Year 12 Mathematic Methods Worksheet

Statistical Interference Test 1

10 questions on Statistical Inference from the Maths C (Maths Methods) national curriculum for Year 12.

Remember you can connect to one of our awesome Mathematics tutors and they'll help you understand where you're going wrong.
They're online 3pm-midnight AET, 6 days a week.

Homework help in a click: yourtutor.com.au

yourtutor

Questions

1. **Statistical inference- definitions**

Statistical inference makes use of information from a sample to draw conclusions (inferences) about the population from which the sample was taken.

A population is the entire collection of things that statisticians want to find out about. It can be people, but it could also be animals, plants, fish or insects. A population is the entire group, which is to be described, and about which conclusions will be drawn.

Suppose scientists were making a study of Cane Toads in Australia in order to try to find a way to control their increase. The population would consist of the entire number of Cane Toads living in Australia. Often, the entire population is too large and it would be too costly to study every single member of the population. Statisticians then study a sample and hope that this sample is representative of the population. From the study of this sample, generalisations about the entire population can be drawn. For every population, many possible samples are possible. Statisticians can take the sample data and calculate such things as the sample mean, median, mode, variance and range. A sample statistic gives information about a corresponding population parameter. The sample mean for a set of data would give information about the overall population mean.

Note: the population mean is called a **parameter**. It refers to the entire population.
The sample mean is called a **statistic** and may or may not be close to the population mean. If time and money permit, it is a good idea to take several samples and then take the mean of the sample statistics.
Suppose a University did a study of Cane Toads found in a 5 kilometre square area of one of Queensland's national parks. This would be a sample of the

Cane Toads. It was found that these Cane Toads could eat lizards, small snakes and mice as well as insects. From this, you could **infer** that *all* Cane Toads ate lizards, small snakes and mice as well as insects. You would be generalising the information from the sample to the population.

Choose all correct statements.

(i) From a census of the entire Australian population, it was found that 99% of Australian homes had a television. This is statistical inference.

(ii) It was found that 65% of students at Newcastle University were supported by their parents. From this it was concluded that 65% of Australian university students were supported by their parents. This is statistical inference.

(iii) A study was done of nesting blue wrens in Lane Cove National park and found that the average number of eggs laid was three. This average is called a parameter.

(iv) From a census of the entire Australian population, it was found that the average number of cars per household was 1.65. This is a parameter.

a) (ii), (iii)

b) (ii), (iv)

c) i), (iv)

d) (i), (iii)

Answer: ____________________

2. Statistical inference: the two-outcome population

Suppose scientists are trialling a new drug to help Asthmatics breath more easily. Such a trial results in success or failure. The proportion of successes X in the population N is represented by $P = X/N$.

Scientists would normally trial the new drug in a sample of people suffering from Asthma. The sample proportion f is defined by $f = x/n$ where x is the number of successes in the sample sized n. Any sample is only one of the infinite number of possible samples that can be taken from a population.

The sampling distribution describes the probabilities associated with a statistic when a random sample is drawn from a population. The sampling distribution is the probability distribution or probability density function of that statistic.

Suppose Kristie wanted to know the effectiveness of a new method of controlling an allergic response to grass mites. She would need to define her population and make sure no-one was included who did not have an allergy to grass mites. Then she would need to take a sample of the population and calculate the number of successes divided by the total population - the probability of success x in the sample number n. Suppose in this case she found $f = 55/100$ where $x = 55$, the number of successes out of a total sample of 100.

Choose all correct statements:

(i) The sample statistic $f = 0.55$ tells Kristie exactly what the probability of success in the total population is.
(ii) The sample statistic $f = 0.55$ gives Kristie an idea of what the probability of success in the total population is.

(iii) If Kristie wanted to get a better idea of the probability of success, she would need to take many samples and average the result.

(iv) The number of successes in the total population is a parameter and this does not vary.

(v) If $f = x/n$ is a statistic from samples, it can vary from sample to sample.

a) (i), (ii), (iii)

b) (iii), (iv)

c) (ii), (iii), (iv)

d) (ii), (iii), (iv), (v)

The Answer________________________

3. Statistical inference: sampling issues and errors

If you take several samples from a population for a study, then your results for each sample are likely to differ. Suppose you and your friends are checking the likelihood of voters to vote for a certain political candidate. You and your friends take different samples of people to survey. The ratio of yes votes *x* to the size the sample *n*, could be expected to vary from sample to sample. It is a variable.

This is called sampling error. It can be reduced by taking a large enough sample if money and time permit. Statisticians can make an estimate of sampling error.

Sampling error results from differences in the sample statistics to the population parameters.

Sampling error can result from sampling bias.

Examples of sampling bias include:

(a) self selection bias, where the selection from the population is influenced by the population itself. If a radio station conducted a poll asking people to ‘phone in their responses' then the people who ‘phone in' are likely to be the people most interested in the topic under question and not really representative of the entire population.

(b) the under-representation or over-representation because of the location of the sample. If you were to take a sample of people in Australia and do a study on skin cancer for example, then your statistic for this sample would be very different to the statistic from a study done in Norway.

(c) accidental or intentional pre-screening of participants in the sample. If you were conducting a survey on the effects of smoking on long term health, if you only asked nurses who took care of cancer patients, their responses would be different to a sample of people from a fitness camp.

(d) the accidental or intentional exclusion of part of the population from the sample. Suppose you conducted a survey on whether or not pedestrian lights should be installed on a main road. If you only asked pedestrians, then you would be excluding the drivers who drove along the road.

There is also non-sampling error. This includes anything which causes the sample statistic to be different to the population parameter. Computer error, entering an incorrect formula into the computer, accidental loss of some of the data can be included in non-sampling error. Statisticians cannot easily quantify non-sampling error.

Question

Bus fares are going to rise and Cindy wants to conduct a survey to see how many of students will decide not to catch the bus and walk to school instead.

She decides to conduct a survey by asking students in the year nine assembly to put up their hands if they think they would rather walk to school.

Choose all correct statements.

(i) Cindy has identified the population she is studying correctly.

(ii) Cindy should have identified students who travelled on the bus and chosen her sample from this population.

(iii) Any statistic Cindy calculates from this sample is likely to have sampling error.

(iv) Cindy's sample excludes older and younger students whose response to the rise in bus fares is likely to be different.

a) (ii), (iii), (iv)

b) (i), (ii), (iii)

c) (i), (ii), (iii), (iv)

d) (i), (iii), (iv)

The Answer______________________

4. Statistical inference: the sample proportion as a random variable

Suppose Elli was conducting a survey of the number of students with blue eyes in her school. There are a fixed number of students and each student has eyes which are either blue or not blue. Elli takes as her sample the students in year nine. She counts the number of students and then determines the number of students with blue eyes. When she counts the number of students with blue eyes, she is working with a binomial random variable. When she works out the proportion of blue eyed students to the total in her sample, this is another random variable. From this sample proportion, Elli can make inferences about the population proportion- the total number of students with blue eyes in the entire school's population.

A random variable represents, in number form, the possible outcomes which could occur for some random experiment.

Let X denote the number of successes out of a sample of n observations. If each observation is a success with probability p independently of the other observations, then X is a binomial random variable with parameters n and p. The proportion of successes in the sample is also a random variable and is computed as a proportion of success = number of successes/total number of observations in the sample.

$f = x/n$

Yan and Elli both conduct the same investigation- the number of students with blue eyes. Yan conducts his survey using as a sample years 7 and 10. Elli uses as her sample the students in year 9. Look at the following results:

Elli

$f = 2/109$

Yan

$f = 1/235$

Choose all correct statements.

(i) Elli's proportion gives a better idea of the population proportion of the school.

(ii) Yan's proportion gives a better idea of the population proportion of the school.

(iii) The proportions give you an idea of the probability of selecting a blue-eyed student if a student is selected at random from the school's population.

(iv) From these proportions you can make the inference that there are not many students with blue eyes in the school.

(v) The random variable represented by *f* varies from sample to sample.

a) (ii), (iii), (iv), (v)

b) (ii), (iii), (iv)

c) (i), (ii), (iii), (v)

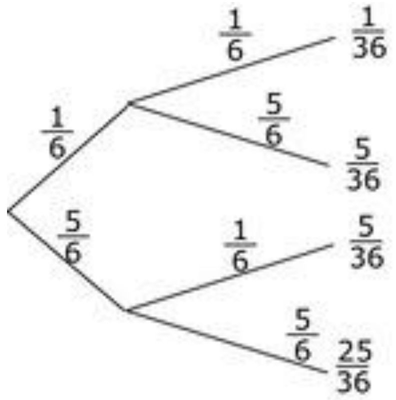
d) (i), (iv), (v)

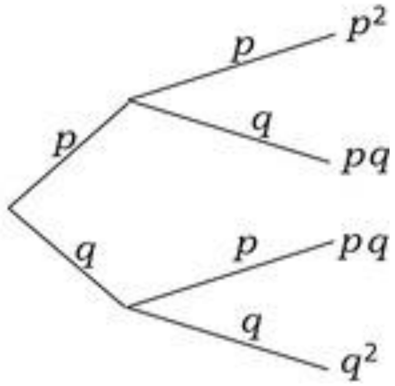
The Answer____________________

5. Statistical inference: the sample proportion and the binomial distribution

When a die is thrown twice, you can get a six or not get a six. The probability of getting a six is independent of whether you get a six or not on the previous throw. The probability of getting the six remains unchanged for each throw.

Look at the tree diagram below. It shows the outcomes of throwing a die twice and the probability of scoring a 6 and not scoring a 6.





$p = 1/6$

$q = 5/6$

Probability (6, 6) = p^2 = 1/36

Probability (one 6) = 2*pq* = 10/36 = 5/18

Probability (no 6) = q^2 = 25/36

Look at the expansion $(p + q)^2$.

$(p + q)^2 = p^2 + 2pq + q^2$

The probabilities are terms in the binomial expansion.

The binomial distribution describes the behaviour of a count variable *x* if the following conditions apply:

(i) The number of observations *n* is fixed. Here, the die is thrown twice.

(ii) Each observation is independent. The probability stays the same for each throw, unaffected by previous outcomes.

(iii) Each observation represents one of two outcomes - success *p* or failure *q*.

(iv) The probability of success, *p* is the same for each outcome. Here, the probability of getting a six is always 1/6.

Throwing a die or flipping a coin are good examples. The binomial distribution can also be applied to manufactured objects; these may be faulty or not faulty.

Suppose a die is thrown 4 times.

Choose all correct statements.

(i) The binomial expansion that represents this is

$p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

(ii) The binomial sum $p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4 = 1$

(iii) If p = the probability of a 6 and q = probability of not getting a 6 then the probability of getting at least 3 sixes = $1 - (6p^2q^2 + 4pq^3 + q^4)$

a) (i), (ii)

b) (i) only

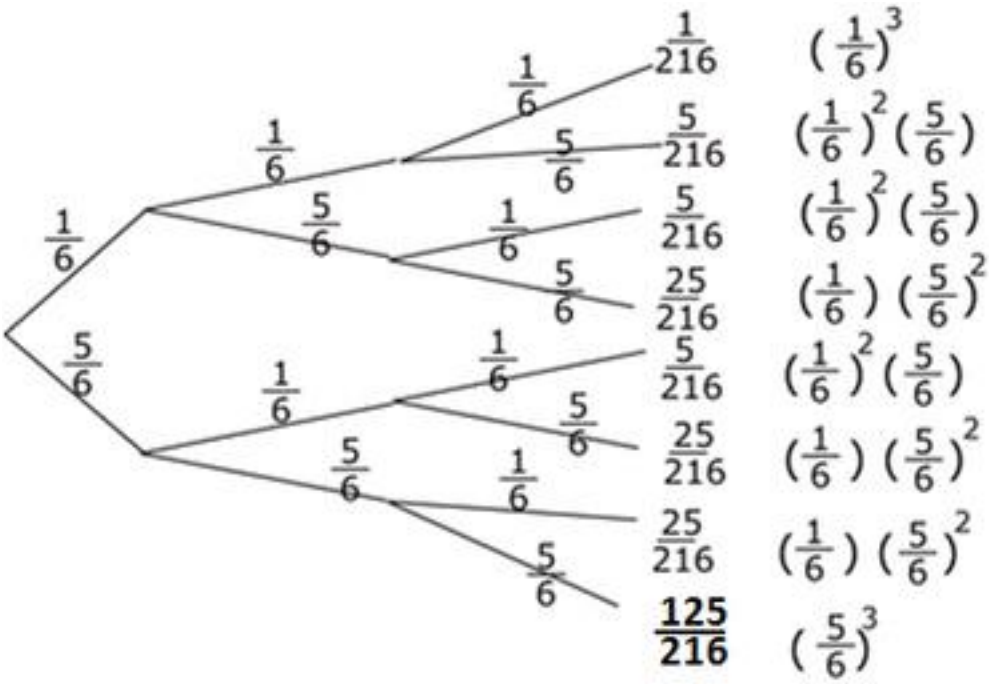
c) (i), (ii), (iii)

d) i), (iii)

The Answer________________

6. Statistical inference: the sample proportion and the binomial distribution

Suppose you define success as a six when you throw a die. You throw the die 3 times.



Using binomial coefficients you can write:

The probability of throwing three sixes = $P(3) = C_3^3 (\frac{1}{6})^3 (\frac{5}{6})^0$

The probability of throwing two sixes = $P(2) = 3\, C_2^3 (\frac{1}{6})^2 (\frac{5}{6})^1$

The probability of throwing one six = $P(1) = 3\, C_1^3 (\frac{1}{6})^1 (\frac{5}{6})^2$

The probability of throwing no sixes = $P(0) = C_0^3 (\frac{1}{6})^0 (\frac{5}{6})^3$

The binomial distribution is the discrete probability distribution of the number of successes in a sequence of *n* independent yes/no experiments, each of which yields success with probability *p*. The binomial distribution is frequently used to model the number of successes in a sample of size *n* drawn with replacement from a population of size *N*. In the example of throwing the die, the proportion of successes, *p* is equal to 1/6 and *n* is equal to 3.
Remember: the binomial distribution describes the behaviour of a count variable *x* if the following conditions apply:

(a) The number of observations *n* is fixed.

(b) Each observation is independent.

(c) Each observation represents one of two outcomes: success *p* or failure *q*.

(d) The probability of success, *p* is the same for each outcome.

Choose which of the following the binomial distribution can be applied to.

(i) A bag contains 10 blue cubes and 15 yellow cubes. You take out 3 cubes one at a time and **do not** replace each one before removing the next cube. The variable is the number of yellow cubes taken out.

(ii) A bag contains 10 blue cubes and 15 yellow cubes. You take out 3 cubes one at a time and then replace each cube before removing the next. The variable is the number of yellow cubes taken out.

(iii) A container holds 10000 bolts and it is known that 5% are defective. A sample of 20 bolts is removed. The variable is the number of faulty bolts.

a) (i), (ii), (iii)

b) (iii) only

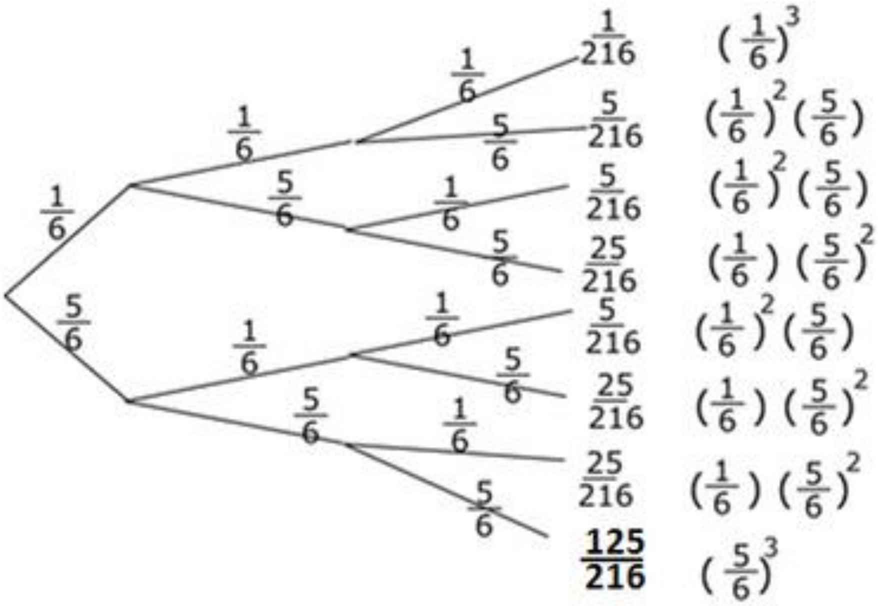
c) (ii) only

d) (i), (iii)

The Answer________________

7. Statistical inference: the sample proportion and the binomial distribution

Suppose you define success as a six when you throw a die. You throw the die 3 times.



Using binomial coefficients you can write:

The probability of throwing three sixes = $P(3) = C_3^3(\frac{1}{6})^3 (\frac{5}{6})^0$

The probability of throwing two sixes = $P(2) = 3\,C_2^3(\frac{1}{6})^2 (\frac{5}{6})^1$

The probability of throwing one six = $P(1) = 3\,C_1^3(\frac{1}{6})^1 (\frac{5}{6})^2$

The probability of throwing no sixes = $P(0) = C_0^3 (\frac{1}{6})^0 (\frac{5}{6})^3$

Note: $C_r^n = \frac{n!}{r!\,(n-r)!}$

For n trials with r successes and $n - r$ failures

$P(x = r) = C_r^n p^r q^{n-r}$ where $q = 1 - p$ and $r = 0, 1, 2, 3, \ldots n$

For example,

$$P(2) = 3C_2^3(\tfrac{1}{6})^2 (\tfrac{5}{6})^1 = \frac{3!}{2!\,(3-2)!} (\tfrac{1}{6})^2 (\tfrac{5}{6})^1 \approx 0.0694$$

Suppose you throw the die 20 times. Calculate the probability of throwing 4 sixes. Write your answer, correct to four decimal places.

Answer______________________

8. Statistical inference: simulation

Think of this example. Before you send someone into space, to a space station or to the moon, ($450 million per launch, each space suit can cost as much as $16 million, millions spent on salary and training of the astronaut) it makes sense to test your equipment for its ability to withstand the stress of the journey. If it doesn't then you have wasted all those millions and lost the astronaut whose expensive training cannot be used for a second trip.

Simulation is used to test the equipment. All the pieces of equipment, the strength of the materials used to construct them, the way they interact together, all these things can be modelled by constructing mathematical equations to represent them. These Mathematical equations can then be entered into a computer to create a model which can be tested simulating various events and likely stresses. This gives scientists an idea of the probability that equipment will survive or fail.

The model must be tested using data that represents the likely conditions experienced in space. Scientists would check the results for anomalies. The simulation's usefulness depends on the ability of the people running the simulations to include all the possible events and stresses the equipment is likely to experience.

Simulation has been used extensively since computers became sufficiently developed. You may like to read about the Monte Carlo methods that rely on repeated random sampling to compute their results. These methods were used in the development of nuclear physics and similar methods are still used today. They rely on the ability of computers to run experiments with simulated data that is too much for even a large team of people to handle. Computers are also getting faster and faster so that the results are known relatively quickly.

Consider this very simple example, the tossing of a coin. You can run this using the random number generator in Excel.

Note! The random numbers generated are not really random; if you know the formula used and the starting point, you can predict what the next number will be. They really should be called pseudo random numbers. But such generators are vital today. They are used, for example, in encryption codes and security code keys. A security code key links into the computer in a building and if its code agrees with the computer, then the person is allowed into the building. The computer and key must be exactly synchronised.

Now, the probability of throwing a head should be 0.5. But what if you simulate this?

Open an excel document.

(a) In column A1, type **Number**

(b) In A2 to A21, put the numbers between 1 and 20.

(c) In B1, type **Random**

(d) In B2, type = RAND() then click on the green tick. A number appears.

(e) Drag down to fill the other cells beside column A.

(f) In C1 type **Heads = 1**

(g) In C2 type = IF(B2<0.5,0,1). Click on the green tick and 1 or 0 should appear.

(h) Drag down to fill the other cells.

Count the number of heads. This is the number of 1s in column C. Is it exactly half the number of results?

Try extending the experiment to say, 100 or 200 by dragging down.

Choose all correct statements.

(i) Even with a small number of experiments, the number of heads is equal to the number of tails.

(ii) Using Excel to simulate the results is faster than actually flipping a coin and counting the results.

(iii) With a large number of experiments, the proportion of heads is usually closer to 0.5.

(iv) After a large number of heads, it is more likely that a tail will occur.

a) (i), (ii), (iv)

b) ii), (iii)

c) (i), (iv)

d) ii), (iv)

Answer______________________

9. Statistical inference: mean, variance and standard error of the sample proportion

Statisticians use their sample statistics to draw inferences about population parameters.

Remember: parameters refer to the entire population. A statistic refers to a sample.

While parameters do not vary, statistics differ from sample to sample. To calculate how the sample statistic differs from the population parameter, you must know the value of the population parameter. Then you can calculate the standard deviation of the statistic.

If you do not know the population parameters, and mostly you do not, then you calculate the standard error. This uses information from the sample or samples and is called standard error to make it clear that it does not use the population parameter.

The aim of gathering sample statistics is to get an idea of the population parameter, but you need to have some idea of the accuracy of your statistic. The standard deviation or if you do not have a population parameter, the standard error, helps give you an idea of the accuracy.

Note that Greek letters are used for the population mean, variance and standard deviation. Roman letters are used for the sample mean, variance and standard deviation. Upper case letters are used for the population under study. Lower case letters are used for the sample.

If you take a sample of the population, the proportion of successes (whatever this is defined as) is represented by *f*. The expectation is you will have *nf/n* number of successes in your sample. The expected value of success or the mean success rate is *f*. This refers to the **SAMPLE**.
The variance of the **SAMPLE** is given by the formula *f(1-f)/n*.

The standard error = $\sqrt{f(1-f)/n}$.

If you measure the mean of multiple samples from a certain population, their means will differ and spread out in a distribution that is approximately a normal, bell-shaped curve. The standard error will give you an idea of the accuracy of the sample mean.

Choose all correct statements.

(i) The aim of taking a sample is to more easily get information about the population parameter.

(ii) The mean of a number of sample means will give you a better idea of the population parameter mean than one single sample.

(iii) It does not matter what size sample you take from a population.

(iv) The standard deviation of the sample statistics will give an idea of the accuracy of the sample mean if you do not know the population parameters.

(v) The standard error is calculated from information gained from the sample.

a) (ii), (iv), (v)

b) (i), (ii), (iii)

c) (i), (ii), (v)

d) (i), (iii), (iv)

Answer________________

10. Statistical inference: the approximate normality of the sampling distribution *f*

You are carrying out a study of a population where the results are in the form of success or failure. It could be a survey of voters for a particular candidate, the flipping of a coin where success is defined as heads, checking a production line for defective goods. You take a sample from the population and determine its mean.

If you are working with binomial random variables, the mean is $(nf)/n$ or f, the sample proportion. If you take multiple samples, then the mean of each sample will be different.

With larger samples, the distribution of the means of the samples will approximate a **normal distribution**. This is the central limit theorem, that the distribution of sample means will approximate a normal distribution.
You can graph the binomial distribution using Excel. Suppose for your experiment you have $f = 0.25$ and the number of trials 5.

(a) Open an Excel document. In cell A1, type in Trials.

(b) In cell B1, type Probability.

(c) In cell A2 type 0 then in cell A3 type = A2+1. Then drag down until you have the numbers 0 to 5.

(d) In cell B2, type = BINOMDIST. A window appears.

Number_s you type A2
Trials you type A7 (this is *n* in the binomial expansion)
Probability_s you type 0.25 (this is the proportion of successes, *f*)
Cumulative you type FALSE since you want the probability density function, not cumulative probability

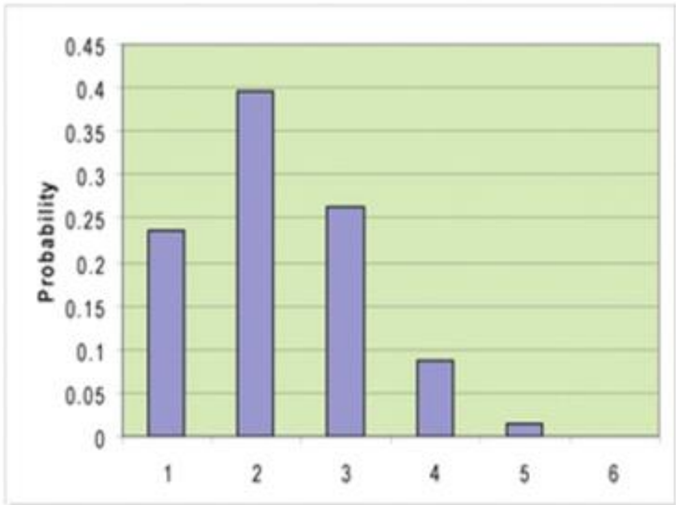
(e) Click on OK.

(f) Drag down to highlight the cells next to column A and click on *fill down*.

Trials	Probability
0	0.237305
1	0.395508
2	0.263672
3	0.087891
4	0.014648
5	0.000977

Note that when you sum the probabilities, the sum is 1, as you would expect.

To graph, highlight the probability numbers and click on the graph symbol. Follow the directions. Your graph should look like this:



This is not the normal bell-shaped curve. Experiment by increasing the number of trials. Remember to change **Trials** where you typed A7 to the cell showing the number of trials Include the $ signs to keep this constant when you fill down.

Look at the following diagrams. Which do you think shows the largest number of trials for the binomial distribution?

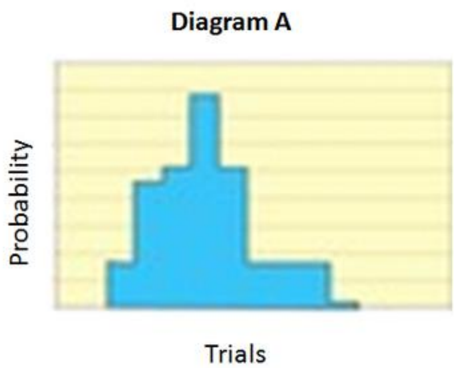
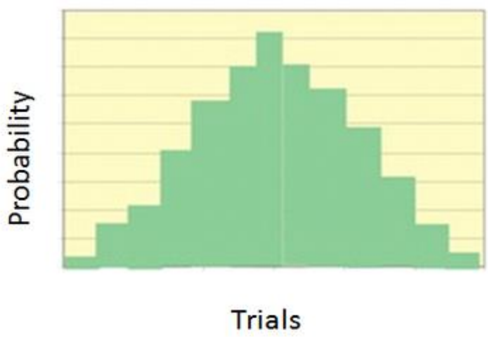


Diagram B

Diagram C

Diagram D



The Answer___________________

The Answers.

Hey! No peeking until you've finished…

Question 1

Answer: b) (ii), (iv)

(i) From a census of the entire Australian population, it was found that 99% of Australian homes had a television. This is statistical inference.

Incorrect. You do not need to make inferences or estimates if you have data on the entire population. This is a parameter.

(ii) It was found that 65% of students at Newcastle University were supported by their parents. From this it was concluded that 65% of Australian university students were supported by their parents. This is statistical inference.

Correct. The information used from the sample is generalised about the entire population of university students in Australia.

(iii) A study was done of nesting blue wrens in Lane Cove National park and found that the average number of eggs laid was three. This average is called a parameter.

Incorrect. Blue wrens nest in many places in Australia, so this is a sample of the total population. A parameter refers to data from the population. This average is the sample average and is called a statistic.

(iv) From a census of the entire Australian population, it was found that the average number of cars per household was 1.65. This is a parameter.

Correct. This average is a parameter because it refers to the population as a whole and is not derived from a sample of that population.

Question 2

Answer: d) (ii), (iii), (iv), (v)

(i) The sample statistic f = 0.55 tells Kristie exactly what the probability of success in the total population is.

Incorrect. The sample might be the same as the parameter for the entire population but it is more likely to be different from it to a greater or lesser extent, depending on whether the sample represents the total population well or not.

(ii) The sample statistic f = 0.55 gives Kristie an idea of what the probability of success in the total population is.

Correct. The sample statistic will hopefully be not too far from the parameter for the entire population.

(iii) If Kristie wanted to get a better idea of the probability of success, she would need to take many samples and average the result.

Correct. The average would be likely to give a much better idea of the parameter for the entire population.

(iv) The number of successes in the total population is a parameter and this does not vary.

Correct. The problem is that the population might be too large or the difficulties of measurement too great to actually calculate it. This is why it must be estimated from samples.

(v) If $f = x/n$ is a statistic from samples, it can vary from sample to sample.

Correct. Suppose you took an infinite number of samples and plotted the statistics. The result would be a bell curve where most of the sample statistics would cluster close to the parameter for the entire population but a small number of statistics would be further away, giving the bell-shaped curve.

Question 3

Answer: a) (ii), (iii), (iv)

(i) Cindy has identified the population she is studying correctly.

Incorrect. It is unlikely that all of the year nine students would be bus travellers. Cindy should have defined her population more carefully and only chosen bus travellers for her sample.

(ii) Cindy should have identified students who travelled on the bus and chosen her sample from this population.

Correct.

(iii) Any statistic Cindy calculates from this sample is likely to have sampling error.

Correct. Students who do not catch the bus are likely to give different answers to students who rely on the bus.

(iv) Cindy's sample excludes older and younger students whose response to the rise in bus fares is likely to be different.

Correct. Older students are likely to be more independent but younger students cannot drive their own cars.

Question 4

Answer: a) (ii), (iii), (iv), (v)

(i) Elli's proportion gives a better idea of the population proportion of the school.

Incorrect. Elli only studied 109 students whereas Yan studied 235.

(ii) Yan's proportion gives a better idea of the population proportion of the school.

Correct. The larger the sample, the more accurate the results are likely to be.

(iii) The proportions give you an idea of the probability of selecting a blue-eyed student if a student is selected at random from the school's population.

Correct. Proportion is another way of looking at probability.

(iv) From these proportions you can make the inference that there are not many students with blue eyes in the school.

Correct. In both samples, the proportion of students with blue eyes is small.

(v) The random variable represented by f varies from sample to sample.

Correct. It is a random variable.

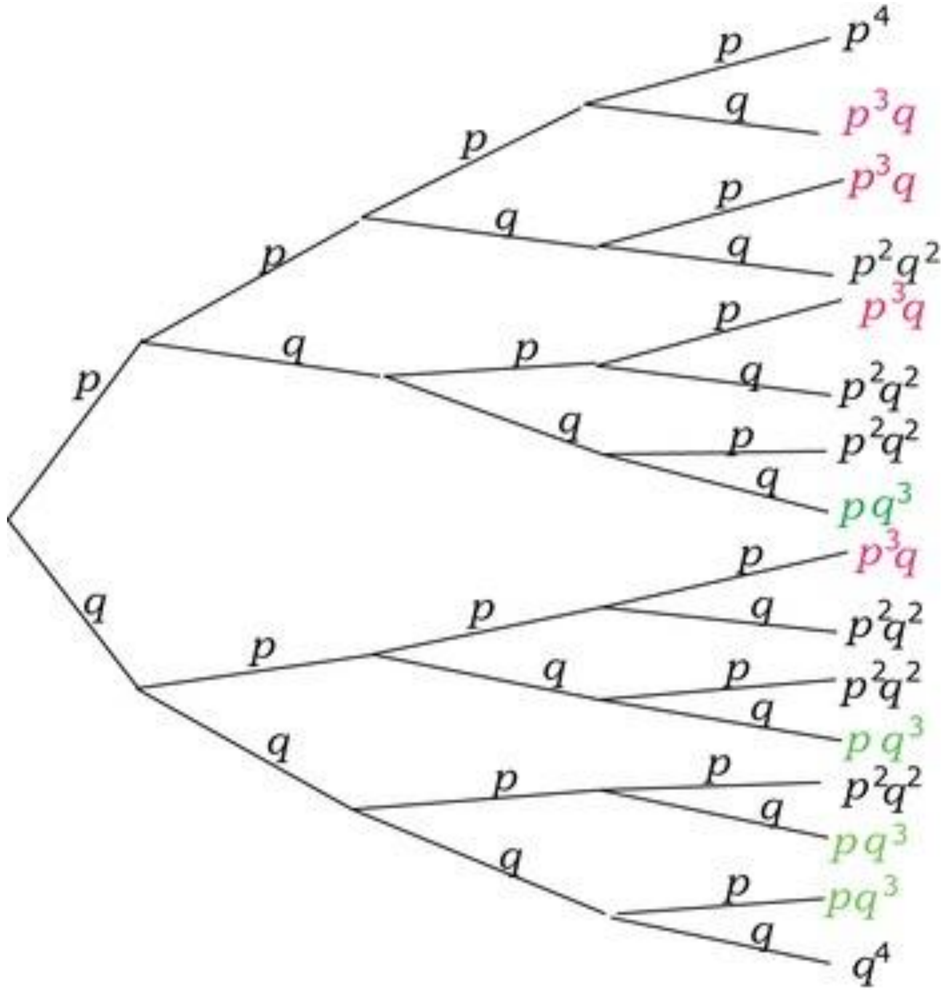
Question 5

Answer: c) (i), (ii), (iii)

(i) The binomial expansion that represents this is

$p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

Correct. Look at the tree diagram and count the like terms.



$p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

(ii) The binomial sum $p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4 = 1$

Correct. The sum of the probabilities is 1, no matter how many experiments are carried out.

(iii) If p = the probability of a 6 and q = probability of not getting a 6 then the probability of getting at least 3 sixes = $1 - (6p^2q^2 + 4pq^3 + q^4)$

Correct.

Question 6

Answer: c) (ii) only

(i) A bag contains 10 blue cubes and 15 yellow cubes. You take out 3 cubes one at a time and **do not** replace each one before removing the next cube. The variable is the number of yellow cubes taken out.

If you remove a cube and do not replace it, this affects the probability of the remaining selections. This means the observations are not independent. It is not part of a binomial distribution.

(ii) A bag contains 10 blue cubes and 15 yellow cubes. You take out 3 cubes one at a time and then replace each cube before removing the next. The variable is the number of yellow cubes taken out.

If you replace the cubes after removing them, the probability of selecting a yellow cube remains unchanged and the results are independent. The binomial distribution applies to this experiment since there is a repetition of independent events.

(iii) A container holds 10000 bolts and it is known that 5% are defective. A sample of 20 bolts is removed without replacement. The variable is the number of faulty bolts.

The binomial distribution does not apply since the bolts are not replaced.

Question 7

Answer: 0.2022

$$P(X = r) = C_r^n p^r q^{n-r} \quad \text{where } n = 20 \;\; p = \tfrac{1}{6} \;\text{ and } q = \tfrac{5}{6} \;\text{ and } r = 4$$

$$P(4) = C_4^{20}\left(\tfrac{1}{6}\right)^4 \left(\tfrac{5}{6}\right)^{16} = \frac{20!}{4!\,(20-4)!}\left(\tfrac{1}{6}\right)^4 \left(\tfrac{5}{6}\right)^{16} \approx 0.2022$$

Question 8

Answer: b) (ii), (iii)

(i) Even with a small number of experiments, the number of heads is equal to the number of tails.

Incorrect. This is unlikely. Several short trials should show this is untrue.

(ii) Using Excel to simulate the results is faster than actually flipping a coin and counting the results.

Question 9

Answer: c) (i), (ii), (v)

(i) The aim of taking a sample is to more easily get information about the population parameter.

Correct.

(ii) The mean of a number of sample means will give you a better idea of the population parameter mean than one single sample.

Correct. You can treat these as any variable and plot them to give a bell curve.

(iii) It does not matter what size sample you take from a population.

Incorrect. A larger sample will usually give you results that are closer to the population parameter.

(iv) The standard deviation of the sample statistics will give an idea of the accuracy of the sample mean if you do not know the population parameters.

Incorrect. The standard deviation can only be calculated if you know a population parameter.

(v) The standard error is calculated from information gained from the sample.

Correct. The standard error is calculated when you do not know the population parameters.

Question 10

Answer: b) The bell-shaped curve of the normal distribution would best fit on diagram D.