Assignment 1

In this assignment you will be analysing (simulated) data on profits from each branch of a chain of fast-food restaurants operating in Iceland.

Like as with most fast-food restaurants staff turnover is fast. the goal of the analysis is to gain insight into whether improving staff retention would improve profitability.

This is not a simple question to answer. While there is variation in both staff turnover/experience and profitability between branchers, adjustments must be made. The task is to use multiple regression analysis to separate the effects of different variables to try to gain insight into what would happen if staff retention, either for managers or all staff, would improve, holding all other variables constant.

The dataset contains profits for all 26 branches of the fast-food chain in the year 2024 along with various other variables:

Region1: Nominal variable with 7 values: “Capital Region”, “Southern Peninsula”, “West”, “North-West and Westfjords”, “Northeast”, “East” and “South”.

Region2: Nominal variable with 2 values: “Capital Region” and “Other”

Urban: Nominal variable with two values, “Yes” and “No”. A branch is considered urban if there are at least 10000 residents in the municipality in which the branch is located.

Branch_size: Nominal variable with three values: “Small”, “Medium” and “Large”. A branch is considered small if it has 15 FTE staff or less, medium with 16-30 FFTE staff and large with more than 30 FTE staff.

AreaPop: Number of people living within a 3 km radius of the branch.

Staff_no: Total number of staff.

Staff_FTE: Full time equivalent number of staff.

Staff_turnover1: Number of employees that left their jobs at the branch in 2024

Staff_turnover2: Number of employees that left their jobs at the branch in 2024, weighted by the number of regular hours they worked before leaving

Staff_mean_experience1: The mean experience of all staff members in months.

Staff_mean_experience2: The mean experience of all staff members in moths, excluding the manager.

Manager_total_experience: Manager’s total experience with the company. Measured in months.

Manager_experience: Manager’s experience in the role of manager. Measured in months.

Profit: Profit in 1000 ISK

Use all the techniques you have learned to explore the data visually. Don't forget to look at the correlation between the predictors. In addition to the visual techniques, you may find various features under "Analysis->Summary" and "Datasets->Aggregate" useful.

Transformations (e.g. logs, polynomials) may be necessary. This can be done in BlueSky under "Variables->Compute" but with this particular dataset it should be safe to do data wrangling in Excel (but that is not something I recommend when working with real data!).

Write up a short report (4-6 pages) explaining your exploration of the data and final model specification. Report the final model and interpret the coefficients.

Note that while the true population parameters are identical for all data sets each group gets their own simulated dataset. There is nothing wrong in discussing the projects with students in other groups, but it would not be a good idea to copy results or conclusions.