

Social Robotics

B. Nebel, F. Lindner, L. Wächter
Summer Semester 2019

University of Freiburg
Department of Computer Science

Exercise Sheet 7

Due: July 1st, 2019

Exercise 7.1 (Hypothesis Testing, ANOVA)

You want to test different movie recommendation algorithms for a smart home assistant robot:

- The *baseline* algorithm always selects a random movie from a database out of 1000 more or less well-known movies of all different genres.
- The *popular* algorithm selects a random movie from a selection of movies which have a high overall rating and which are popular at the moment.
- The *data-mining* algorithm uses machine learning on the user's social network data to match the user's preferences as closely as possible.
- The *improved* data-mining algorithm also incorporates the feedback provided by the user.

In our experiment, each participant got 50 movies recommended during a time span of three months and had to rate each movie on a scale between 1 and 10 directly after watching it. Then, the ratings were averaged into a single score assessing how satisfied the participant was with the recommendations.

You can find the (fictional) experimental data in the file `ex07_experiments.csv`.

- (a) Formulate three hypotheses about the relative quality of the algorithms (and the corresponding null hypotheses) and set a significance level. At least one of the hypotheses should be about a relation between at least three of the conditions.
- (b) Decide on the statistics you plan to use (t or F) for testing your hypothesis, as well as on the order of your tests and the alpha-adjustment technique (Bonferroni, Holm, Frane). Give a short explanation for your decisions.
- (c) Use R to perform the tests accordingly.
- (d) Report your results. Visualize.
- (e) Choose one of your previous hypotheses for which you have used an ANOVA. Why could it be problematic to calculate a t-test instead?