MATH 474 Probability and Statistics
Spring 2026

HOMEWORK 2

Due on Canvas at 8:59pm on Wednesday, January 28.
Please write down the time you spent on this homework on the top

Problem 1 (14 points). (Exercises 6.1.1-6.1.6, 6.2.2-6.2.3)

(a) Will the sample mean always correspond to one of the observations in the sample?

(b) Will exactly half of the observations in a sample fall below the mean?

(c) Will the sample mean always be the most frequently occurring data value in the sample?

(d) For any set of data values, is it possible for the sample standard deviation to be larger than the sample mean? If so, give an example.

(e) Can the sample standard deviation be equal to zero? If so, give an example.

(f) When will the median of a sample be equal to the sample mean?

(g) When will the median of a sample be equal to the mode?

Problem 2 (10 points). (Exercise 6.3.4) An article in *Technometrics* ["Validation of Regression Models: Methods and Examples" (1977, Vol. 19(4), p. 425)] presented the following data on the motor fuel octane ratings of several blends of gasoline:

88.5	98.8	89.6	92.2	92.7	88.4	87.5	90.9	94.7	88.3	90.4	83.4	87.9
92.6	87.8	89.9	84.3	90.4	91.6	91.0	93.0	93.7	88.3	91.8	90.1	91.2
90.7	88.2	94.4	96.5	89.2	89.7	89.0	90.6	88.6	88.5	90.4	84.3	92.3
92.2	89.8	92.2	88.3	93.3	91.2	93.2	88.9	91.6	87.7	94.2	87.4	86.7
88.6	89.8	90.3	91.1	85.3	91.1	94.2	88.7	92.7	90.0	86.7	90.1	90.5
90.8	92.7	93.3	91.5	93.4	89.3	100.3	90.1	89.3	86.7	89.9	96.1	91.1
87.6	91.8	91.0	91.0									

(a) Calculate the sample mean, variance, standard deviation, median and quartiles.

(b) Construct histograms with 8 and 16 bins for the data. Compare the histograms. Do both histograms display similar information?

Problem 3 (8 points). (Exercise 6.3.9) An important variation of a histogram for categorical data is the Pareto chart. This chart is widely used in quality improvement efforts, and the categories usually represent different types of defects, failure modes, or product/process problems. The categories are ordered so that the category with the largest frequency is on the left, followed by the category with the second largest frequency, and so forth. These charts are named after Italian economist Vilfredo Pareto, and they usually exhibit "Pareto's law"; that is, most of the defects can be accounted for by only a few categories. Suppose that the following information on structural defects in automobile doors is obtained: dents, 4; pits, 4; parts assembled out of sequence, 6; parts undertrimmed, 21; missing holes/slots, 8; parts not lubricated, 5; parts out of contour, 30; and parts not deburred, 3. Construct and interpret a Pareto chart.

Problem 4 (10 points). (Exercises 6.1.11, 6.4.5) The following data are the joint temperatures of the O-rings (°F) for each test firing or actual launch of the Space Shuttle rocket motor (from Presidential Commission on the Space Shuttle Challenger Accident, Vol. 1, pp. 129–131):

$$84, 49, 61, 40, 83, 67, 45, 66, 70, 69, 80, 58, 68, 60, 67, 72, 73, 70,$$
$$57, 63, 70, 78, 52, 67, 53, 67, 75, 61, 70, 81, 76, 79, 75, 76, 58, 31.$$

(a) Compute the sample mean, standard deviation, median, lower and upper quartiles.

(b) Set aside the lowest observation (31°F) and recompute the quantities in parts (a). Comment on your findings. How different are the other temperatures from 31°F?

(c) Construct a box plot of the full data set and comment on the possible presence of outliers.

Problem 5 (8 points). (Exercises 6.4.3 and 6.4.8) The "cold start ignition time" (in seconds) of an automobile engine is being investigated by a gasoline manufacturer. For one formulation of gasoline, the following times were obtained for a test vehicle:

$$1.75,\ 1.92,\ 2.62,\ 2.35,\ 3.09,\ 3.15,\ 2.53,\ 1.91. \tag{1}$$

A second formulation of gasoline was tested in the same vehicle, producing:

$$1.83,\ 1.99,\ 3.13,\ 3.29,\ 2.65,\ 2.87,\ 3.40,\ 2.46,\ 1.89,\ 3.35. \tag{2}$$

(a) For each formulation, compute the median and the lower and upper quartiles.

(b) Using your results from (b), construct side-by-side box plots and write a brief interpretation comparing the two formulations.