Behavioral Data Science: Homework Assignment 4
Miscellaneous and Project Pre-Study

April 9, 2025

All answers and solutions to non-programming questions should be submitted to LMS as a **legible** write-up (either fully digital or a scan). The use of LLMs (e.g., ChatGPT) is **explicitly discouraged**, unless specified otherwise. All code should be committed to and merged into the `main` branch of your team's GitHub repository, unless specified otherwise. Your LMS submissions should contain a single ZIP file named according to the pattern:

- BDSl_Assignment[#]_[TeamMember1Initials]_[TeamMember2Initials]

Problem 1: True-False Questions (5 points)

Mark all statements which are **FALSE**.

1. Path coefficients in structural equation models (SEMs) are analogous to regression coefficients from ordinary least squares (OLS) regression.

2. Latent variables in SEM can serve as both predictors and outcomes.

3. Sample size does not affect the reliability of SEM estimates if the model is properly specified.

4. An underidentified SEM model cannot produce meaningful results.

5. Goodness-of-fit indices like RMSEA, CFI, and TLI can quantify how well the model reproduces the latent variables.

6. Correlations among observed variables undermine the estimation of SEM parameters.

7. The variance of a Wiener process with scale coefficient $\sigma = 1$ at time t is t^2.

8. The standard Drift-Diffusion Model (DDM) assumes that evidence about a dominant alternative accumulates in discrete chunks over time.

9. The Euler-Maruyama method can only be used to simulate linear stochastic differential equations (SDEs).

10. The drift diffusion model has three core parameters: (relative) starting point, boundary separation, and drift rate.

Problem 2: Reflection (6 points)

Describe and elaborate on three personal takeaways related to the topics of the course.

Problem 6: Project Pre-Study (8 points)

What problem are you considering? What is the type of data? What is the modeling task (e.g., regression, classification, latent variable modeling, parameter estimation...)? What existing models have been applied to tackle the problem? What would be an adequate model in your case? What would be its strengths and limitations? How would you criticize the model(s) in your intended project? What is the metric of success?