Econometrics
Seminar 4: Hypothesis testing

Consider the dataset *AmesHousing.csv*. The description of the dataset is given in *AmesHousing_data_description.txt*

1. Estimate a linear regression model of *SalePrice* on *Lot.Frontage*, *Lot.Area*, *Year.Built*, $X1st.Flr.SF$, $X2nd.Flr.SF$, *BsmtFin.SF.1*, *BsmtFin.SF.2*, *Bsmt.Unf.SF*. Can we instead estimate a linear regression model of *SalePrice* on *Lot.Frontage*, *Lot.Area*, *Year.Built*, $X1st.Flr.SF$, $X2nd.Flr.SF$, *Gr.Liv.Area*, *BsmtFin.SF.1*, *BsmtFin.SF.2*, *Bsmt.Unf.SF*? Why or why not?

2. Test the statistical significance of the variable *Lot.Frontage*. Write down the corresponding hypotheses and the test statistics which should be used. Explain how the test works. Write down the critical region for significance level 5%. Run the test and formulate the conclusion.

3. Estimate a new model dropping *Lot.Frontage*. Analyze the change in R^2 and R^2-adjusted.

4. Interpret the value of the intercept in the estimated model.

5. Analyse the model for multicollinearity. Which diagnostic measures would you use?

6. Test the residuals for normality.

7. Test the model for heteroskedasticity. Specify the test used and explain how it works.

8. Test whether *Lot.Frontage* and *Lot.Area* are **jointly** significant. What test to use? Write down the hypotheses, test statistics and critical region for significance level 5%. Run the test. Formulate conclusions.

9. Test whether the effect of $Year.Built$ on *SalePrice* is positive, ceteris paribus. State your intuitive expectation. Formulate the hypotheses. Test them. Interpret the result.

10. Test whether the expected increase in SalePrice for each additional square foot of the First Floor is the same as for each additional square foot of the Second Floor. What model would you consider?