Beckfoot

Maths	AS Statistics	Year 12

enjoy learn succeed

Collecting Data

1	**Know the sampling techniques** You need to know the definition of the term simple random sample. You also need to understand the uses of different types of sampling such as opportunity sampling, systematic sampling, stratified sampling, quota sampling, cluster sampling and self-selected samples.
2	**Be aware of sources of bias** Bias may be introduced through an inappropriate sampling method (for example, one which is likely to exclude certain groups of people) or through the method of data collection (for example, through biased questions or through questioning in situations in which truthful answers may not be given).
3	**Remember that different samples may lead to different conclusions** For example, if you use the mean and variance of a sample to estimate the mean and variance of the population, different samples will give different results.

Data Presentation (single variable)

1	**Choose the right diagram for the data** When working with a large data set in spreadsheet form, you can generate all sorts of diagrams quickly and easily. However, this does not mean that all diagrams are appropriate for the data you are working with.
2	**Remember that the vertical axis for a histogram is frequency density** In a histogram, the frequency represented by each bar is found from the area of the bar, not the height. So for a particular class, frequency = class width x frequency density.
3	**Make sure that you choose a sensible number of classes when grouping data** This depends on how much data you have to group
4	**Make sure you know the difference between measures of central tendency and measures of spread** Measures of central tendency (averages) include the mean, median, and mode.You might use these to compare, for example, the heights of two sets of children to see which group on average are taller. Measures of spread include the range, interquartile range, variance and standard deviation. You might use these to compare, for example, the heights of two sets of children to see which group had a greater variation in height.
5	**Know how to use your calculator to work out statistical measures** Your calculator should have functions for working out the mean and standard deviation of a set of data. Make sure you know how to input the data and how to carry out the calculations.

Data Presentation (bivariate)

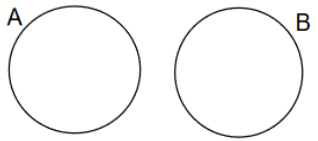
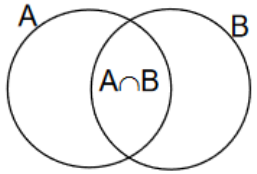
1	**Remember that correlation does not imply causation** If there is correlation between two sets of variables, it may be the case that one variable causes the other, but this is not necessarily the case. For example, a third variable might affect both variables.
2	**Look out for outliers** Just as an outlier can distort the mean in single variable data, an outlier in bivariate data can distort the value of the correlation coefficient. As with single variable data, you should consider whether the outlier could be an error, and whether it should be removed from the data.

Beckfoot

Maths	AS Statistics	Year 12

enjoy learn succeed

Probability

1	**Make sure your answer is sensible** If you get a probability which is greater than 1, you **must** have made an error! When using a tree diagram, it's useful to check that the probabilities of all the possible answers add up to 1
2	**Use the correct notation to help you explain your answers** Correct notation will make your life easier and will help an examiner to give you the marks you deserve
3	**Remember that you can only add probabilities to find P(A ∪ B) if the events are mutually exclusive** For example, if you want to find the probability that a student chosen at random studies either Maths or English, you need to take into account that some students might study both Maths and English – these events are not mutually exclusive .P(A ∪ B) = P(A)+P(B) only if A and B are mutually exclusive events.  P(A ∪ B) = P(A) + P(B) - P(A ∩ B) ALWAYS because when A and B are mutually exclusive, P(A∩B) = 0 
4	**Remember that you can only multiply probabilities to find P(A ∩ B) if the events are independent** To take a very simple example, suppose you want to find the probability that when you throw one dice you get a number that is both an odd number and a prime number. P(odd) ½ and P(prime) ½ (2, 3 and 5 are prime) but P(odd and prime) is not ½ x ½ = ¼ as there are two odd prime numbers, 3 and 5, so the probability is 2/6 = 1/3 . A number being odd, and a number being prime, are not independent.
5	**Make sure you understand clearly the difference between P(A ∩ B) and P(A ∪ B)** P(A ∩ B) is the probability that both event A **and** event B occur. It is equal to 0 for mutually exclusive events. P(A ∪ B) is the probability that event A **or** event B **or both** events A and B occur.
6	**Recognise there may be several different methods of solving a probability question** Think about whether using a sample space diagram, a Venn diagram or a tree diagram might be helpful
7	**Read the question carefully to ensure you have answered the correct problem** For example, there is a difference between the event of a train being late once in two journeys and the event being late on the first journey and not on the second

Probability Distributions

1	**Always check that the probabilities add up to 1** If they don't, you have made a mistake in your calculations, or there are other possible outcomes which you have not included.
2	**Illustrate the data using a vertical line chart** A vertical line chart is more appropriate than a bar chart for discrete data like this.
3	**Use tables to list results where appropriate** This ensures you include all possible outcomes

Beckfoot

Maths | AS Statistics | Year 12

enjoy learn succeed

Binomial Distribution

1	**Read questions very carefully** Be careful with the wording in the question: many students make careless errors misinterpreting the question. Make sure that you recognise the difference between "more than" and "at least": e.g. "more than 10" means "at least 11" – think about it!
2	**Use the correct notation and define your variables** The correct notation is important to help you explain your answers. • Define your variable X clearly at the start of the question. • Write down values of n and p clearly. • Show what you are trying to calculate: this is much better than a list of fractions and decimals that appears to anybody else randomly ordered! e.g. $P(X> 1) = 1 –(P(X= 0) + P(X= 1))$
3	**Remember to use the binomial coefficient** Make sure you include the binomial coefficient ${}_nC_r$ in your method
4	**Take care when finding the probability for a range of values** Be careful with inequalities – write out your working carefully . e.g. $P(X \geq 4) = 1 – P(X \leq 3)$ $= 1 – 0.6477$ $= 0.3523$
5	**Check the conditions before using the binomial distribution** Obviously not all probability questions can be solved using the binomial distribution. If you are in doubt about whether you should be using the binomial distribution check the conditions: • random samples of a fixed size, n • the probability of success, denoted by p, is constant (hence q= 1 – p is also constant). • the trials are independent
6	**Take care with accuracy when using decimals** If using decimals, work to at least 3 significant figures. If possible use exact numbers until the end of the calculation.

Hypothesis Testing

1	**Use the correct notation for stating hypotheses** Set up the hypothesis test carefully, using the correct notation. First state the definition of p. There is often a mark given for this. e.g. Let p be the probability of getting a head. H_0: p = ½ H_1: p < ½ [NOT H_0= 0.2, or P(X= 0.2)]
2	**Remember to test a region of probabilities** Always work out a region of probabilities (a tail), rather than a point. $P(X \leq 3)$ not $P(X = 3)$, for example.
3	**Use the alternative hypothesis for deciding the region** Use the alternative hypothesis to help you decide on the region. If H_1: p < ½ then you will calculate $P(X \leq r)$ If H_1: p > ½ then you will calculate $P(X \geq r)$
4	**Show your calculations clearly** Show clearly what you are trying to calculate. This is much better than a list of fractions and decimals that appears to anybody else randomly ordered! e.g. $P(X \geq 4) = 1 – P(X \leq 3)$ $= 1 – 0.6477$ $= 0.3523$
5	**Make sure that you compare the probability with the significance level** You must do this explicitly. For example, $P(X \leq 2) < 0.05$, so reject H_0.
6	**Once you have accepted/ rejected H_0, give your conclusion in words** Decide whether you are to accept or reject H_0 but then put a final conclusion in words, answering what was requested in the question. Do not state that this "proves" anything but use wording like "the evidence suggests that..."
7	**Use at least 3 significant figures when using decimals** If using decimals work to at least 3 significant figures, to avoid rounding errors. Where possible use exact numbers until the end of the calculation.
8	**Make sure that you compare the observed value with the critical region** You must do this explicitly. For example "The critical region is $X \leq 2$. The observed value is 3, which does not lie in the critical region."
9	**Read the question carefully** Always read the wording in the question carefully. Do not be put off by the data, which may indicate an increase or decrease. In practice we will be setting up the hypothesis test before we have collected the data.Make sure you do not get fooled by the question! Read it carefully to determine whether it is a 1-tail or 2-tail test.