**Introduction to Programming with Python**

**Homework 6**

***This homework does not need to be submitted:***

***You may work on it with whomever you wish***

1. (100 points) **Some NumPy Practice**
2. In IDLE or the Python development environment of your choice, create a Python code file **hw6.1.py**. At the top of the file, add comments giving the name of the file and your homework group members as authors. Next, import the **numpy** module with the conventional abbreviation, **np**. Create a 1-dimensional **ndarray** object named **a1** containing the sequence of values 6, 7, 8, 9, and 10; then, display **a1** using the **print()** function.
3. The **observations.csv** file contains a record of information for each patient at a doctor’s office during a day. Each record gives the *age* in years, *height* in inches, and *weight* in pounds of the patient. Read from the **observations.csv** file into a 2-dimensional **ndarray** named **aobs**. Display **aobs**, the *shape* of **aobs**, and the *number of dimensions* of **aobs**.
4. Display just the first 8 patient records of **aobs**. (*Hint:* display a slice of rows.)
5. Display just the last 4 patient records of **aobs**.
6. Display the middle 10 patient records of **aobs**.
7. Display just the *ages* of all the patients. (*Hint:* display all rows, and one column.)
8. Display the *heights* and *weights* of all the patients. (*Hint:* display all rows, and a slice of columns.)
9. Display the *heights* and *weights* of the middle 10 patients. (Hint: display a slice of columns from a slice of rows.)
10. A *Boolean index* is more flexible in selecting rows/columns from a 2-dimensional **ndarray**. For example, this statement prints the *heights* of all the patients using a slice for rows and a Boolean index for columns:

**print(aobs[:, [False, True, False]])**

Display the *ages* and *weights* of the first 10 patients.

1. Create a Boolean index for the *rows* of **aobs**, where the Boolean value is **True** if the patient’s height is >= 70 inches (5 ft. 10 in.), and **False** otherwise. Display this Boolean index. (*Hint*: see slide 13 of the Week 6 Part 2 Lecture materials.)
2. Using the Boolean index from (j), select just the *rows* from **aobs**, where the patient’s height is >= 70 inches (5 ft. 10 in.). Display the selected rows. (Hint: see slide 10 of the Week 6 Part 2 Lecture materials.)
3. NumPy provides many statistical functions that work on **ndarray**s: **mean()**, **min()**, **max**(), **std()** (standard deviation), **corrcoef()** (correlation coefficients), etc. For example, given:

**a2 = np.array([1, 2, 3, 4, 6])**

then **a2.mean()** is **3.2**. Define **a2** as shown above, and display the mean, min, max, and

standard deviation for **a2**.

1. The *correlation coefficient* describes how correlated (or not) two data sets are. A given data set is perfectly correlated with itself, that is, has correlation 1.0. The **ndarray**s **a1** (from part (a)) and **a2** (above) are highly but not perfectly correlated:

**print(np.corrcoef(a1, a2))**

displays: **[[1. 0.98639392]**

**[0.98639392 1. ]]**

1. Display the mean, min, max, and standard deviation of the *ages* of all patients in **aobs**.
2. Display the mean, min, max, and standard deviation of the *heights* of all patients in **aobs**.
3. Display the mean, min, max, and standard deviation of the *weights* of all patients in **aobs**.
4. What is the correlation between *age* and *height* for all patients in **aobs**? (Is this surprising? Make a comment in your code.)
5. What is the correlation between *age* and *weight* for all patients in **aobs**? (Is this surprising? Make a comment in your code.)
6. What is the correlation between *height* and *weight* for all patients in **aobs**? (Is this surprising? Make a comment in your code.)