ZNOTES // A-LEVEL SERIES

visit www.znotes.org

Updated to 2019 Syllabus

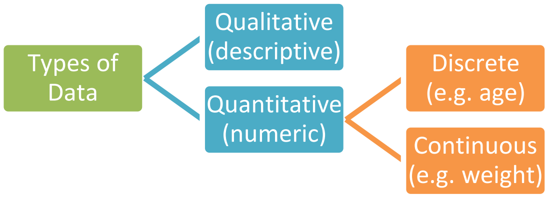
CIE A-LEVEL MATHS 9709 (S1)

FORMULAE AND SOLVED QUESTIONS FOR STATISTICS 1 (S1)

TABLE OF CONTENTS

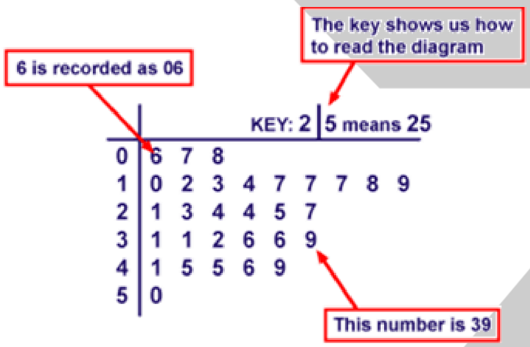
1. Representation of Data

1.1 Types of Data



1.2 Stem-and-Leaf Diagrams:

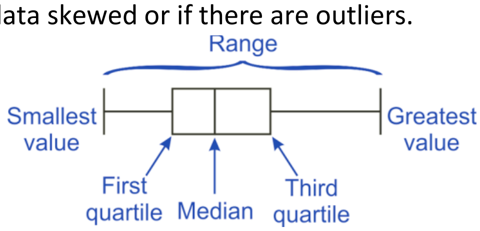
- Used to represent data in its original form.
- Each piece of data split into 2 parts; stem & leaf.
- Leaf can only by 1 digit and should be written in ascending order
- Always include a key on your diagram.



- **Advantage:** contains accuracy of original data

1.3 Box-and-Whisker Plots

- Five figure summary:
  - Lowest and highest values
  - Lower and upper quartiles
  - Median
- Mean & standard deviation most useful when data roughly symmetrical & contains no outliers
- Median and interquartile range typically used if data skewed or if there are outliers.



- **Advantage:** easily interpreted and comparisons can easily be made.

1.4 Histograms

- A bar chart which represents continuous data
- Bars have no space between them
- Area of each bar is proportional to frequency

$$Frequency = Frequency\ Density \times Class\ Width$$

- For open ended class width, double the size of previous class width and use this
- If range '$0 - 9$' then class width is '$-0.5 \leq x \leq 9.5$'

1.5 Cumulative Frequency Graphs

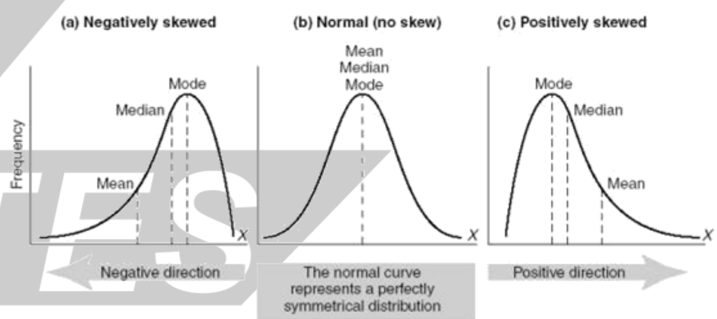
- Upper quartile $= 75\%$
- Lower quartile $= 25\%$

$$Interquartile\ Range = Upper\ Quartile - Lower\ Quartile$$

- When finding median & quartiles, draw in vertical and horizontal dashed lines.
- Join points together with straight lines unless asked to draw a cumulative frequency curve

1.6 Skewness

- **Symmetrical:** Median line lies in the middle of the box (i.e. UQ – median = median – LQ)
- **Positively skewed:** median line lies closer to LQ than UQ (i.e. UQ – median > median – LQ)
- **Negatively skewed:** median line lies closer to UQ than to the LQ (i.e. UQ – median < median – LQ)



2. Measure of Location

2.1 Mode

- Most common or most popular data value
- Only average that can be used for qualitative data
- Not suitable if the data values are very varied
- **Modal class:** class with highest frequency density

2.2 Median

- Middle value when data ordered
  - If n odd, median $= {}^{1}/_{2}\,(n+1)^{th}$ value
  - If n even, median $= {}^{1}/_{2}\,n^{th}$ value
- Not affected be extreme values

Estimating Median from Grouped Frequency Table:

x	Frequency f	Cumulative Frequency
$10-20$	4	4
$20-25$	8	12
$25-35$	5	17
$35-50$	3	20

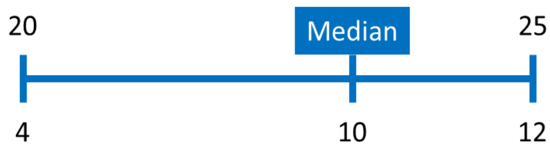
Solution:

Use cumulative frequency to find the middle value i.e.

$$20 \div 2 = 10$$

∴ you are finding the 10th value

The 10th value lies between 20 and 25



$$(12-4) : (25-20)$$
$$(12-10) : (25-Median)$$
$$25 - Median = \frac{12-10}{12-4} \times (25-20)$$
$$Median = 23.75$$

2.3 Mean

- Sum of data divided by number of values

$$\bar{x} = \frac{\sum x_i}{n} \qquad \text{or} \qquad \bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

- Important as it uses all the data values
- Disadvantage: affected by extreme values
- If data is grouped – use mid-point of group as x
- **Coded mean:** if being used to calculate standard deviation, can be used as is else:

$$\bar{x} = \frac{\sum(x-a)}{n} + a$$

3 Measure of Spread

3.1 Standard Deviation

- Deviation from the mean is the difference from a value from the mean value
- The standard deviation is the average of all of these deviations
- If coded mean and sums given, use as it is, standard deviation not altered

3.2 Variance of Discrete Data

$$\frac{1}{n}\sum(x_i - \bar{x})^2 \qquad \text{or} \qquad \frac{1}{n}\sum x_i^2 - \bar{x}^2$$

Standard deviation is the square root of that

3.3 Variance in Frequency Table

$$\frac{\sum(x_i-\bar{x})^2 f_i}{\sum f_i} \qquad \text{or} \qquad \frac{\sum {x_i}^2 f_i}{\sum f_i} - \bar{x}^2$$

{W04-P06} **Question 4:**

The ages, x years, of 18 people attending an evening class are summarised by the following totals:

$$\sum x = 745, \sum x^2 = 33\,951$$

i. Calculate the mean and standard deviation of the ages of this group of people.

ii. One person leaves group and mean age of the remaining 17 people is exactly 41 years. Find age of the person who left and standard deviation of the ages of the remaining 17 people.

Solution:

Part (i)

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} \qquad \bar{x} = \frac{\sum x}{n}$$

$$\sigma = 13.2 \qquad \bar{x} = 41.4$$

Part (ii)

The total age of the 18 people

$$\sum x = 745$$

Find the total age of the 17 people

$$\sum x = 41 \times 17 = 697$$

Subtract the two to get the age

$$745 - 697 = 48 \text{ years}$$

Calculating the new standard deviation

Find the $\sum x^2$ of the 17 people

$$\sum x^2 = 33\,951 - 48^2 = 31\,647$$

Find the standard deviation

$$\sigma = \sqrt{\frac{31\,647}{17} - (41)^2} = 13.4$$

{S13-P62} **Question 2:**

A summary of the speeds, x kilometres per hour, of 22 cars passing a certain point gave the following information:

$$\sum(x-50) = 81.4 \qquad \text{and} \qquad \sum(x-50)^2 = 671.0$$

Find variance of speeds and hence find the value of $\sum x^2$

Solution:

Finding the variance using coded mean

$$\text{Variance} = \frac{671.0}{22} - \left(\frac{81.4}{22}\right)^2 = 16.81$$

Find the actual mean

$$\sum x = 81.4 + (22 \times 50) = 1181.4$$

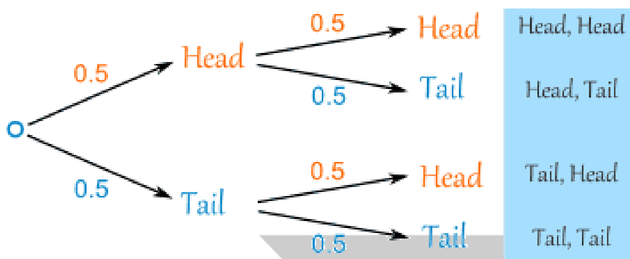
Put this back into variance formula

$$16.81 = \frac{\sum x^2}{n} - \left(\frac{1181.4}{22}\right)^2$$
$$\therefore \sum x^2 = 2900.5 \times 22$$
$$\sum x^2 = 63811$$

4 PROBABILITY

4.1 Basic Rules

- All probabilities lie between 0 and 1
- $P(A) =$ The probability of event A
- $P(A') = \ 1 - P(A) =$ The probability of not A
- To simplify a question represent info in tree diagram:



{S08-P06} **Question 7:**

A die is biased so that the probability of throwing a 5 is 0.75 and probabilities of throwing a 1, 2, 3, 4 or 6 are all equal. The die is thrown thrice. Find the probability that the result is 1 followed by 5 followed by any even number

Solution:

Probability of getting a 1

$$1 - 0.75 = 0.25$$

5 numbers $\therefore 0.25 \div 5 = 0.05$

Probability of getting a 5 $= 0.75$

Probability of getting an even number; can be 2, 4 or 6 $\therefore$

$$0.05 \times 3 = 0.15$$

Total probability

$$0.05 \times 0.75 \times 0.15 = 0.00563$$

4.2 Mutually Exclusive Events

- 2 events which have no common outcomes
- **Addition Law of MEEs:**

$$P(A \text{ and } B) = 0$$

4.3 Conditional Probability

- Calculation of probability of one event given that another, connected event, had occurred
- **Multiplication Law of Connected Events:**

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

4.4 Independent Events

- Events that aren't connected to each other in any way
- **Multiplication Law for IEs:**

$$P(A \text{ and } B) = P(A) \times P(B)$$

{S07-P06} **Question 2:**

Jamie is equally likely to attend or not to attend a training session before a football match. If he attends, he is certain to be chosen for the team which plays in the match. If he does not attend, there is a probability of 0.6 that he is chosen for the team.

i. Find probability that Jamie is chosen for team.
ii. Find the conditional probability that Jamie attended the training session, given that he was chosen for the team

Solution:

Part (i)

Probability attends training and chosen

$$0.5 \times 1 = 0.5$$

Probability doesn't attend and chosen

$$0.5 \times 0.6 = 0.3$$

Total probability

$$0.3 + 0.5 = 0.8$$

Part (ii)

$$P(Attends|Chosen) = \frac{P(Attends \text{ and } Chosen)}{P(Chosen)}$$

$$P(Attends|Chosen) = \frac{0.5}{0.8} = 0.625$$

Old Question: **Question 7:**

Events A and B are such that $P(A) = 0.3, P(B) = 0.8$ and $P(A \text{ and } B) = 0.4$. State, giving a reason in each case, whether A and B are

i. independent
ii. mutually exclusive

Solution:

Part (i)

A and B are not mutually exclusive because:

$P(A \text{ and } B)$ does not equal 0

Part (ii)

A and B are not independent because:

$P(A) \times P(B)$ does not equal 0.4

{S11-P63} **Question 4:**

Tim throws a fair die twice and notes the number on each throw. Events A, B, C are defined as follows.

A: the number on the second throw is 5
B: the sum of the numbers is 6
C: the product of the numbers is even

By calculation find which pairs, if any, of the events A, B and C are independent.

Solution:

Probability of Event A = P(Any Number) × P(5)

$$\therefore P(A) = 1 \times \frac{1}{6} = \frac{1}{6}$$

Finding the probability of Event B

Number of ways of getting a sum of 6:

5 and 1   1 and 5   4 and 2   2 and 4   3 and 3

$$\therefore P(B) = \left(\frac{1}{6}\times\frac{1}{6}\right)\times 5 = \frac{5}{36}$$

Finding the probability of Event C

One minus method; you get an odd only when odd multiplies by another odd number:

$$1 - P(C) = \frac{1}{2}\times\frac{1}{2}$$

$$\therefore P(C) = \frac{3}{4}$$

For an independent event, $P(A \text{ and } B) = P(A) \times P(B)$

$$P(A \text{ and } B) = P(1 \text{ and } 5) = \frac{1}{36}$$

$$\neq P(A) \times P(B)$$

$$P(A \text{ and } C) = P[(2,5) + (4,5) + (6,5)] = \frac{3}{36}$$

$$\neq P(A) \times P(C)$$

$$P(B \text{ and } C) = P[(2,4) + (4,2)] = \frac{2}{36}$$

$$\neq P(B) \times P(C)$$

$\therefore$ none are independent.

5 Permutations and Combinations

5.1 Factorial

- The number of ways of arranging n unlike objects in a line is $n!$

Total arrangements for a word with repeated letters:

$$\frac{(Number\ of\ Letters)!}{(Repeated\ Letter)!}$$

If more than one letter repeated, multiply the factorial of the repeated in the denominator

Total arrangements when two people be together:

- Consider the two people as one unit

Example:

In a group of 10, if A and B have to sit next to each other, how many arrangements are there?

Solution:

$$(9!)\times(2!)$$

2! is necessary because A and B can swap places

- If question asks for two people **not** to be next to each other, simply find total arrangements (10!) and subtract the impossible i.e. (9!)×(2!)

Total arrangements when items cannot be together:

Example:

In how many ways can the letters in the word SUCCESS be arranged if no two S's are next to one another?

Solution:

U E C_1 C_2 & S_1 S_2 S_3

	U		E		C_1		C_2	

S has 5 different places in can be placed into.

From previous note, we must divide by repeated letters

No. of Arrangements $= \frac{4!}{2!} \times \frac{5\times4\times3}{3!} = \mathbf{120}$

5.2 Combination

- The number of ways of selecting r objects from n unlike objects is:

$$^nC_r = \frac{n!}{r!\,(n-r)!}$$

- Order **does not** matter

5.3 Permutations

- The number of ordered arrangements of r objects taken from n unlike objects is:

$$^nP_r = \frac{n!}{(n-r)!}$$

- Order **matters**

6 Probability Distribution

- The probability distribution of a discrete random variable is a listing of the possible values of the variable and the corresponding probabilities
- Total of all probability always equals 1
- Can calculate unknowns in a probability distribution by summing them to equal 1

{S05-P06} **Question 3:**

A fair dice has four faces. One face is coloured pink, one is orange, one is green and one is black. Five such dice are thrown and the number that fall on a green face is counted. The random variable X is the number of dice that fall on a green face. Draw up a table for probability distribution of X, giving your answers correct to 4 d.p.

Solution:

This is a binomial distribution where the probability of success is $\frac{1}{4}$ and the number of trials is 5

$$P(X = x) = nC_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{n-x}$$

The dice are rolled five times thus the number of green faces one can get ranges from 0 to 5

Use formula to obtain probabilities e.g. $P(X = 1)$,

$$P(X=1) = 5C_1\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^4 = 0.3955$$

Thus draw up a probability distribution table

x	0	1	2	3	4	5
$P(X=x)$	0.2373	0.3955	0.2637	0.0879	0.0146	0.0010

7 Binomial Distribution

Conditions:

- Only 2 possible outcomes & are mutually exclusive
- Fixed number of n trials
- Outcomes of each trial independent of each other
- Probability of success at each trial is constant

$$P(X=x) = {}^nC_x \times p^x \times q^{(n-x)}$$

Where $p=$ probability of success
$q=$ failure $=(1-q)$
$n=$ number of trials

- A binomial distribution can be written as:

$$X \sim B(n,p)$$

{W11-P62} **Question 6:**

In Luttley College; 60% of students are boys. Students can choose exactly one of Games, Drama or Music on Friday afternoons. 75% of the boys choose Games, 10% choose Drama and remainder choose Music. Of the girls, 30% choose Games, 55% choose Drama and remainder choose Music. 5 drama students are chosen. Find the probability that at least 1 of them is a boy.

Solution:

First we calculate the probability of selecting a boy who is a drama student; a conditional probability:

$$P(S) = \frac{P(Boy|Drama)}{P(Drama)}$$

$$P(S) = \frac{P(Boy)\times P(Drama)}{(P(Boy)\times P(Drama)) + (P(Girl)\times P(Drama))}$$

$$P(S) = \frac{0.6\times 0.1}{(0.6\times 0.1)+(0.4\times 0.55)} = \frac{3}{14}$$

We can calculate the probability there is at least 1 boy present from 5 drama students using a binomial distribution with 5 trials and P of success $=\frac{3}{14}$

Find probability of 0 and subtract answer from 1:

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

$$P(X\geq 1) = 1 - 5C_5\times\left(\frac{3}{14}\right)^0\times\left(\frac{11}{14}\right)^5$$

$$P(X\geq 1) = 0.701$$

8 Discrete Random Variables

8.1 Probability Distribution Tables

- To calculate the expected value of a random variable or its mean:

$$E(x) = \mu = \sum x_i p_i$$

- To calculate the variance of a random variable, first calculate the expected value of a random variable squared

$$E(x^2) = \sum (x_i)^2 \times p_i$$

- Finally to calculate the variance

$$\sigma^2 = \sum (x_i-\mu)^2 p_i = \sum x_i^2 p_i - \mu^2$$

{W11-P63} **Question 3:**

A factory makes a large number of ropes with lengths either 3m or 5m. There are four times as many ropes of length 3m as there are ropes of length 5m. One rope is chosen at random. Find the expectation and variance of its length.

Solution:

From information given, calculate probabilities

$$P(3m\ Rope) = \frac{4}{5} \qquad P(5m\ Rope) = \frac{1}{5}$$

Calculate expectation/mean

$$E(x) = \sum x_i p_i = \left(3\times\frac{4}{5}\right) + \left(5\times\frac{1}{5}\right) = 3.4$$

Calculate expectation squared

$$E(x^2) = \sum (x_i)^2\times p_i = \left(3^2\times\frac{4}{5}\right) + \left(5^2\times\frac{1}{5}\right) = 12.2$$

Calculate the variance

$$\sigma^2 = \sum x_i^2 p_i - \mu^2 = 12.2 - (3.4^2) = 0.64$$

8.2 Binomial Distribution

$$X\sim B(n,p)$$

- To calculate the expected value of a random variable or its mean with a binomial distribution:

$$E(x) = \mu = np$$

- To calculate the variance:

$$\sigma^2 = np(1-p)$$

{S11-P63} **Question 6:**

The probability that Sue completes a Sudoku puzzle correctly is 0.75. Sue attempts 14 Sudoku puzzles every month. The number that she completes successfully is denoted by X. Find the value of X that has the highest probability. You may assume that this value is one of the two values closest to the mean of X.

Solution:

Calculate the mean of X

$$E(X) = 14 \times 0.75 = 10.5$$

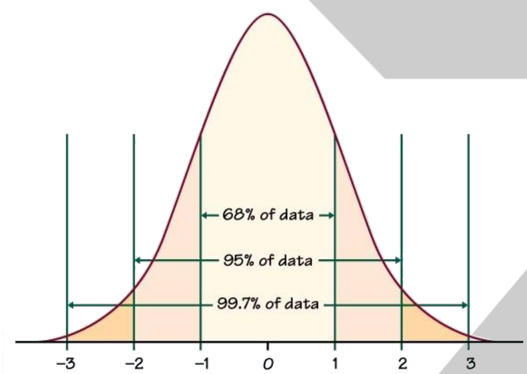
Successful puzzles completed has to be a whole number so can either be 10 or 11.

$$P(10) = 14C_{10} \times 0.75^{10} \times 0.25^4 = 0.220$$
$$P(11) = 14C_{11} \times 0.75^{11} \times 0.25^3 = 0.240$$

Probability with 11 is higher ∴

$$X = 11$$

9 The Normal Distribution



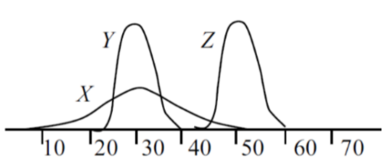
{W13-P61} **Question 1:**

It is given that $X \sim N(30, 49)$, $Y \sim N(30, 16)$ and $Z \sim N(50, 16)$. On a single diagram, with the horizontal axis going from 0 to 70, sketch 3 curves to represent the distributions of X, Y and Z.

Solution:

For X, plot center of curve at 30 and calculate $\sigma = \sqrt{49}$

Plot $3 \times \sigma$ to the left and right i.e. $30 - 21 = 9$ and $30 + 21 = 51$. Follow example for the other curves.



9.1 Standardizing a Normal Distribution

To convert a statement about $X \sim N(\mu, \sigma^2)$ to a statement about $N(0,1)$, use the **standardization** equation:

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

9.2 Finding Probabilities

Example

For a random variable X with normal distribution

$$X \sim N(20, 4^2)$$

Find the probability of $P(X \leq 25)$

Standardize the probability

$$Z = \frac{25 - 20}{4} = 1.25$$

Search for this value in normal tables

$$\Phi(1.25) = 0.8944$$

Find the probability of $P(X \geq 25)$

Change from greater than to less than using:

$$P(Z \geq a) = 1 - P(Z \leq a)$$
$$P(X \geq 25) = 1 - P(X \leq 25)$$

Using the probability from above

$$P(X \geq 25) = 1 - 0.8944 = 0.1057$$

Find the probability of $P(X \leq 12)$

Standardize the probability

$$Z = \frac{12 - 20}{4} = -2$$

Change from negative value to positive by:

$$P(Z \leq -a) = 1 - P(Z \leq a)$$
$$P(Z \leq -2) = 1 - P(Z \leq 2)$$

Search for 2 in the normal tables

$$P(Z \leq -2) = 1 - 0.9773 = 0.0228$$

Find the probability of $P(10 \leq X \leq 30)$

Split inequality into two using:

$$P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$$
$$P(10 \leq X \leq 30) = P(X \leq 30) - P(X \leq 10)$$

Standardize values

$$= P(Z \leq 2.5) - P(Z \leq -2.5)$$

Convert negative value to positive

$$= P(Z \leq 2.5) - \left(1 - P(Z \leq 2.5)\right)$$

Search for 2.5 in the normal tables

$$= 0.9938 - (1 - 0.9938) = 0.9876$$

9.3 Using Normal Tables Given Probabilities

{S12-P61} **Question 6:**

The lengths of body feathers of a particular species of bird are modelled by a normal distribution. A researcher measures the lengths of a random sample of 600 feathers and finds that 63 are less than 6 cm long and 155 are more than 12 cm long.

i. Find estimates of the mean and standard deviation of the lengths of body feathers of birds of this species.

ii. In a random sample of 1000 body feathers from birds of this species, how many would the researcher expect to find with lengths more than 1 standard deviation from the mean?

Solution:

Part (i)

Interpreting the question and finding probabilities:

$$P(X < 6) = 0.105 \qquad P(X > 12) = 0.258$$

For $X < 6$, the probability cannot be found on the tables which means it is behind the mean and therefore we must find 1 − and assume probability is negative

$$-P(X < 6) = 0.895$$

Using the standardization formula and working back from the table as we are given probability

$$\frac{6-\mu}{\sigma} = -1.253$$

Convert the greater than sign to less than

$$P(X > 12) = 1 - P(X < 12)$$
$$P(X < 12) = 1 - 0.258 = 0.742$$

Work back from table and use standardization formula

$$\frac{(12-\mu)}{\sigma} = 0.650$$

Solve simultaneous equations

$$\sigma = 3.15 \quad \text{and} \quad \mu = 9.9$$

Part (ii)

Greater than 1sd from μ means both sides of the graph however area symmetrical ∴ find greater & double it

Using values calculated from (i)

$$P(X > (9.9 + 3.15) = P(X > 13.05)$$

Standardize it

$$\frac{13.05 - 9.9}{3.15} = 1$$

Convert the greater than sign to less than

$$P(Z > 1) = 1 - P(Z < 1)$$

Find probability of 1 and find $P(Z > 1)$

$$P(Z > 1) = 1 - 0.841 = 0.1587$$

Double probability as both sides taken into account

$$0.1587 \times 2 = 0.3174$$

Multiply probability with sample

$$0.3174 \times 1000 = 317 \text{ birds}$$

9.4 Approximation of Binomial Distribution

- The normal distribution can be used as an approximation to the binomial distribution
- **For a binomial to be converted to normal, then:**

For $X \sim B(n, p)$ where $q = 1 - p$:

$$np > 5 \quad \text{and} \quad nq > 5$$

- If conditions are met then:

$$X \sim B(n, p) \quad \Leftrightarrow \quad V \sim N(np, npq)$$

9.5 Continuity Correction Factor (e.g. 6)

Binomial	Normal
$x = 6$	$5.5 \le x \le 6.5$
$x > 6$	$x \ge 6.5$
$x \ge 6$	$x \ge 5.5$
$x < 6$	$x \le 5.5$
$x \le 6$	$x \le 6.5$

{S09-P06} **Question 3:**

On a certain road 20% of the vehicles are trucks, 16% are buses and remainder are cars. A random sample of 125 vehicles is taken. Using a suitable approximation, find the probability that more than 73 are cars.

Solution:

Find the probability of cars

$$1 - (0.16 + 0.2) = 0.64$$

Form a binomial distribution equation

$$X \sim B(125, 0.64)$$

Check if normal approximation can be used

$$125 \times 0.64 = 80 \quad \text{and} \quad 125 \times (1 - 0.64) = 45$$

Both values are greater than 5 so normal can be used

$$X \sim B(125, 0.64) \quad \Leftrightarrow \quad V \sim N(80, 28.8)$$

Apply the continuity correction

$$P(X > 73) = P(X \ge 73.5)$$

Finding the probability

$$P(X \ge 73.5) = 1 - P(X \le 73.5)$$

Standardize it

$$Z = \frac{73.5 - 80}{\sqrt{28.8}} = -1.211$$

As it is a negative value, we must one minus again

$$1 - (1 - P(Z < -1.211) = P(Z < 1.211)$$

Using the normal tables

$$P = 0.8871$$

CIE A-LEVEL MATHEMATICS//9709

WWW.ZNOTES.ORG