Examiners' Report/
Principal Examiner Feedback

Summer 2010

GCE

Statistics S1 (6683)

Edexcel Limited. Registered in England and Wales No. 4496750
Registered Office: One90 High Holborn, London WC1V 7BH

A PEARSON COMPANY

If you have any subject specific questions about the content of this Examiners' Report that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

Ask The Expert can be accessed online at the following link:

http://www.edexcel.com/Aboutus/contact-us/

Summer 2010

Publications Code UA024764

Statistics Unit S1
Specification 6683

Introduction

In general the candidates found the paper accessible and there were some excellent scripts. Areas of the paper causing the most difficulty were histograms, the interpretation of results in terms of the question set, calculation of the standard deviation and the normal distribution. Q7 was challenging for many candidates.

A considerable number of candidates failed to consider the accuracy of their answers. Many candidates rounded to 2 significant figures in Q1 and again in the regression equation in Q6, where rounding off lost accuracy in the final answer.

As in previous years, calculation of the standard deviation continues to defeat a considerable number of candidates in spite of the availability of calculators to check the answer. Also, comments still tend to be verbose and interpretation weak. Comments are expected to be in the context of the question and, at this level, the use of statistical terms and reference to the quantities calculated in the question are required.

Report on individual questions

Question 1

Typically candidates successfully used the correct formula in order to calculate the product moment correlation coefficient in part (a). However, a number of candidates lost the accuracy mark by only giving a rounded answer to 2 decimal places. Providing an interpretation of their value of the correlation coefficient was less straightforward. Most frequently candidates made general remarks and described the correlation as positive without relating this to the context of the question. Of those who did attempt to provide an interpretation, many failed to appreciate that it was the attendance at the matches being compared to the total number of goals scored and not the number of home matches that were played.

Part (c) was answered well overall and correct answers were often justified by accompanying statements which indicated that linear coding does not affect the product moment correlation coefficient. Some candidates, however, seemed unaware of this fact and a common mistake was to divide their original product moment correlation coefficient by 100. In addition many candidates failed to recognise the significance of them being asked to write down their answer and chose to perform a full calculation in order to obtain the product moment correlation coefficient, which sometimes led to processing errors.

Question 2

Overall there were very few errors made when candidates completed their tree diagrams. A small number of candidates repeated their probabilities of 2/3 (for obtaining a head) and 1/3 (for obtaining a tail) on the second branches for the fair coin. Occasionally the 5/12 and 7/12 probabilities were placed on the wrong branches and, in a few instances, quantities rather than probabilities were used. The vast majority of candidates were able to calculate the probability that Shivani selects a head correctly, or at least follow through the correct method from their tree diagrams, with few errors seen.

In contrast the quality of candidates' attempts at part (c) was extremely varied. Very few candidates quoted the correct formula despite it being given in the formula booklet, and of those who did, few realised that the numerator should be 5/12 x 2/3. The numerator was quite often seen as 5/12 alone, and a number of candidates failed to recognise that their denominator should be their answer to part (b), leading in some cases to a repeated fraction in the numerator and denominator. $P(H/R)$ was sometimes calculated instead of $P(R/H)$.

The final part of the question was attempted fairly successfully overall. Indeed, many of the candidates who had erred in previous parts of the question were able to gain some credit, as most could identify at least one of $(5/12)^2$ or $(7/12)^2$. The special case pertaining to no replacement was occasionally seen.

Question 3

Finding the correct value of a in the first part of the question proved to be relatively straightforward for most candidates. Few errors were seen although some candidates provided very little in the way of working out and did not always make it explicit that they were using the fact that the sum of the probabilities equals one. Similarly, most candidates were able to obtain the correct value of $E(X)$, though not many deduced this fact by recognising the symmetry of the distribution.

The majority opted to use the formula to calculate $E(X)$, which resulted in processing errors in some cases. Common errors seen in calculating $Var(X)$ included forgetting to subtract $[E(X)]^2$ from $E(X^2)$ or calculating $E(X^2) – E(X)$, although on the whole the correct formula was successfully applied.

Most candidates were able to correctly apply $Var(aX+b) = a^2Var(X)$ to deduce $Var(Y) = 4Var(X)$, although $Var(Y) = 6 – 2Var(X)$ was a typical error. Quite a number of candidates attempted to calculate $E(Y^2) – [E(Y)]^2$ with varying degrees of success. Occasionally, candidates divided their results in part (b), part (c) and part (d) by 5.

The final part of the question proved to be the most challenging of all and was often either completely omitted or poorly attempted with little or no success. Only a minority of candidates knew they would need to equate $6-2X$ to X in order to obtain the corresponding values of X and of those who did, only a small number scored full marks, as candidates were generally unable to identify the correct values of X.

Question 4

Overall this question proved to be quite challenging for candidates and incorrect interpretation of the Venn diagram lost many candidates marks. In spite of this, most candidates had no trouble proving the given probability in part (a).

In part (b), however, quite a number of candidates neglected one of the four components of the numerator, usually the 3, and 11/30 was consequently an extremely common wrong answer. Other wrong answers included 9/30, 13/30 and 16/30. Some candidates chose to use the addition rule, which was generally written down correctly, although quite often P(A) was given as 4/30 and P(B) as 5/30, giving rise to P($A \cup B$) = 7/30.

In contrast, the majority of candidates were able to deduce that P($A \cap C$) = 0 and quite a few gave explanations as part of their answer, such as 'there is no overlap', or 'no intersection' and some even discussed the idea of mutual exclusivity. A small proportion of candidates had the right idea but failed to give a probability, giving their answer as 'nobody' or in a few cases 'the empty set'. However, not all of the candidates realised that mutually exclusive events have a probability of 0 of occurring together and some mistakenly thought that P($A \cap C$) equalled P(A)P(C) here.

Answers to part (d) were extremely varied. Most candidates did not recognise that a conditional probability was required and consequently did not obtain the correct denominator. Common wrong answers were 6/30, 6/20 and 3/20. A significant number attempted to perform some complex calculations in which they tried unsuccessfully to use the formula for conditional probability. Very few candidates used the Venn diagram to calculate the probability directly.

Testing for independence was generally performed successfully overall, with the majority of candidates carrying out suitable tests. However, some candidates did find this challenging and often the wrong probabilities were compared and some incorrect probabilities were obtained. A number of candidates appeared to be confusing independence with mutual exclusivity. Some candidates merely provided a comment on the perceived nature of independence without performing any calculations at all. Of those candidates who were successful, the most common approach was to test whether P($B \cap C$) = P(B)P(C), although there were a few cases where P($A \cap C$) was compared with P(A)P(C) by mistake. Rather worryingly, a surprisingly high number of candidates failed to recognise 3/30 and 1/10 as equivalent fractions and thus concluded that the events were not independent.

Question 5

Finding the midpoints of the given groups was predominantly carried out correctly with very few errors seen. In contrast, attempts at finding the width and height of the 26 –30 group were extremely varied, with most candidates finding this particularly challenging, especially in finding the height. In the majority of cases, candidates obtained the wrong width and height, mostly with no clear strategy, although these did multiply together to make 20.8 in some cases. Calculation of the mean was carried out successfully on the whole, although there were some apparent misconceptions, with quite a few candidates merely summing the midpoints (without multiplying by the frequency) and dividing this by 56.

The standard deviation proved to be more problematic, with frequent mistakes in both the formula and in their calculations. Some candidates used the sum of the f^2x's and others the sum of the $(fx)^2$ or the sum of the fx's all squared in their formula. Very few candidates calculated s. Most candidates were able to use the correct interpolation technique to obtain the median, although many lost the accuracy mark through their use of 21 as the lower class boundary (which was relatively common) and /or 4 as the class width. Quite a few candidates worked with 28.5. A few candidates tried to apply the correct formula to the wrong class interval, however. Some candidates appeared to have a limited understanding of the class boundaries and failed to recognise the continuous nature of the data.

The majority of candidates were able to carry out a suitable test to determine the skewness of the data correctly. This mostly involved comparing $Q_3 - Q_2$ to $Q_2 - Q_1$ (with or without explicit substitutions), although the wrong conclusion was often drawn, either following on from a previous error in evaluating the median or from a lack of understanding of what their result was showing. A few students evaluated 3(mean-median)/standard deviation. Quite often the result of their test was described in words not figures, for example Q_2 is closer to Q_3 than Q_1. Some candidates merely attempted to describe the skewness without carrying out any test.

Question 6

The vast majority of candidates produced accurate scatter diagrams and on the rare occasion that there was a point missing it was predominantly point *D*. Explaining exactly why a linear regression model was appropriate proved to be difficult for candidates overall. Most candidates seemed to have the general idea but did not express this in the required terms and consequently very few earned this mark. Comments tended to be much more general about why linear regression is carried out and most talked about correlation being high without explaining that the points lie close to a line.

On the whole the correct formulae were used in calculations of S_{dd} and S_{fd}, with most candidates earning the method mark at the very least. The same was true in the calculations of b and a overall, although a common mistake was to calculate S_{ff} and go onto use that in the calculation of b. Premature approximation cost many candidates accuracy marks. Interpretations of the value of b were considerably varied, with relatively few candidates gaining this mark and some opted to omit this part altogether. Most candidates failed to relate their value to the context of the question and often tended to discuss b merely in terms of being the gradient. As a consequence, despite having the right kind of idea and correctly understanding the concept of the gradient, frequently candidates failed to gain this mark due to missing out the relevant units, mixing up the units or not quoting the actual value of b.

Very few candidates were able to formulate the correct equation with the correct units in part (f), and the majority found this particularly challenging, either omitting this part or resorting to evaluating the lines at the data points rather than equating and solving the equations. Often no clear strategy was apparent and a common mistake was to equate their equation to 5. There was clearly confusion over t and d and even out of those who were able to solve the required equation or inequality, not many found the value of t or range of t in km, as most tended to give their answer in terms of d. Occasionally the intersection point was evaluated using their graph after the lines had been plotted.

Question 7

This question proved to be quite challenging for a high proportion of candidates. A significant number either made no attempt at the question or offered very little in the way of creditable solutions, with many unable to progress beyond part (a). Time issues may have been a contributing factor in some cases.

The majority of candidates however, were able to earn some credit at least in part (a), for their standardisation, although whilst this was often completely correct, a fairly common mistake was to give 1-0.8944 = 0.1056 as their final answer.

Many students did not recognise that they needed to actually use the normal distribution in part (b) and part (c), giving rise to extremely poor attempts by numerous candidates. Of these, many merely gave 45 and 15 as their quartiles, whilst others calculated ¾ of some value as their upper quartile (for example ¾ x 60) and ¼ of the same value as their lower quartile. Alternatively, of those who understood that they were required to use the normal distribution, most attempts were successful, though there were some instances of their setting their standardisation equal to a probability, usually 0.75 or $P(Z < 0.75)$, and not a z-value. Unfortunately 0.68 was used fairly frequently as the z value. The majority of candidates were however able to follow through their value of the upper quartile to find their lower quartile using symmetry, though some performed a second calculation involving standardisation. Some candidates miscalculated their lower quartile as 1/3 of their upper quartile.

Despite previous errors most candidates tended to be successful in substituting their values correctly into at least one of the given formulae. However, a few seemed unaware of the order of the operations.

The final part of the question also proved difficult for many candidates with some running into trouble as a consequence of previous errors in part (b), part (c) and part (d) and others providing no attempt at all. Indeed, for numerous candidates, incorrect values for h and k led to probabilities of 0 being calculated from results such as $P(Z > 7)$ and thus many creditable attempts lost marks through earlier inaccuracies.

Grade Boundary Statistics

The table below give the lowest raw marks for the award of the stated uniform marks (UMS).

Module		Grade	A*	A	B	C	D	E
		Uniform marks	90	80	70	60	50	40
AS	6663 Core Mathematics C1			59	52	45	38	31
AS	6664 Core Mathematics C2			62	54	46	38	30
AS	6667 Further Pure Mathematics FP1			62	55	48	41	34
AS	6677 Mechanics M1			61	53	45	37	29
AS	6683 Statistics S1			55	48	41	35	29
AS	6689 Decision Maths D1			61	55	49	43	38
A2	6665 Core Mathematics C3		68	62	55	48	41	34
A2	6666 Core Mathematics C4		67	60	52	44	37	30
A2	6668 Further Pure Mathematics FP2		67	60	53	46	39	33
A2	6669 Further Pure Mathematics FP3		68	62	55	48	41	34
A2	6678 Mechanics M2		68	61	54	47	40	34
A2	6679 Mechanics M3		69	63	56	50	44	38
A2	6680 Mechanics M4		67	60	52	44	36	29
A2	6681 Mechanics M5		60	52	44	37	30	23
A2	6684 Statistics S2		68	62	54	46	38	31
A2	6691 Statistics S3		68	62	53	44	36	28
A2	6686 Statistics S4		68	62	54	46	38	30
A2	6690 Decision Maths D2		68	61	52	44	36	28

Grade A*

Grade A* is awarded at A level, but not AS to candidates cashing in from this Summer.

- For candidates cashing in for GCE Mathematics (9371), grade A* will be awarded to candidates who obtain an A grade overall (480 UMS or more) *and* 180 UMS or more on the total of their C3 (6665) and C4 (6666) units.
- For candidates cashing in for GCE Further Mathematics (9372), grade A* will be awarded to candidates who obtain an A grade overall (480 UMS or more) *and* 270 UMS or more on the total of their best three A2 units.
- For candidates cashing in for GCE Pure Mathematics (9373), grade A* will be awarded to candidates who obtain an A grade overall (480 UMS or more) *and* 270 UMS or more on the total of their A2 units.
- For candidates cashing in for GCE Further Mathematics (Additional) (9374), grade A* will be awarded to candidates who obtain an A grade overall (480 UMS or more) *and* 270 UMS or more on the total of their best three A2 units.

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publications@linneydirect.com
Order Code UA024764 Summer 2010

For more information on Edexcel qualifications, please visit www.edexcel.com/quals

Edexcel Limited. Registered in England and Wales no.4496750
Registered Office: One90 High Holborn, London, WC1V 7BH