AI+X: Introduction to data science

AIX0008 (Summer 2022)

Patrice Koehl
koehl@cs.ucdavis.edu

July 8, 2022

Homework 2 - For July 11th, 2022

Exercise 1 *(10 points)*

Consider a dataset of 12 numbers:

5 7 4 150 5 7 3 5 5 8 100 5

a) Find the mean and median. Which is larger?

b) Find the skew using the formula $\sum_{i=1}^{N} \frac{(X(i)-\mu)^3}{\sigma^3}$ where $X(i)$ are the different values, μ is the mean, σ is the standard deviation, and the sum runs over all N numbers (here $N = 12$).

Exercise 2 *(10 points)*

A sample of 30 distance scores measured in yards has a mean of 10, a variance of 9, and a standard deviation of 3.

a) You want to convert all your distances from yards to feet, so you multiply each score in the sample by 3. What are the new mean, variance, and standard deviation?

b) You then decide that you only want to look at the distance past a certain point. Thus, after multiplying the original scores by 3, you decide to subtract 4 feet from each of the scores. Now what are the new mean, variance, and standard deviation?

Exercise 3 *(10 points)*

You recorded the time in seconds it took for 8 participants to solve a puzzle. These times appear below. However, when the data was entered into the statistical program, the score that was supposed to be 22.1 was entered as 21.2. You had calculated the following measures of central tendency: the mean, the median, and the following measures of variability: the standard deviation and the variance. Which of these measures will change when you correct the recording error?

15.2 18.8 19.3 19.7 20.2 21.8 22.1 29.4

Exercise 4 *(10 points)*

The table below represents the percentages of the votes cast for the Democratic candidate in the U.S. presidential years 1952-2008:

Year	California	Oregon	Washington
1952	42.7	38.9	44.7
1956	44.3	44.8	45.4
1960	49.6	44.7	45.4
1964	59.1	63.7	62.0
1968	44.7	43.8	47.2
1972	41.5	42.3	38.6
1976	47.6	47.6	46.1
1980	35.9	38.7	37.3
1984	41.3	43.7	42.8
1988	47.6	51.3	50.1
1992	46.0	42.5	45.1
1996	51.1	47.2	49.8
2000	53.4	47.0	50.2
2004	54.3	51.3	52.8
2008	61.0	56.7	57.7

Fill in the following table that provides a five-number summary for the Democrat percentages in the three states:

State	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
California					
Oregon					
Washington					

Exercise 5 *(10 points)*

You encounter a problem on an exam with only answer choices:

a) Option 1

b) Option 1 or Option 2

c) Option 2 or Option 3

You do not know what those options are, as the question has been omitted, but you know that only one answer (a, b, or c) is possible. Can you find that answer? Explain your reasoning.

Extra credit *(5 points)*

You are on an island inhabited by three types of people: knights (always make true statements), knaves (always make false statements) and spies (sometimes make true statements and sometimes make false statements). You encounter three people, A, B, and C and you know that one is a knight, one is a knave, and one is a spy. Determine, if possible, what A, B, and C are if they address you in the way described. If you cannot determine what these three people are, can you draw any conclusions?

A says "C is a knave"

B says "A is the knight"

C says "I am the spy"