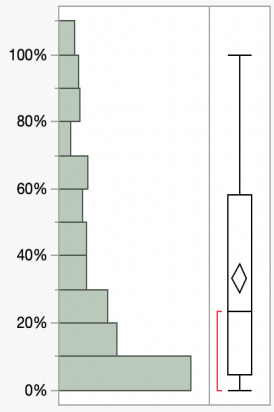
The data does not what bases were the countries selected. Understanding inclusion/exclusion would be crucial. There may be limitations in the accuracy of measurements. The data does not include a time frame on the source of the data, also the dataset has some mission values on Gwh production.

Some variables that might influence the continent of the country might influence renewable energy adoption. The location is an influence due to the availability and type of resources available.

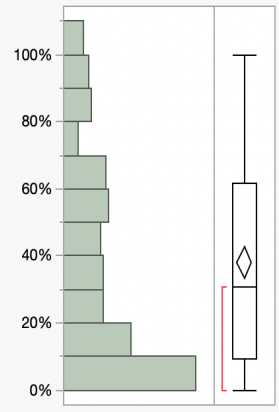
Explore the data

Distributions

All renewables 2020



all renewables 2015



Quantiles

100.0%	maximum	100%
99.5%		100%
97.5%		100%
90.0%		87%
75.0%	quartile	58%
50.0%	median	24%
25.0%	quartile	5%
10.0%		0%
2.5%		0%
0.5%		0%
0.0%	minimum	0%

Summary Statistics

Mean	0.3346009
Std Dev	0.3197284
Std Err Mean	0.0219074
Upper 95% Mean	0.3777852
Lower 95% Mean	0.2914166
N	213
N Missing	0
Median	0.235

Quantiles

100.0%	maximum	100%
99.5%		100%
97.5%		100%
90.0%		89%
75.0%	quartile	62%
50.0%	median	31%
25.0%	quartile	9%
10.0%		2%
2.5%		0%
0.5%		0%
0.0%	minimum	0%

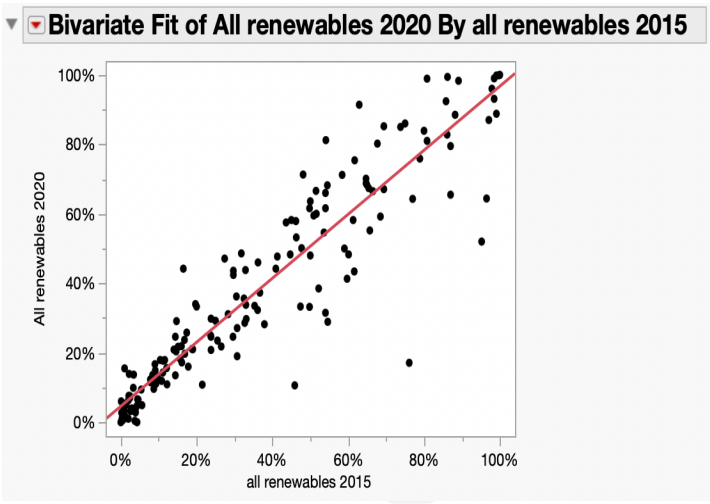
Summary Statistics

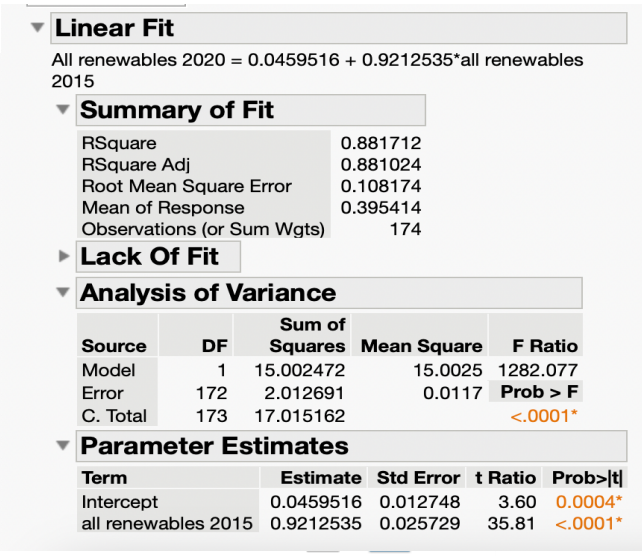
Mean	0.3793333
Std Dev	0.3196533
Std Err Mean	0.0242329
Upper 95% Mean	0.4271634
Lower 95% Mean	0.3315032
N	174
N Missing	39
Median	0.307

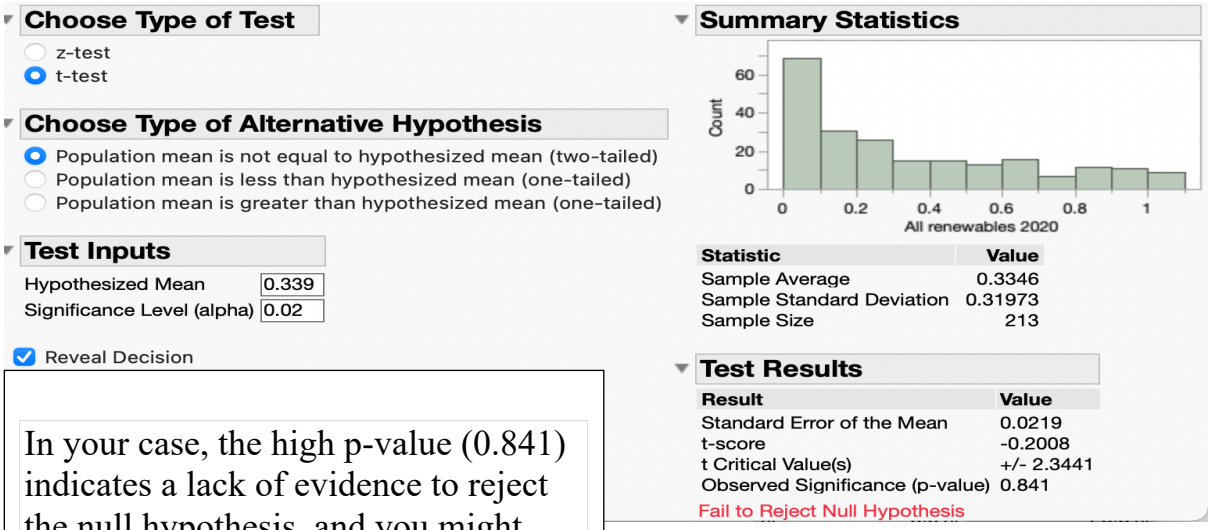
In this report we are looking at two variables that is all renewables is 2020 and 2015. The median and the mode for both are close to each other which implies that in 2020 more countries had lower shares of renewables as compare 20215 but this could also be because of more missing values in 2015 hence influences the stats. The statistics meaning of these changes depends on variety of factors maybe changes In country policy for renewable resources or lack of resources or even economic factors for that matter.
The median is less affected by the outliers so we can catch on the median in this case. The median is relatively close between 2020 and 2015 , hence the central tendency of the data hasn't changed dramatically over years. Also the sample size is 213 , providing a reasonable dataset for analysis.

Analysing the data

When you see the linear equation we can see that there is a positive linear relationship, between all renewables in 2020 and all renewables in 2015 and the coefficient of 0.09212 indicates. In the linear fit below we see that our slope is positive hence all our points are near the normal. The key metrics under consideration include median, and regression analysis to understand the trends and relationships within the dataset. The R square value is 0.8521, indicating that approximately 85.21% of the variance in the dependent variable is explained by our model. It reveals a strong relationship between renewable energy production in 2020 and 2015. The High R—R-squared and significant F ratio indicate that the model is a right fit.







In your case, the high p-value (0.841) indicates a lack of evidence to reject the null hypothesis, and you might interpret this as no significant difference between renewable energy production in 2015 and 2020 based on the statistical test you conducted.

Conclusion

The population in this context would be all the countries and their renewable energy production. This dataset is probably derived from various sites and hence strong likelihood that the study can generalize to different countries.

When we did t t-test our t-score and p-value both were high which indicated that there was no significant difference between all renewables in 2015 →2020 and that there isn't enough evidence to reject the null hypothesis.

By Dhruvi Naik , Divya Padala , P.Gowtham Sai , Nimit Taswala s