Problem 1

At a pharmaceutical production company, a **sample** of pills is taken and their dimensions are measured. The widths are recorded in mm in the stem and leaf plot below:

0 l 6, 7, 7, 7, 8, 8, 9, 9, 9, 9

1 l 0, 0, 0, 0, 2

What is the standard deviation of the sample? Give your answer to 3 decimal places.

Solution:

First solve for the sum of squares:

$$SS = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$SS = 1179 - \frac{17161}{15} = 34.9333$$

Now the standard deviation can be solved for (N-1 is used since we are looking at a sample of the entire population):

$$\sigma = \sqrt{\frac{SS}{N-1}}$$

$$\sigma = \sqrt{\frac{34.93}{15-1}} = 1.580$$

<u>Problem 2</u>

At a caramel production company, the flow rate of the caramel through the pipes varies greatly depending on the product temperature. The different flow rates for a specific recipe are recorded below:

1.1, 1.2, 1.5, 1.5, 1.7, 1.7, 1.7, 1.8, 1.9, 2.0, 2.3

In order for the product to be economically feasible, the z-score of the third quartile must be less than 1. Is this product economically feasible? (Answer "yes" or "no")

Solution:

First we must find the position of the third quartile:

$$N_3 = 0.75(N + 1)$$

$$N_3 = 0.75(11 + 1) = 9$$

The third quartile is the number in the 9^{th} position: 1.9

Now we can solve for the z-score:

$$z = \frac{X - \mu}{\sigma}$$

$$\mu = 1.673$$

$$\sigma = 0.3438$$

$$z = \frac{1.9 - 1.673}{0.3438} = 0.660$$

The z-score is less than 1, therefore the product is economically feasible: yes.

Problem 3

In an attempt to measure the effects of acid rain, researchers measured the pH of water collected from rain in Ingham County, Michigan. It was found that the collected pH values follows a normal distribution, with median pH of 5.38. Two additional data points are then added; pH = 5.21 and 5.40.

(a) Calculate the median of the new sample set.

(b) Given that the old sample size is 10, calculate the mean of the new sample set, giving your answers to 2 decimal places.

Solution:

(a) Before adding data points: 50th percentile location:

$$m_0 = \frac{N+1}{2}$$

After adding the new data points, the new 50th percentile location:

$$m_n = \frac{N+1+2}{2} = m_o + 1$$

The current m_o is pointing to pH = 5.21, therefore the value of median remains at 5.38. The additional data points do not affect the median

(b) Since this is a normal distribution, mean and median are equal. Therefore before adding the data points, $\bar{x}$ can be expressed as:

$$\overline{x_1} = \frac{\sum x}{n} = 5.38$$

After adding the two data points, $\overline{x_2}$ can be expressed as:

$$\overline{x_2} = \frac{\sum x + 5.21 + 5.40}{n+2} = \frac{5.38x10 + 5.21 + 5.40}{12} = 5.37$$

Problem 4

In X-ray fluorescence spectroscopy, film self-absorption and attenuation length can affect the detected signal response significantly. An experiment is performed and data on the detected signal of iodine X and the incident angle to the normal Y. It is known that there is a linear relationship between X and Y. Estimate the missing value in the table by performing an regression of X and Y. Please give your answer to two decimal place.

Y	X
40	825
42	830
49	-
46	895
44	890
48	910

Solution:

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{N} = 3790750 - \frac{(4350)^2}{5} = 6250$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{N} = 9720 - \frac{72900}{5} = 40$$

$$SP = \sum xy - \frac{\sum x \sum y}{N} = 191870 - \frac{4350 \times 220}{5} = 470$$

$$r = \frac{SP}{\sqrt{SS_{xx} SS_{yy}}} = \frac{470}{\sqrt{6250 \times 40}} = 0.94$$

$$b = r\sqrt{\frac{SS_{xx}}{SS_{yy}}} = 11.75$$

$$a = \bar{X} - b\bar{Y} = 870 - 11.75 \times 44 = 353$$

$$X = 353 + 11.75Y$$

$$X = 353 + 11.75 \times 49 = 928.75$$

Problem 5

In a nanoparticle synthesis experiment, a nanoparticle tracking analysis machine measures the radius distribution of the synthesized nanoparticles in nm. A stem-and-leaf plot is generated to show this data. From a TEM image, you measure a single nanoparticle to be 4.5 nm. Using the information provided in the stem-and-leaf plot below, what is the approximate percentile rank of the particle's radius? Please give your answer to three decimal places.

01|1
02|2, 4, 5, 6, 7
03|3, 3, 4, 6, 6, 7, 8, 9, 9
04|3, 4, 6, 7, 7
05|2, 3, 3, 3, 4, 5, 5, 6, 6
06|1, 2, 3, 5,
07|2, 5

Solution:

The total number of samples is 35. With the new nanoparticles, the number of samples is 36. Number of radius that with value smaller than 4.5 nm is 17. So the percentile rank of 4.5 is:
$\frac{17}{36} \approx 0.472$

Problem 6

Graphene is a promising material for the development of new electronic devices. In general, only graphite with less than 10 monolayers of carbon atoms can be defined as graphene. If you use adhesive tape for graphite sample preparation, the number of monolayers of carbon atoms in the prepared graphite sample is a random variable with $\mu = 20, \sigma = 40$. Suppose you repeat the experiment 100 times. Of these 100 trials, what is the probability that the mean value of the number of monolayers is less than 10? Please give your answer to four decimal places.

Solution:

Based on the central limit theorem, for that 100 samples, the sample average $\bar{X}$ follows approximately the normal distribution with $\mu = 20$, and $\sigma' = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{100}} = 4$. So $P(\bar{X} < 10) = P\left(z < \frac{10-20}{4}\right) = P(z < -2.5) = 0.0062$

Problem 7

The viscosity of a water decreases with temperature. Below is a sample of 10 data points collected over a temperature range of 10 °C to 80 °C.

Temperature T (°C)	Viscosity μ (mPa s)
10	1.3059
20	1.0016
30	0.7972
40	0.6527
45	0.5958
50	0.5465
55	0.5036
60	0.466
70	0.4035
80	0.354

A model for how the dynamic viscosity changes with temperature is the exponential model,

$$\mu(T) = \mu_0 e^{-bT}$$

where the temperature T is in Kelvin. Using the data above, determine the coefficients μ_0 and b using linear regression. Give all answers to three significant figures.

Ans:

μ_0 = 223

b = 0.0185

Solution:

First, need to convert all temperature to Kelvin degree.

$$\ln(\mu) = -bT + \ln(\mu_0)$$

We define Y = $\ln(\mu)$ and X=T. Now we have linear relationship, with slop β = -b and the intercept A = $\ln(\mu_0)$

$\beta = \frac{SP}{SS_x} = \frac{\sum XY - \frac{\sum X \sum Y}{10}}{\sum X^2 - \frac{(\sum X)^2}{10}} = \frac{-79.315}{4290} = -0.0185$, so b = 0.0185.

To solve A, $\ln(\mu_0) = \overline{\ln(\mu)} + b\bar{T} = -0.4902 + 319 * 0.01848 = 5.405$, so μ_0= exp(5.405) ≈ 223