Topic	Description
Module	Data Science I
Module Code	MECH-B-4-MLDS-MLDS1-ILV
Semester	SS 2025
Lecturer	Daniel T. McGuiness, Ph.D
ECTS	5
SWS	2
Lecture Type	ILV
Teaching UE	
Coursework Name	A Assignment
Work	Individual
Suggested Private Study	10 hours
Submission Format	Submission via SAKAI
Submission Deadline	June 10th 23:59
Late Submission	Not accepted
Resubmitting Opportunity	No re submission opportunity

No lecture time is exclusively devoted to the aforementioned assignment.

- Please make sure your calculations are legible and easy to follows.
- If no clear indication of derivation is given, no points will be awarded.
- There is no need for text in this assignment.
- You will only be awarded grades on your derivation and solution to the questions given in this assignment.

A portion of the mark for every assignment will be, where applicable, based on style. Style, in this context, refers to organisation, flow, sentence and paragraph structure, typographical accuracy, grammar, spelling, clarity of expression and use of correct IEEE style for citations and references. Students will find *The Elements of Style (3rd ed.)* (1979) by Strunk & White, published by Macmillan, useful with an alternative recommendation being *Economist Style Guide (12th ed.)* by Ann Wroe.

Question	Maximum Point	Received Point
Probability	50	
Statistics	50	
Sum	100	

[Q1] Probability 50

1. The following measurements were recorded for the drying time, in hours, of a certain brand of paint.

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

   Assume that the measurements are a simple random sample.

   a. What is the sample size for the above sample? (5)

   b. Calculate the sample mean for these data. (5)

2. In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there? (10)

3. Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function:

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere} \end{cases}$$

   a. Please verify $f(x)$ is a density function. (5)

   b. Find $P(0 < X \leq 1)$. (5)

4. A food manufacturer is aware that the weight of the product in the box varies slightly from box to box. In fact, considerable historical data have allowed the determination of the density function that describes the probability structure for the weight (in kilograms).

   Letting X be the random variable weight, the density function can be described as

$$f(x) = \begin{cases} \frac{2}{5}, & 23.75 \leq x \leq 26.25 \\ 0, & \text{elsewhere} \end{cases}$$

   a. Please verify $f(x)$ is a valid density function. (5)

   b. Determine the probability that the weight is smaller than 24 kilograms. (5)

   c. The company desires that the weight exceeding 26 kilograms be an extremely rare occurrence. What is the probability this rare occurrence does actually occur? (10)

[A1] Probability 50

1a. There are fifteen samples ($n = 15$) in the dataset. (5)

1b. The mean for the data-set is as follows: (5)

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_{\mathrm{j}} = 3.78667 \quad \blacksquare$$

2. The solution is as follows: (10)

$$P(n,\ k) = P(25,\ 3) = \frac{25!}{(25-3)!} = 13800 \quad \blacksquare$$

3a. We integrate the function across the given range: (5)

$$F(x) = \int_{-1}^{2} \frac{x_2}{3}\,\mathrm{d}x = \left.\frac{x^3}{9}\right|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1 \quad \blacksquare$$

3b. We integrate the equation as follows: (5)

$$F(x) = \int_{-1}^{2} \frac{x_2}{3}\,\mathrm{d}x = \left.\frac{x^3}{9}\right|_{0}^{1} = \frac{1}{9} + \frac{0}{9} = \frac{1}{9} \quad \blacksquare$$

4a. We integrate the function across the given range: (5)

$$F(x) = \int_{23.75}^{26.25} \frac{2}{5}\,\mathrm{d}x = \left.\frac{2x}{5}\right|_{23.75}^{26.25} = \frac{2}{5}\cdot(2.5) = 1 \quad \blacksquare$$

4b. We solve the question as follows: (5)

$$P(X \le 24) = \int_{23.75}^{24} \frac{2}{5}\,\mathrm{d}x = \left.\frac{2x}{5}\right|_{23.75}^{24} = \frac{2}{5}\cdot(0.25) = 0.1 \quad \blacksquare$$

4c. We solve the question as follows: (10)

$$P(X \ge 24) = \int_{26}^{26.25} \frac{2}{5}\,\mathrm{d}x = \left.\frac{2x}{5}\right|_{26}^{26.25} = \frac{2}{5}\cdot(0.25) = 0.1 \quad \blacksquare$$

[Q2] Statistics 50

1. Find the maximum likelihood estimate for the parameter μ of a normal distribution with known variance $\sigma^2 = \sigma_o^2$ (10)

2. Apply the maximum likelihood method to the normal distribution with $\mu = 0$ (10)

3. Five independent measurements of the point of inflammation (flash point) of Diesel oil (D-2) gave the values (in F)

   144 147 146 142 144

   Assuming normality. determine a 99% confidence interval for the mean. (20)

4. Find a 95% confidence interval for the mean μ of a normal population with standard deviation 4.00 from the sample: (10)

   30 42 40 34 48 50

[A2] Statistics 50

1. We obtain the likelihood function:

$$L = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-\mathrm{h}}$$

$$\text{where} \qquad h = \frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(x_{\mathrm{j}} - \mu\right)^2 .$$

   Taking logarithms, we have

$$\ln L = -n \ln \sqrt{2\pi} - n \ln \sigma - h.$$

   The first equation for the parameters is $\dfrac{\partial \ln L}{\partial \mu} = 0$, written out:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{\partial h}{\partial \mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}\left(x_{\mathrm{j}} - \mu\right) = 0,$$

$$\text{therefore} \qquad \sum_{j=1}^{n} x_{\mathrm{j}} - n\mu = 0.$$

   The solution is the desired estimate $\hat{\mu}$ for μ: we find (10)

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} x_{\mathrm{j}} = \overline{x}\,.$$

2. The second equation for the parameter is $\partial \ln L / \partial \sigma = 0$, written out

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} - \frac{\partial h}{\partial \sigma} = -\frac{1}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{n}\left(x_{\mathrm{j}} - \mu\right)^2 = 0.$$

   Replacing μ by $\hat{\mu}$ and solving for σ^2, we obtain the estimate: (10)

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(x_{\mathrm{j}} - \overline{x}\right)^2 \quad \blacksquare$$

3. The solution is as follows:

a) $\gamma = 0.99$ is required based on 99% confidence level.

b) $F(c) = \frac{1}{2}(1+\gamma) = 0.99$ and looking at the reference table with $n-1 = 4$ d.f., which gives $c = 4.60$.

c) Calculating the mean and the variance gives $\overline{x} = 144.6$ and $s = 3.8$,

d) $k = \sqrt{3.8} \times 4.60/\sqrt{5} = 4.01$. Therefore the confidence interval is: (20)

$$\text{CONF}_{0.99}\{140.5 \leq \mu \leq 148.7\} \quad \blacksquare$$

4. We start by first gathering the information. As the confidence interval is set as $\gamma = 0.95$ and we have 6 data points, it means our d.f. is 5. Based on the table we gather c as 1.96. The sample mean is:

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_{\text{j}} = 40.66666$$

Our range is

$$k = \frac{c\sigma}{\sqrt{n}} = \frac{1.96 \cdot 4}{\sqrt{6}} = 3.20067$$

Which gives our confidence interval as: (10)

$$\text{CONF}_{0.99}\{40.667 - 3.2 \leq \mu \leq 40.667 + 3.2\} \quad \blacksquare$$