

CARIBBEAN EXAMINATIONS COUNCIL

**REPORT ON CANDIDATES' WORK IN THE
CARIBBEAN ADVANCED PROFICIENCY EXAMINATION
MAY/JUNE 2007**

STATISTICAL ANALYSIS

**STATISTICAL ANALYSIS
CARIBBEAN ADVANCED PROFICIENCY EXAMINATION**

MAY/JUNE 2007

INTRODUCTION

The revised Statistical Analysis syllabus was followed this year for the third time. Two hundred and seventy candidates registered for this examination.

This is a one-unit course comprising three papers. Papers 01 and 02 were examined externally while Paper 03 was examined internally by the teacher and moderated by CXC. Contributions from Papers 01, 02 and 03 to the unit were 40 per cent, 40 per cent and 20 per cent respectively.

Paper 01 consisted of Sections 1, 2, and 3 corresponding to Modules 1, 2, and 3 respectively. There were fifteen compulsory short-response questions. Five questions were in each section worth a maximum of 40 marks. Candidates could obtain a maximum of 120 marks on this paper.

There were six extended-response (essay) questions on Paper 02 with two questions in each of the three sections. The maximum marks for each section was 40 with a total of 120 marks for the paper.

Paper 03 consisted of a project from any area of the syllabus. A candidate could obtain a maximum of 60 marks on Paper 03.

GENERAL COMMENTS

Sixty-eight per cent of the candidates obtained acceptable grades, Grades I to V. An average standard of work was seen from many candidates on this paper. Some of them appeared to be well-prepared and answered the questions competently, showing the necessary working.

COMMENTS ON CANDIDATES' PERFORMANCE ON EACH QUESTION

PAPER 01

Section A

Module 1

Question 1

This question tested candidates' ability to:

- (i) Identify the characteristics of the box and whisker plot, frequency polygon, pie chart, stem and leaf plot, bar chart and histogram
- (ii) distinguish among the properties of the mean, mode, median
- (iii) identify the shape of a distribution.

Most candidates responded well to this question. One common error was to state mode and median instead of mean and median.

Question 2

This question tested candidates' ability to distinguish between discrete and continuous variables, as well as, to determine mode, mean, median, and n^{th} percentile.

This question seemed appropriate for most candidates as the result showed that they had little difficulty in answering the question.

Question 3

In this question, candidates were required to estimate the median and mode by calculation.

This question was poorly done by the majority of candidates. Most were able to recall the formula but could not apply the knowledge. Very few candidates used the method of linear interpolation to successfully find the median.

Question 4

This question tested candidates' knowledge of various aspects knowledge on various aspects of Module 1. It was generally well done.

Question 5

This question tested candidates' ability to construct a Cumulative Frequency Curve and to estimate the interquartile range. It was generally well done. However, some candidates need practice in drawing smooth curves.

Module 2Question 6

This question tested candidates' knowledge of finding the probability and standard deviation of the binominal model. A large number of candidates were unable to identify the parameters of the Binomial – n.p. In addition, some candidates were unable to convert from percentage to decimal.

Candidates encountered problems writing the interval of 3 d" x d" 6 instead they wrote $3 < x < 6$.

Question 7

This question tested candidates' ability to construct the probability function table and to calculate $E(x)$. Many candidates were unable to grasp the idea of $F(n)$ – Cumulative Distribution Function. The Arithmetic mean was incorrectly used as $E(x)$ in a few cases.

Question 8

This question tested candidates' ability to find probabilities of independent events.

This question was poorly done. Many candidates were unable to grasp the concept of independence, thus the data were interpreted incorrectly. The few candidates who used the binominal to answer the question were successful.

Question 9

In this question, candidates were asked to solve problems involving probabilities of the normal distribution using the z-score. Many candidates rounded off too early thus ending up with the incorrect value. (Rounding off should be two to three significant figures as required by the examiner). Candidates were unable to identify variance. Many failed to answer the part of the question where a percentage was required.

Question 10

This question tested candidates' ability to use the normal approximation to calculate probabilities with given intervals. In most cases, the continuity correction was generally ignored and if done, it was incorrectly done. Most candidates were unable to identify the parameter for the normal approximation especially the standard deviation. Generally, candidates were competent in using the standard table.

Module 3

Question 11

In this question, candidates were asked to determine probabilities and degrees of freedom from both the X^2 table and the t distribution. This question was well done. Many candidates were able to read the value from the table.

Question 12

This question tested candidates' ability to identify the independent variable. This question was generally well done but a few candidates had problems identifying the independent variables.

Question 13

This question tested candidates' ability to calculate $E(\bar{x})$ and $\text{Var}(\bar{x})$. This question was reasonably well done. Some candidates were unable to determine the application of the normal distributions as it applies to the central limit theorem.

Candidates in general were unable to recall the correct formula for $\text{Var}()$.

Question 14

In this question, candidates were asked to explain the term confidence interval in the context of a population mean and to construct a 96 per cent confidence interval for proportion from a large sample. Candidates did not have an accurate concept of the meaning of confidence interval.

Many candidates used the formula for the mean rather than the population proportion.

Question 15

In this question, candidates were asked to state the null hypotheses, identify the critical region and state the conclusion of the test.

Part (a) was generally well done by candidates. In part (b), many candidates incorrectly used the z-score rather than the t -test. Part (c) was poorly done. However, there were situations where candidates were able to make a conclusion based on the erroneous use of the z-test.

PAPER 02

Module 1Question 1

This question tested candidates' understanding of:

- Population and sample
- Simple random, stratified random numbers, systematic random and quota sampling
- The use of the table of random numbers
- The process of using stratified random and systematic sampling to obtain a sample

For part (a) some candidates were unable to give a clear distinction between *population* and *sample*. These candidates were unable to clearly define “population”. An example of an appropriate response is: *A population is the entire group of items/people under investigation while a sample is a subset of the items/people drawn from a population.* Some candidates gave a geographical definition of population.

Part (b) required that candidates to **describe** one similarity and one difference between stratified random and quota sample. Candidates lost marks by **stating** a similarity or difference. For example, simply stating “*Stratified random sampling is more time-consuming than quota sampling*” was worth one of two marks. Candidates who went on to elaborate, for example “*in a stratified random sample proportionate simple random samples are taken from the subgroups of the population, while in a quota sample the only requirement is that predetermined numbers come from each subgroup*” were awarded full marks. Instructors need to demonstrate to the candidates what is required when the question asks to ‘state’, ‘describe’ or distinguish’.

For part (c) (i), a few candidates thought that the P, Q, R and S were the first letter for the sampling method. In addition, sampling method P was not clearly identified as quota sampling while method R was identified by some candidates as ‘table of random numbers and not simple random sampling’.

Part (c) (ii) was well done. However in part (c) (iii), owing to premature approximation, some candidates got 203 and not the correct answer of 205. Instructors must encourage candidates to work with the fraction if the equivalent decimal is recurring.

For part (c) (iii) some candidates lost marks for listing all 16 numbers and not the FIRST FOUR as required. Some candidates considered only the number between 01 and 16 and not 01 and 656. Candidates need more practice in using the table of random numbers to obtain a sample from a population of 100 or more.

Part (d) was well done. However a few candidates had correctly written $\frac{656}{16}$ simplified this to an incorrect number. Candidates need to pay more attention to accuracy.

Question 2

This question tested candidates' ability to:

- Read a stem-and-leaf diagram
- Determine quartiles from ungrouped data
- Calculate range and interquartile range
- Calculate trimmed mean
- Interpret the shape of a frequency distribution in terms of skewness
- Interpret the findings obtained from data.

Most candidates were able to obtain the values of p and q on part (a) (i). However some candidates had difficulty obtaining accurate answers for r , s , and t . For example, when determining the value of t some

candidates had $t = 15 \frac{3}{4}$ Value = 16^{th} value = 87. A few candidates had $t = 15^{\text{th}}$ Value = 86.75.

Also, instructors need to encourage candidates to use n for grouped data and $n+1$ for ungrouped data.

For part (a) (i) (a), a few candidates wrote Range = $Q_3 - Q_1$. Some candidates lost marks for writing range as an interval (42 to 92). For part (a) (i) (b), some candidates found semi-interquartile range instead of interquartile range. For part (a) (i) (c), many candidates identified that candidates performed better in Mathematics than in English, however they were unable to support their answer.

Many candidates confused negatively skewed with positively skewed in part (b). For part (c), some candidates discarded only one value from the lower-end and only one value from the upper end. More than half of the candidates who attempted this question obtained $x = 81$ and not the correct answer of $x = 1$.

Module 2

Question 3

This question tested candidates' knowledge of Probability Theory.

Part (a) (i) was very well done by most candidates. Throughout this question, some candidates assumed the events were independent and used the multiplication rule/law. For example, in part (ii), many candidates wrote incorrectly: $P(\text{male and cricket}) = P(\text{male}) \times P(\text{cricket})$. For part (iii), candidates recognized that the addition rule/law was to be used. However they wrote incorrectly: $P(\text{movies or female}) = P(\text{movies}) + P(\text{female}) - P(\text{movies}) \times P(\text{female})$. The same error was made in part (iv). Also in part (iv), a few candidates interpreted the questions as $P(\text{male/cricket})$.

For part (b) (i), many candidates worked out $P(A) \times P(B)$, however they did not compare the value obtained with $P(A \cap B)$ to arrive at the conclusion. For part (b) (ii), many candidates did not relate the expressions to the events and spoke in general about the sets. For example, some candidates define events as "*A prime intersect B prime*" and not "*an adult chosen at random is female and/who prefers watching motives*".

Question 4

This question tested candidates' knowledge of:

- Tree diagram
- Probability
- Continuous random variables

For part (a) (i), candidates lost marks for drawing on their tree diagram more than the required two draws of the marbles. Some candidates ignored the fact that the marbles were chosen without replacement and hence calculated incorrect probabilities. For part (a) (ii) the appropriate branch was accurately identified by most candidates. However some candidates added the probabilities instead of multiplying. Some candidates found the probability of selecting a red marble and the probability of selecting a white marble and hence gave two answers for this question. For part (a) (iii), most candidates recognized that they had to calculate the probability of getting a red marble on both draws and the probability of getting a white marble on both draws. However, after finding the probabilities, they had no idea what to do with these probabilities.

For part (b), candidates knew they had to find the area but some candidates did not equate the area with one. Many candidates had difficulty finding the area under the graph.

Module 3

Question 5

This question tested candidate's knowledge on the Chi-Squared Distribution.

For part (a), some candidates mixed up the null and the alternative hypothesis.

Part (b) was well done by most candidates. Some candidates experienced difficulty in calculating the missing values of the column E while a few candidates could not interpret $\sum \frac{O^2}{E} - N$.

In part (c), some candidates used $n-1$ to calculate the degrees of freedom. A few candidates used the wrong table to determine the critical value. Some candidates said the critical region was $\chi^2 = 5.991$. Instructors need to emphasize that the critical region is not a point.

For part (d), candidates needed to be clear regarding their conclusion. Instructors need to encourage candidates to write the three criteria for the conclusion in the context of the problem.

Question 6

This question tested candidates' knowledge of:

- Sampling Distribution
- Central Limit Theorem
- Hypothesis Testing

For part (a) (i), most candidates knew that they had to standardize, however some candidates divided by σ and not $\frac{\sigma^2}{n}$. In addition, in an attempt to standardize, some candidates did $\mu - \bar{x}$ and not $\frac{\mu - \bar{x}}{\frac{\sigma^2}{n}}$. Some candidates used a rounded off denominator in the standardization resulting in an inaccurate value. In determining the probability, some candidates did not subtract from one. Most candidates responded accurately to part (a) (ii).

For part (b) (i) most candidates responded accurately to this item, only a few candidates gave one tail as their response. For part (b) (ii), most candidates recognized that a z -test was to be used, however, they did not state "Reject H_0 ". Some candidates who recognized that there were two critical regions did not obtain the correct critical point. A few candidates used diagrams to support their answers. For part (b) (iii) most candidates did not state the three criteria expected for this item. Instructors must ensure that candidates connect the conclusion to the decision rule or test criteria.

Comments on Internal Assessment

The overall presentation and quality of the samples submitted was satisfactory. The projects showed sufficient evidence of candidates' individual work and were appropriate for the objectives in the Unit. Appropriate graphs, diagrams and tables were used to illustrate the data collected. There was generally clear agreement between marks awarded by the teachers and those by the CXC moderator.