Problem Set 1 Solutions

Prof. Oke

CEE 260/MIE 273: Probability & Statistics in Civil Engineering *09.24.2021*

Problem 1 *(6 points)*

Respond "T" (*True*) or "F" (*False*) to the following statements. Use the boxes provided. Each response is worth 1 point.

(i) F — The standard deviation of a random sample is a measure of its central tendency.

(ii) T — In a perfectly symmetric distribution, the median is equal to the mean.

(iii) F — A right-skewed distribution has a long left tail.

(iv) T — In a boxplot (introduced in 1970 by John Tukey), the whiskers are conventionally selected as the values closest to within $Q_1 - 1.5IQR$ and $Q_3 + 1.5IQR$. In this case, the distance between the whiskers is $4IQR$.

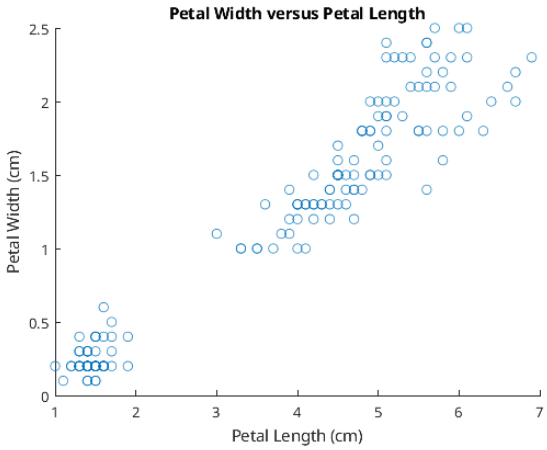
(v) F — The mean of a sample dataset can be determined from its boxplot.

(vi) T — Given A and B are two mutually exclusive events. $A \cap B = \varnothing$.

Problem 2 *(2 points)*

How would you describe the relationship between the variables shown in each of the following plots? Indicate the correct answer in each case (by circling, or otherwise).

[1] **(a)**

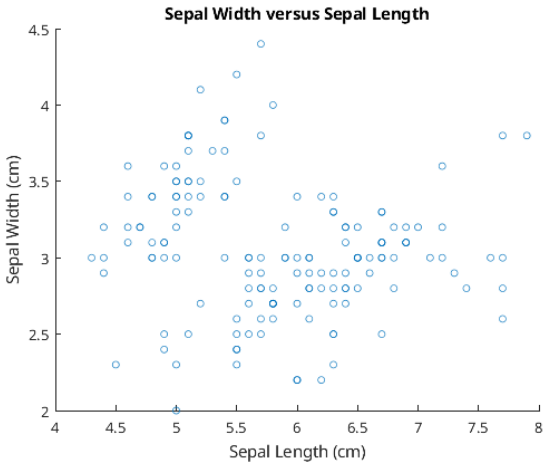


(i) **Approximately linear**

(ii) Nonlinear

(iii) No discernible relationship

[2] **(b)**



(i) Approximately linear

(ii) Nonlinear

(iii) **No discernible relationship**

Problem 3 *(10 points)*

You are given a sample of 43 beak sizes (X) measured from a certain bird species on an island. The sample mean $\overline{X}$ is 10.77 mm and the sample variance $s^2 = 1.048$. The median m is 10.65 mm.

(a) Find the sample standard deviation s of the dataset. *[2]*

$$s = \sqrt{Var} = \sqrt{1.048}$$
$$= \boxed{1.0237}$$

(b) Find the sampling error (also known as the standard error of the mean, SE). *[2]*

$$SE = \frac{s_X}{\sqrt{n}}$$
$$= \frac{1.0237}{\sqrt{43}} = \boxed{0.156}$$

(c) Find the coefficient of variation δ_X. *[2]*

$$\delta_X = \frac{s_X}{\overline{X}}$$
$$= \frac{1.0237}{10.77} = \boxed{0.0951}$$

(d) The first and third quartiles of the sample are given by $Q_1 = 10.2175, Q_3 = 11.22$.

(i) What is the second quartile? $\boxed{Q_2 = m = 10.65}$ *[1]*

(ii) Find the interquartile range (IQR) of the sample. *[1]*

$$IQR = Q_3 - Q_1$$
$$= 11.22 - 10.2175 = \boxed{1.0025}$$

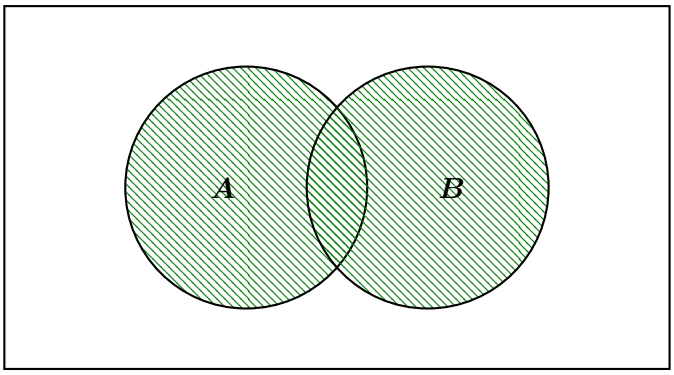
(e) The maximum value of the sample is 13.49 mm. If a boxplot is constructed with the whiskers determined by $Q_1 - 1.5IQR$ and $Q_3 + 1.5IQR$, would you expect to see any outliers in the plot? Explain briefly. *[2]*

Yes. The "maximum" as indicated by the upper whisker would be $Q_3 + 1.5IQR = 11.22 + 1.5(1.0025) = 12.724 < 13.49$. Thus, the true maximum of the sample would appear as an outlier in the boxplot.

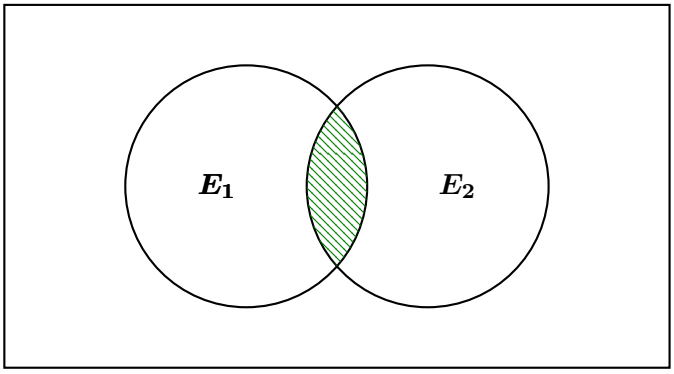
Problem 4 *(6 points)*

Shade the area corresponding to the given events in the following Venn diagrams.

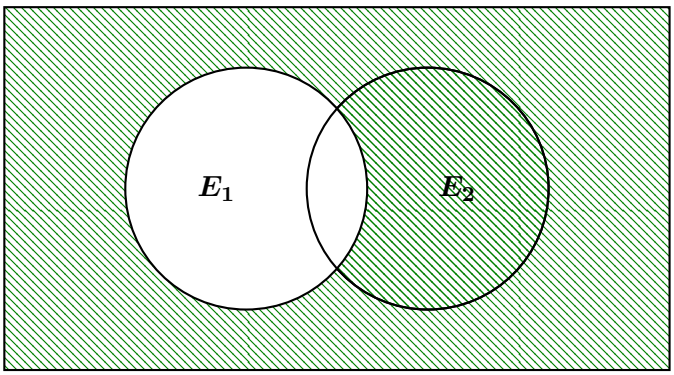
[1] **(a)** $A \cup B$



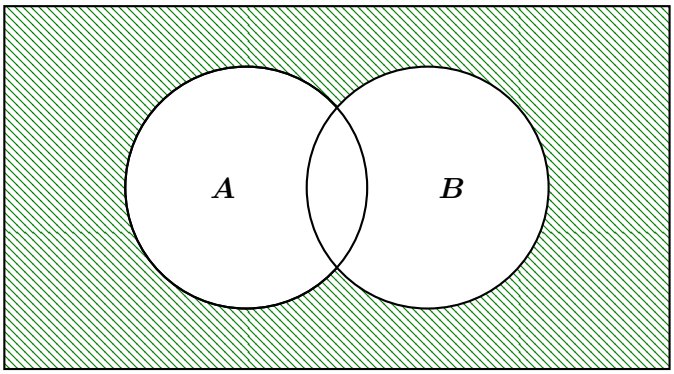
[1] **(b)** $E_1 \cap E_2$



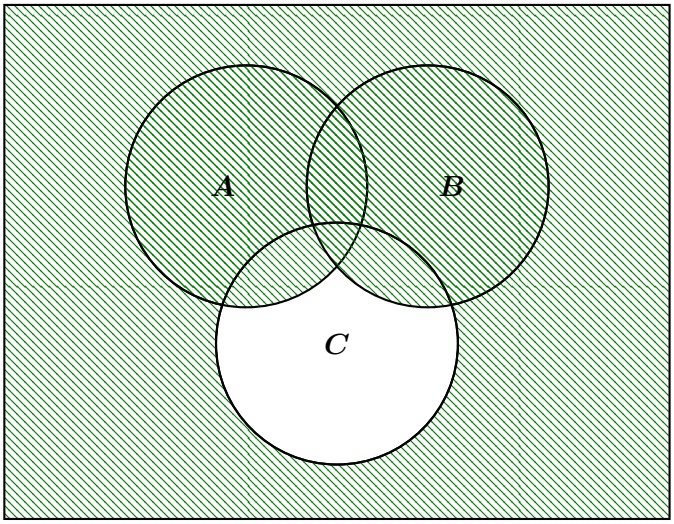
[1] **(c)** $\overline{E_1}$ (Note that $\overline{E_1} \equiv E_1^c$, i.e. the complement of E_1.)



(d) $\overline{A \cup B}$ *[1]*



(e) $A \cup B \cup \overline{C}$ *[1]*



(f) $\overline{AB} \cap C$ *[1]*