Probability Tutorial Answers 2021

Note that the final answers are in bold.

Answer to Question 1:

In an archery competition, the team consisting of John, David and Mike will win 1st prize if at least two of them hit the bullseye with their last arrows. From past experience, they know that the probability that John, David and Mike will hit the bullseye on their last arrow is 1/5, 1/6 and ¼ respectively.

i. Complete the table below to show all the ways in which they could win 1st prize.

	Way 1	Way 2	Way 3	Way 4
John	√	**√**	**x**	**√**
David	√	**x**	**√**	**√**
Mike	×	**√**	**√**	**√**

Where

√=Hit

×=Miss

ii. Hence or otherwise find the probability that they will win the competition.

The probabilities attached to the events in the table above are as follows where the likelihood of a miss= 1- likelihood of hit.

	Way 1	Way 2	Way 3	Way 4
John	√ (1/5)	√ (1/5)	X (4/5)	√ (1/5)
David	√ (1/6)	X (5/6)	√ (1/6)	√ (1/6)
Mike	× (3/4)	√ (1/4)	√ (1/4)	√ (1/4)

The probability of winning via Way 1: John hits and David hits and Mike misses

=(1/5)*(1/6)*(3/4)

=0.025

The probability of winning via Way 2: John hits and David misses and Mike hits

=(1/5)*(5/6)*(1/4)

=0.04166666666

The probability of winning via Way 3: John misses and David hits and Mike hits

=(4/5)*(1/6)*(1/4)

=0.0333333333

The probability of winning via Way 4: John hits and David hits and Mike hits

=(1/5)*(1/6)*(1/4)

=0.008333

The probability of winning the game by way 1,2, 3 or 4

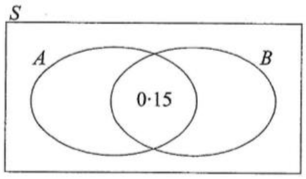
=p(winning by way 1) or p(winning by way 2) or p(winning by way 3) or p(winning by way 4)

=0.025 +0.04166666666 +0.0333333333 +0.008333

=13/120 or 0.108

Answer to Question 2:

Two events are such that P (A) = 0.2, P(A ∩ B) = 0.15 and P(A' ∩ B) = 0.6

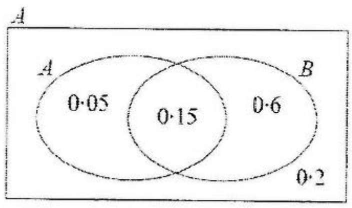


i) Find the probability that neither A nor B happens

P(A ∩ B') = 0.2-.015=0.05

So the diagram looks as follows:

(a)



The probability that A nor B happens = 1- P(A) – P(B) + P(A ∩ B)
= 1 -0.2 -.75 +.15

Or = 1- 0.05 -.15 -.6

= 0.2 or 20%

ii) Find the conditional probability P(A|B)

The probability that A happens given that B happens is

$$\frac{P(A \cap B)}{P(B)}$$

$$=\frac{0.15}{0.75}$$

=0.2 or 20%

iii) State whether A and B are independent events and justify your answer

An event is independent of another event if P(A|B)= P(A) i.e. the fact that B has happened doesn't mean that A is more or less likely to happen that would be otherwise and vice versa.

P(A|B)= P(A∩ B) /P(B) =0.15/0.75=.2 = P(A)

P(B|A)= P(A∩ B) /P(A) =0.15/0.2=.75 = P(B)

A and B are independent events as P(A|B)= P(A)=0.2

Answers to Question 3:

A certain basketball player scores 60% of the free-throw shots she attempts. During a particular game, she gets six free throws.

a) What assumption(s) can be made in order to regard this as a sequence of Bernoulli trials?
The assumptions that can be made to regard it as a sequence of Bernoulli trials are:
That the trials are independent
The probability of success is the same each time

b) Based on such assumption(s), find, correct to three decimal places, the probability that:

i. She scores on exactly four of the six shots
X is the number of scores
N is the number of trials (which is 6)
P is the probability of success (which is 60% or 0.6)
Q is the probability of failure (which is 40% or 0.4)
P(X=4) = (p)^(4)*(q)^2
P(X=4) = (6) * (0.6)^(4) * (0.4)^(2) (where the first part is a permutation *6 C 4*)
(4)
P(X=4) = (6!) * 0.1296 * 0.16
(4!)(6-4)!
P(X=4) = 720* 0.020736
24*(2)
P(X=4) = 15*0.020736
P(X=4) = 0.311
The probability that she scores exactly four of the six shots is **0.311**

ii. She scores for the second time on the fifth shot.
= (4C1) * (p)^(1) * q^(3)* (p) (where the first part is a permutation *4 C 1*)
= (4C1) * (0.6) * (0.4)^(3) *(0.6)
= 4 * 0.0374 *0.6
=0.092
The probability that she for the second time on the fifth shot is **0.092**

Question 4

In a competition Mary has a probability of 1/20 of winning, a probability of 1/10 of finishing in second place, and a probability of 1/4 of finishing in third place. If she wins the competition she gets €9000. If she comes second she gets €7000 and if she comes third she gets €3000. In all other cases she gets nothing. Each participant in the competition must pay €2000 to enter.

(a) Find the expected value of Mary's loss if she enters the competition.

The expected value E of Mary's loss is the average loss that she would make were she to enter this competition a very large number of times.

$$\text{E(loss)} = 2000 - \frac{9000}{20} - \frac{7000}{10} - \frac{3000}{4}$$

$$= 2000 - 450 - 700 - 750$$

$$= 100$$

(b) Each of the 3 prizes in the competition above is increased by the same amount (€x) but the entry fee is unchanged. For example, if Mary wins the competition now, she would get €(9000 + x). Mary now expects to break even. Find the value of x.

$$\text{E(loss)} = 2000 - \frac{9000+x}{20} - \frac{7000+x}{10} - \frac{3000+x}{4}$$

$$= 2000 - 450 - 700 - 750 - x(.05 + .10 + .25)$$

$$= 100 - .40x$$

$$= 0 \qquad \text{(since she now expects to break even)}$$

$$x = 250$$

Check: $(.05 + .10 + .25)x = 100$

Question 5

(a) A security code consists of six digits chosen at random from the digits 0 to 9. The code may begin with zero and digits may be repeated. For example 0 7 1 7 3 7 is a valid code.

(i) Find how many of the possible codes will end with a zero.

Only the first five digits can be varied, the sixth being fixed at 0.

The first digit has 10 possibilities. For any particular first digit, there are 10 possibilities for the second, and so on.

Hence there are 10^5 possible codes, i.e. 100,000.

(ii) Find how many of the possible codes will contain the digits 2 0 1 8 together and in this order.

The 2018 sequence can begin in the first, second or third position of the six-digit code.

For each of these there are two other digits to be filled, with 10^2 possibilities, i.e. 100.

Hence there are 3× 100, i.e. 300 possible codes.

(b) Find a, b, c and d if $\frac{(n+3)!(n+2)!}{(n+1)!(n+1)!} = an^3 + bn^2 + cn + d$, where a, b, c and d $\in \mathbb{N}$.

$$\frac{(n+3)!(n+2)!}{(n+1)!(n+1)!} = \frac{(n+3)!}{(n+1)!} \times \frac{(n+2)!}{(n+1)!}$$

$$= (n+3)(n+2) \times (n+2)$$

$$= (n+3)(n^2 + 4n + 4)$$

$$= n^3 + 7n^2 + 16n + 12$$

a = 1, b = 7, c = 16, d = 12

Answer to Question 6:

In Galway, rain falls in the morning on 1/3 of the school days in the year.

When it is raining the probability of heavy traffic is ½.

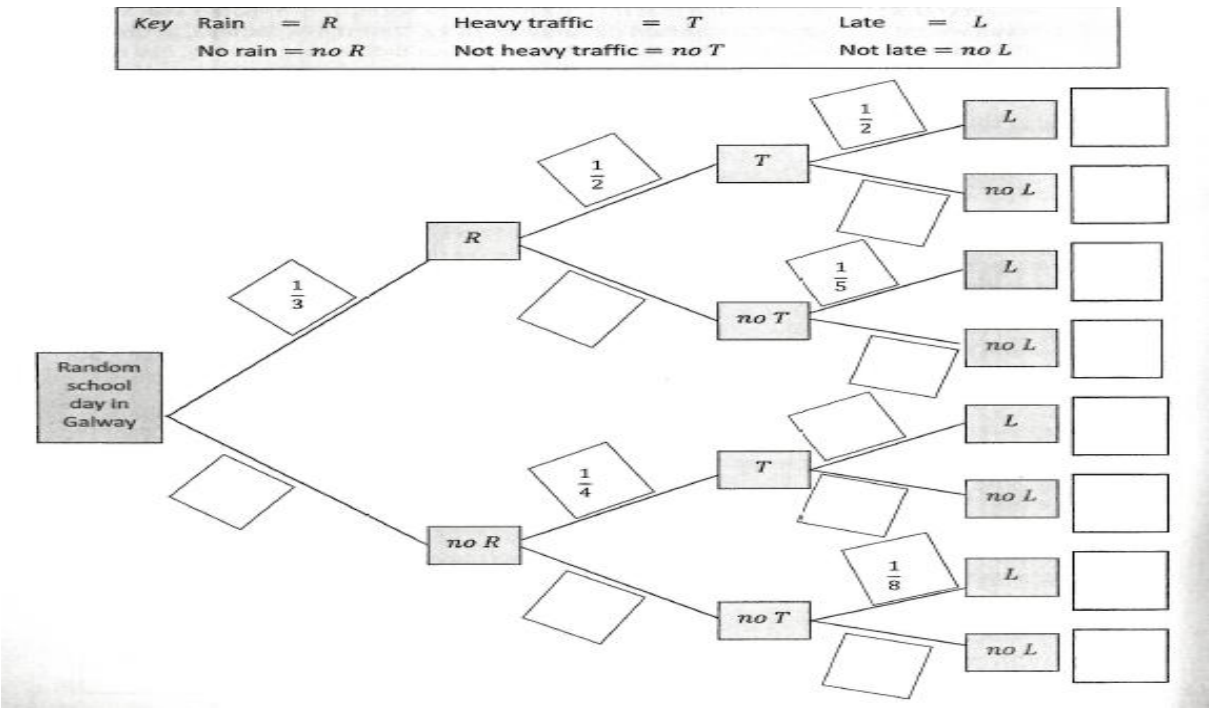
When it is not raining, the probability of heavy traffic is ¼.

When it is raining and there is heavy traffic, the probability of being late for school is ½.

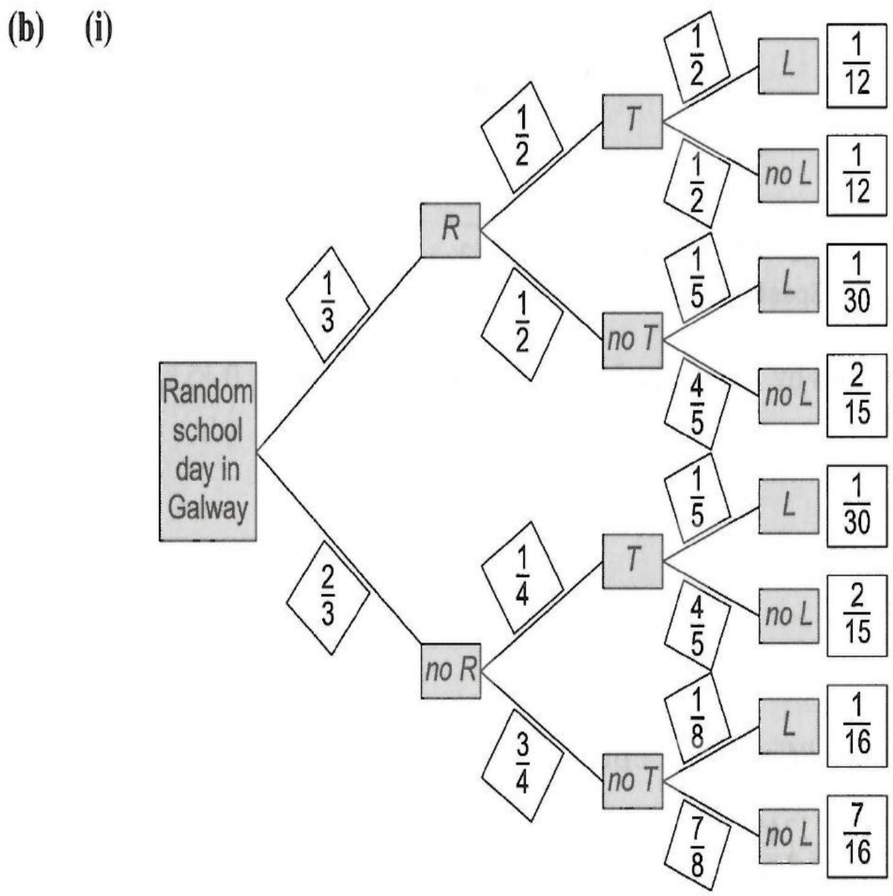
When it is not raining and there is no heavy traffic, the probability of being late for school is 1/8.

In any other situation the probability of being late for school is 1/5.

Some of this information is shown in the tree diagram below.



i) Write the probability associated with each branch of the tree diagram **and** the probability of each outcome into the blank boxes provided.
Give each answer in the form of a/b, where a, b are natural numbers.



ii) On a random school day in Galway, find the probability of being late for school.

The probability of being late for school is

Summing the probabilities of being late from the above diagram:

1/12 + 1/30 + 1/30 + 1/16 = **17/80** or 0.2125

iii) On a random school day in Galway, find the probability that it rained in the morning, given you were late for school.

P(Rain|Late) = P (Rain ∩ Late)/ P(Late)

From the diagram above

P (Rain ∩ Late) = 1/12 +1/30 = 7/60

P (Late) = 17/80 (from part ii)

P(Rain|Late) = (7/60) / (17/80)= **28/51** or 0.549

Answer to Question 7:

(a) A random variable X follows a normal distribution with mean 60 and standard deviation 5.

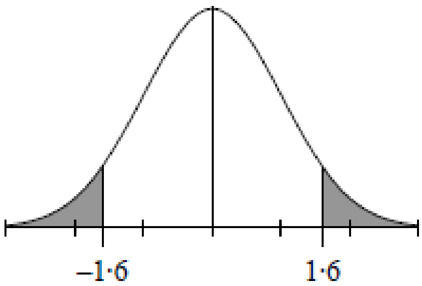
(i) Find $P(X \le 68)$.

$$P(X \le 68) = P\left(Z \le \frac{68-60}{5}\right) = P(Z \le 1\cdot 6) = 0\cdot 9452$$

(ii) Find $P(52 \le X \le 68)$.

$$P(52 \le X \le 68) = P\left(\frac{52-60}{5} \le Z \le \frac{68-60}{5}\right)$$

$= P(-1\cdot 6 \le Z \le 1\cdot 6)$



$P(Z \le -1\cdot 6) = P(Z \ge 1\cdot 6)$

$= 1 - P(Z \le 1\cdot 6)$

$= 1 - 0\cdot 9452 = 0\cdot 0548$

$P(-1\cdot 6 \le Z \le 1\cdot 6) = P(Z \le 1\cdot 6) - P(Z \le -1\cdot 6)$

$= 0\cdot 9452 - 0\cdot 0548 = 0\cdot 8904$

OR

$$P(52 \le X \le 68) = P\left(\frac{52-60}{5} \le Z \le \frac{68-60}{5}\right)$$

$= P(-1\cdot 6 \le Z \le 1\cdot 6)$

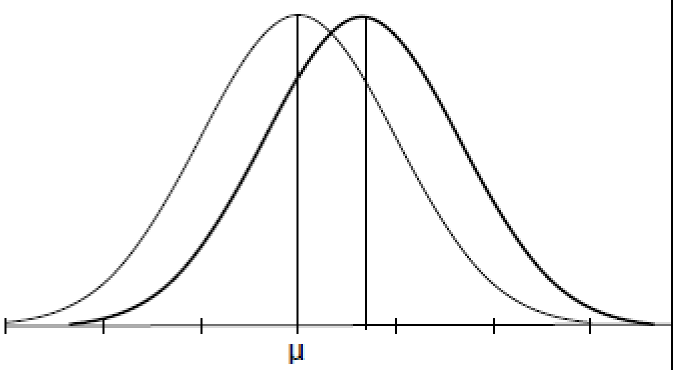
$= 1 - 2P(Z \ge 1\cdot 6)$

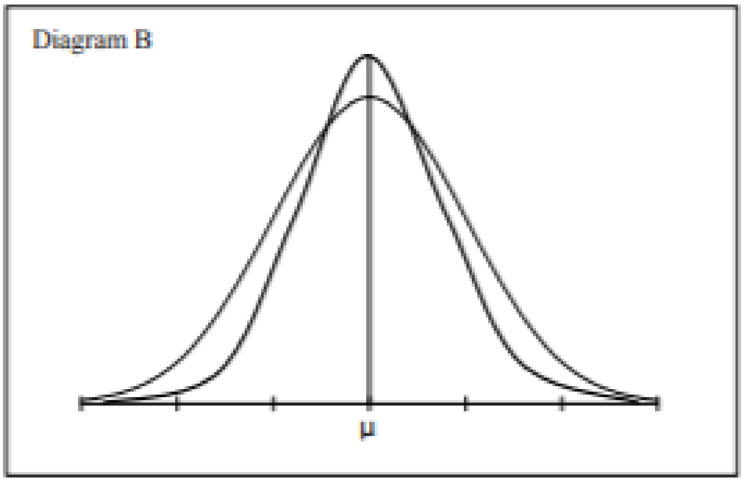
$= 1 - 2(1 - P(Z \le 1\cdot 6))$

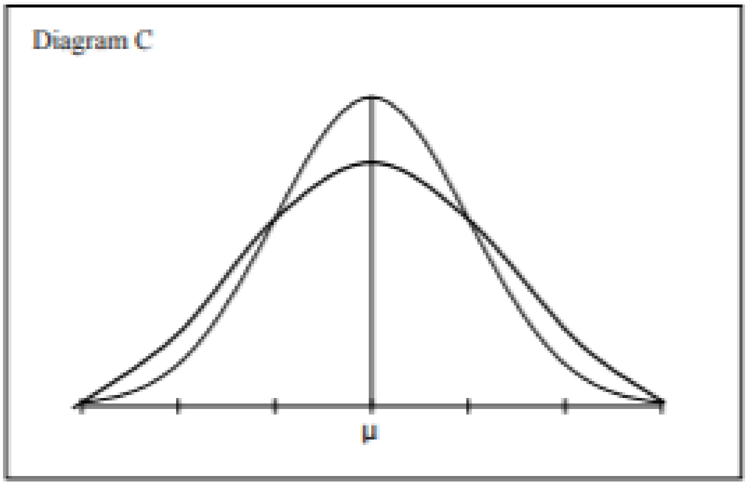
$= 1 - 2(1 - 0\cdot 9452) = 1 - 2(0\cdot 0548) = 1 - 0\cdot 1096 = 0\cdot 8904$

b)

Diagram A







Answer to Question 8:

(a) (i) Looking at the ranges of times in the three histograms it is evident that they are in the order: Run, Cycle, Swim.

(ii) From the histogram, about half of the data is to the left of 25.

Focusing on the median, we need to find a cumulative frequency of 112.
Reading the frequencies from the histogram from left to right: 3 + 16 + 25 +47 = 91.
Reading the frequencies from the histogram from right to left: 1 + 1 + 6 + 12 + 14 + 33 = 67.
The central frequency is 224 – 91 – 67 = 66 (which looks correct).
The median is within the 24 to 26 range. It is 21 from the left and 45 from the right.
We are told that the the mean and median are approximately equal.

$$\text{Mean} \approx \text{median} \approx 24 + 2 \times 21/66 \approx 25 \text{ minutes}$$

(iii) Using the histogram for the swim, rounding the mean to 18 and taking the mid-point of each range:

Time t	11	13	15	17	19	21	23	25	27	31	33
Frequency F	1	8	32	74	60	23	13	5	3	3	1
$F(t-18)^2$	49	200	288	74	60	207	325	245	243	507	225

The standard deviation is the the square root of the variance.

Sample variance = $\frac{\Sigma F(t-18)^2}{n-1}$ = $\frac{2423}{223}$ = 10.9 Formulae, pages 33, 34

(This is not the same as the sample variance of 10.017 given in the summary statistics presumably because the latter was calculated from the exact times and exact sample mean.)

The sample standard deviation is 3.3. ($\sqrt{10.017}$ = 3.2)

Alternatively, the shape of the histogram is not very different from a normal distribution. If it were normal we would expect 95% of the values within two standard deviations of the mean. We have to exclude .05 × 224 = 11 swim times at the extremes, leaving times roughly between 14 and 28 minutes. We then expect 28 – 14 = 14 to be four times the standard deviation, giving a standard deviation of 3.5.

(iv) The times have been recorded to 1000th of a minute so it may be assumed that the software did not find a mode because no two competitors had the same exact score.

(b) For each pair of events, if a candidate has a good time in one then that candidate tends to have a good time in the other, and vice versa. In statistical jargon, the times in the two events are positively correlated. The correlation is strongest between run and cycle times.

(c) Brian's time for the run (y) can be estimated from the best-fit line based on either swim-time or cycle-time. We will use all the information at our disposal by taking the average of these two estimates.

Y_1 = 0.53 × 17.6 + 15.2 = 24.528

Y_2 = 0.58 × 35.7 + 0.71 = 21.416

$$Y = \frac{24.528+21.416}{2} = 23.0$$

(d) The **empirical rule** states that in any normal distribution approximately 95% of the distribution lies within two standard deviations of the mean.

95% of the athletes took between 88.1 – 2 × 10.3 = 67.5

and 88.1 + 2 × 10.3 = 108.7

minutes to complete the race.

(e) To use the Tables on pages 36 and 37 of "formulae and tables" we have to calculate the standard normal variable equivalent to x = 100 minutes.

On page 34 of the Formulae we have the standard normal variable z where

$z = \frac{x-\mu}{\sigma}$ where μ = mean and σ = standard deviation

$$z = \frac{100-88.1}{10.3} = 1.16$$

In the Tables on page 37 we find that the probability of $z \le 1.16$ is .8729. We can use this to estimate the number of athletes who completed the race in less than 100 minutes:

$$224 \times .8729 = 196$$

Answer to Question 9:

We are given $\bar{x} = 90.45$ s = 20.73 n = 100

(s is the standard deviation calculated from the sample, which will be taken as an approximation of the standard deviation, σ, of x)

Hence $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20.73}{\sqrt{100}} = 2.073$$

N is sufficiently large for the distribution of $\bar{x}$ to be approximately normal.

(a) We don't know the mean, μ, of x.

The standard normal variable corresponding to $\bar{x} = 90.45$ is

$$\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{90.45 - \mu}{2.073}$$

From the Tables page 37 we note that

$$P(Z \le 1.96) = .9750$$

Hence $P(Z \ge 1.96) = .025$ so $P(\frac{90.45 - \mu}{2.073} \ge 1.96) = .025$

Also $P(Z \le -1.96) = .025$ so $P(\frac{90.45 - \mu}{2.073} \le -1.96) = .025$

Hence $P(90.45 - 1.96 \times 2.073 \le \mu \le 90.45 + 1.96 \times 2.073) = .95$

$P(86.39 \le \mu \le 94.51) = .95$

The required confidence interval is $86.39 \le \mu \le 94.51$.

If we took a large number of samples and did this each time μ would be within the interval 95% of the time.

(b) Null hypothesis (H_0): $\mu = 94$
Alternative hypothesis (H_1): $\mu \ne 94$

The standard normal variable corresponding to $\bar{x} = 90.45$ assuming H_0 is

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{90.45 - 94}{2.073} = -1.71$$

From the Tables page 37

$$P(Z \le 1.71) = .9564$$

$$P(Z \ge 1.71) = P(Z \le -1.71) = 1 - P(Z \le 1.71) = .0436$$

$$P(Z \geq 1.71) \;+\; P(Z \leq -1.71) \;=\; .0872$$

The probability of getting a sample average as extreme as 90.45 when H_0 is true is 8.72%, which is greater than 5% so we can't reject the null hypothesis.

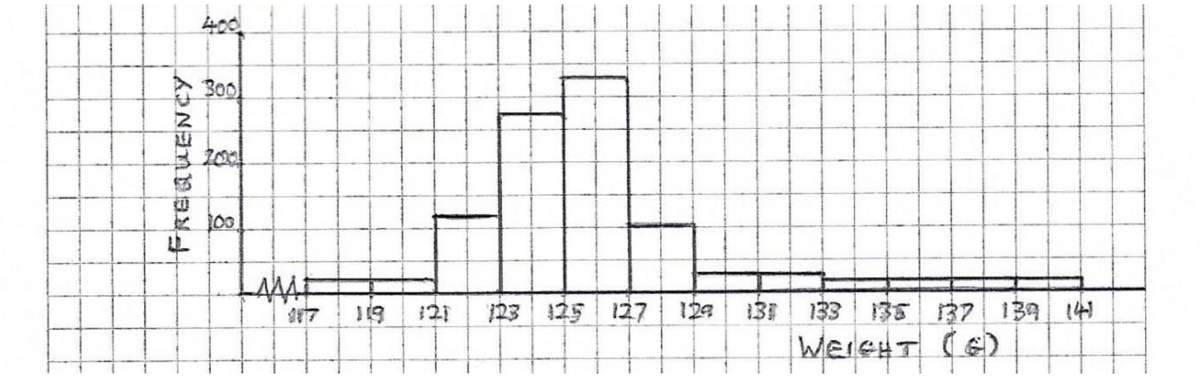
Note: We could have reached our conclusion more directly using the empirical rule since the observed value is within two standard deviations of the mean so it is not significant at a 5% level.

(c) The p-value of the test in (b) above is the probability that the average spend of a sample of 100 shoppers will differ from the mean 94 by at least 3.55 in either direction, which we have calculated as 8.72%.

Answer to Question 10:

(a) The weight groups are not all of equal width. We have to rearrange the data before drawing a histogram:

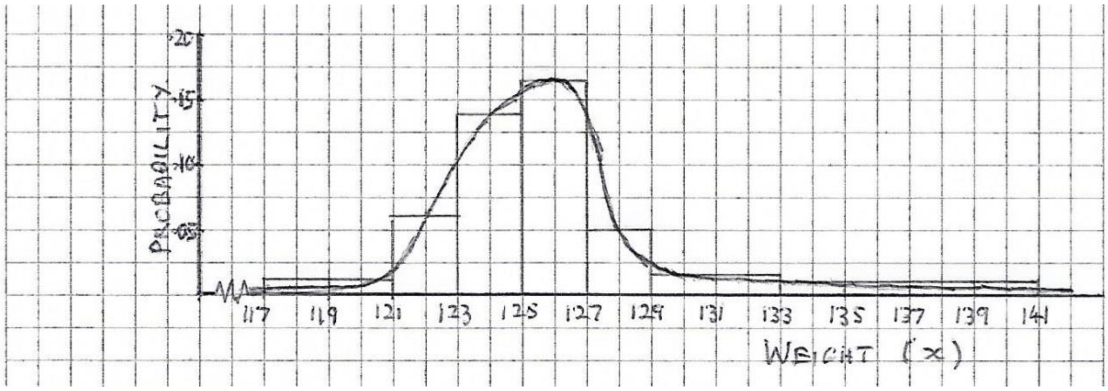
Weight x≤	119	121	123	125	127	129	131	133	135	137	139	141
Frequency	33	34	138	274	327	102	23	22	12	12	12	11



(b) The probability distribution has the same shape as the histogram but the scale on the y-axis has to be such that the area covered by the distribution is 1. This will ensure that the sum of the probabilities of all possible outcomes will be 1.

A frequency of 1 in the histogram covers an area of 2 since each cell has a width of 2. The total frequency is 1,000 so the histogram covers an area of 2,000. Therefore we need to divide the weight scale by 2,000 to turn the histogram into a probability distribution.

In practice any weight is possible so we can draw a continuous distribution by sight.



(c) The probability of $x_1 \le x \le x_2$ is the area under the curve between x_1 and x_2.

(d) (i) The mean of the distribution of average weights $\bar{x}$ is the same as the original mean,

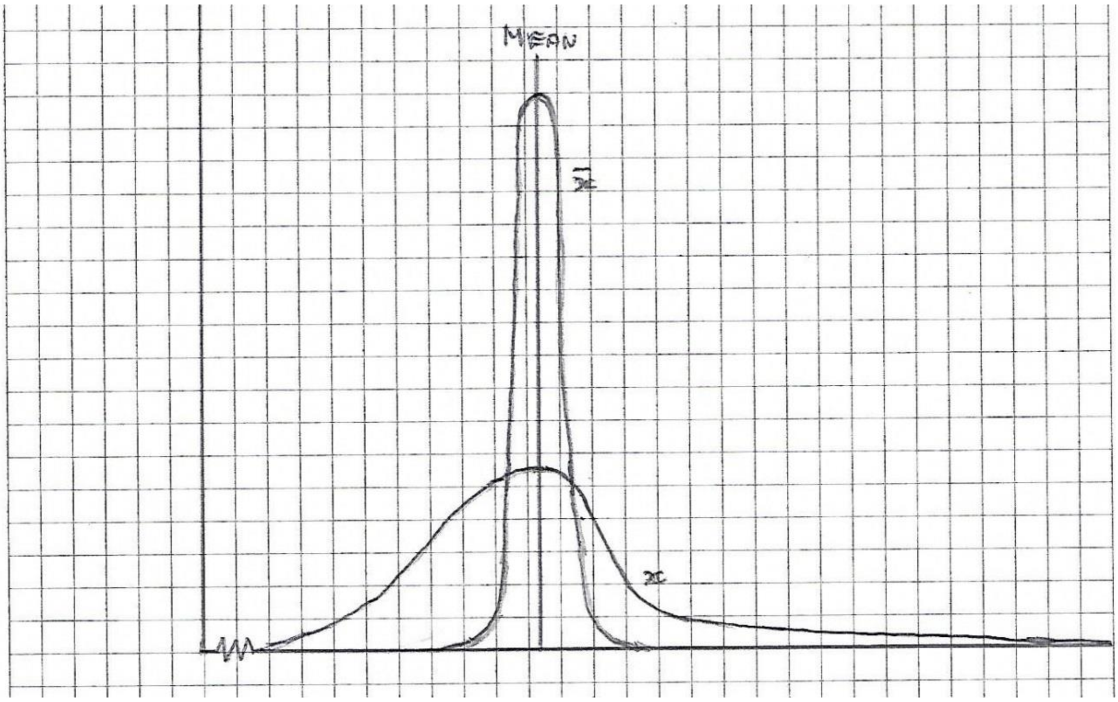
$$\mu_{\bar{x}} = \mu_x = 125.38$$

(ii) The standard deviation of the distribution of average weights is the original standard deviation divided by the square root of the sample size (Formulae, page 34),

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.66}{\sqrt{50}} = .52$$

(iii) $\bar{x}$ is approximately normally distributed (the distribution of any sample mean is approximately normal when the sample size is large enough, i.e. $n \ge 30$).

(iv)



The distribution of $\bar{x}$ has a symmetrical shape around the mean of x and it is less disperse because its standard deviation is the standard deviation of x divided by √n.

(e) We find the standard normal variable, z, which is equivalent to $\bar{x}$ = 124.32:

$$z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{124.32-125.38}{0.52} = -2.04$$

To use the Tables on pages 36 and 37 we use

$P(Z \le -2.04) = P(Z \ge 2.04)$ since the normal distribution is symmetrical about the mean 0

$= 1 - P(Z \le 2.04) = 1 - .9793 = .0207$

We are asked what is the probability of a deviation from the mean as large as observed, so we must add the probability that $z \ge 2.04$ to the probability that $z \le -2.04$. Using the symmetrical property of the normal distribution, the required probability is .0414.

(f) Since D can be either positive or negative, we start by halving the 0.5% probability of error which the management will accept, finding the z value for which

$P(Z \ge z) = 1 - P(Z \le z) = 1 - .0025 = .9975$

From the Tables, page 37, this probability corresponds to z = 2.81.

This is the deviation in terms of the standard normal variable and we need to transform it to the original variable $\bar{x}$. We use

$$z = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} \qquad 2.81 = \frac{D}{.52} \qquad D = 1.46$$

If a sample of 50 packets has an average weight of less than or equal to 123.92 or greater than or equal to 126.84 the machine should be stopped for maintenance.