**Middle East Technical University**

**Department of Statistics**

**Spring 2020-2021**

**STAT 364 – PROJECT INFORMATION**

|  |  |  |  |
| --- | --- | --- | --- |
| **Step** | **Deadline \*** | **Deadline Hour** | **Details** |
| 1. pick your teams | 2 April 2021 | 15:40 | 1 |
| 2. pick data | 16 April | 15:40 | 2 |
| 3. research questions and proposed methods | 23 April | 15:40 | 3 |
| 4. Final report | 11 June | 15:40 | 4 |
| 5. Presentations | Will be announced, but could be on 17 and/or 24 June in lecture |  | 5 |

\* You can complete the steps and email Serenay before the deadlines.

**1.** Construct your project teams of 1, 2, 3, 4 or at most 5 people. Pick one of your team members as a corresponding member. That corresponding person will be responsible from the communication between team members and Serenay. However, all other works of the project should be done by all members of the team. The corresponding person should send the names of the team members to Serenay by email before deadline.

**2.** Find a data set to work on.

**Do not use time series data, or any data that is repeatedly collected over time** from the same subjects, since you don’t know how to analyze such data yet.

Your data should have **at least** one categorical, one discrete and one continuous variables. It should have one response and **at least** 5 independent variables to start with. During analysis, you can reduce the number of independent variables if you find they are irrelevant. The data should have **at least** 50 subjects. If the data set you find is very big, you can use a **randomly** **chosen** subset of it.

The corresponding person should submit the following information to Turnitin by deadline:

the name of the data set, its short description (in your own words), the list of the response and independent variables (each variable should be classified as categorical, discrete or continuous), and the source (e.g. the link of the web page)

You may go through the following websites while looking for an appropriate dataset.

[Rdatasets](https://vincentarelbundock.github.io/Rdatasets/datasets.html)

[AwesomePublicDatasets](https://github.com/awesomedata/awesome-public-datasets)

[FiveThirtyEight](https://github.com/fivethirtyeight)

[OpenIntro](https://www.openintro.org/data/)

We will check that no 2 teams pick the same data set. We have posted the data sets that were used previously in this and other related courses (for example Stat 356/412) at ODTUClass. Please do not use one these datasets. Also, you cannot use the same dataset for 2 courses in this semester. Please submit the info to Turnitin as soon as possible, so that you can change the data set if the similarity % with student projects at METU is too high, suggesting it was used in our department previously.

**3.** Write research questions and related methods that you will use to answer these questions. Also, include which variable is usually going to be used as a response. The corresponding person should email these to Serenay by deadline.

For example:

Research question: “Which variables are likely to best predict yield?”

Proposed method: Scatter plots of yield vs soil nutrients

Response: yield

**4.** Prepare and submit a final report. One common report for the group should be submitted to ODTUClass by the corresponding person. This report should include everything: introduction, analysis, results (don’t forget the interpretations and discussions), conclusion, outputs, codes if R is used, steps if SPSS or Minitab is used, turnitin report and more if you want. This report should be **at most** 10 pages of A4 written in Times New Roman font 12 with 1.5 space between lines. You can use both front and back of these 10 pages.

**5.** Prepare a presentation (probably, in PowerPoint). You may also try other presentation methods such as Beamer, xaringan or slidy presentations in RMarkdown. More directions will be given on this in June.

**Important note:** Please note that, while writing project reports, articles, proposals, thesis etc. it is not OK to directly copy-paste a material even if you provide a reference. If you are using more than 3 consecutive words from a source, you need to give it in “ ” and provide the reference. During a presentation, also do not copy-paste someone else’s sentences. Of course you can copy-paste from your own report.