Project Plan Computational Statistics

Leila Maillefer, Fabia Schreyer, Loan Strübi

April 2025

Project Plan: Impact of House Style on Sale Price

Objective

We aim to evaluate the influence of the architectural style and building type on the sale price of houses in the Ames dataset. The goal is to determine whether specific combinations of styles and types are systematically more valued, and to identify which factors drive price differences while controlling for other important variables such as living area or overall quality.

1. Classical Statistical Inference

- Descriptive statistics: distribution of `SalePrice`, `HouseStyle`, `BldgType`.

- Boxplots of `SalePrice` by `HouseStyle` and `BldgType`.

- Hypothesis tests: e.g., are 2-story houses significantly more expensive than 1-story houses?

- Confidence intervals for the average sale price by house style.

2. 2^k Factorial Design and ANOVA

- Binary variables selected:

  - `HouseStyle`: 2Story vs 1Story
  - `BldgType`: 1Fam vs Twnhs
  - `OverallQual`: High quality (>7) vs others
  - `CentralAir`: Yes vs No

- Construct a 2^k factorial design with 4 factors.

- Perform ANOVA to analyze main effects and interactions on `SalePrice`.

3. Regression and ANOVA Analysis

- Multiple linear regression model:

  - Model 1: SalePrice $\sim$ GrLivArea + OverallQual + HouseStyle + BldgType
  - Model 2: includes interaction term: GrLivArea * HouseStyle

- Analyze coefficients, R^2, RMSE, and multicollinearity (VIF).

- Compare models using ANOVA.

4. Time Series Analysis

- Aggregate data by month or quarter using MoSold and YrSold.

- Create time series of average SalePrice for one or two styles (e.g., 1Story vs 2Story).

- Fit SARIMA models to detect trend and seasonality.

Expected Outcomes

This project will provide a comprehensive view of how architectural style and building type influence housing prices. It will also showcase how classical and computational statistical methods can be combined to support data-driven decision-making in real estate.