

Statistics Statements

True? False? Or Sometimes True?

Teachers Notes and Solutions:

A – True

Primary data is data which has been collected by me – *or on my behalf* – specifically for the purpose of my investigation. Data that I find on the internet or am given by someone else is called secondary data.

B – False

Although first impressions suggest that the google section is the largest, this is due to the warped perspective of the 3D chart and perfectly illustrates the need to avoid 3D graphs when trying to accurately illustrate information graphically.

C – False

This could provoke some interesting discussion! There are instances where the mean is the 'best' average, but this is not because it is calculated. The mean is different from the other averages in that it uses every data point – but this can also skew results by the inclusion of extreme values.

D – False

By the nature of the word average, it is impossible for everyone to end up with an above average score! The average always gives some indication of the mid-value, so there will always be approximately half of the group above this and approximately half below. If you used the same test twice, a good second test might be one where everyone scored above the average of the first.

E – False

This is a common misconception. Probabilities are not simply calculated on the number of options (in this case 2) but also on the likelihood of those options occurring. For example if it has rained today it may increase the likelihood that it will rain tomorrow. Similarly the likelihood of it raining tomorrow on the 4th April may be significantly higher than the likelihood of it raining tomorrow in June.

F – True

The relative frequencies of the events can be used to estimate the probability of each outcome occurring:

$$P(\text{Home win}) = 35/(35+5+10) = 35/50 = 0.7$$

$$P(\text{Away win}) = 5/(35+5+10) = 5/50 = 0.1$$

$$P(\text{Draw}) = 10/(35+5+10) = 10/50 = 0.2$$

Thus from the data the probability of a home win is estimated at 0.7 and not 0.33.

G – True

The probability of getting a head when I throw a coin is 1 (there is one head) out of 2 options (head or tails). These are both equally likely outcomes (for a fair coin) thus $P(\text{head})=1/2=0.5$. Similarly, the probability of getting a 6 on a dice is $1/6$ or 0.17. Since 0.5 is bigger than 0.17, it can be said that it is true as the probability is greater.

H – False

105 and 135 mark the extremes of the 'box' in the box and whisker plot displayed. Boxes are drawn such that 50% of the data is contained in the 'box' thus 50% and not 75% of the students weigh between 105 and 135kg.

I – False

The line within the box on the box and whisker diagram represents the median of the data, thus it can be said that the median is 60g and not the mean. A more thorough answer recognises that the distribution of weights is skewed, thus the mean will be less than the median ie. less than 60g.

J – False

The pie charts show that a higher *proportion* of respondents like hip hop in 2008 as compared with 2002. Since the number of people represented by the graphs is not shown, it is not possible to say there are more people who like it in 2008 – it could be that there were less people surveyed and so the actual number of people represented by the hip hop section is fewer than in 2002.

K – False

The number 2 is repeated in the original set of numbers – but not in the calculation of the median. It should read:

Ordered numbers: 2,2,3,5,7,10

Finding the middle: 2,2,3,5,7,10

Median = $(3+5)/2 = 4$

L – Sometimes True

The pie chart is constructed using the same proportions as the data in the bar graph – but it could also represent a much larger or smaller data set given the same proportions. Thus in this case, they are probably taken from the same data set but it is not necessarily true that they always are.

M – False

The way a sample is chosen should always be considered so that it is representative of the population under consideration. For example, if I was researching shopping habits and stood outside Tesco asking people where they shop even if I asked a thousand people, the sample would be biased toward people that shop at tesco. In this case I would be better to ask fewer people, but selected properly at random.

N – False

In some cases these will be the same but it is not usual or by any means always the case. [In fact the values are the same only if you have a perfectly symmetrical distribution.]

O – True

The graph shows that as x increases, y decreases, thus the correlation between the variables is negative.

P – False

The possible combinations when throwing 2 dice are:

(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)

(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6)

(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)

The combination (6,6) occurs only once in the 36 options – thus the likelihood of it occurring is 1/36. This is the same as the likelihood of getting any double number. All other combinations appear twice – for example – (1,2) and (2,1) so the likelihood of these is 2/36 or 1/18 which is more likely than getting a double.