**STAT 01515****-Applied Multivariate DATA Analysis-SUMMER 2025**

**Final Project Guidelines**

For your final project, you will produce a paper presenting an application of **multivariate statistical methods** to one or more datasets of interest.

Your paper should include the following components:

* **Description of research questions/issues**  
  (These may be scientific or statistical in nature.)
* **Description of data**
* **Presentation of statistical analysis**, including:
  + **Methods**: What analyses were performed and why? If you encounter challenges during the analysis, describe how you addressed them.
  + **Results**: Do **not** include raw computer output. A small number of well-designed tables and graphics are appropriate.
  + **Conclusion**: Communicate your findings in a way that is understandable to a broader audience.

The paper must not exceed **10 pages**, including figures and tables. Begin your report with an appropriate title that clearly reflects your topic and analysis.  
You may work in **teams of two**.

**Due Dates**

* **Proposal**: *Monday, June 23rd (in class)*  
  Briefly describe your data source, central research questions(s), and the statistical methods you plan to use. The proposal should be **no more than one page**.
* **Interim Report**: *Tuesday, July 8th*  (**5 %**)
* **Final Paper and Presentation**: *Monday, July 21st*   
  Submit your paper and deliver an in-class presentation.

**Grading [25%]**

Grading will be based on the quality of your writing, analysis, and the context in which your work is presented. Considerations include:

* Is the report well organized, clearly written, and easy to follow?
* Were appropriate analyses selected and correctly executed?
* Are your conclusions well-supported by numerical and graphical evidence?

**Statistical Methods**

By the end of Week 4, we will have covered:

* Hypothesis testing
* Dimension reduction
* Classification
* Factor analysis

Our next topic will be **cluster analysis**. You are encouraged to use one or two of these methods in your project.

Alternatively, you may:

* Explore a multivariate method not covered in this course, or
* Compare the performance of multiple statistical methods on your dataset(s)

These are all valid and valuable approaches.

**Data and Topic Selection**

Choose a topic and dataset ideally aligned with your **field of study**.  
If you prefer to use publicly available datasets, the **UCI Machine Learning Repository** (<http://archive.ics.uci.edu/ml/>) offers many preprocessed multivariate datasets.

Below is a list of additional data sources you might explore:

* [America's Best Colleges - U.S. News & World Reports](http://www.usnews.com/usnews/edu/college/rankings/rankindex_brief.php)
* [American FactFinder](http://factfinder.census.gov/servlet/BasicFactsServlet)
* [The Baseball Archive](http://www.baseball1.com/)
* [The Bureau of Justice Statistics](http://www.albany.edu/sourcebook/)
* [The Bureau of Labor Statistics](http://www.bls.gov/)
* [The Bureau of Transportation Statistics](http://www.bts.gov/)
* [The Census Bureau](http://www.census.gov/)
* [Data and Story Library (DASL)](http://lib.stat.cmu.edu/DASL/)
* [Data Sets, UCLA Statistics Department](http://www.stat.ucla.edu/data/)
* [DIG Stats](http://www.cvgs.k12.va.us/DIGSTATS/)
* [Economic Research Service, US Department of Agriculture](http://www.ers.usda.gov/Briefing/)
* [Energy Information Administration](http://www.eia.doe.gov/index.html)
* [Eurostat](http://epp.eurostat.ec.europa.eu/)
* [Exploring Data](http://exploringdata.cqu.edu.au/)
* [FedStats](http://www.fedstats.gov/)
* [The Gallop Organization](http://www.gallup.com/)
* [International Fuel Prices](http://www.gtz.de/en/themen/umwelt-infrastruktur/transport/10285.htm)
* [Journal of Statistics Education Data Archive](http://www.amstat.org/publications/jse/)
* [Kentucky Derby Race Statistics](http://www.kentuckyderby.com/2003/derby_history/derby_statistics/)
* [National Center for Education Statistics](http://www.ed.gov/NCES/)
* [National Center for Health Statistics](http://www.cdc.gov/nchs/)
* [National Climatic Data Center](http://www.ncdc.noaa.gov/oa/ncdc.html)
* [National Geophysical Data Center](http://www.ngdc.noaa.gov/)
* [National Oceanic and Atmospheric Administration](http://www.noaa.gov/)
* [Sports Data Resources](http://www.amstat.org/sections/SIS/Sports%20Data%20Resources/)
* [Statistics Canada](http://www.statcan.ca/start.html)
* [StatLib---Datasets Archive](http://lib.stat.cmu.edu/datasets/)
* [UK Government Statistical Service](http://www.statistics.gov.uk/)
* [United Nations: Cyber SchoolBus Resources](http://www.un.org/cyberschoolbus/index.asp)