CEG2002: Statistics for Civil Engineers
Exercises 1: solutions

QUESTION 1

(a) Stem-and-leaf plot to show average flows through the Tombigbee Dam:

2	34
3	5667
4	0022236
5	479
6	01

n = 18, stem unit = 10, leaf unit = 1

Comment: vaguely symmetrical dataset, which peaks in the interval 40-50.

(b) Mean: 43.17 cubic metres per second
Standard deviation: 11.31 cubic metres per second

(c) Modal flow = 42 cubic metres per second (just look at the plot above!)

QUESTION 2

(a) Mean: 53.67 cubic metres per second
Standard deviation: 16.55 cubic metres per second

Comparing means it looks as though the Arklatex has the highest average flow, though the standard deviation is quite a bit bigger for flows through this Dam, so flows through the Arklatex are obviously more variable.

(b) The mean and standard deviation are both influenced by unusual, or "outlying" observations, and a flow of 175 is certainly unusual for this dataset. Using the mean and standard deviation here would result in over-estimated measures of location and spread.

(c) Placing the values in rank order, we get:

26 39 42 48 58 60 64 65 81 175

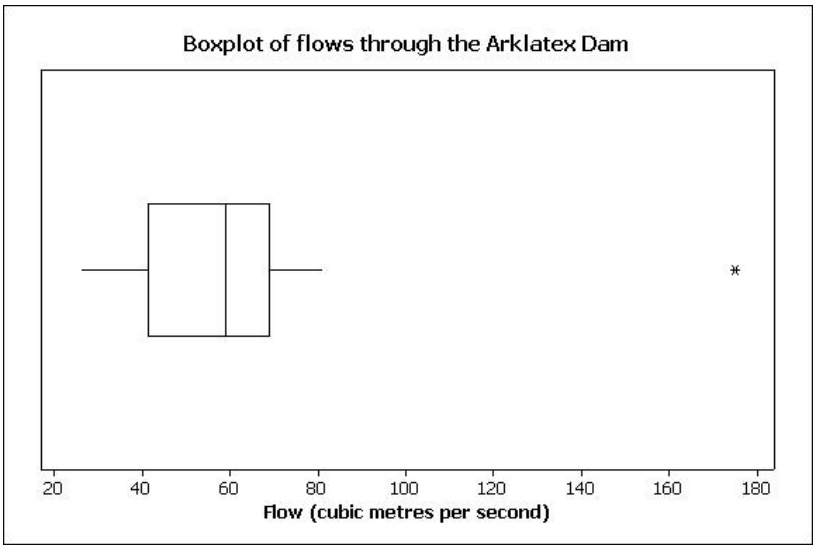
Thus, the median = 59 cubic metres per second

$Q1$ is in position 2.75, and $Q3$ is in position 8.25. Thus,

$Q1$ = 41.25
$Q3$ = 69

Thus, IQR = 27.75 cubic metres per second.

(d) Box and whisker plot for flows through the Arklatex Dam:



Comment: very asymmetric distribution (positive skew), with the "box" being to the far left-hand-side of the "whiskers" line. However, within the middle 50% of the data, there is slight negative skew, with the median line being closer to the upper quartile than the lower. There is at least one outlier (the observation with the maximum value).

QUESTION 3

(a) Since the data are given in a grouped frequency table, we assume each observation takes the midpoint of its group. Doing so gives:

Rainfall (mm)	Frequency	Mid-point (m)
$0\leq x < 10$	4	5
$10\leq x < 20$	15	15
$20\leq x < 30$	22	25
$30\leq x < 40$	25	35
$40\leq x < 50$	20	45
$50\leq x < 60$	16	55
$60\leq x < 70$	8	65
$70\leq x < 80$	0	75
$80\leq x < 90$	0	85
$90\leq x < 100$	1	95

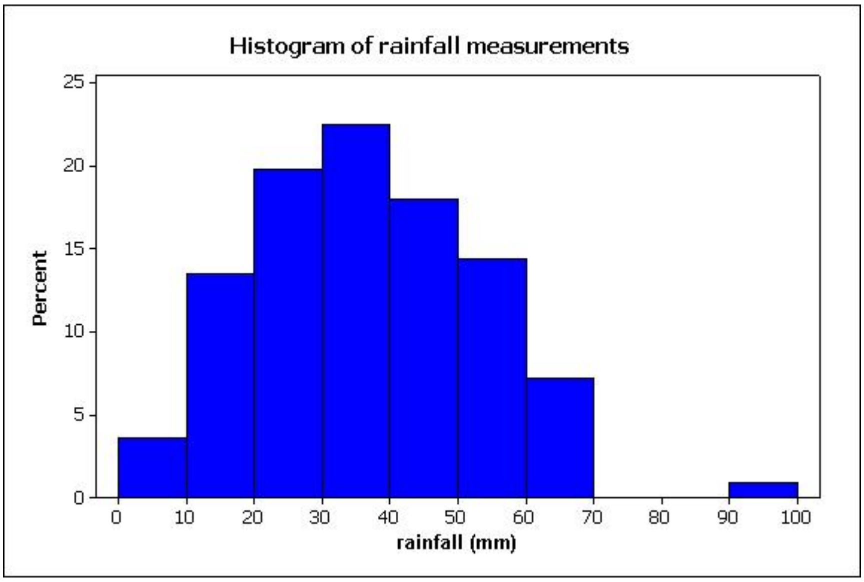
Thus, $\bar{x} \approx \dfrac{(4\times 5)+(15\times 15)+\ldots+(1\times 95)}{111} \approx 36.62mm$

Similarly, $s \approx \sqrt{\dfrac{\sum fm^2 - n\bar{x}^2}{n-1}} \approx 17.36mm$

(b) The answers in part (a) are estimates because we assume that each observation takes the midpoint of its group – that's the best we can do!

(c) Modal class: $30\leq x < 40$

(d) Percentage relative frequency histogram:



Comments: It is slightly asymmetric (slight positive skew). The observation in the range $90 \leq x < 100$ may be an outlier.

QUESTION 4

Let X be the number of beaches which pass. Since we have a fixed number of trials (12), there are two possible outcomes for each trial, and we have a constant probability of 'success' (95% =0.95), we can say that

$$X \sim \text{Bin}(12, 0.95).$$

We need

$$P(X > 9) = P(X \geq 10) = P(X = 10) + P(X = 11) + P(X = 12)$$

Now

$$P(X = 10) = \binom{12}{10} 0.95^{10}(1 - 0.95)^{12-10} = 66 * .59874 * .0025 = .09879.$$

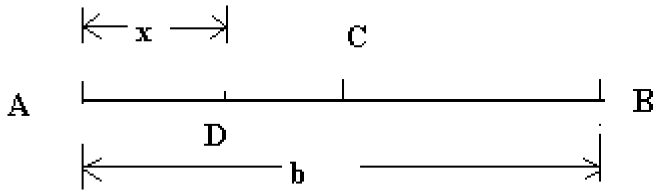
Similarly,

$$P(X = 11) = \binom{12}{11} 0.95^{11}(1 - 0.95)^{12-11} = .3413.$$
$$P(X = 12) = \binom{12}{12} 0.95^{12}(1 - 0.95)^{12-12} = .5404.$$

Thus,

$$\begin{aligned} P(X > 9) &= P(X = 10) + P(X = 11) + P(X = 12) \\ &= .09879 + .3413 + .5404 = .9804 \end{aligned}$$

QUESTION 5



$X \sim U(0,b)$, and from the question, we know that

$$AD = x,\ BD = b - x,\ \text{and } AC = b/2$$

The conditions that *AD, BD* and *AC* form a triangle are

$$\begin{cases} AD + BD \geq AC \\ AD + AC \geq BD \\ BD + AC \geq AD \end{cases}$$

or

$$\begin{cases} x + (b - x) \geq b/2 \\ x + b/2 \geq b - x \\ (b - x) + b/2 \geq x \end{cases}$$

From the above inequalities, it is easy to know that

$$b/4 \leq x \leq 3b/4$$

The probability is therefore

$$P(b/4 \leq x \leq 3b/4) = P(x \leq 3b/4) - P(x \leq b/4)$$

$$= \frac{3b/4 - 0}{b - 0} - \frac{b/4 - 0}{b - 0} = 3/4 - 1/4 = 1/2$$

QUESTION 6

If $Z \sim N(0, 1)$, then, from tables,

(a) $P(Z < 1.75) = 0.9599$
(b) $P(Z > 2.3) = 1 - P(Z < 2.3) = 1 - 0.9893 = 0.0107$
(c) $P(-2 < Z < 2) = P(Z < 2) - P(Z < -2) = 0.9772 - 0.0228 = 0.9544$
(d) $P(-3 < Z < 3) = P(Z < 3) - P(Z < -3) \approx 0.9986 - 0.0014 = 0.9972$, or nearly certain!

QUESTION 7

X: hourly maximum wind speed (mph)

$X \sim N(40, 12^2)$

(a) $P(X < 30) = P\left(Z < \frac{30 - 40}{12}\right) = P(Z < -0.833) = 0.2033$

(b) $P(X > 55) = 1 - P(X < 55)$

$= 1 - P\left(Z < \frac{55 - 40}{12}\right)$

$= 1 - P(Z < 1.25)$

$= 1 - 0.8944 = 0.1056$

(c) $P(15 < X < 45) = P(X < 45) - P(X < 15)$

$$= P\left(Z < \frac{45-40}{12}\right) - P\left(Z < \frac{15-40}{12}\right)$$

$= P(Z < 0.42) - P(Z < -2.08)$

$= 0.6628 - 0.0188$

$= 0.644$

(d) $P(Z > z) = 0.1 \Rightarrow z = 1.28$

Now

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

i.e

$$1.28 = \frac{X - 40}{12}$$

So, rearranging, we get $X = 55.36$ mph.

QUESTION 8

X: number of bacteria (per litre)

(a) $X \sim \text{Po}(5)$

(b) $P(X = 4) = \dfrac{e^{-5} \times 5^4}{4!} = 0.1755$

(c) $P(X \geq 1) = 1 - P(X = 0)$

$= 1 - \dfrac{e^{-5} \times 5^0}{0!}$

$= 1 - 0.00674$

$= 0.99326$

(d) Y: number of bacteria (**per 5 litre**) $\rightarrow Y \sim \text{Po}(25)$

$P(Y = 27) = \dfrac{e^{-25} \times 25^{27}}{27!} = 0.071$