# Fall 2024 Semester Project Detailed Instructions

**Due Date:**I *must* receive your project no later than 11:59 PM on the due date in the syllabus. Given the early advanced notice and the amount of in-class time devoted to this project,there will be***no extensions granted****.* Projects turned in after the deadline will receive a grade of 0.

**Document Format:** You must submit your project report in *a single file* through Canvas. The acceptable formats are Microsoft Word (\*.xlsx) or PDF (\*.pdf) – no exceptions. The submission page is in the Semester Project module. *Projects submitted in multiple parts, in a format other than Word or PDF, or via email/hardcopy will be rejected.*

**Style Requirements:** The Semester Project module contains a sample project that would receive a 100% grade. Your report should be formatted similarly is a simplified APA7.

* The first page of your report must be a title page containing the title “Semester Project” (in bold), your name, name of school, the course and section number, and the submission date.
* Page numbers in the top right
* Use a font suitable for an official business document. Any standard typeface is acceptable as long as it is readable and presents a professional appearance (Calibri and Times New Roman are good examples). The size should be no smaller than 12 point.
* Do not include any borders, illustrations, or watermarking on any page (title page included).
* Embed all graphics directly into your project file. I will not accept separate files containing graphics.

**Data Set:** All students will use the same data set: hybrid\_reg.csv. The data set is located in the Canvas Module. The variables of interest are ***Height*** and ***Weight***.

**Technology Requirements:** Except where required to build graphs or charts, all numerical calculations *must* be performed using StatCrunch. ***Do not use a graphing calculator, Excel, standard normal tables, or any other method for your numerical calculations.***

**Graphics Requirements:** All graphics must be constructed using StatCrunch, Excel, or other computer-based graphics program. Hand-drawn plots, cell phone pictures of graphics, etc., are not acceptable. All graphics must include an informative title and (except for boxplots) correct labels for both axes. Orient all boxplots horizontally.

**Rounding Rules:** In Sections 1-4, round all calculated sample statistics to one further decimal place than the data values. In Section 5, round response variable values to one decimal place and all other regression calculations to four decimal places. In all sections, round p-values to three decimal places. Add trailing zeroes to any rounded value as needed.

**Required Content:** Organize your report in five separate sections using the following numbers and titles. The required elements for each section are as follows:

PART I

*Section 1 – Descriptive Statistics.*

1. For each Height and Weight, find the mean, range, variance, standard deviation, and five-number summary. Display these numbers in a format that is easy to understand.
2. Construct a regular boxplot for each variable. For each boxplot, include a brief statement containing an assessment of whether the data appear to be symmetric, left-skewed, or right-skewed.
3. For each variable, construct a modified boxplot and use it to identify potential outliers. ***If any exist***, list them by value.

*Section 2 – Visual Data Assessment.* For each variable of interest – Height and Weight – create a grouped frequency histogram with exactly 7 intervals (towers). Use the data from Section 1. Use the minimum value of each as the “Start at:”. You will need to find the interval width by taking the range of each respective data and dividing by 8. Always round up or the largest data point(s) will not fit and a new interval will be created. For each histogram, include a paragraph that answers each of the following questions:

1. Is the histogram symmetric, left-skewed, or right-skewed?
2. How many peaks does the histogram have, and in which class(es) are they located (must include the correct lower and upper bounds for each class listed)?
3. Does the histogram have any gaps between classes? If so, where are they?

PART II

*Section 3 – Confidence Intervals.*

Construct a 90% confidence interval for the mean *μ* of each variable (two intervals total). Use algebraic format for each interval. State the distribution you used for each interval ( or normal).

Summarize your Data – Include a statement for each data set resembling:

Based on the sample we can be \_\_\_% confident the population \_\_\_\_\_\_\_\_ lies between #### and #####.

*Section 4 – Hypothesis Test.* Conduct two formal hypothesis tests.

The first hypothesis will test of the claim that, , the mean Height is less than 6’7” (notice will have to convert that to total inches for this test). Use. Include the following in your written summary of the results:

1. Your null and alternate hypotheses in the proper format using standard notation.
2. The type of distribution you used ( or normal).
3. The p-value and its logical relationship to ( or).
4. Your decision regarding the null hypothesis: reject or fail to reject.
5. A statement interpreting your decision: reject/fail to reject (or support/fail to support) the original claim that the mean Height is less than 6’7”. The final statement must be similar to:

*Based on the sample there is/is not sufficient evidence to support/not support the claim that the mean Height is less than 6’7”*

The second hypothesis will test of the claim that, , the mean Weight is 224 pounds. Use. Include the following in your written summary of the results:

1. Your null and alternate hypotheses in the proper format using standard notation.
2. The type of distribution you used ( or normal).
3. The p-value and its logical relationship to ( or).
4. Your decision regarding the null hypothesis: reject or fail to reject.
5. A statement interpreting your decision: reject/fail to reject (or support/fail to support) the original claim that the weight is 224 pounds. The final statement must be similar to:

*Based on the sample there is/is not sufficient evidence to support/not support the claim that the mean Weight is 224 pounds.*

*Section 5 – Correlation/Regression Analysis.*

1. Construct the least-squares equation *(must be in proper algebraic format for full credit)*.
2. Create a scatter plot with Height as the predictor and Weight as the response. The plot must include an informative title, correct labels for both axes, and a plot of the least-squares equation from 5a.
3. Use the coefficient of determination to identify the percentage of the variation in Weight explained by the variation in Height.