[Marks]

(5) 1. In a study of the effect of a certain artificial hormone on cattle in Canada , 500 cattle are selected at random and for each cow, the ratio of weight(in kilograms) to height(in centimeters) is measured. From this group, an average ratio of 4.5 is computed. Match the terms in column I with the statistical terms in column II. Note that not all terms in column II will be used.

Column I	**Column II**
_______ The average ratio for all cattle in Canada	(a) Data (one)
_______ The ratio for a cow	(b) Data (set)
_______ The computed average ratio 4.5	(c) Experiment
_______ Ratio 4.8 measured for a cow	(d) Parameter
_______ The selected 500 cattle	(e) Population
	(f) Sample
	(g) Statistic
	(h) Variable

Solution:

Column I	**Column II**
(d) The average ratio for all cattle in Canada	(a) Data (one)
(h) The ratio for a cow	(b) Data (set)
(g) The computed average ratio 4.5	(c) Experiment
(a) Ratio 4.8 measured for a cow	(d) Parameter
(f) The selected 500 cattle	(e) Population
	(f) Sample
	(g) Statistic
	(h) Variable

[Marks]

(3) 2. True or False (write down **T** or **F** at the front of each statement)

(a) ______ Given any two events A and B, $P(A \cup B) = P(A) + P(B)$.

(b) ______ Two events A and B can be mutually exclusive and independent at the same time when both $P(A) \neq 0$ and $P(B) \neq 0$.

(c) ______ If a teacher adds 5 marks to everyone's test score, the standard deviation will remain unchanged.

Solution:

(a) **F** Given any two events A and B, $P(A \cup B) = P(A) + P(B)$.

(b) **F** Two events A and B can be mutually exclusive and independent at the same time when both $P(A) \neq 0$ and $P(B) \neq 0$.

(c) **T** If a teacher add 5 marks to everyone's test score, the standard deviation will remain unchanged.

3. Consider the following data:

x	10	11	12	15	20	25	30
f	3	2	4	4	2	2	1

(2) (a) Find the mean, median, mode and P_{40}.

Solution:

$$\overline{x} = \frac{3 \cdot 10 + 2 \cdot 11 + 4 \cdot 12 + 4 \cdot 15 + 2 \cdot 20 + 2 \cdot 25 + 1 \cdot 30}{3 + 2 + 4 + 4 + 2 + 2 + 1} = \frac{280}{18} = 15.5555$$

$$\widetilde{x} = \frac{12 + 15}{2} = 13.5$$

mode = no mode

As for P_{40}, $k = 40$. So $\frac{nk}{100} = \frac{18 \cdot 40}{100} = 7.2$. So P_{40} is the number at the 8-position of the sorted data, i.e. $P_{40} = 12$.

(3) (b) Find the standard deviation.

Solution:

$$s = \sqrt{\frac{\sum x^2 f - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}} = \sqrt{\frac{4968 - \frac{78400}{18}}{18 - 1}} = 6.0022$$

Question 3 continues on the next page.

[Marks]

(2) (c) Find the midquartile and interquartile range.

Solution: For Q_1, $\frac{nk}{100} = \frac{18 \cdot 25}{100} = 4.5$. So Q_1 is the fifth number, i.e. $Q_1 = 11$. For Q_3, $\frac{nk}{100} = \frac{18 \cdot 75}{100} = 13.5$. So Q_3 is the fourteenth number, i.e. $Q_3 = 20$. Therefore

$$\text{midquatile} = \frac{Q_3 + Q_1}{2} = \frac{11 + 20}{2} = 15.5$$
$$\text{interquartile} = Q_3 - Q_1 = 20 - 11 = 9$$

(2) (d) Find the range and midrange.

Solution:

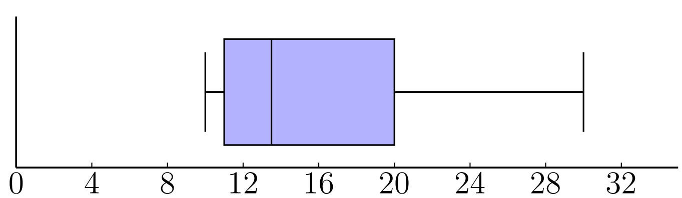
$$\text{midrange} = \frac{L + H}{2} = \frac{30 + 10}{2} = 20$$
$$\text{range} = H - L = 30 - 10 = 20$$

(5) (e) Write down the 5-number summary and construct a box-and-whisker plot.

Solution: The 5-number summary is

L	Q_1	$\tilde{x}$	Q_3	H
10	11	13.5	20	30

The box-and-whiskers plot is



(10) 4. At a certain college, some students were asked how many books they read each year. The result is shown in the following table.

Number of Books (Class limits)	Frequency f	Class Mark (x)	Cumulative Frequency	Cumulative Relative. Frequency
$0 - 2$	30			
$2 - 4$	22			
$4 - 6$	55			

Question 4 continues on the next page.

[Marks]

$6 - 8$	20			
$8 - 10$	5			

(a) Complete the table above.

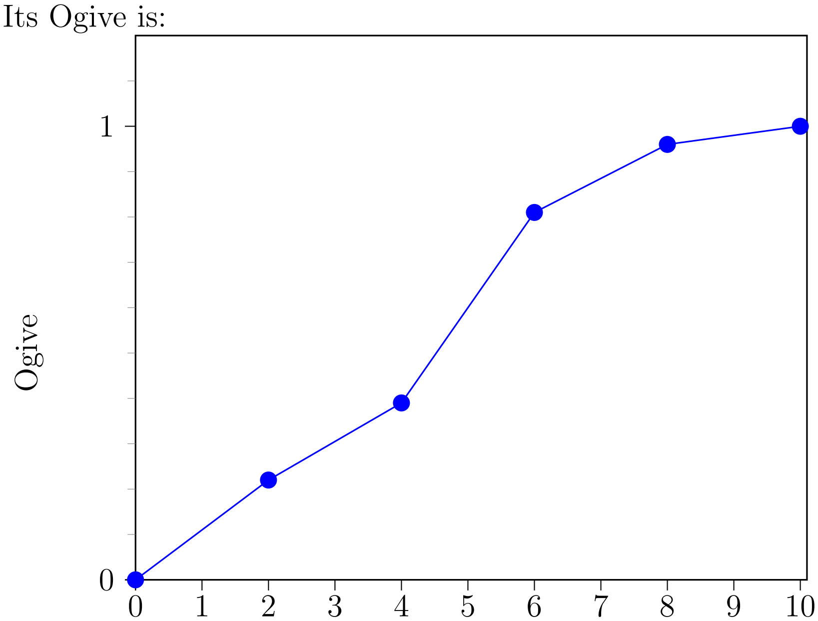
Solution:

Distances (Class limits)	Frequency f	Class Mark (x)	Cumulative Frequency	Cum. Rel.. Freq.
$0 - 2$	30	1	30	0.2273
$2 - 4$	22	3	52	0.3939
$4 - 6$	55	5	107	0.8106
$6 - 8$	20	7	127	0.9621
$8 - 10$	5	9	132	1

(b) Draw an ogive for the distribution.

Solution:

Its Ogive is:



(3) 5. A population has mean $\mu = 40$ and standard deviation $\sigma = 3$. According to Chebyshev's theorem, what interval must contain at least 75% of the data?

[Marks]

Solution: From Chebyshev's theorem,

$$1 - \frac{1}{k^2} = 75\% = 0.75$$

So

$$k = 2$$

therefore

$$-2 \leq \frac{x-\mu}{\sigma} \leq 2; \quad -2 \leq \frac{x-40}{3} \leq 2;$$
$$-6 \leq x - 40 \leq 6; \quad 34 \leq x \leq 46$$

(6) 6. Let A and B be two events. Suppose $P(A) = 0.3$, $P(B|A) = 0.1$, and $P(A|B) = 0.3$. Find the following:

(a) $P(\overline{A})$

Solution:

$$P(\overline{A}) = 1 - P(A) = 1 - 0.3 = 0.7$$

(b) $P(A \cap B)$

Solution:

$$P(A \cap B) = P(A)P(B|A) = 0.3 \cdot 0.1 = 0.03$$

(c) $P(B)$

Solution: Since $P(A|B) = \frac{P(A \cap B)}{P(B)}$,

$$P(B) = \frac{P(A \cap B)}{P(A|B)} = \frac{0.03}{0.3} = 0.1$$

(d) $P(A \cup B)$

Solution:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.1 - 0.03 = 0.37$$

(e) Are A and B independent? Justify your answer.

Solution: Yes because $P(A) = P(A|B)$.

Question 6 continues on the next page.

[Marks]

(f) $P(A \cap \overline{B})$

Solution: Since A and B are independent,

$$P(A \cap \overline{B}) = P(A)P(\overline{B}) = 0.3 \cdot (1 - 0.1) = 0.27$$

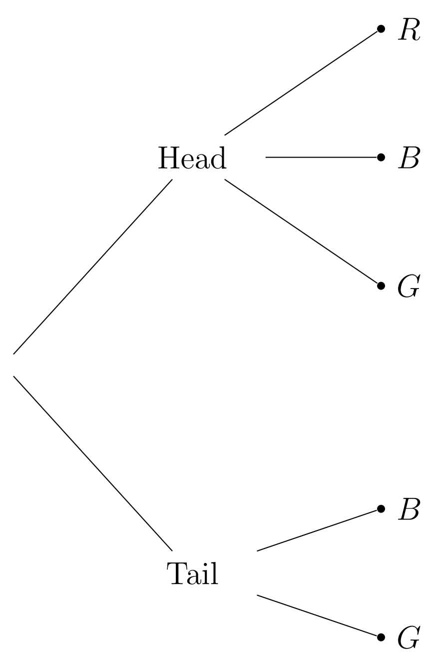
(3) 7. The odds *in favour* of event A are $1:7$. The odds *against* event B are $3:9$. If events A and B are mutually exclusive, find the probability $P(A \cup B)$.

Solution: Since A and B are mutually exclusive,

$$P(A \cup B) = P(A) + P(B) = \frac{1}{8} + \frac{9}{12} = \frac{7}{8} = 0.875$$

(6) 8. An experiment involves an unfair coin and two boxes A and B. The coin lands on heads 30% of the time. Box A has two black marbles, three red marbles and five green marbles. Box B has three black marbles and four green marbles. The coin is flipped once. If the coin lands on heads, one marble from box A is selected at random. If the coin lands on tails, one marble from box B is selected at random.

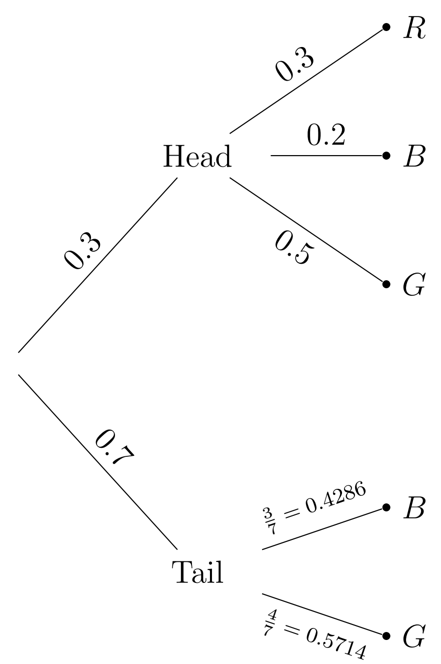
(a) In the following tree diagram that represents this situation(where R, G and B mean red, green and black marbles respectively), label the branches to find the probability of each outcome.



Solution:

Question 8 continues on the next page.

[Marks]



(b) Find the probability that the marble selected is black.

Solution:

$$0.3 \cdot 0.2 + 0.7 \cdot \frac{3}{7} = 0.36$$

(c) Find the conditional probability that the coin lands on heads provided a green marble is selected.

Solution:

$$P(\text{head}|\text{green}) = \frac{P(\text{head} \cap \text{green})}{P(\text{green})} = \frac{0.3 \cdot 0.5}{0.3 \cdot 0.5 + 0.7 \cdot \frac{4}{7}} = \frac{3}{11} = 0.2727$$

(3) 9. We have a set of words

$$\{\text{a, an, math, bin, leader, slip, lord, world, college, level, slow}\}.$$

If four words are to be selected from this set at random without replacement, find the probability that exactly one word contains the letter " i " and exactly two words contain the letter " a ".

Solution:

$$\frac{C_2^1 \cdot C_4^2 \cdot C_5^1}{C_{11}^4} = \frac{60}{330} = \frac{2}{11} = 0.1818$$

(4) 10. Suppose every time you meet a new person, the probability that person smokes is 0.2.

Question 10 continues on the next page.

[Marks]

(a) What is the probability that exactly two out of six persons you meet smoke?

Solution:

$$\binom{6}{2}(0.2)^2(0.8)^4 = 0.24576$$

Or one can use the table in the formula sheets.

(b) What is the probability that at most two out of six persons you meet smoke?

Solution:

$$\binom{6}{0}(0.2)^0(0.8)^6 + \binom{6}{1}(0.2)^1(0.8)^5 + \binom{6}{2}(0.2)^2(0.8)^4 = 0.90112$$

(3) 11. Find the values of a and b so that the following table determines a probability distribution with mean $\mu = 11$. (Hint: you need to solve a linear system.)

x	9	12
$P(x)$	$\frac{a}{3}$	$\frac{b}{3}$

Solution: The sum of all probabilities is 1:

$$\frac{a}{3} + \frac{b}{3} = 1$$

On the other hand, the mean

$$\mu = 9 \cdot \frac{a}{3} + 12 \cdot \frac{b}{3},$$

i.e.

$$3a + 4b = 11$$

Therefore we have a linear system

$$\begin{cases} \frac{a}{3} + \frac{b}{3} = 1 \\ 3a + 4b = 11 \end{cases}$$

After solving it, one obtains

$$a = 1, \quad b = 2$$

(3) 12. The number of misspellings in an article follows the Poisson distribution. Suppose on average, there are 3 misspellings in every 2000 words, find the probability that one finds 5 misspellings in 2000 words?

Solution: On average, there are 3 misspellings for a given period (here 2000 words), i.e. $\lambda = 3$. Therefore

$$P(x = 5) = \frac{3^5 e^{-3}}{5!} = 0.100$$

[Marks]

(3) 13. It is known that 2% of a certain population have Type AB blood. Suppose 100 people from this population are randomly selected. The number of people x among the selected people that have Type AB blood follows the Binomial distribution. Use the Poison distribution to estimate $P(x = 3)$.

Solution: In order to use the Poison distribution to estimate the binomial distribution, One uses

$$\lambda = np = 100 \cdot 0.02 = 2$$

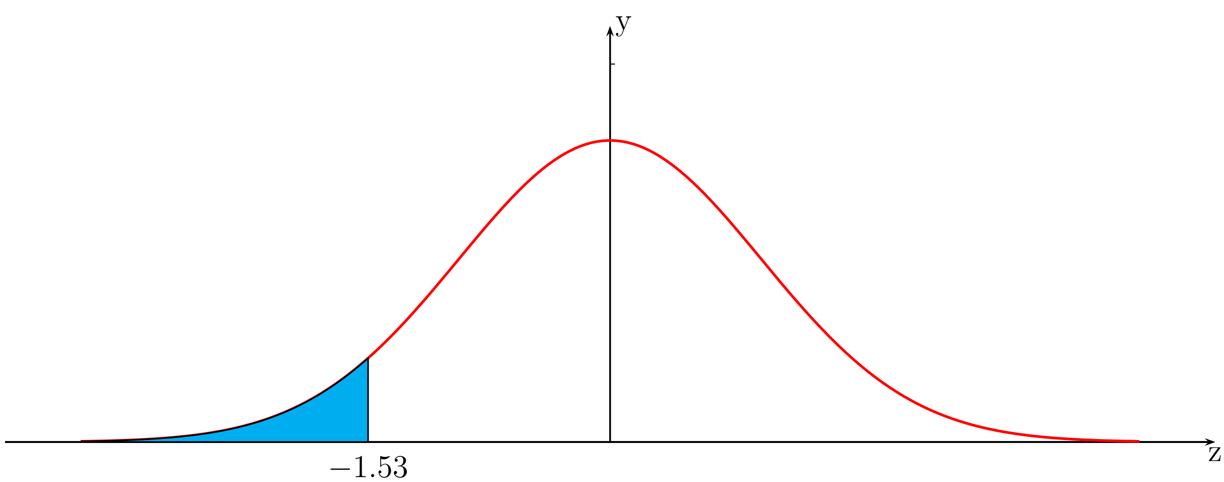
Therefore

$$P(x = 3)(\text{binomial}) \approx P(x = 3)(\text{Poisson}) = \frac{2^3 e^{-2}}{3!} = 0.1804$$

(4) 14. Draw a sketch and evaluate the following probabilities where z is drawn from a standard normal distribution.

(a) $P(z < -1.53)$

Solution:



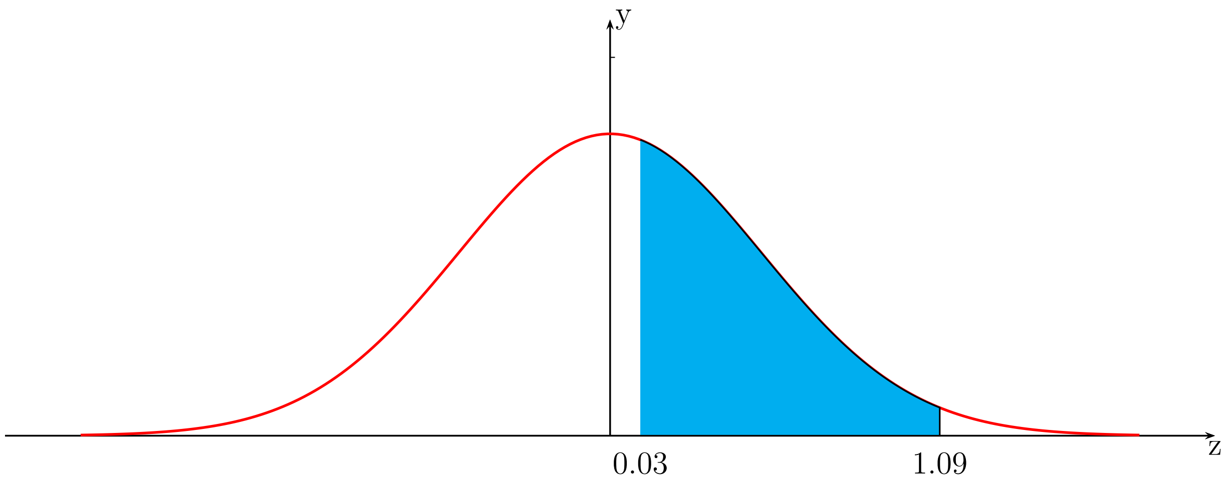
Therefore

$$P(z < -1.53) = 1 - 0.4370 = 0.0630$$

(b) $P(0.03 < z < 1.09)$

Solution:

[Marks]



Therefore

$$P(0.03 < z < 1.09) = 0.3621 - 0.0120 = 0.3501$$

(2) 15. Find a such that $P(-a < z < a) = 0.758$.

Solution: By symmetry, $P(0 < z < a) = 0.379$, hence $a = 1.17$.

Starting from question 16 , your answers should keep three decimal places.

(2) 16. Suppose the variable x is normally distributed with $\mu = 17$ and $\sigma = 0.5$. Evaluate the probability that $P(x > 16)$.

Solution:

$$P(x > 16) = P\left(z > \frac{16-\mu}{\sigma}\right) = P\left(z > \frac{16-17}{0.5}\right) = P(z > -2) = 0.5 + 0.4772 = 0.9772$$

(2) 17. Suppose the variable x is normally distributed with mean $\mu = 60$. If $P(x < 65) = 0.6064$, find the standard deviation σ.

Solution:

$$P(x < 65) = P\left(z < \frac{65-60}{\sigma}\right) = 0.6064$$

Therefore

$$P\left(0 < z < \frac{65-60}{\sigma}\right) = 0.6064 - 0.5 = 0.1064$$

[Marks]

and hence

$$\frac{65-60}{\sigma} = 0.27$$

so

$$\frac{5}{\sigma} = 0.27; \quad \sigma = \frac{5}{0.27} = 18.5185$$

(4) 18. Two out of five adult smokers acquired the habit by age fourteen. Four hundred smokers are randomly selected. Use a normal distribution to estimate the probability that 170 or fewer acquired the habit by age fourteen.

Solution: Here $\mu = np = 400 \cdot 0.4 = 160$ and $\sigma = \sqrt{npq} = \sqrt{400 \cdot 0.4 \cdot 0.6} = 9.7980$. Therefore

$$\begin{aligned} P(0 \le x \le 170)(\text{binomial distribution}) &= P(-0.5 < x < 170.5)(\text{normal distribution}) \\ &= P\left(\frac{-0.5-160}{9.7980} < z < \frac{170.5-160}{9.7980}\right) = P(-16.3809 < z < 1.07) \\ &= 0.5 + 0.3577 = 0.8577 \end{aligned}$$

(4) 19. A researcher wants to estimate, by using a 98% confidence interval, the average time students spend on social media each week. At least how many students should he consider in order to estimate this average within 2 hours? Assume the population is normally distributed with $\sigma = 6$ hours per week.

Solution: Here $E = 2$ hours and $\alpha = 0.02$. Therefore the sample size for 98% confidence is

$$\left(z(\alpha/2) \cdot \frac{\sigma}{E}\right) = \left(z(0.01) \cdot \frac{6}{2}\right)^2 = (2.33 \cdot 3)^2 = 48.86 \approx 49$$

(8) 20. The age (in months) at which a child learns how to count to ten has a normal distribution with a mean of 24 months and a standard deviation of 6 months.

(a) If we consider one child taken at random, what is the probability that he will learn counting to ten before 19 months old?

Solution:

$$P(0 < x < 19) = P(-4 < z < -0.83) = 0.5 - 0.2967 = 0.2033$$

(b) Now suppose we consider a group of 15 children. What is the probability that the average learning age to count to ten is between 19 months and 29 months?

[Marks]

Solution: This is SDSM. So $\mu_{\overline{x}} = 24$ and $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{15}} = 1.5492.$

$$P(19 < \overline{x} < 29) = P(-3.23 < z < 3.23) = 0.4994 \cdot 2 = 0.9988$$

(4) 21. A random sample of 78 students were interviewed and 58 said they would vote for John Doe as student body president. Let p represent the proportion of all students at this college who will vote for John. Find the 90% confidence interval for p.

Solution: Here $p' = \frac{58}{78} = 0.7436$ and hence $q' = 1 - p' = 0.2564$. Also $\alpha = 0.1$. Therefore

$$z(\alpha/2)\sqrt{\frac{p'q'}{n}} = 1.65 \cdot 0.04944 = 0.08158$$

So the confidence interval is

$$(0.7436 - 0.08158, 0.7436 + 0.08158) = (0.662, 0.825)$$

(4) 22. A random sample of 12 evenings at the Shannon household showed the family watched an average of 7 shows each evening. The sample standard deviation was 0.4. Find a 90% confidence interval for the population mean number of shows watched each night. Assume the number of shows watched follows a normal distribution.

Solution: We use t-distribution, where $df = 12 - 1 = 11$, $\alpha = 0.1$ and $s = 0.$. Therefore $t(df, \alpha/2) = t(11, 0.05) = 1.8$ and $\frac{s}{\sqrt{n}} = \frac{0.4}{\sqrt{12}} = 0.1155$. So the confidence interval is

$$(7 - 1.8 \cdot 0.1155, 7 + 1.8 \cdot 0.1155) = (6.792, 7.208)$$

This is the end of the examination.