Econometrics
Week 1: main notions

- *Classification of Variables*

- *Data visualisation*

- *Summarizing statistics*

- *Pearson correlation coefficient; Spearman correlation coefficient; Corrected contingency coefficient of Pearson*

Seminar 1: Descriptive statistics

Imaging that you're a real estate analyst preparing a market report. Consider dataset *AmesHousing.csv*. The description of the dataset is given in *AmesHousing_data_description.txt*

1. **Basic SalePrice stats.** *(This part must be completed by all teams.)*

   (a) Classify variables *SalePrice*, *LotArea*, *HouseStyle*, *OverallQual*, *MSSubClass*, *YearBuilt* by scales of measurement; as qualitative or quantitative.

   (b) Plot the histogram for *SalePrice*. Experiment with different number of bins.

   (c) Compute quantiles of levels 0.1,0.2,..., 0.9 for *SalePrice*.

   (d) Compute main summary statistics for *SalePrice*.

2. **Univariate analysis.** *(Team 1 is responsible for this task.)*

   (a) Compute all dispersion measures for *SalePrice*.

   (b) Plot the boxplot for *SalePrice*.

   (c) Compute skewness.

   (d) Identify the observation with median SalePrice (or closest to it) and one observation that is an outlier according to your boxplot. Write down all their characteristics in a table format.

3. **Compare groups.** *(Team 2 is responsible for this task.)*
   Group your data by *House styles*;

   (a) Compute summary statistics for *SalePrice* for different groups.

   (b) Compute all dispersion measures.

   (c) Plot the boxplots for *SalePrice* for different groups.

   Repeat for grouping by *BldgType*.

4. **Paired data analysis.** *(Team 3 is responsible for this task.)*

   (a) Plot the scatterplots for *SalePrice* and *LotArea*,*YearBuilt*, *TotalBsmtSF*, *GrLivArea*, *GarageArea*

   (b) Compute Pearson correlation coefficients; Spearman correlation coefficients.

   (c) Group the data by three groups $SalePrice < 150000$, $150000 \leq SalePrice \leq 215000$ and $Saleprice \geq 215000$. Using contingency table and corrected contingency coefficient of Pearson explore the relation between these groups and SaleCondition.

5. **Feature Engineering.** *(Team 4 is responsible for this task.)*
   Create a new variable: $PricePerSqFt = SalePrice/GrLivArea$. This standardized metric helps compare homes of different sizes.

   (a) Plot the histogram for *PricePerSqFt*

   (b) Compute main summary statistics for *PricePerSqFt*

   (c) Compute all dispersion measure for *PricePerSqFt*

   (d) Plot the boxplot for *PricePerSqFt*

   (e) Compute skewness for *PricePerSqFt*

   (f) Plot the scatterplots for *PricePerSqFt* and *LotArea*,*YearBuilt*, *TotalBsmtSF*, *GrLivArea*, *GarageArea*

   (g) Compute Pearson correlation coefficients; Spearman correlation coefficients.