**Year-Cause + Gender**

**As a result, I see how the trends of mortality from different causes for each gender have changed over the years. (Later, I also look at the differences for each age group).**

**Assumingly at the end I get two columns of graphs: rows are different numerical variables about death, columns are countries. Each graph includes trends for different mortality reasons (for different genders)**

For every reason group:

1. \*\*Descriptive Analysis:\*\* Start with a descriptive analysis to understand the main characteristics of your dataset. Use the `describe()` and `info()` functions in pandas to get an overview of statistics and data information. You can identify the main trends, such as the mean, standard deviation, minimum and maximum, as well as the distribution of mortality in different years and by gender.

(do not forget to include the percentage of deaths from this cause in the study for certain reasons)

* + Make an overview of statistics (functions info and describe)
  + Write comments to the statistics review
  + To see and determine whether the distribution of the number of deaths and the mortality rate in general is normal or determine other distributions
  + Compare mortality statistics between different sexes. To do this, you can use t-tests, variance analysis (ANOVA), or nonparametric tests to determine if there are statistically significant differences between groups (gender).
  + View the trends of the number of deaths, age stand and mortality rate by year and by gender (histograms, scatter plots, and boxplots)

Knowledge of the distribution of quantities in data is important when analyzing and interpreting data for several reasons:

1. \*\*Descriptive statistics:\*\* Distribution allows us to get a general idea of the data. This includes understanding averages, data spread, asymmetry, and other characteristics.
2. \*\*Interpretation of data:\*\* The distribution can give an idea of how the data is distributed across different categories or groups. For example, it can show how the mortality rate changes over the years or differs between different population groups.
3. \*\*Forecasting and modeling:\*\* Knowledge about the distribution of data is important for building statistical models and forecasting. For example, the normal distribution can be used to build linear models, and the Poisson distribution can be used to analyze events such as deaths.
4. \*\*Selection of statistical methods:\*\* The choice of statistical methods depends on the form of data distribution. For example, t-tests can be used for normally distributed data, and nonparametric methods can be used for asymmetric data.
5. \*\*Detection of outliers and anomalies:\*\* The distribution can help identify unusual values in the data, such as outliers, which may require additional analysis.
6. \*\*Decision-making:\*\* Understanding the distribution can help in making more informed decisions based on the data. For example, when determining measures to reduce mortality.

Test for abnormal distribution:

If your data does not have a normal distribution and you want to determine statistically significant differences between the mean values of two independent groups, then the Wilcoxon-Mann-Whitney U-test\*\* may be the best choice. This test is a nonparametric analogue of the t-test for independent samples and can be used when the data does not correspond to a normal distribution and does not satisfy the assumptions of parametric tests.

The Wilcoxon-Mann-Whitney criterion allows you to compare the medians of two data groups, and it is more resistant to deviations from the normality of the data. Hypotheses in this test are formulated as follows:

- Null hypothesis (H0): The average values of the two groups do not differ (there are no statistically significant differences).

- Alternative hypothesis (H1): The average values of the two groups differ (there are statistically significant differences).

If the p-value obtained using the Wilcoxon-Mann-Whitney Test is less than the selected significance level (usually 0.05), then the null hypothesis is rejected, and you can conclude that there are statistically significant differences between the groups.

Thus, the Wilcoxon-Mann-Whitney criterion is a good choice when the data does not correspond to a normal distribution.

1. \*\*Correlation Analysis:\*\* Investigate the correlation between the number of deaths and the mortality rate. This may help to identify if there is a relationship between these two variables.