Final Project Proposal: Divorce Prediction

Importance of the topic

Family is the first unit of society. It is the factory that builds and raises every individual. Thus, taking care of the family and enhancing the bonds between family members would be to their benefit. Divorce could cause severe harm to individuals that might reflect their whole life. That's why studying the causes or the factors that could lead to divorce will be a great help to understand and coop with these factors to decrease the probability of divorce.

Research Question: Which factors, such as communication, financial stress, trust, and infidelity, most strongly predict the likelihood of divorce.

Relation to the course

In this project, I am going to use multiple methods and tools taught in the course. I will assign a research hypothesis and test whether it is accepted or rejected. I am also going to use PCA (Principal Component Analysis) to combine correlated features into new principal components. I will use logistic regression to predict the occurrence of divorce.

Project Plan

- H_0 (Null Hypothesis): Financial stress does not affect the likelihood of divorce. H_1 (Alternative Hypothesis): Financial stress increases the likelihood of divorce.
- Dataset: The dataset includes 5,000 records with 22 features related to couples' demographics, relationship quality, financial status, and emotional well-being. Key variables include communication score, trust score, financial stress level, and infidelity occurrence, along with a binary target variable (divorced).
- Statistical Methods: The analysis will use logistic regression to evaluate the significance and predictive power of each factor on divorce outcomes. Additionally, PCA (Principal Component Analysis) may be applied to identify relationships among variables and reduce dimensionality.
- Expected Outputs: Correlation heatmaps and visualizations showing relationships between predictors and divorce likelihood. Interpretations identifying which variables have the strongest effect on divorce probability.

Sources and Tools

- Dataset Source: Divorce prediction dataset (CSV file) downloaded from Kaggle https://www.kaggle.com/datasets/vanpatangan/divorce-prediction/data

- Programming Tools: Python packages: pandas, numpy, matplotlib, seaborn, scikit-learn, statsmodels
- Relevant Literature: Intention to Divorce and Its Determinants in the Young Newly Marrieds: A Field-Based Cross-sectional Study from Iran
  https://brieflands.com/journals/semj/articles/148536?utm_source=chatgpt.com