OCR
Oxford Cambridge and RSA

Accredited

AS Level Further Mathematics A
Y532 Statistics
Sample Question Paper

Date – Morning/Afternoon
Time allowed: 1 hour 15 minutes

OCR supplied materials:
- Printed Answer Booklet
- Formulae AS Level Further Mathematics A

You must have:
- Printed Answer Booklet
- Formulae AS Level Further Mathematics A
- Scientific or graphical calculator

INSTRUCTIONS

- Use black ink. HB pencil may be used for graphs and diagrams only.
- Complete the boxes provided on the Printed Answer Booklet with your name, centre number and candidate number.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.**
- Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.
- The acceleration due to gravity is denoted by $g\,\mathrm{m\,s^{-2}}$. Unless otherwise instructed, when a numerical value is needed, use $g = 9.8$.

INFORMATION

- The total number of marks for this paper is **60**.
- The marks for each question are shown in brackets **[ ]**.
- **You are reminded of the need for clear presentation in your answers.**
- The Printed Answer Booklet consists of **12** pages. The Question Paper consists of **4** pages.

1 Two music critics, P and Q, give scores to seven concerts as follows.

Concert	1	2	3	4	5	6	7
Score by critic P	12	11	6	13	17	16	14
Score by critic Q	9	13	8	14	18	16	20

(i) Calculate Spearman's rank correlation coefficient, r_s, for these scores. **[4]**

Concert	1	2	3	4	5	6	7
Score by critic P	12	11	6	13	17	16	14
Score by critic Q	9	13	8	14	18	16	20

Rank P	3	2	1	4	7	6	5
Rank Q	2	3	1	4	6	5	7
P - Q	1	-1	0	0	1	1	-2

$\Sigma d^2 = 1+1+1+1+4 = 8$

$r_s = 1 - \frac{6 \times 8}{7 \times 48} = 0.857$

(ii) Without carrying out a hypothesis test, state what your answer tells you about the views of the two critics. **[1]**

They strongly agree with each other.

2 The probability distribution of a discrete random variable W is given in the table.

w	0	1	2	3
$P(W=w)$	0.19	0.18	x	y

Given that $E(W)=1.61$, find the value of $\text{Var}(3W+2)$. **[7]**

total probability = 1

$0.19+0.18+x+y=1$

$x+y=0.63$

$y=0.63-x$ —①

$E(w)=0\times0.19+1\times0.18+2x+3y$

$1.61=0.18+2x+3y$

$1.43=2x+3y$ —②

sub ① into ②:

$1.43=2x+3(0.63-x)$

$1.43=2x+1.89-3x$

$x=0.46$

$y=0.63-0.46=0.17$

So $\text{Var}(w)=E(w^2)-[E(w)]^2$

$=0.18+(0.46\times2^2)+(0.17\times3^2)-1.61^2$

$=0.18+1.84+1.53-2.5921$

$=0.9579$

$\text{Var}(3w+2)=3^2\text{Var}(w)$

$=9\times0.9579=8.6211$

3 Carl believes that the proportions of men and women who own black cars are different. He obtained a random sample of people who each owned exactly one car. The results are summarised in the table below.

	Black	Non-black	Total
Men	69	71	140
Women	30	55	85
Total	99	126	225

Test at the 5% significance level whether Carl's belief is justified. **[8]**

H_0 : car colour is independant of gender

H_1 : car colour is not independant of gender.

	Black		Non-black		
	O_i	E_i	O_i	E_i	
Men	69	61.6	71	78.4	140
Women	30	37.4	55	47.6	85
	99		126		225

Expected values : $\text{Men, black} = \frac{99 \times 140}{225} = 61.6$

$\text{Men, non-black} = \frac{126 \times 140}{225} = 78.4$

$\text{Women, black} = \frac{99 \times 85}{225} = 37.4$

$\text{Women, non-black} = \frac{126 \times 85}{225} = 47.6$

$$X_1^2 = \frac{(169 - 61.61 - 0.5)^2}{61.6} + \frac{(171 - 78.41 - 0.5)^2}{78.4} + \frac{(130 - 37.41 - 0.5)^2}{37.4} + \frac{(155 - 47.61 - 0.5)^2}{47.6}$$

$$= 0.7729 + 0.6073 + 1.2730 + 1.0002 = 3.6534$$

Critical value = 3.841

3.6534 < 3.841 so do not reject H_0

Insufficient evidence to suggest car colour is dependant on gender.

4 **(i)** Four men and four women stand in a random order in a straight line. Determine the probability that no one is standing next to a person of the same gender. **[3]**

Either:

M W M W M W M W or W M W M W M W M

Within each of these arrangements, the men can swap places in 4! ways, and the women can swap places in 4! ways.

So there are $2 \times 4! \times 4! = 1152$ different valid combinations.

There are 8! combinations altogether regardless of positions.

$\therefore$ Probability $= \frac{1152}{8!} = 0.0286$

(ii) x men, including Mr Adam, and x women, including Mrs Adam, are arranged at random in a straight line. Show that the probability that Mr Adam is standing next to Mrs Adam is $\frac{1}{x}$. **[3]**

They can either be stood with Mrs Adams on the left and Mr Adams on the right, or with Mr Adams on the left and Mrs Adams on the right. Two possibilities.

If you treat Mr and Mrs Adams as one item, there are now $2x-1$ items to arrange.

$\therefore$ Probability $= \frac{2 \times (2x-1)!}{(2x)!} = \frac{2}{2x} = \frac{1}{x}$

5 **(i)** The random variable X has the distribution $\text{Geo}(0.6)$.

(a) Find $\text{P}(X \geq 8)$. **[2]**

(b) Find the value of $\text{E}(X)$. **[1]**

(c) Find the value of $\text{Var}(X)$. **[1]**

a. $X \sim \text{Geo}(0.6)$

$P(X \geq 8) = 0.4^7 = 0.00164$

b. $\frac{1}{0.6} = \frac{5}{3}$

c. $\frac{1-0.6}{0.6^2} = \frac{10}{9}$

(ii) The random variable Y has the distribution $\text{Geo}(p)$. It is given that $\text{P}(Y < 4) = 0.986$ correct to 3 significant figures. Use an algebraic method to find the value of p. **[3]**

$P(Y < 4) = 1 - (1-p)^3$

$0.986 = 1 - (1-p)^3$

$(1-p)^3 = 0.014$

$1 - p = 0.241$

$p = 0.759$ (3sf)

6 Sabrina counts the number of cars passing her house during randomly chosen one minute intervals. Two assumptions are needed for the number of cars passing her house in a fixed time interval to be well modelled by a Poisson distribution.

(i) State these two assumptions. **[2]**

(ii) For each assumption in part **(i)** give a reason why it might not be a reasonable assumption for this context. **[2]**

i. Cars must pass independantly of other cars, at a constant rate.

ii. Cars may go past in groups after being stopped at traffic lights. Amount of cars that go past will be different at different times day, so rate is not constant.

Assume now that the number of cars that pass Sabrina's house in one minute can be well modelled by the distribution $\text{Po}(0.8)$.

(iii) (a) Write down an expression for the probability that, in a given one minute period, exactly r cars pass Sabrina's house. **[1]**

(b) Hence find the probability that, in a given one minute period, exactly 2 cars pass Sabrina's house. **[1]**

a. $X \sim Po(0.8)$

$$P(X=r) = \frac{e^{-0.8} \times 0.8^2}{r!}$$

b. when $r=2$, $P(X=2) = \frac{e^{-0.8} \times 0.8^2}{2!}$

$= 0.144$

(iv) Find the probability that, in a given 30 minute period, at least 28 cars pass Sabrina's house. **[3]**

for 30 minute period, $\lambda = 30 \times 0.8 = 24$

$X \sim Po(24)$

$P(X \geq 28) = 1 - P(x \leq 27)$

$= 1 - 0.76774$

$= 0.232$ (3SF)

(v) The number of bicycles that pass Sabrina's house in a 5 minute period is a random variable with the distribution $\text{Po}(1.5)$. Find the probability that, in a given 10 minute period, the total number of cars and bicycles that pass Sabrina's house is between 12 and 15 inclusive. State a necessary condition.

[4]

bikes cars

$\lambda = 2 \times 1.5 + 10 \times 0.8$

$\lambda = 11$

$Y \sim Po(11)$

$P(12 \leq Y \leq 15) = P(Y \leq 15) - P(Y \leq 11)$

$= 0.9074 - 0.5792$

$= 0.3282$ (4dp)

The distributions must be independant.

7 The discrete random variable X is equally likely to take values 0, 1 and 2.
$3N$ observations of X are obtained, and the observed frequencies corresponding to $X = 0$, $X = 1$ and $X = 2$ are given in the following table.

x	0	1	2
Observed frequency	$N-1$	$N-1$	$N+2$

The test statistic for a chi-squared goodness of fit test for the data is 0.3. Find the value of N.

[4]

The expected value for each is $\frac{3N}{3} = N$

$\sum \frac{(O-E)^2}{E} = \frac{(N-1-N)^2}{N} + \frac{(N-1-N)^2}{N} + \frac{(N+2-N)^2}{N}$

$= \frac{1}{N} + \frac{1}{N} + \frac{4}{N}$

$= \frac{6}{N}$

$\frac{6}{N} = 0.3 \Rightarrow N = 20$

8 The following table gives the mean per capita consumption of mozzarella cheese per annum, x pounds, and the number of civil engineering doctorates awarded, y, in the United States in each of 10 years.

x	9.3	9.7	9.7	9.7	9.9	10.2	10.5	11.0	10.6	10.6
y	480	501	540	552	547	622	655	701	712	708

source: www.tylervigen.com

(i) Find the equation of the regression line of y on x. **[2]**

You are given that the product moment correlation coefficient is 0.959.

(ii) Explain whether this value would be different if x is measured in kilograms instead of pounds. **[1]**

i. $y = 157x - 988$

ii. It would not be different, r is not affected by linear coding.

It is desired to carry out a hypothesis test to investigate whether there is correlation between these two variables.

(iii) Assume that the data is a random sample of all years.

(a) Carry out the test at the 10% significance level. **[6]**

(b) Explain whether your conclusion suggests that manufacturers of mozzarella cheese could increase consumption by sponsoring doctoral candidates in civil engineering. **[1]**

let ρ be the population correlation coefficient.

a. $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$cv = 0.5494$

$0.9586 > 0.5494$ so reject H_0

Result is significant, there is evidence of correlation between consumption of mozzarella and number of civil engineering doctorates awarded.

b. No, correlation does not imply causation.