REGRESSÃO E ANÁLISE DE VARIÂNCIA - 2024

João Gomes

FCUL DEIO

Problems - Preferably, case studies

- Software: **Excel** and **R**.
- Report: pdf
- Case Studies datasets:
  - **https://www.kaggle.com/**
  - **http://www.statsci.org/data/multiple.html**
  - **https://users.stat.ufl.edu/~winner/datasets.html**

Problem 1. Analysis of Variance.

Case Study – Analysis and Report - **delivery for 4th of April**. **You have to solve a problem using analysis of variance (balanced two-way) as a statistical tool - use the different topics covered in the lessons (2 to 5)! Try to imagine that you are publishing a scientific paper.Site:** Link **With 79 explanatory variables describing (almost) every aspect of houses in Ames, Iowa, the challenge is to predict the final price (SalePrice) of each house using only "kitchenqual" and "centralair" as explanatory variables.**

SalePrice: the property's sale price in dollars.

CentralAir: Central air conditioning

N No
Y Yes

KitchenQual: Kitchen quality

Ex Excellent
Gd Good
TA Typical/Average
Fa Fair
Po Poor

1.1. To get a sense of the overall problem, first read the overview. Carry out the following steps before analysing:

1. Select only the variables of interest.
2. Transform the response variable so that you have Saleprice in tens of thousands of dollars.
3. Consider only the levels "Gd", "Fa" and "TA" in the variable "kitchenqual", transforming the level "Ex" to "Gd".
4. In order to have a balanced analysis, you randomly select 6 observations from the different groups.

1.2. Using the appropriate graphical representations indicate your opinion on the importance, or not, of each of the factors and also the importance of the interaction between them.

1.3. Perform a global test that allows you to say that there are differences between the means of the groups.

1.4. Verify the suggestions made in 1.2 using 3 simple tests (5% significance level).

1.5. Plan and carry out a full set of orthogonal contrasts (design this study based on the previous item).

1.5.1. Is the sum of partial SS equal to SS_{TRAT}? Comment.

1.5.2. Compare the average of the F's obtained in each contrast with the F_ratio of 1.3 and comment on the result obtained.

1.5.3. Summarises the results obtained in the different contrasts .

1.6. Find a 95% confidence interval for the mean saleprice of houses with a good quality kitchen and central air.

1.7. If any of the tests carried out in 1.4 are not significant, present a suitable alternative model and repeat question 1.6.