

INF632 Homework 2

Basic Statistical Tests

Assigned: January 29.

Due: February 19 at 11:59pm. Submissions turned in after this time and still within two weeks of this deadline will be automatically docked 50% of the possible points.

Submission: Share a direct link to your repo / folder on canvas.nau.edu by the deadline.

Points: EE499 students, this is worth 15% of your final grade. EE599 students, this is worth 10% of your final grade.

Background:

Wearable devices have seen great utility in behavioral health science. Though they also provide very different data from what is typically collected in such fields, many of the behavioral health statistical methods apply to data from these sources too.

Assignment:

Your assignment will be to write several functions to perform statistical tests on a collection of data sets and then use those functions to evaluate the dataset given. You may use Octave/MATLAB, Python, R, or any of these in Notebooks, but you must write the functions to compute and perform the statistics outlined below from scratch. You shouldn't need any libraries beyond basics for reading CSVs.

Functions

Harmonic Mean

then maths

Should accept N complete datasets

Pooled Standard Deviation

more maths

T-Test

maths

ANOVA

anova

Repeated Measures ANOVA

Application of your Functions

Using the functions you have developed from above, and the data provided, compute the following and answer the questions posed.

0.0.1 Daily Steps

How many steps per day, on average do the subjects walk? Use the harmonic and arithmetic mean. Are they different? Why?

0.0.2 Group Variance

What is the variance of the group?
(across subjects, pooled sd)

0.0.3 Comparing the Devices

Does the Fitbit report the same step measures as the ActiGraph? (t-test)

0.0.4 Weekend Warriors

Are the subjects equally active across each day of the week? (as determined by daily steps in DoW, anova)

0.0.5 Seasonality

In the two year data set, you'll find daily step totals. Across the two years, were all months traveled equally? (repeated measures anova)

Expectations and Grading Rubric:

Great news, you get a break from IEEE format submissions on this homework! This time I need to see your code, with comments that make it clear what you are doing. If you choose to complete this assignment in a Notebook (e.g. Jupyter Notebook), you can show your answers to the “Application of your Functions” in markdown cells after each step. If you choose not to use a Notebook, prepare a markdown file (.md) for each of your responses to the “Application of your Functions” questions. Your submission will be graded as follows:

| Area | Frac. of Points | UG: Meets Expectations | UG:Exceeds, G:Meets Expectations | G:Exceeds Expectations |
|-----------------------|-----------------|---|---|---|
| Code Requirements | 40% | The code takes inputs and provides outputs as specified | The code also accepts optional input and makes use of defaults. | The code can self identify characteristics that change how the calculations will be performed. |
| Code Clarity | 40% | Minimal comments, but sufficient to understand the basics of what is happening. | The code make good use of comments, including outlining some assumptions and where errors could happen (or how they are being handled). | The comments make reading the code so easy that one doesn't even have to know the programming language to understand the steps. |
| Analysis and Response | 20% | The analysis is performed correctly and the findings are accurate. | The analysis and findings are correct, and the explanation shows an understanding of what was done and why. | The analysis and findings are correct, the explanation shows depth of thought, and further analysis is suggested in detail. |