CARIBBEAN EXAMINATIONS COUNCIL

REPORT ON CANDIDATES' WORK IN THE CARIBBEAN ADVANCED PROFICIENCY EXAMINATION MAY/JUNE 2005

STATISTICAL ANALYSIS

STATISTICAL ANALYSIS

CARIBBEAN ADVANCED PROFICIENCY EXAMINATION

MAY/JUNE 2005

INTRODUCTION

The revised Statistical Analysis syllabus was followed this year for the first time. Two hundred and twenty one candidates registered for this examination of which 200 sat all the required papers.

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 with 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.

GENERAL COMMENTS

Eighty-five per cent of the candidates obtained acceptable grades, Grades I – 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.

It is recommended that candidates:

(i) use the third decimal place when looking up Z-tables;
(ii) write definitions, interpretations and explanations

COMMENTS ON CANDIDATES' PERFORMANCE ON EACH QUESTION

PAPER 01

SECTION A

Module 1

This question tested the candidates' ability to distinguish between

(i) discrete and continuous data
(ii) qualitative and quantitative data

Many candidates had problems distinguishing discrete and continuous data, but were generally able to distinguish between qualitative and quantitative data

Question 2

(a) In this question candidates were required to distinguish between

(i) a population and a sample
(ii) a census and a sample survey

Many candidates used the geographical definition of a population rather than the statistical definition.

Most candidates were able to define a sample. Similar problems were seen with the census and the sample survey.

(b) Candidates generally had problems explaining or giving reasons for sampling. They often repeated one reason in different forms.

Question 3

This question tested the candidates' ability to estimate the mean and median of a grouped frequency distribution.

Most candidates were able to use the mid-points of the classes to calculate an estimate of the mean.

Many candidates could not remember the formula for the median and this formula was often quoted incorrectly.

Candidates attempted to find class boundaries although the end points of intervals were the boundaries. Most candidates could not answer part (b).

Question 4

This question tested the candidates' ability to read and interpret a box-and-whisker diagram.

Many candidates found the interquartile range rather than the semi-interquartile range, some candidates found the mid-range.

Many candidates could not describe the shape of the distribution.

Question 5

This question tested the candidates' ability to choose an appropriate measure of central tendency and to use pie charts to illustrate data.

(a) Many candidates correctly chose the mode, but quite a few candidates chose the mean.

(b) Most candidates were able to calculate the sector angle.

(c) Many candidates did not attempt this part of the question. Of those who attempted the question, very few showed any knowledge of comparing pie charts.

Most of those candidates used the ratio of 'r' rather than the ratio of the 'r^2'.

Module 2

Question 6

This question tested the candidates' ability to read probabilities from a Venn Diagram.

(a) (i) many candidates gave an answer of C only rather than C

(ii) many candidates gave the answer as

$P(B) + P(C)$ rather than $P(B) + P(C) - P(B \cap C)$

(iii) for P(B′) many candidates, as in (i), read the value for B only

(b) While candidates were able to say that A and C were mutually exclusive, many of them could not say why.

Question 7

Candidates generally did not know the properties of the binomial distribution. Many candidates quoted the formula for calculating probabilities.

Question 8

(a) Most candidates were able to read off the values of p and q from the table.

(b) Many of the candidates who attempted to use the formula

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ used the formula incorrectly, often omitting to subtract $P(A \cap B)$.

(c) Many of the candidates who attempted to use the formula

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

used the formula incorrectly.
Most of these candidates wrote P(head Ç score >5) as P(head) x P(score >5)

Question 9

(a) (i) Most candidates got this question wrong as they found $P(X < 2)$ rather than $P(X = 2)$.

(ii) Candidates rewrote this question to read

$P(Z < 1.24)$

(iii) Most candidates were able to write this as 1 ¾ f(1.162), but as in (ii) there seems to be a problem finding the probability with a 3rd decimal place involved. This question was generally well done.

Question 10

Most candidates were able to calculate the parameters np, and npq for the normal approximation. The use of the continuity correction gave some problems. Again, the use of the third decimal place when using the probability tables was a problem.

Question 11

(a) This question tested whether candidates could write the distribution of the sample mean. Most candidates did this correctly – the most problems came where candidates did not use and where they could not write the variance

as $\frac{\pi^2}{h}$

(b) Most candidates did this correctly.

Question 12

(a) Very few candidates could not write these conditions.

(b) Candidates either omitted H_0, H_1 or u when writing the hypotheses.

Some candidates interchanged the inequality -. Some wrote $\bar{x}$ rather than m in the hypothesis.

(c) This was generally well done.

Question 13

(a) Candidates generally did not write this definition correctly and often could not express the -95 probability.
This was generally done well, but students had problems with $Z_{0.025} = 1.96$ and with .

Candidates also wrote the interval with the larger figure first.

Many candidates did not know how to do this question.

Question 14

Many candidates, while being able to write the value of r, could not write the interpretation.

Most candidates have problems explaining the 'gradient', or even expressing the meaning.

Question 15

(a) Most candidates gave 0.99 as the required probability.

(b) Candidates have problems with the direction of the inequality and with the sign of the magnitude. There were also some problems with the magnitude – candidates used 2.576 instead of 2.326.

Candidates did not use the continuity correction in calculating the test statistic.

PAPER 02

Module 1

Question 1

This question was based on the following:

- defining a simple random sample and a stratified random sample;
- using a table of random numbers to obtain a simple sample of size 5;
- describing the "lottery" technique in a given situation;
- explaining one distinct advantage of obtaining a sample by quota, simple random sample and stratified random sample in a given situation.

Most candidates were unable to give clear and full definitions of a simple random sample and a stratified random sample. Examiners were hoping to read the following: a simple random sample is one so selected that each member of the population has an equal chance of inclusion and in a stratified random sample, the population is divided into strata and proportionate simple random samples taken from each stratum.

Part (c) (i) was done satisfactorily by the majority of the candidates. The modal score was 5.

For Part (c) (ii) even though candidates gave non-routine responses for c (ii) this part was not well done.

Question 2

Part (a) (i) was well done with only a few candidates writing the frequency of the modal score as the answer. Parts (a) (ii) and (a) (iii) were well done.

However for Part (a) (iv), some candidates used the various formulae for variance incorrectly.

Parts (a) (v) to (a) (viii) were done satisfactorily. The majority of candidates wrote incorrect responses for Parts (b) (i) and (ii). Valid responses for (b) (i) and (b) (ii) may be "none of these values reflect an actual mark scored" and "variance unaffected by the addition of five marks to each student's score".

Question 3

For Part (a) (i) some candidates had difficulty in calculating the area under the given graph. However, many demonstrated knowledge that the area under the graph of a p.d.f. is equal to 1.

Part (a) (ii) was not well done.

For Part (b) (ii), most candidates were able to state at most one property, that is
(P(Y = y) = 1. In (b)(ii) (a) the determination of F(6) was challenging to most candiates.

F(6)= (P(Y = y) where y = 0,2,4. However, (b) (ii) (b) was well done.

For (b) (iii) the majority of candidates responded poorly to this item. Candidates were not able to interpret what is meant by "Y_2 is 4 more than the value of Y_1". Examiners were hoping to see the following "$P(Y_2 = 4 + Y_1) = P(Y_1 = 0, Y_2 = 4) + P(Y_1 = 4, Y_2 = 8)$. The modal score was zero.

Question 4

This question was based on the construction and use of a tree diagram in solving simple problems in probability including conditional probability.

For Part (a), the majority of the candidates drew and labeled correctly tree diagrams and as a result they scored full marks.

Part (b) (i) was satisfactorily done by most candidates. Examiners were pleased to see candidates making use of the concept of mutually exclusive events but only a little more than 50% of them were able to complete Part (b) (ii) correctly.

Part (b) (iii) was not well done. Using P(pineapple juice or a juice with honey added) = P(pineapple juice without honey) + (PC juice with honey added) will result in

answer 0.7175 or $\frac{287}{400}$

Most candidates demonstrated good conceptual understanding of conditional probability but solving the required conditional probability was rarely done correctly.

Question 5

This question was based on the hypothesis testing for a population mean from a large sample.

In Part (a), some candidates stated incorrectly that a one-tailed test was appropriate and their reasoning stemmed from comparing the sample mean to the population mean.

For Part (b), a few candidates wrote their hypotheses in words instead of using symbols as requested.

For Part (c), over 50% of the candidates received full credit for this part.

For Part (d) those candidates who illustrated their answers with sketches assisted the examiners immensely in awarding scores. On the whole, most candidates performed satisfactorily.

Part (e) was not well done.

Parts (f) – (g) were well done.

Question 6

For Part (a), the majority of the candidates were able to plot the scatter diagram correctly and, therefore, they earned full marks.

For Parts (b) and (c) those candidates who used the formulae given in the list of formulae and statistical tables had no great difficulty in calculating r(º -0.849) a (=86.1) and b = (-1.62). Unfortunately, a subset of these candidates used n = 12 instead of the correct value of 10. It was commendable that some candidates used

the formula $r = \frac{S_{xy}}{S_x S_y}$ but penalized themselves by getting into cumbersome calculations.

It is recommended that candidates make full use of the list of formulae and statistical tables.

In Part (d), the majority of the candidates were able to do this part well either by substituting x = 13 in their regression equations or reading off from their graphs. Furthermore, a few candidates wrote valid comments on the reliability of the requested estimate.

COMMENTS ON INTERNAL ASSESSMENT

The projects showed sufficient evidence of candidates' individual work and were appropriate for the objectives in the Unit. Appropriate graphs, figures and tables were used to illustrate the data collected. However, some candidates did not clearly describe the method of data collection. In many cases, the analyses done were at the CSEC level. Analysis of data should involve concepts learnt at the CAPE level.

Most teachers did not make comments on the STATI-5 Form or in the project.

Specific Comments

Missing components – many projects did not include a conclusion; variables were not identified and issues of reliability, validity or worthwhileness of findings were not raised. A few projects had no title.

List of references – candidates continue to use an acceptable style of referencing especially for internet references. An example of an acceptable style of referencing a website may be

> United Nations, 2005, World Population Prospects:
> The 2004 Revision Population Database (Online)
> 03/03/04, http://esa.un.org/unpp.

Paper 03B

One candidate wrote this paper which consisted of three questions which collectively span the entire syllabus.

Question 1

This question tested topics in Linear Regression (Module 3) and Data Analysis (Module 1).

The candidate was only able to score on Parts (b) (i) and (ii) of the question.

Question 2

In this question, the candidate was required to determine probabilities from a contingency table (Module 2) and carry out some aspects of a x2-test (Module 3). The candidate was successful in answering only Parts (a) (i) and (iii) of the

question.

Question 3

This question focused on Probability Theory and the normal approximation to the binomial distribution. The candidate received no credit for this question.