# Descriptive Statistics

## **Correlation of Fuel Prices**

## The figure above shows the relationship between gasoline and diesel prices in the examined sample. The points lie almost along a straight line, indicating a very strong linear correlation. The value of the correlation coefficient is 0.974, which suggests that the prices of the two fuels move almost perfectly together.

## This high correlation supports the conclusion that gasoline and diesel prices are largely influenced by the same factors. Consequently, in the analysis, modelling the evolution of gasoline prices, the diesel price is not included as a direct explanatory variable (or vice versa), as it would cause strong multicollinearity.

## **Data on Population, Income, and Fuel Prices**

### **Fuel Prices**

Gasoline prices are highly concentrated (standard deviation of only around 13 HUF), with values clustering tightly around 580–585 HUF per liter. This indicates a strongly homogeneous pricing structure, likely reflecting either centralized price influence or a highly competitive market environment.

The distribution of diesel prices closely mirrors that of gasoline, reinforcing the observation that the two fuel types move together in price (as already evidenced in the earlier scatterplot). The average difference between gasoline and diesel prices is approximately +6–7 HUF, which is typical for the Hungarian market due to tax structure differences.

The near equality of the mean and median suggests a symmetric distribution. The peaking histogram and price clustering around identical levels both point to a market that is either centrally regulated or subject to strong price competition.

Both distributions are extremely concentrated, with nearly all prices falling within a narrow range (580–590 HUF/liter). Only a few observations exceed 600 HUF or fall below 560 HUF, these outliers can likely be explained by geographic or brand-specific factors.

In summary, gasoline and diesel prices exhibit strong co-movement, and the market appears highly homogeneous. Minor price differences are more likely explained by secondary factors such as local income levels, geographic distance, or competitive intensity.

**Population**

The population distribution is strongly right-skewed: there are many small settlements (dense cluster on the left) and relatively few large cities (long tail on the right). This pattern is typical for Hungary. The skew distorts the mean, making the median a more robust measure of central tendency.

The mean, which is approximately 45,000 people, is considerably higher than the median ( around 23,000), confirming the skewness. The large standard deviation (around 60,000) further indicates significant heterogeneity.

Given this asymmetry, it might be useful to log-transfrom the Population variable in regression models to better handle extreme values. Smaller settlements typically exhibit less competition, which may lead to slightly higher fuel prices.

**Income**

The income distribution is multimodal, displaying several peaks where observations are concentrated, interspersed with gaps. Values range approximately between 475,000 and 780,000 HUF, reflecting substantial regional disparities. This is supported by both the large standard deviation (around 86,000 HUF) and the close proximity of the mean and median.

Income is a potentially a strong explanatory variable for gasoline prices, as its high variance provides meaningful differentiation across regions. Thus, variation in income levels could plausibly account for part of the observed differences in fuel prices.

Based on the descriptive statistics, fuel prices both gasoline and diesel fluctuate within a relatively narrow range, indicating a homogeneous and competitive market structure.  
The average price difference between gasoline and diesel is only about 6–7 HUF per liter.  
Observed price disparities are therefore unlikely to stem from a general lack of competition, but rather from localized factors.

In contrast, the Population and Income variables exhibit high variability, suggesting considerable economic and demographic heterogeneity across the sampled areas.  
This supports the assumption that these variables may partly explain differences in gasoline prices. In particular, demand pressures associated with higher income levels, as well as the lower competitive intensity typical of smaller settlements, are likely to play a role in subsequent modeling.

### **Relationship between the Fuel Prices and the other numerical variables**

### **Income and Fuel Prices**

The figures illustrate the relationship between income levels and fuel prices (gasoline and diesel). Each point represents an individual fuel station, while the colored line denotes the estimated linear regression trend.

Gasoline prices exhibit substantial dispersion across different income levels. The regression line is nearly flat, indicating no meaningful linear relationship between income and gasoline prices. In other words, fuel is not systematically more expensive in high-income regions, nor cheaper in low-income ones.

This suggests that national pricing policies or the homogenizing effects of market competition largely offset local income differences. Major fuel chains typically determine prices at the national level, which minimizes the influence of local demand and supply conditions.

Diesel prices display a very similar pattern: the data points are widely scattered, and the trendline shows only a marginal positive slope, essentially flat. This implies a very weak, or statistically insignificant correlation between income and diesel prices. No systematic price differentials are observed across income brackets. These results prove the theoretical expectation that fuel prices are primarily driven by macroeconomic factors rather than local economic development.

Overall, neither gasoline nor diesel prices exhibit a statistically significant association with municipal income levels. The flat regression lines and dispersed data points suggest that income variation does not explain fuel price formation. Consequently, pricing dynamics are primarily influenced by national or global determinants, while local demand (potentially proxied by income) does not have enough explanatory power.

### **Population and Fuel Prices**

The charts illustrate the relationship between settlement population size and local fuel prices (gasoline and diesel). Each point corresponds to an individual station, and the colored line represents the linear trend reflecting the general direction of association.

The points are widely dispersed, and no clear linear pattern is shown. The trendline exhibits a slightly negative slope in both cases, suggesting that bigger settlements tend to have lower fuel prices. However, the slope is very small, implying that the relationship is weak and likely it will not be significant.

The weak negative association may reflect competitive market effects. Larger settlements typically have more fuel stations, which intensifies price competition, while smaller settlements often operate as local monopolies, allowing for higher price margins. Nevertheless, the strong national price homogenization constrains population-related price variation.

For gasoline, the trendline is horizontal, indicating independence from population size. For diesel, the negative slope is slightly more visible, hinting that price competition may be somewhat stronger in larger urban markets. In both cases, however, the effect remains weak, and population alone should not be considered a strong determinant of price levels.

Overall, the results indicate only a very mild negative association between settlement size and fuel prices. Larger settlements may exhibit slightly lower prices due to intensified competition, but this effect is probably not statistically significant. The spatial dispersion of data points further implies that fuel prices are mostly determined by national-level factors, such as taxes, or crude oil prices, rather than by settlement-specific characteristics.

## **Ratio of Settlement types in the sample**

The figure shows the distribution of observations across settlement types. The majority of sampled fuel stations are located in urban environments, reflecting the spatial concentration of the Hungarian fuel market.

Cities account for approximately 43.2 % of the sample, while capitals of counties represent an additional 21.7 %. Districts of the capital city comprise 14.4 %, highlighting Budapest’s substantial market weight. Rural areas, like villages (12.1 %) and large villages (4.1 %), collectively take up only about 16 % of the sample.

This indicates that the analyzed data predominantly originate from urbanized regions, while rural and less populated areas are under-represented. The distribution partly reflects Hungary’s actual settlement structure, as the majority of the population resides in urban areas. However, this pattern is also relevant for interpreting fuel price formation.

Urban areas typically exhibit higher demand and stronger competition, which may have a price-equalizing effect, whereas smaller settlements, with fewer stations and greater distances often experience more concentrated market power, potentially leading to higher prices.

To summarise, the sample is city-centric. Consequently, including *Settlement Type* as a control variable in the econometric model will be useful to gain insight on urban–rural differences in fuel pricing.

## **Ratio of the Largest Companies in the sample**

The figure presents the market share of major fuel retailers represented in the sample.

The data clearly indicate a highly concentrated market structure.

MOL is by far the dominant player, operating roughly one-third of the sampled stations.  
It is followed by Shell (15.3 %) and OMV (14.4 %), while Orlen, Mobil Petrol, and Avia hold much smaller shares.

These figures confirm that the Hungarian fuel market is dominated by a few large, nationwide networks, with smaller companies maintaining only marginal presence. Such concentration likely influences pricing strategies.

Competition among large operators may be relatively moderate, whereas smaller firms operating in more localized markets may adjust prices more flexibly in response to local demand or cost fluctuations.

### **Relationship between Fuel Prices and the Nearest Competing Brand**

The plots show the relationship between fuel prices (gasoline and diesel) and the distance to the nearest competitor of a different brand. The horizontal axis shows the distance (in km) to the nearest competing station, while the vertical axis represents the fuel price. Colored lines indicate the fitted linear trends.

The data are widely scattered with no visible trend. The regression lines show a slightly positive slope, implying that stations further from competitors tend to charge higher prices. However, the relationship is weak, with most observations concentrated in the 0–5 km range.

The weak positive association can be explained: where competitive pressure is lower, stations can sustain higher margins, resulting in higher prices; where nearby competitors exist, price competition intensifies and exerts downward pressure. Nevertheless, the low correlation suggests that while this effect persists, it is secondary, since fuel prices are primarily shaped by national-level determinants and company-wide pricing strategies rather than by purely local competition.

For gasoline, the regression line is nearly flat, implying negligible sensitivity to competitor proximity. For diesel, the slope is slightly steeper, suggesting that competition may have a somewhat stronger influence on diesel pricing. In both cases, however, the relationship remains weak and likely statistically insignificant.

## **Fuel Prices on County Level**

The county-level figures display the geographic distribution and dispersion of fuel prices, offering insight into possible regional patterns of price formation.

### **Diesel prices**

The first figure presents a boxplot of county-level diesel prices. Each box represents the interquartile range (middle 50 % of prices), with the horizontal line indicating the median. Individual points represent station-level observations, showing within-county scattering.

Diesel prices are broadly homogeneous across counties, with no huge differences in median values. Upper-range prices are slightly higher in counties such as Bács-Kiskun, Baranya, and Tolna, while lower ranges are observed in Nógrád and Vas.

Outliers within counties likely correspond to stations pricing higher or lower due to specific local market conditions.

Price deviation in Budapest is minimal, suggesting stronger competition and more uniform pricing policies.

Overall, spatial differences in diesel prices are minor on a national level, though peripheral areas tend to exhibit slightly higher prices.

**Gasoline**

The second figure shows a very similar pattern.

Gasoline prices also fall within a narrow range, indicating that national pricing rules or competitive pressures strongly standardize prices.

Some counties (e.g., Zala, Borsod-Abaúj-Zemplén) exhibit wider dispersion, suggesting localized market deviations or a distinct player’s behavior. Median price differences across counties are minimal, only a few forints, thus statistically insignificant.

Overall, both gasoline and diesel prices are highly consistent across counties, confirming that Hungary’s fuel market is centralized and competition-regulated. Geographical location alone has limited influence on pricing. Outliers are attributable mainly to the local market, or other corporate strategies.

## **Distribution of Fuel Prices based on Settlement Type**

The boxplots above show the distribution of diesel and gasoline prices across different settlement categories. The median price (thick horizontal line within each box) does not vary substantially between categories, suggesting that settlement type alone might not strongly explain price differentials.

In the capital districts, both diesel and gasoline prices are slightly higher on average, likely reflecting stronger demand, higher operating costs, and the prevalence of premium retail networks.

In villages and large villages, prices are marginally lower but exhibit greater dispersion, indicating more heterogeneous market structures (a mix of small independent stations and large-chain outlets).

Towns and county-level cities show stable and uniform price levels, comparable in median to the capital, but with smaller variance, implying dominance by larger, standardized networks.

Outliers, in other words extreme price observations occur mainly in smaller settlements, likely reflecting local pricing policies driven by distance, transport costs, or limited competition.

Summarising, settlement type effect shows a moderate but non-negligible influence, meaning that prices are slightly higher in the capital and more volatile in rural areas.  
This aligns with the expectation that urban markets exhibit more stable, competitive price structures, while smaller settlements experience greater price variability.