INSTITUTE OF ACTUARIAL
& QUANTITATIVE STUDIES

Subject: Probability and Statistics - I

Chapter: Unit 1

Category: Practice Questions

1. CT3 April 2006 Q1

The atom and leaf plot below gives the surrender values (to the nearest €1,000) of 40 endowment policies issued in France and recently purchased by a dealer in such policies in Paris. The stem unit is €10,000 and the leaf unit is €1,000.

5	3
5	6
6	02
6	5779
7	122344
7	556677899
8	1123444
8	567778
9	024
9	6

Determine the median surrender value for this batch of policies.

Ans: median = €77,500.

2. CT3 September 2006 Q1

A bag contains 8 black and 6 white balls. Two balls are drawn out at random, one after the other and without replacement.

Calculate the probabilities that:

(i) The second ball drawn out is black.

(ii) The first ball drawn out was white, given that the second ball drawn out is black.

Ans:

(i) $8/14 = 0.571$

(ii) $6/13 = 0.462$

3. CT3 September 2006 Q2

Let A. and B denote independent events.

Show that A and B', the complement of event B, are also independent events.

4. CT3 September 2006 Q3

Consider 12 independent insurance policies, numbered 1, 2, 3,... , 12, for each of which a maximum of 1 claim can occur. For each policy, the probability of a claim occurring is 0.1.

Find the probability that no claims arise on the group of policies numbered 1, 2, 3, 4, 5 and 6, and exactly 1 claim arises in total on the group of policies numbered 7, 8, 9, 10, 11, and 12.

Ans: 0.188.

5. CT3 September 2007 Q1

Data collected on claim amounts (£) for two postcode regions give the following values for n (the number of claims), x (the mean claim amount) and s (the sample standard deviation) of the claim amounts.

	Region I	Region 2
n	25	18
x	120.2	142.7
s	58.1	62.2

Calculate, to one decimal place, the mean and sample standard deviation of the claim amounts for both regions combined.

Ans: Mean = 129.6
SD = 60.2

6. CT3 April 2008 Q1

The number of claims which arose during the calendar year 2005 on each of a group of 80 private motor policies was recorded and resulted in the following frequency distribution:

Number of claims x	0	1	2	3	4
Number of policies f	64	12	3	0	1

For these data:

$$, \Sigma fx = 22, \Sigma fx^2 = 40$$

Calculate the sample mean and standard deviation of the number of claims per policy.

Ans: Mean = 0.275
SD = 0.656

7. CT3 April 2008 Q2

Data on a sample of 29 claim amounts give a sample mean of £461.5 and a sample standard deviation of £618.8.
One claim amount of £3,657.50 is Identified as an outlier and after investigation Is found to be in error.
Calculate the revised sample mean and standard deviation if this erroneous amount is removed.

Ans: Mean = £347.4
SD = £72.6

8. CT3 April 2008 Q3

The following sample contains claim amounts (£) on a particular class of insurance policies:

1,717	1,595,	1,764	1,464	1,854	1.560-	1,698.
1,614	1,524.	4,320	1,626	1,440	1.602	

(i) Determine the mean and the median of the claim amounts.

(ii) State, with reasons, which of the two measures considered above you would prefer to use to estimate the central point of the claim amounts.

Ans :

(i) Mean = 1829.08
Median= 1614

(ii) The median should be preferred, as it is not sensitive to the extreme observed claim of £4320.

9. CT3 April 2008 Q4

Consider two events A and B, such that P(A) = 0.3 and P(A∩B) = 0.1.
Find the minimum and maximum possible values of the conditional probability, P(A\B).

Ans: Maximum = 0.125
Minimum = 1

10. CT3 April 2008 Q5

An insurance company covers claims from four different non-life portfolios, denoted as G1, G2, G3 and G4
The number of policies included in each portfolio is given below:

Portfolio	G1	G2	G3	G4
Number of Policies	400	700	1300	600

It is estimated that the percentages of policies that will result in a claim in the following year in each of the portfolios are 8%, 5%, 2% and 4%, respectively.
Suppose a policy is chosen at random from the group of 30,000 policies comprising the four portfolios after one year and it is found that a claim did arise on this policy during the year.

Calculate the probability that the selected policy comes from Portfolio G3.

Ans: $P(B_3 | C) = 0.222$

11. CT3 September 2008 Q1

The mean of a sample of 30 claim amounts arising from a certain kind of insurance policy is
£5,200. Six of these claim amounts have mean £8,000 while ten others have mean £3,100. Calculate the mean of the remaining claim amounts in this sample.

Ans: 5500

12. CT3 September 2008 Q2

Five years ago a financial institution issued a specialized type of Investment bond and Investors had the option to cash In after 1, 2, 3, 4 or 5 years. The following table gives a frequency distribution showing the numbers of those Investors who cashed in at each stage.

Duration (length of time held before being cashed in)				
1 year	2 years	3 years	4 years	5 years
130	151	97	64	98

Calculate the sample mean and standard deviation of the duration of these bonds before being cashed in.

Ans:
Mean = 2.72 years
SD = 1.42 years

13. CT3 April 2009 Q1

A random sample of 12 claim amounts (in units of £1,000) on a general insurance portfolio is given by:

14.9 12.4 19.4 3.1 17.6 21.5 15.3 20.1 18.8 11.4 46.2 16.2

For these data:

$\Sigma x = 216.9$, $\Sigma x^2 = 5{,}052.13$, sample mean $\bar{x} = £18{,}075$
sample median = £16,900, sample standard deviation $s = £10{,}143$.

Calculate the sample mean, median, and standard deviation of the sample (of size 10) which remains after we remove the claim amounts 3.1 and 46.2 from the original sample (you should show intermediate working and/or give justifications for your answers).

Ans :
Mean = 16.76 i.e. £16,760
Median = £16,900
SD = 3.31837 i.e. £3,318.37

14. CT3 April 2009 Q2

Consider three events A, B, and C for which A and C are independent, and B and C are mutually exclusive. You are given the probabilities P(A) = 0.3, P(B) = 0.5, P(C) = 0.2 and P(A∩B) = 0.1.
Find the probability that none of A, B, or C occurs.

Ans: P(none occur) = 0.16

15. CT3 April 2009 Q3

The random variable X has probability density function

$$f(x) = k(1 - x)(1 + x),\ 0 < x < 1,$$

where k is a positive constant.

(i) Show that k = 1.5
(ii) Calculate the probability P(X > 0.25)

Ans :
(ii) 0.633

16. CT3 September 2009 Question 1

In a sample of 100 households in a specific city, the following distribution of number of people per household was observed:

Number of people x	1	2	3	4	5	6	7
Number of households f_x	7	f_2	20	f_4	18	10	5

The mean number of people per household was found to be 4.0. However, the frequencies for two and four members per household ($f2$ *and* $f4$ respectively) are missing.

(i) Calculate the missing frequencies $f2$ *and* $f4$

(ii) Find the median of these data and hence comment on the symmetry of the data.

Ans:

(i) $f_2 = 6$ and $f_4 = 34$

(ii) Median is equal to the midpoint between the 50th and 51st ordered observations, i.e.
median= 4. 1. We have mean = median, suggesting that the distribution of these data is roughly symmetric.

17. CT3 September 2009 Q2

Two tickets are selected at random, one after the other and without replacement, from a group of six tickets, numbered 1, 2, 3, 4, 5, and 6.

(i) Calculate the probability that the numbers on the selected tickets add up to 8.
(ii) Calculate the probability that the numbers on the selected tickets differ by 3 or more.

Ans:
(i) 0.133
(ii) 0.4

18. CT3 April 2010 Q1

The mean height of the women in a large population is 1.671 m while the mean height of the men in the population is 1.758m. The mean height of all the members of the population is 1.712m.
Calculate the percentage of the population who are women.

Ans: 52.9%

19. CT3 April 2010 Q2

Consider a group of 10 life insurance policies, seven of which are on male lives and three of which are on female lives. Three of the 10 policies are chosen at random (one after the other, without replacement).

Find the probability that the three selected policies are all on male lives.

Ans: 0.292

20. CT3 October 2010 Q1

The marks of a sample of 25 students from a large class in a recent test have sample mean 57.2 and standard deviation 7.3. The marks are subsequently adjusted: each mark is multiplied by 1.1 and the result is then increased by 8.

Calculate the sample mean and standard deviation of the adjusted marks.

Ans: Mean = 70.92
SD = 8.03

21. CT3 October 2010 Q2

In a survey, a sample of 10 policies is selected from the records of an insurance company. The following data give, in ascending order, the time (in days) from the start date of the policy until a claim has arisen from each of the policies in the sample.

297	301	31	317	355	379	40	419	432	463

Some of the policies have not yet resulted in any claims at the time of the survey, so the times until they each give rise to a claim are said to be censored. These values are represented with a plus sign in the above data.

(i) Calculate the median of this sample.

(ii) State what you can conclude about the mean time until claims arise from the policies in this sample.

Ans :

(i) Median = 367 days

(ii) We know that the last two observations have minimum values 432 and 463. Using these two values the sample mean would be equal to 3679/10 = 367.9. So, the sample mean is at least equal to 367.9 days.

22. CT3 April 2011 Q1

The numbers of claims which have arisen in the last twelve years on 60 motor policies (continuously in force over this period) are shown (sorted) below:

0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1
1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3
3 3 3 3 3 3 3 4 4 4 4 4 5 5 5 5 6 6 6 7

Derive:

(i) The sample median, mode and mean of the number of claims.

(ii) The sample interquartile range of the number of claims.

(iii) The sample standard deviation of the number of claims.

Ans:

(i) Median = 2

Mode = 1

Mean = 2.18

(ii) IQR = 2

(iii) SD = 1.8

23. CT3 April 2011 Q7

An Insurance company distinguishes between three types of fraudulent claims:
Type 1: legitimate claims that are slightly exaggerated
Type 2: legitimate claims that are strongly exaggerated
Type 3: false claims

Every fraudulent claim is characterized as exactly one of the three types. Assume that the probability of a newly submitted claim being a fraudulent claim of type 1 is 0.1. For type 2 this probability is 0.02, and for type 3 it is 0.003.

(i) Calculate the probability that a newly submitted claim is not fraudulent.

The insurer uses a statistical software package to identify suspicious claims. If a claim is fraudulent of type 1, it is identified as suspicious by the software with probability 0.5. For a type 2 claim [his probability is 0.7, and for type 3 it is 0.9.

Of all newly submitted claims, 20% are identified by the software as suspicious.

(ii) Calculate the probability that a claim that has been identified by the software as suspicious is:

(a) a fraudulent claim of type 1,

(b) a fraudulent claim of any type.

(iii) Calculate the probability that a claim which has NOT been identified as suspicious by the software is in fact fraudulent

Ans :

(i) 0.877

(ii) a) = 0.25, b)= 0.3335

(iii) 0.0704

24. CT3 October 2011 Q1
The first 20 claims that were paid out under a group of policies were for the following amounts (in units of £1,000):

3.2	2.1	6.3	4.0	3.8	4.4	6.5	7.8	2.8	5.2
7.0	8.1	4.4	5.8	1.7	2.8	5.0	3.2	3.7	4.4

For these data $\sum x = 92.2$.
(i) Calculate the mean of these 20 claim amounts.

The next 80 claims paid out had a mean amount of £5,025.

(ii) Calculate the mean amount for the first 100 claims.

Ans:

(i) 4.61 or £4,610.

(ii) £4,942

25. CT3 April 2012 Q1
The following 24 observations give the length of time (in hours, ordered) for which a specific fully charged laptop computer will operate on battery before requiring recharging.

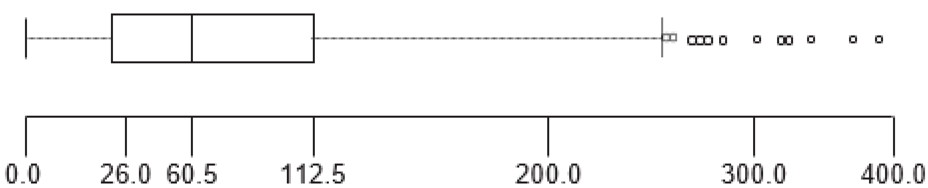
1.2	1.4	1.5	1.6	1.7	1.7	1.8	1.8	1.9	1.9	2.0	2.0
2.1	2.1	2.1	2.2	2.3	2.4	2.4	2.5	3.1	3.6	3.7	4.5

The owner of this computer is about to watch a film on his fully charged laptop,

Calculate from these data the longest showing time for a film that he can watch, so that the probability that the battery's lifetime will be sufficient for watching the entire film is 0.75.

Ans: This is asking for the first quartile = 1.75 hours

26. CT3 April 2012 Q2
Data were collected on the time (in days) until each of 200 claims is settled by the insurer in a certain insurance portfolio. A box plot of the data is shown below.



Time lo settlement (days)
(i) Calculate the median and the interquartile range of the data using the plot.
(ii) Comment on the distribution of the data as shown in the plot.

Ans:
(i) Median = 86.5 days
(ii) The distribution is skewed to the right and a number of values appear to be outliers.

27. CT3 April 2012 Q3
Two students are selected at random without replacement from a group of 100 students, of whom 64 are male and 36 are female.

Calculate the probability that the two selected students are of different genders

Ans: 0.465

28. CT3 October 2012 Q1
Calculate the mean, the median and the mode for the data in the following frequency table.

Observation	0	1	2	3	4
Frequency	20	54	58	28	0

Ans: Mean = 1.5875
Median = 2
Mode = 2

29. CT3 October 2012 Q2
The following data are sizes of claims (ordered) for a random sample of 20 recent claims submitted to an insurance company:

174	214	264	298	335	368	381	395	402	442
487	490	564	644	686	807	1092	1328	1655	2272

(i) Calculate the Interquartile range for this sample of claim sizes.
(ii) Give a brief interpretation of the interquartile range calculated in part (i).

Ans:
(i) IQR = 395
(ii) The length of the interval containing the central half of the claim sizes is 395. 1

30. CT3 October 2012 Q3

Let X be a discrete random variable with the following probability distribution:

X	0	1	2	3
P(X=x)	0.4	0.3	0.2	0.1

Calculate the variance of Y, where Y = 2X + 10.

Ans: Variance of Y = 4

31. CT3 April 2013 Q1

The following data represent the number of claims for twenty policyholders made during a year.

0	0	0	0	0	0	0	1	1	1
1	1	1	1	1	1	2	2	2	3

Determine the sample mean, median, mode and standard deviation of these data.

Ans: Mean = 0.9
Median= 1
Mode = 1
SD = 0.8522

32. CT3 April 2013 Q3

A discrete random variable X has a cumulative distribution function (CDF) with the following values:

Observation	10	20	30	40	50
CDF	0.5	0.7	0.85	0.95	1

Calculate the probability that X takes a value:

(i) larger than 10.
(ii) less than 30.
(iii) exactly 40.
(iv) larger than 20 but less than 50.
(v) exactly 20 or exactly 40.

Ans:
(i) larger than 10. = 0.5
(ii) less than 30 = 0.7
(iii) exactly 40 = 0.1
(iv) larger than 20 but less than 50 = 0.25
(v) exactly 20 or exactly 40 = 0.3

33 CT3 October 2013 Q1

The stem and leaf plot below shows 40 'observations of an exchange rate.

1.21	9
1.22	4569
1.23	2679
1.24	3467889
1.25	011222345677778
1.26	00346688
1.27	
1.28	1

For these data, $\sum x = 50.000$.

(i) Find the mean, median and mode.

(ii) State, with reasons, which measure of those considered in part (i) you would prefer to use to estimate the central point of the observations.

Ans:

(i) Mean = 1.25
Median = 1.252
Mode = 1.257

(ii) Mean. Distribution is roughly symmetrical with no outliers so no reason to use anything else.

34 CT3 October 2013 Q7

A motor insurance company has a portfolio of 100,000 policies. It is distinguished between three groups of policy holders depending on the geographical region in which they live. The probability p of a policyholders submitting at least one claim during a year is given in the following table together with the number, n, of policyholders belonging to each group. Each policyholder belongs to exactly one group and it is assumed that they do not move from one group to another over time.

Group	A	B	C
P	0.15	0.1	0.05
n (in 1000s)	20	20	60

It is assumed that any individual policyholder submits a claim during any year independently of claims submitted by other policyholders. li is also assumed that whether a policyholder submits any claims in a year is independent of claims in previous years conditional on belonging to a particular group.

a. Show that the probability that a randomly selected policyholder will submit a claim in a particular year is 0.08.

b. Calculate the probability that a randomly selected policyholder will submit a claim in a particular year given that the policyholder is not in group C.

c. Calculate the probability for a randomly selected policyholder to belong to group A given that the policyholder submitted a claim last year.

d. Calculate the probability that a randomly selected policyholder will submit a claim in a particular year given that the policyholders submitted a claim in the previous year. It is assumed that the insurance company does not know to which group the policyholder belongs.

e. Calculate the probability that a randomly selected policyholder will submit a claim in two consecutive years.

Ans:
b. 0.05
c. 0.375
d. 0.1
e. 0.008

35 CT3 April 2014 Q2

A set of data has mean 62 and standard deviation 6.

Derive a linear transformation for these data that will result in the new data having mean 50 and standard deviation 12.

Ans:
We want to find a and b for y = a + bx such that

$$\bar{y} = a + b\bar{x} = 50 \text{ and } s_y = |b| s_x$$

These give b = 2 and a = 50 – 124 = -74 or
b = – 2 and a = 50 +124 = 174

36 CT3 April 2014 Q3

Sixty per cent of new drivers in a particular country have had additional driving education. During their first year of driving, new drivers who have not had additional driving education have a probability 0.09 of having an accident, while new drivers who have had additional driving education have a probability 0.05 of having an accident.

(a) Calculate the probability that a new driver does not have an accident during their first year of driving.

(b) Calculate the probability that a new driver has had additional driving education, given that the driver had no accidents in the first year.

Ans:

a. 0.934

b. 0.610

37 CT3 September 2014 Q1

A sample of marks from an exam has median 49 and interquartile range 19. The marks are rescaled by multiplying by 1.2 and adding 6.

Calculate the new median and interquartile range.

Ans: New Median = 64.8

IQR = 22.8

38 CT3 September 2014 Q2

Consider an insurer that offers two types of policy: home insurance and car insurance. 70% of all customers have a home insurance policy, and 80% of all customers have a car insurance policy. Every customer has at least one of the two types of policies.

Calculate the probability that a randomly selected customer:

a. does not have a car insurance policy.

b. has car insurance and home insurance.

c. has home insurance, given that he has car insurance.

d. does not have car insurance, given that he has home insurance.

Ans:

a. 0.20 = 20%

b. 0.5

c. 0.625

d. 0.2857

39 CT3 April 2015 Q1

Two groups of students sat the same exam. The marks in the first group of 64 students had an average of 52 and a standard deviation of 9, The marks in the second group of 42 students had an average of 45 and a standard deviation of 8.

Calculate the average and standard deviation of the combined data set of 106 students.

Ans:
Average = 49.23
SD = 9.243

40 CT3 April 2015 Q7
A continuous random variable X has the cumulative distribution function $F_X(x)$ given by:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{8}x^3, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases}$$

a. Determine the probability density function of X.
b. Calculate P(0. 5 < X < 1).

Let Y = $\sqrt{X}$

c. Determine the cumulative distribution function and the probability density function of Y.
d. Calculate the expected values of X and Y.

Ans:
a. $f(x) = \frac{3}{4}x^2$
b. $\frac{7}{64} = 0.1093$
c. $F_x(y^2) = \frac{1}{8}y^6$, $f_y(y) = \frac{6}{8}y^5$
d. $E(X) = \frac{3}{2}$, $E(Y) = 1.2122$

41 CT3 September 2015 Q1
A random sample of 20 claim amounts (x, in £) made on a certain type of travel insurance policy (type A) was selected and gave the following data summaries:
$\sum x = 11{,}860$
$\sum x^2 = 8{,}438{,}200$
sample mean = 593
sample standard deviation = 271.95.

It was later discovered that two of these 20 claims were made on a different incorrect type of policy, and the corresponding amounts were £770 and £510.

These claims are going to be replaced by two claims made on the correct type of policy, with corresponding amounts £1,000 and £280.

(i) Determine the sample mean and standard deviation of the revised sample.
(ii) Comment on how your answers compare with the original sample mean and standard deviation.

Ans:
(i) Mean= 593
SD = 292.95
(ii) Since the sum of the two new claims is the same as those replaced, the mean is the same. However, the SD has increased as the two new claims are further away from the mean as compared to the two claims in the first sample.

42 CT3 April 2016 Q1
A university director of studies records the number of students failing the examinations of several courses. The data are presented in the following stem-and leaf plot where the stems are with units 10 and the leaves with units 1:

0 | 224
0 | 59
1 | 03
1 | 57889
2 |
2 |
3 | 144
3 | 5

(i) Determine the range of the data.

(ii) Determine the median number of students failing the examinations of these courses.

(iii) Determine the mean number of students failing the examinations of these courses.

Ans:
(i) Range = 33
(ii) Median = 16
(iii) Mean = 16.625

43 CT3 April 2016 Q3
A random variable Y has probability density function
$f(y) = \frac{\theta}{y^{\theta+1}}$, y > 1 where θ > 0 is a parameter.

(i) Show that the probability density function of Z = ln(Y) is given by $\theta e^{-\theta z}$ and determine its range.

Ans:

(i) $f(z) = \theta e^{-\theta z}$ and

range = since y> 1 $\Rightarrow$z> 0.

44 CT3 April 2016 Q4

A manufacturing company is analysing its accident record. The accidents fall into two categories:

· Minor – dealt with by first aider. Average cost £50.
· Major – hospital visit required. Average cost £1,000.

The company has 1,000 employees, of which 180 are office staff and the rest work in the factory.
The analysis shows that 10% of employees have an accident each year and 20% of accidents are major. It is assumed that an employee has no more than one accident in a year.

(i) Determine the expected total cost of accidents in a year. On further analysis it is discovered that a member of office staff has half the probability of having an accident relative to those in the factory.

(ii) Show that the probability that a given member of office staff has an accident in a year is 0.0549.

(iii) Determine the probability that a randomly chosen employee who has had an accident is office staff

Ans:

(i) Expected total cost of accidents = £240 *1000 * 0.1 = £24,000
(ii) P(A|O) = 0.0549
(iii) P(O|A) = 0.099

45 CT3 September 2016 Q1

Consider the following sample with 20 observations x_i:

1	1	5	7	9	11	11	14	14	19
20	21	23	28	28	31	39	41	43	47

$$\sum_{i=1}^{20} x_i = 413 \; and \; \sum_{i=1}^{20} x_i^2 = 12,311$$

(i) Calculate the mean of this sample.

(ii) Calculate the standard deviation of this sample.

(iii) Calculate the median of this sample.

(iv) Calculate the interquartile range of this sample.

Ans:
(i) Mean = 20.65
(ii) SD = 14.11
(iii) Median = 19.5
(iv) IQR = 19.5

46 CT3 September 2016 Q2

A sample of data has a distribution that has a single mode and is strongly positively skewed. An analyst computes three measures of location for these data: the mean, the median and the mode.

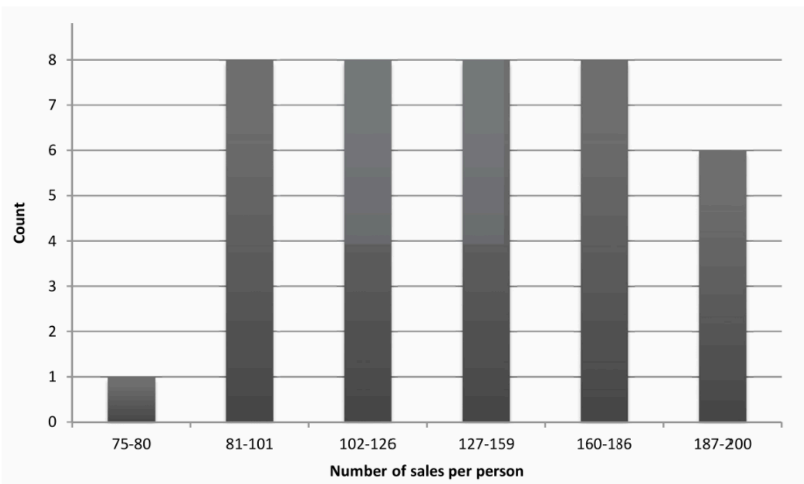
(i) State the largest of these three location measures.

(ii) Suggest, with a reason, which would be the best measure of location.

Ans:
(i) Mean
(ii) Median – mode could be at the lowest point of the distribution and the mean could be raised by a long tail.

47 CT3 April 2017 Q1

A company is collecting data on its sales. It has 39 employees and records the number of sales that each one made in a month. The data are summarized by grouping the sales employees according to the number of sales they made and presented in the following chart:



Determine the mean and standard deviation of the number of sales made by a sales employee.

Ans: Mean = 138.63
SD = 37.83

48 CT3 April 2017 Q4

An insurance company calculates car insurance premiums based on the age of the policyholder according to three age groups: Group A consists of drivers younger than 22 years old; Group B consists of drivers 22–33 years old; and, Group C consists of drivers older than 33 years.

Its portfolio consists of 10% Group A policyholders, 38% Group B policyholders and 52% Group C policyholders.

The probability of a claim in any 12-month period for a policyholder belonging to Group A, B or C is 13%,
3% and 2%, respectively.

(i) Calculate the probability that a randomly chosen policyholder from this portfolio will make a claim during a 12-month period.

One of the company's policyholders has just made a claim.

(ii) Calculate the probability that the policyholder is younger than 22 years.

Ans:
(i) 0.0348
(ii) 0.374

49 CT3 September 2017 Q1

The number of cans of fizzy drinks consumed by teenagers each day is the subject of an empirical study. The following data have been recorded.

cans per day	0	1	2	3	4	5
number of teenagers	25	30	26	20	14	10

Assume that no teenager drinks more than five cans per day.

(i) Calculate the mean, median and mode for this sample.

(ii) Comment on the symmetry of the observed data, using your answer to part (i) and without making any further calculations.

Ans:
(i) Mean = 1.984
Median = 2
Mode = 1

(ii) The mean is almost identical to the median which is an indication for symmetrical data as 50% of data are below the mean. On the other hand, the mode is different. However, the frequency of the mode, 1, and the median, 2, are almost identical. Again, this is no strong sign for asymmetry.

50 CT3 September 2017 Q4

An airline is analysing the punctuality of its scheduled flights. It measures departures and arrivals and classifies them as early, on time or late. From its records no flights depart early and 85% depart on time. Arrivals are early 10% of the time and late 20% of the time.

(i) Determine the probability that a flight arrives on time.

Further examination shows that:

- None of the flights that depart late are early arrivals.
- 10% of the flights that depart on time are late arrivals.

(ii) Determine the probability that a flight will arrive early if its departure is on time.

(iii) Show that the probability that a flight both departs on time and arrives on time is 0.665.

(iv) Determine the probability that a flight will arrive on time if it departs late.

(v) Determine the probability that if a flight arrives late it departed on time.

Ans:
(i) 0.7
(ii) 0.1176
(iii) -
(iv) 0.233
(v) 0.425

51 CT3 April 2018 Q1

A scientist collects the following data sample on the number of plants grown on newly fertilised plots of land.

Number of plants	***Frequency***
1	2
2	6
3	3
4	8
5	1

(i) Calculate the mean, median and mode of the sample.
(ii) Calculate the standard deviation of the sample.

Ans:
(i) Mean = 3
Median = 3
Mode = 4
(ii) SD = 1.170

52 CT3 April 2018 Q2
Consider the following data, and the corresponding sums derived from the data:

x_i: 10.0 6.9 11.4 12.6 10.3 12.4 9.8

$$\sum x_i = 73.4;\ \sum x_i^2 = 792.22;\ \sum x_i^3 = 8{,}750.972.$$

(i) Determine the third moment about the mean for these data.

(ii) (a) Write down the mathematical definition of the coefficient of skewness of a set of data.
(b) Determine the coefficient of skewness for the data above

Ans:
(i) -4.188

(ii) (a) $\dfrac{\sum(x_i-\bar{x})^3/n}{\left(\sum\frac{(x_i-\bar{x})^2}{n}\right)^{\frac{3}{2}}}$

(b) −0.723

53 CT3 September 2018 Q1
A data set of 20 observations has mean 45 and standard deviation 25.4. The data set is reviewed and one observation which was incorrectly recorded as 130 is now corrected to 30.

Determine the mean and standard deviation of the corrected data.

Ans: Mean = 40
SD = 15.825

54 CS1A April 2019 Q4
Alice and Bob are playing a game of dice. Two fair six-sided dice are rolled. Consider the following events:

A = 'sum of two dice equals 3'
B = 'sum of two dice equals 7'
C = 'at least one of the dice shows a 1'.

(i) Show that $P(C) = 11/36$.

(ii) Calculate $P(A|C)$.

(iii) Calculate $P(B|C)$.

(iv) Determine whether A and C are independent.

(v) Determine whether B and C are independent.

Ans:
(ii) 2/11
(iii) 2/11
(iv) They are not independent
(v) They are not independent

55 CS1A September 2020 Q2
A pair of fair six-sided dice is rolled once.

(i) Identify which **one** of the following options gives the probability that the sum of the two dice is seven:

A1	1/36
A2	1/6
A3	1/12
A4	1//3

(ii) Identify which **one** of the following options gives the probability that at least one dice shows three:

A1	25/36
A2	1/36
A3	11/36
A4	5/36

(iii) Identify which **one** of the following options gives the probability that at least one dice shows an odd number:

A1	1/4
A2	3/4
A3	1/2
A4	1/12

Ans:
(i) 1/6
(ii) 11/36
(iii) 3/4

56 CS1A September 2020 Q3
The following data are available on three television factories that produce all the televisions used in a country.

Factory	***% of total production***	***Probability of defect (Def)***
A	0.35	0.020
B	0.40	0.015
C	0.25	0.010

A television is selected at random and found to have a defect (Def).

(i) Write the expression to correctly calculate the probability that the selected television was made in factory B.
(ii) Calculate, by using your answer to part (i), the probability that the selected television was produced by Manufacturer B.

Ans:
(i) Using Bayes' theorem:

$$=\frac{P(defective|made\ in\ factory\ B)\times P(made\ in\ factory\ B)}{P(defective|factory\ A)P(factory\ A)+P(defective|factory\ B)P(factory\ B)+P(defective|factory\ C)P(factory\ C)}$$

(ii) 0.38710

57 CS1A September 2022 Q1
From national statistics, it is known that 7% of all drivers in a country are young drivers. It is also known that 18% of all drivers involved in road accidents are young drivers (less than 25 years old). Define the two events:

A: a randomly chosen driver is involved in a road accident.
Y: a randomly chosen driver is a young driver.

(i) Determine the conditional probability $P[A \mid Y]$ as a function of $P[A]$.

(ii) Comment on the result from part (i).

Ans:

(i) $\frac{0.18}{0.07}P(A)$

(ii) Therefore, the probability for young drivers to be involved in an accident is almost three times (18/7) as high as it is for drivers in general.