ME 453: Data Science in Manufacturing Quality Control
Homework 1

Assigned: September 25, 2023 Due: October 13, 2023

Homework guidelines:

1. The total number of points is 100 and 130 for 3-credit-hour and 4-credit-hour sections, respectively. The assigned points are given beside questions. To get full credit you must SHOW ALL OF YOUR WORK.

2. Problems marked by "[4 credit only]" should be completed by students in the 4-credit-hour section. Students in the 3-credit-hour section are welcome to work on these problems and the earned points will be used to compensate for lost points in homework assignments.

3. A complete submission features the following items: (a) a brief report including all figures and results, and explanations of necessary steps taken to obtain them; and (b) and the source code (Python is recommended). Item (a) can be scanned copies of handwritten or typeset reports. Both items shall be submitted through Canvas.

Review the following topics discussed in lecture:

1. Course overview and introduction (lecture videos 1–2).

2. Modeling process quality (lecture videos 3–4).

3. Work session: Python basics (lecture video 5).

4. Inferences about quality (lecture videos 6–7).

Problem 1 (20 points + 5 points)

The time to failure in hours of an electronic component subjected to an accelerated life test is shown in Table 1. To accelerate the failure test, the units were tested at an elevated temperature (read down, then across).

127	124	121	118
125	123	136	131
131	120	140	125
124	119	137	133
129	128	125	141
121	133	124	125
142	137	128	140
151	124	129	131
160	142	130	129
125	123	122	126

Table 1: Electronic Component Failure Time

Use software to complete the following problems.

(1) Calculate the **sample** average and standard deviation. (6 points)
Note: `np.std` and many other packages use population standard deviation by default. You will have to manually correct that, or write a function on your own.

(2) Construct a histogram. (5 points)
Note: A rule of thumb is to use $\sqrt{n}$ bins for a set of data with n entries.

(3) Find the sample median and the lower and upper quartiles. (9 points)
Note: Your answer may be slightly different depending on your package of choice.

(4) [4 credit only] Generate a time-series plot for the test data. Are there samples that appear to be different with others? Justify your answer. Note that this is an open-ended question. Any answer that is well justified will receive full credit. (5 points)

Problem 2 (15 points)

The net contents in ounces of canned soup is a random variable with probability distribution:

$$f(x) = \begin{cases} 4(x - 11.75) & 11.75 \leq x \leq 12.25 \\ 4(12.75 - x) & 12.25 \leq x \leq 12.75 \end{cases}$$

Find the probability that a can contains less than 12 ounces of product.

Problem 3 (25 points + 15 points)

The diameters of aluminum alloy rods produced on an extrusion machine follow a normal distribution. It is known that the standard deviation of this distribution is 0.0001 in. A random sample of 25 rods has an average diameter of 0.5046 in.

(1) Test the hypothesis that mean rod diameter is 0.5025 in. Assume a two-sided alternative and use $\alpha = 0.05$. (8 points)

(2) Find the p-value for this test. (8 points)

(3) Construct a 95% two-sided confidence interval on the mean rod diameter. (9 points)

(4) [4 credit only] The specification limits of the rod diameter are 0.5025 ± 0.0005. What is the expected defective rate? (8 points)

(5) [4 credit only] Following (4), if we are able to adjust the process mean (i.e., the average diameter produced by the extrusion machine) to the nominal design value 0.5025, what will the expected defective rate be? (7 points)

Problem 4 (40 points + 10 points)

See the attached excel file `Welding data.csv`. An ultrasonic metal welder was used to join copper sheets under 3 kinds of machine configurations (A, B, and C). Five features were extracted from online sensing signals. Use Python to complete the following tasks.

Part I. Global analysis (25 points)

(1) How many welds were produced during this period? (5 points)

(2) Construct a table containing Pearson's correlation coefficient for each arbitrary pair of features. Display your table as a heatmap and show the coefficient on each block of the heatmap. Also show the scale of the heatmap by adding the colorbar. (10 points)
Hint: take a look at `pandas.DataFrame.corr` and matplotlib examples.

(3) Which two features have the strongest correlation? Draw a scatter plot for them. (5 points)

(4) Plot histograms for all features. (5 points)

Part II. Different machine configurations (15 points + 10 points)

(6) Plot each feature in a single time-series graph. (5 points)

(7) Calculate the mean, median, range, and variance of all features for each machine configuration. (10 points)

(8) [4 credit only] Assuming normal distributions, for each feature, determine whether machine configurations B and C have the same mean value as machine configurations A. Justify your answer. (10 points)