[Marks]

(8) 1. In a study of the effect of manure on a certain species of plants, 200 such plants are randomly selected and their heights are measured. From this group, an average height of 245cm is computed. Match the terms in column I with the statistical terms in column II.

Column I	**Column II**
_______ All plants of this species	(a) Data (one)
_______ The height of one plant of this species	(b) Data (set)
_______ The process used to select the 200 plants and determine their average height	(c) Experiment
_______ The 200 measured heights	(d) Parameter
_______ The computed 245cm	(e) Population
_______ The average height of all plants of this species	(f) Sample
_______ 250cm measured for one plant	(g) Statistic
_______ The 200 plants	(h) Variable

Solution:

(e) All plants of this species	(a) Data (one)
(h) The height of one plant of this species	(b) Data (set)
(c) The process used to select the 200 plants and determine their average height	(c) Experiment
(b) The 200 measured heights	(d) Parameter
(g) The computed 245cm	(e) Population
(d) The average height of all plants of this species	(f) Sample
(a) 250cm measured for one plant	(g) Statistic
(f) The 200 plants	(h) Variable

(8) 2. Consider the following ordered list of data: $1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5$

Find the **mean**, **median**, **mode**, **midrange**, **range**, and **standard deviation**.

[Marks]

Solution:

$$\sum x = 35,\ n = 15,\ \text{ so } \overline{x} = \frac{\sum x}{n} = \frac{35}{15} = \frac{7}{3} = 2.333$$
$\tilde{x} = 2$

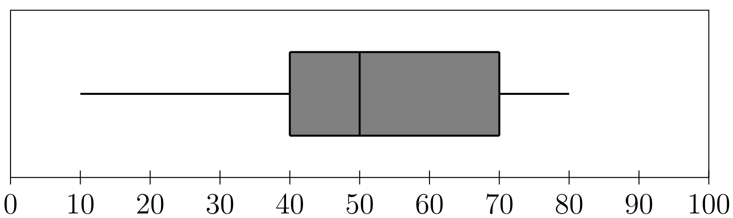
mode $= 1$

midrange $= \dfrac{1+5}{2} = 3$

range $= 5 - 1 = 4$

$$SS(x) = \sum x^2 - \frac{(\sum x)^2}{n} = 105 - \frac{35^2}{15} = \frac{70}{3},\ s^2 = \frac{SS(x)}{n-1} = \frac{\frac{70}{3}}{14} = \frac{5}{3},\ s = 1.291$$

(6) 3. The distribution of a sample variable x is summarized in the following box-and-whisker display:



(a) State the median.

Solution:

$$\tilde{x} = 50$$

(b) State the midquartile.

Solution:

$$\frac{Q_1 + Q_3}{2} = \frac{40 + 70}{2} = 55$$

(c) State the midrange.

Solution:

$$\frac{H + L}{2} = \frac{80 + 10}{2} = 45$$

(d) State the range.

Solution:

$$H - L = 80 - 10 = 70$$

(e) State the interquartile range.

Question 3 continues on the next page.

[Marks]

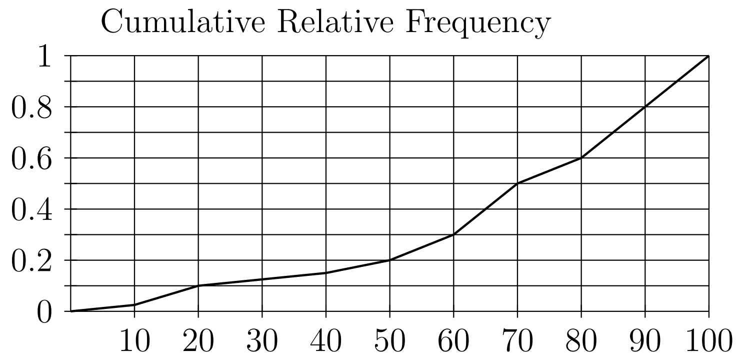
Solution:

$$Q_3 - Q_1 = 70 - 40 = 30$$

(f) If a data point x is selected at random from the sample, what is the probability that $x \geq 40$?

Solution: Since $Q_1 = 40$, from the definition of quartiles, the probability that $x \geq 40$ is 75%.

(3) 4. The distribution of a sample variable x is summarized in the following ogive:



(a) State the median of this data set.

Solution:

70

(b) State P_{30}, the 30th percentile.

Solution:

60

(c) If a data point x from this sample were selected at random, what is the probability that $x \geq 80$?

Solution:

$$1 - 60\% = 40\%$$

(3) 5. A continuous variable x is distributed with mean $\mu = 13$ and standard deviation $\sigma = 2$. According to Chebyshev's Theorem, what is the minimum proportion of the population for which $10 \leq x \leq 16$?

Solution: The z-score is

$$\frac{10-13}{2} \leq z \leq \frac{16-13}{2}, \quad -1.5 \leq z \leq 1.5$$

So from Chebyshev's theorem, the minimum proportion of the population is

$$1 - \frac{1}{1.5^2} = 55.55\%$$

[Marks]

(7) 6. Events A and B are independent. $P(A) = 0.4$ and $P(A \cap B) = 0.2$. Find the following:

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

Solution:

$$P(\overline{A}) = 1 - P(A) = 1 - 0.4 = 0.6$$

(b) $P(A|B)$

Solution: Since A and B are independent,

$$P(A|B) = P(A) = 0.4$$

(c) $P(B)$

Solution: Since A and B are independent,

$$P(A \cap B) = P(A)P(B),$$

therefore

$$P(B) = \frac{P(A \cap B)}{P(A)} = \frac{0.2}{0.4} = 0.5$$

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

Solution:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - 0.2 = 0.7$$

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

Solution:

$$P(A \cap \overline{B}) = P(A) - P(A \cap B) = 0.4 - 0.2 = 0.2$$

(4) 7. The odds *in favour* of event A are $2 : 7$. The odds *against* event B are $4 : 9$. If events A and B are mutually exclusive find the probability $P(A \cup B)$ of either A or B occurring.

Solution: Since A and B are mutually exclusive,

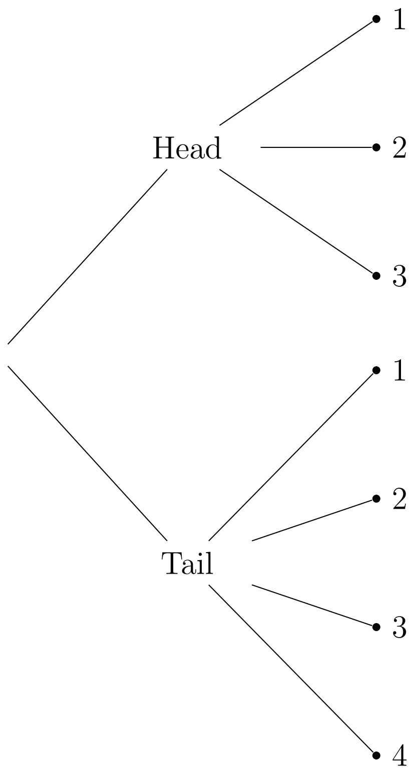
$$P(A \cup B) = P(A) + P(B) = \frac{2}{9} + \frac{9}{13} = 0.915$$

(6) 8. A fair coin is flipped. If the coin lands on heads then one of the numbers $\{1, 2, 3\}$ is selected at random. If the coin lands on tails, then one of the numbers $\{1, 2, 3, 4\}$ is selected at random.

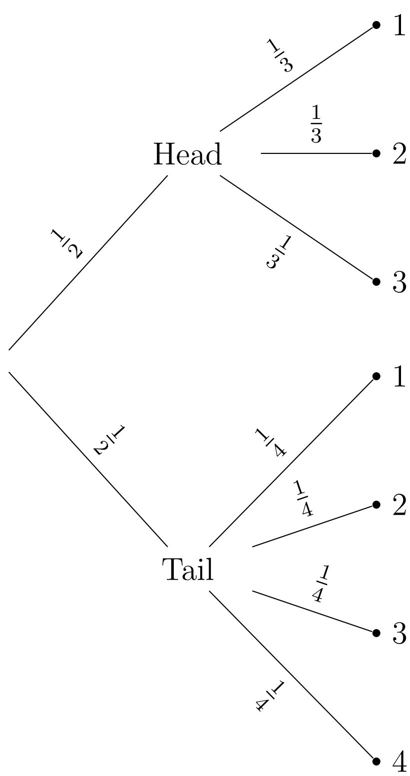
Question 8 continues on the next page.

[Marks]

(a) In the following tree diagram that represents this situation, 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 number selected is even.

Solution:

$$\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}\cdot\frac{1}{4}=\frac{5}{12}=0.417$$

(c) Find the conditional probability that the coin landed on heads, given that the number selected was even.

Solution:

$$P(\text{head}|\text{even number})=\frac{P(\text{head}\cap\text{even number})}{P(\text{even number})}=\frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{5}{12}}=\frac{2}{5}=0.4$$

(3) 9. Three characters from the set $\{a, b, c, d, e, f\}$ are to be selected at random without replacement. Find the probability that at least one vowel is chosen.

Solution:

$$\frac{C_2^1\cdot C_4^2+C_2^2\cdot C_4^1}{C_6^3}=\frac{2\cdot 6+1\cdot 4}{20}=0.8$$

(3) 10. Each time you go to a particular restaurant, the probability that you are attended by your favourite server is 1/3. What is the probability that you favourite server attends you on exactly three out of your next five visits to the restaurant?

Solution:

$$\binom{5}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2=0.165$$

(3) 11. Find the value of y so that the following table determines a probability distribution.

x	1	2
$P(x)$	$\frac{y}{2}$	$\frac{y}{3}$

Solution:

$$\frac{y}{2}+\frac{y}{3}=1$$

hence $y=\frac{6}{5}$.

(3) 12. Find the value of y so that the following probability distribution has mean $\mu = 11$.

[Marks]

x	y	$y+6$
$P(x)$	$\frac{1}{3}$	$\frac{2}{3}$

Solution:

$$\frac{1}{3}y + \frac{2}{3}(y+6) = 11$$
$$y + 4 = 11$$
$$y = 7$$

(3) 13. A unfair coin lands on heads 20% of the time. Find the number of flips for which the standard deviation of the number of heads is 20.

Solution: We know that $p = 0.2$, $q = 0.8$ and $\sigma = 20$. Since $\sigma = \sqrt{npq}$,

$$20 = \sqrt{n \cdot 0.2 \cdot 0.8}, \;\; n = \frac{20^2}{0.16} = 2500.$$

(5) 14. The number of defective products on an assembly line is given by a Poisson distribution. If, on average, one product turns out to be defective each day, find the probability that on a given week exactly five products will be defective?

Solution: On average, there are seven products that are defective each week, i.e. $\lambda = 7$. Therefore

$$P(x=5) = \frac{7^5 e^{-7}}{5!} = 0.128$$

(3) 15. A college chooses seven people to form a committee from nine science students and seven art students. In how many ways can this committee be formed if four members are chosen from science students and three members are chosen from art students?

Solution:

$$\binom{9}{4} \cdot \binom{7}{3} = 126 \cdot 35 = 4410$$

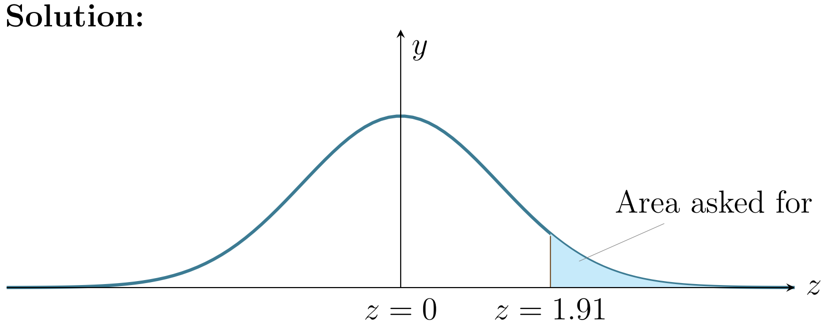
(3) 16. Draw a picture, shade the area which is represented, and evaluate:

(a) $P(z > 1.91)$

Question 16 continues on the next page.

[Marks]

Solution:

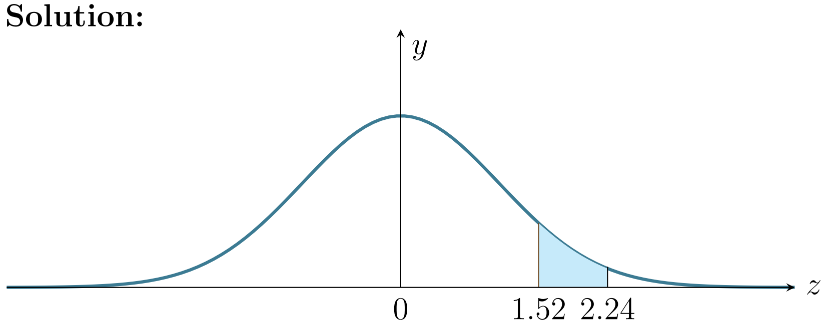


Hence

$$P(z > 1.91) = 0.5 - P(0 < z < 1.91) = 0.5 - 0.4719 = 0.0281$$

(b) $P(1.52 < z < 2.24)$

Solution:

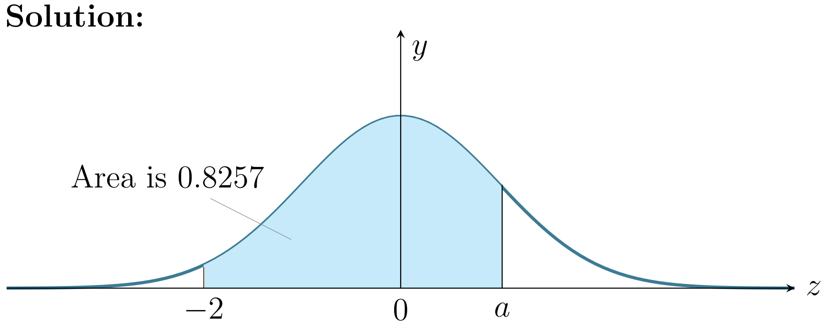


Therefore

$$P(1.52 < z < 2.24) = P(0 < z < 2.24) - P(0 < z < 1.52) = 0.4875 - 0.4357 = 0.0518$$

(c) Find a such that $P(-2 < z < a) = 0.8257$.

Solution:



Clearly $a > 0$. Since $P(-2 < z < a) = 0.8257$ and $P(-2 < z < P(0 < z < 2) = 0.4772$,

$$P(0 < z < a) = P(-2 < z < a) - P(-2 < z < 0) = 0.8257 - 0.4772.3485$$

So $a = 1.03$

(4) 17. (a) Suppose a random variable x follows a normal distribution with $\mu = 20$ and $\sigma = 11$. What is the probability that $x > 32.21$?

Question 17 continues on the next page.

[Marks]

Solution:

$$
\begin{aligned}
P(x > 32.21) &= P\left(z > \frac{32.21 - 20}{11}\right) \\
&= P(z > 1.11) \\
&= 0.5 - P(0 < z < 1.11) \\
&= 0.5 - 3665 \\
&= 0.1335
\end{aligned}
$$

(b) Suppose a random variable x follows a normal distribution with $\sigma = 20$ and $P(x > 30) = 0.0735$, find μ.

Solution: One know $P\left(z > \frac{30-\mu}{\sigma}\right) = 0.0735$, hence

$$P\left(0 < z < \frac{30 - \mu}{\sigma}\right) = 0.5 - 0.0735 = 0.4265$$

We also know $P(< z < 1.45) = 0.4265$, so $\frac{30-\mu}{\sigma} = 1.45$, hence $30 - \mu = 29$, therefore $\mu = 1$.

Starting from question 18, your answers should keep three decimal places.

(5) 18. According to a study published in March 2016, "overall, 17.4% of people in the United States diagnosed with lung cancer survive five years after the diagnosis." In a random sample of 50 patients who are diagnosed with lung cancer, what is the probability that 30% or more will survive five years after the diagnosis?

Solution: This is a binomial distribution. We define the success to be surviving five years after the diagnosis. Then $p = 17.4\% = 0.174$ and $q = 0.826$. Then $\mu = np = 50 \cdot 0.174 = 8.7$ and $\sigma = \sqrt{npq} = \sqrt{50 \cdot 0.174 \cdot 0.826} = 2.681$. We meed to find the probability that at least $30\% \cdot 50 = 15$ patients in the random sample of 50 patients will survive five years after the diagnosis. Since the number of trials (the size of the sample) is 50, we use normal distribution to estimate:

$$
\begin{aligned}
P(x \geq 15)(\text{binomail distribution } B(50, 0.174)) &\approx P(x > 14.5)(\text{normal distribution } N(8.7, 2.681)) \\
&= p\left(z > \frac{14.5 - 8.7}{2.681}\right) \\
&= p(z > 2.163) \\
&= 0.5 - 0.4846 \\
&= 0.0154
\end{aligned}
$$

(5) 19. Suppose that the average height of female students in Quebec CEGEPs is 165cm and height follows normal distribution with standard deviation 15cm. Find the probability that the mean of a random sample of 20 female students from Quebec CEGEPs is between 160cm and 170cm.

[Marks]

Solution: Let x be the random variable of the height of a female student in Quebec CEGEPs. This is SDSM. we use $\sigma_{\overline{x}} = \sigma/\sqrt{n}$ and $\mu_{\overline{x}} = \mu$. So

$$\begin{aligned} P(160 < x < 170) &= P\left(\frac{160-165}{15/\sqrt{20}} < z\frac{170-165}{15/\sqrt{20}}\right) \\ &= P(-1.49 < z < 1.49) \\ &= 0.8638 \end{aligned}$$

(5) 20. It is estimated that in a certain area the proportion of families whose income is above \$150000 is 10%. We want to know if such estimation is correct. Determine the sample size needed if we want our estimate within ±3% with 95% confidence. Assume family incomes are normally distributed.

Solution: Here $E = 0.03$, $p' = 0.1$ and $q' = 0.9$. Also $\alpha = 1-95\% = 0.05$. Since $E = z(\alpha/2)\sqrt{\frac{p'q'}{n}}$, one has

$$0.03 = z(0.025)\sqrt{\frac{0.1 \cdot 0.9}{n}}$$
$$0.03 = 1.96\sqrt{\frac{0.1 \cdot 0.9}{n}}$$

Hence

$$n = \left(\frac{1.96}{0.03}\right)^2 \cdot 0.09 = 384.16 \approx 385(\text{rounded up}).$$

(5) 21. An agriculture scientist measures the diameters of a random sample of 15 seedless watermelons of a new species. The average diameter of this sample is 36.6cm and the standard deviation of this sample is 3.2cm.

(a) Give the point estimate of the mean diameter μ.

Solution:

36.6 cm

(b) Estimate 99% confidence interval for the mean diameter μ of watermelons of this species. Assume diameters are normally distributed.

Solution: This is the inference of μ with σ unknown. Hence one needs t-distribution. One has $\alpha = 1 - 99\% = 0.01$.

$$E = t(14, \alpha/2)\left(\frac{s}{\sqrt{n}}\right) = t(14, 0.005)\left(\frac{3.2}{\sqrt{15}}\right) = 2.98 \cdot \frac{3.2}{\sqrt{15}} = 2.462$$

[Marks]

Therefore, the confidence interval is

$$(36.6 - 2.462, 36.6 + 2.462) = (34.138, 39.062)$$

(5) 22. From historical data, we know the standard deviation of the height of a certain product is 8cm. A random sample of 30 such products is selected and it is found that the average height of this sample is 198 cm. Find the 98% confidence interval for μ, the mean height of the product. Assume the height is normally distributed.

Solution: This is the inference of μ with σ known. One has $\alpha = 1 - 98\% = 0.02$. Therefore the maximum error of estimate

$$E = z(0.01)\left(\frac{\sigma}{\sqrt{n}}\right) = 2.33 \cdot \frac{8}{\sqrt{30}} = 3.403$$

Therefore the required confidence interval is

$$(198 - 3.403, 198 + 3.403) = (194.597, 201.403)$$

This is the end of the examination.