**STAT 7500 – Statistical Programming**

**Homework #1**

Complete the following tasks:

1. Open SAS and enter the following code in the editor window, exactly as written.

**data** my\_first\_data;

input name $ id;

mydatalines;

Manish 577

Arlene 499

Dzvinka 140

;

**run**;

1. Run the code above and check the log window for errors. State what error was found.
2. Fix the code and verify that it runs correctly.
3. Run a PROC PRINT to see the dataset that you created. Please include a printout of the data.

For this and future homeworks using SAS, please choose one of the following options to turn in:

* **A single file**, in Word or some other word processing program, that includes relevant information copied from the SAS editor(s) and the SAS output/reviewer file(s). In addition, please make sure to include your responses to specific inquiries in the assignment. Please also include your entire SAS (editor window) file as an Appendix. On the next page, I offer an example of what this format might look like. **Note - this is the preferred method of submitting homework.**
* **Multiple files** that include your SAS editor(s), the SAS output/reviewer file(s), and a place where you write responses to specific inquiries in the assignment. For this assignment, this refers to the line in problem #2 that asks you to “Comment on what these outputs indicate to you”. This could be either in a separate document that clearly identifies which problem you are responding to (i.e., #2 for this assignment) or embedded in the appropriate places as comments in your SAS editor(s). Note, Blackboard should accept multiple files from you, but if there is a problem, you should try zipping your files together and submitting a single zipped file.

**Homework Submission Example**

On the next page is an example of a homework submission document meeting the criteria from option 1. Of note, anything in **yellow and bold** is part of the problem assigned. This assignment was made up and to give you a sense of the *format* that I am looking for. You are not necessarily supposed to understand all the code (yet).

Student Name

Statistical Programming

Homework #1 - Sept 1, 2021

1. **Read in the health data file entitled “heart health study.csv”, which can be found on the website.**

I downloaded the file hhstudy2.csv and saved it to my local machine. I then read it in using the following code:

**data** hhs;

infile "C:\ Stat Prog - Fall 2021\SAS\HHSTUDY2.csv";

input id ht smoke;

**run**;

To check the data, I ran a PROC PRINT and verified that the data was read in properly, as it had 50 observations and 3 variables. The output is not included here. (professor’s note - even though this was not specifically asked for in the problem, it’s a good idea to check your data after you enter it)

1. **Calculate the mean height for smokers (*smoke* = 1) and non-smokers (*smoke* = 0). Interpret the results that you get.**

**proc** **means** data=hhs noobs;

var ht;

class smoke;

**run**;

I added “no obs” option to stifle the number of observations being printed to make it look nicer. The following output was produced (prof note – this was done by highlighting output and pasting it here):

The MEANS Procedure

| **Analysis Variable : ht** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **smoke** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| **0** | 29 | 161.3103448 | 4.7667258 | 152.0000000 | 172.0000000 |
| **1** | 21 | 163.5714286 | 6.4928532 | 150.0000000 | 175.0000000 |

From this, we see that smokers were taller than non-smokers, being 163.6 cm on average compared to 161.3 cm on average. In addition, there was more variability in heights for the smokers see both from the higher standard deviation (6.5 vs. 4.8 respectively) as well as the lower minimum (150 vs. 152) and higher maximum (175 vs. 172). In addition, we see that the majority of people in the data were non-smokers (29 out of the 50 or 58%).

Appendix – Entire SAS Code

**data** hhs;

infile "C:\Stat Prog - Fall 2021\SAS\HHSTUDY2.csv";

input id educ wt ht age smoke physact glucose chol sysbp pondidx;

**run**;

**proc** **print** data=hhs;

**run**;

**proc** **means** data=hhs;

var ht;

class smoke;

**run**;