Statistical analysis of electric energy production of Romania and trend forecasting

**Introduction**

**Parameters for statistical Analysis**

In order to gain a better understanding in the insight of the trends in our energy production data we use the following statistical measures.

*Mean*

The central tendency of our data is measured using the mean (or the expected value)

*Standard deviation (Std)*

The amount of dispersion of our data is measured using the standard deviation. A low value of the standard deviation shows that the data values tend to be close to the expected value of the set, a higher value indicates that the data values are spread out on a bigger interval

*Skew*

For short- and medium-term predictions the expected value is less likely to be achieved, so the take this into account we also want to analyse the extremes of the data set by calculating the skew.

The skewness can pe positive (right skewness) negative (left skewness) or zero (bell curve)

In general, the more skewed the data, either positive or negative, the less accurate the future predictions will be.

A distribution is considered fiary simmetical if the skewness is between -0.5 and 0.5, moderately skewed if the skewness is between -1 and -0.5 or between 0.5 and 1, and hihly skewed if the skewness if less than -1 or greater than 1.

*Kurtosis*

We also consider how light or heavy tailed the data distribution is by calculating the kurtosis. A positive kurtosis means that our data distribution has lighter/tinner tail than a normal curve with the same mean and std and thus a higher peak. A negative kurtosis translates in a heavyer than normal tailed distribution and thus a lower peak.

Most often, kurtosis is measured against the normal distribution.  If the kurtosis is close to 0, then a normal distribution is often assumed.  These are called mesokurtic distributions.  If the kurtosis is less than zero, then the distribution is light tails and is called a platykurtic distribution.  If the kurtosis is greater than zero, then the distribution has heavier tails and is called a leptokurtic distribution.

In general, if the kurtosis is between -0.5 and 0.5 a the data distribution can be assumed to be normal and is called a mesokurtic distribution, ff the kurtosis Is less than -0.5 the distribution is called platykurtic, and if the kurtosis is greater than 0.5 the distribution is called leptokurtic.

*Range*

The minimum and maximum values of our data is also noted as to begin from a interval with our future predictions of energy production

*Variation (Coefficient of variation)*

It is well known that big values of standard deviation in relation to the average value may lead to limitations in the quality of forecasts. So, we use the coefficient of variation to rank the reliability of our future predictions. From our knowledge, even if the coefficient of variation is used in manufacturing and finance (citari) to better understand the risk-return trade-off, there is no direct interpretation for it in the statistical analysis of energy production, so we are going to assume that a coefficient of variation over 0.5 is associated with energy production methods that are highly dependant on day to day condition and are highly influenced by fenomena that are unpredictable in the long term such as the weather in general, cloud positioning, and wind speeds.

**Plots**

The following section consists of plots of electrical energy production by method of production on the interval 7th March 2019 – 9th March 2022, together with a visual analysis of the plots.

Coal - approx. constant value on the interval March 2019 - December 2020. We notice a 50% Decrease starting in March 2020, correlating perfectly with the 2020 Covid 19 Market crash and the time when the oil prices went negative, this drop continues on the interval March 2020 - September 2020 and is followed by a 250% increase in the subsequent period. The Coal energy production has not fully recovered since then.

Hydro – We can notice a periodicity in the lows and highs of the data. Highs corresponding to months in which snow usually melts in Romania (February, March and April). Lows corresponding to summer months and autumn months. We notice an anomaly in 2021 where the low is period is longer than that of the previous year, this is due to less rain this year

Chart

Description automatically generatedChart

Description automatically generated

Fig 1: Coal Fig 2: Hydro

Oil & Gas – We notice a periodicity in the lows and highs that is opposite to the one in Hydro. The highs are in the winter months where hydro production in not that great. The lows are in the summer and spring period where melting snow.

Nuclear – the only nuclear power plant in our county can either function at 1400MW or at 700MW.

This can be clearly seen in the data plot. There seems to be a strong negative correlation correlation between the peaks of the Hydorelectric production and the lows of the nuclear production. Considering the before discussed rationement at the hydro section we can say that in times which correspond to rainy season the power output of the nuclesr powerplant will be at 700MW while, for the reast of the year it will be at 1400MW

Chart

Description automatically generatedChart, histogram

Description automatically generated

Fig 3: Oil&Gas Fig 4: Nuclear

Windpower production seems to ride in the winter months because. This is logic because of the storong arctic winds coming form Siberia

Solar. It is evident that the power generatd by solar means will be higher in the summer than in the winter. This behaviour can be clearly seen in the figure

Chart, line chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Fig 5: Wind Fig 6: Solar

Sold- Negative values of sold mean that electrical energy was exported and positive values mean that energy was imported from outside the country.

Again we notice a negative correlation between the production ofm hydro electricity and imported/exported electricity. The peaks of production in hydro energy correspond to the negative lows (i.e., electrical energy was exported) and the lows of hydro electricity correspond to the zones in the sold chart where demand of electrical energy from outside the country is high.

Biomass

Chart

Description automatically generated with medium confidenceA picture containing pencil

Description automatically generated

Fig 6: Biomass Fig 8: Sold

**Results of Statistical Analysis and Histogram analysis**

The selected results of the daily distribution of energy production grouped by method of production are presented in Table 1.

The highest 3 means are hydro, nuclear and coal, lowest mean is biomass, solar and sold, but if we exclude sold because it allows negative values, we get that the lowest 3 means are those of biomass.

This means that most of the energy in Romania is generated by means of hydro, nuclear and coal.

Negative skew right

Positive skew left

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Std | Skew | Kurtosis | Variation | Range |
| Coal | 1270 | 315.2 | (-)0.231 | 0.103 | 0.248 | 333 - 2168 |
| Hydro | 1853 | 724.3 | 0.661 | 0.101 | 0.390 | 159 - 4200 |
| Oil & Gas | 1150 | 431.5 | (-)0.02 | (-)0.877 | 0.375 | 210 - 2068 |
| Nuclear | 1299 | 236.54 | (-)2.12 | 2.68 | 0.182 | 626 - 1436 |
| Wind | 766 | 672.96 | 0.972 | 0.013 | 0.879 | 0 - 2756 |
| Solar | 154 | 225.67 | 1.304 | 0.366 | 1.468 | 0 - 839 |
| Biomass | 59 | 14.02 | (-)0.345 | (-)0.676 | 0.240 | 19 - 87 |
| Sold\* | 248 | 700.09 | (-)0.248 | (-)0.458 | 2.818 | (-)2073 - 2025 |

Table 1: Results of the statistical analysis by method of production

\*Negative values of sold represent export of electrical energy and positive values represent import.

By looking the histogram of the energy produced by coal in Figure 9 we observe that the highest frequency of occurrence is at 1250, a value close to the mean of 1270 calculated in Table 1. This means that the data distribution is skewed. This fact is confirmed by the coefficient of skewness calculated as -0.231, which is associated with a fairy symmetrical data distribution that is a bit skewed to the right (higher values).

The coefficient of kurtosis was calculated as 0.103, which is between the interval of -0.5 and 0.5, so our distribution fairly resembles a normal one and is called a mesokurtic distribution.

The standard deviation in relation with the mean, calculated through the coefficient of variation is below 0.5, corresponding with data that is distributed near the mean and is not highly influenced by long term unpredictable phenomena.

We now look at Figure 10, the histogram of the energy produced through hydroelectric means, and we observe that the highest frequency of occurrence is at 1750 MW, far off from the expected value of 1853 calculated in Table 1. Again, we look at the calculated skewness coefficient of 0.661 confirms our visual findings that the data distribution for hydroelectric power generation is moderately skewed to the left (lower values). The kurtosis values is between that of -0.5 and 0.5 so our distribution fairy resembles a normal one and is called a mesokurtic distribution.

The coefficient of variation is below 0.5 so we can say that out data distribution is packed near the mean value is not highly affected by long term unpredictable phenomena.

**Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated**

Fig 9: Coal Fig 10: Hydro

In figure 11, the histogram of Oil& Gas, we remark that the highest frequency of occurrence is at 1150, exactly as the value that was calculated in Table 1. So our distribution is almost symmetrical, fact that is confirmed by the skew of -0.02, so we can say that our data distribution is not skewed at all.

The kurtosis value is -0.877 which is smaller than -0.5 so our distribution is heavier tailed than normal distribution with the same std and thus has a lower peak and is called a platykurtic distribution.

The coefficient of variation is 0. 375, which is below 0.5, and comparable to the coefficient of variation obtained for Hydro, so we say that our data is packed near the expected value and is not highly affected by long term unpredictable phenomena.

For figure 12 we see that the highest frequency of occurrence is at 1400, quite different from the expected value calculated in Table 1, so we expect our data distribution to be skewed. Our expectations are confirmed by the skew of -2.12 which means that our data distribution is highly skewed to the right (higher values).

The kurtosis value is that of 2.68 which is greater than 0.5, so our if lighter tailed that a normal distribution with the same standard thus has a higher peak and is called a leptokurtic distribution.

The coefficient of variation is obtained for nuclear is 0.182, which is below 0.5 and the lowest from all the one calculated in Table 1, this not only show that the majority of the data is packed near the expected value and that it is not highly affected by long term unpredictable phenomena but it also shows that nuclear energy has the most constant output from all the other means of energy production

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Fig 11: Oil&Gas Fig 12: Nuclear

Coal – By looking the the histogram in figure 9 we obsereve that the highest frequency of occurrence is at 1250 , a value close to the mean of 1270 calculated in Table 1. This means that the distribution of the energy production values by coal skewed. This is confirmed by the calculated skewness coffieicent of -0.231 associated with a fairly symmetrical data but nontherless a small skewness to the right (higher values).

The kurtosis value is that of 0.103 and is between the interval of -0.5 and 0.5, so our distribution fairly resembles a normal one and is called a Mesokurtic distribution.

The standard deviation in relation with the mean, calculated through the coefficient of variation is below 0.5, corresponding to a distribution that is more packed near the mean value and is not highly influenced by long term unpredictable phenomena.

By looking at the histogram in figure 13 we observe that the highest frequency of occurrence is at 0 while the mean value calculated in Table 1 is at 766. This means that out distribution is skewed, which can be clearly be seen visually, but as a confirmation we have the calculated skew of 0.972, value which is associated with a distribution skewed to the left (lower values).

The kurtosis value is that of 0.013 and is between -0.5 and 0.5, so our distribution fairly resembles a normal one and is called a mesokurtic distribution.

The standard deviation reported to the mean, calculated through the coefficient of variation is 0.879, greater than 0.5, corresponding to a distribution that is not packed near the mean and is highly influenced by long term unpredictable phenomena, in this case the local wind speed.

We now switch to figure 14, the histogram of solar, we remark that the highest frequency of occurrence is at 0 and the expected value calculated in Table 1 is 154. So our distribution is skewed, fact that is confirmed by the skew of 1.304 which is associated with a distribution skewed to the left (lower values).

The kurtosis value is that of 0.366, which is between -0.5 and 0.5 so our distribution fairly resembles a normal one and is called a mesokurtic distribution. The coeffieicnt of variation is 1.468, greater than 0.5, corresponding to a distribution that is not packed near the expected value and is highly influenced by long term unpredictable phenomena, in this case the general weather and the position of the clouds.

Chart, histogram

Description automatically generatedA picture containing histogram

Description automatically generated

Fig 13: Wind Fig 14: Solar

For figure 15 we see that the highest frequency of occurrence is at 65, while the expected value calculated in Table 1 is 59, so our data distribution is skewed, fact which is confirmed by the skew value of -0.345which is associated to a data distribution skewed to the right(higher values).

The kurtosis value is that of 0.366 whick between -0.5 and 0.5 so our distribution fairly resembles a normal one and is called a mesokurtic distribution.

The coefficient of variation obtained for Biomass is 0.24, which is below 0.5m this shows that the majority of the data is distributed near the mean(asa sa scriu la toate!!!!!!!!) and is not highly affected by long term unpredictable phenomena

We now look at figure 16 of the histogram of data of sold, and we see that the highest frequency of occurrence is at 500 and the mean value calculated in Table 1 is 248 so our data is skewed. The skewness is confirmed by the skew of -0.248, associated with a data distribution that is skewed to the right (higher values).

The kurtosis value was calculated as -0.458 which is between -0.5 and 0.5 but still extremely close to the limit value of -0.5 so our data distribution is at the limit between a mesokurtic and a platykurtic distribution.

The coefficient of variation obtained for sold is 2.818, which is above 0.5, this shows that data is not mainly distributed near the mean and is highly affected by long term unpredictable phenomena, in this case other all the other parameters that influence the rest of the energy production. This is normal as energy is imported or exported depending on the energy production in that day, which tends to follow a season trend but is also highly influenced by long term unpredictable phenomena.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Fig 15: Biomass Fig 16: Sold

**Trend forecasting method**

**Conclusions**

**References**

[**https://www.sistemulenergetic.ro**](https://www.sistemulenergetic.ro)