MODEL ANSWERS

Write your name here

Surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number

Candidate Number

Further Mathematics

Advanced
Further Mathematics Option 1
Paper 3: Further Statistics 1
Further Mathematics Option 2
Paper 4: Further Statistics 1

Sample Assessment Material for first teaching September 2017

Time: 1 hour 30 minutes

Paper Reference
9FM0/3B
9FM0/4B

You must have:
Mathematical Formulae and Statistical Tables, calculator

Total Marks

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for algebraic manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- There are 7 questions in this question paper. The total mark for this paper is 75.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over

S54441A

1/1/1/1/

S 5 4 4 4 1 A 0 1 2 3

Pearson

Answer ALL questions. Write your answers in the spaces provided.

1. Bacteria are randomly distributed in a river at a rate of 5 per litre of water. A new factory opens and a scientist claims it is polluting the river with bacteria. He takes a sample of 0.5 litres of water from the river near the factory and finds that it contains 7 bacteria. Stating your hypotheses clearly test, at the 5% level of significance, whether there is evidence that the level of pollution has increased.

(5)

Let the random variable X= the number of bacteria per 0.5 litres of water.

$X \sim Po(2.5)$

$H_0: \lambda = 2.5$ $H_1: \lambda > 2.5$ $\alpha = 5\%$

$P(X \geq 5) = 1 - P(X \leq 4)$
$= 1 - 0.8912$
$= 0.1098 > 0.05$

$P(X \geq 6) = 1 - P(X \leq 5)$
$= 1 - 0.9580$
$= 0.0420 < 0.05$

Critical region is $X \geq 6$

Since the observed value of 7 is within the critical region, there is sufficient evidence to reject H_0 at the 5% significance level. There is evidence that the level of pollution has increased.

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

2. A call centre routes incoming telephone calls to agents who have specialist knowledge to deal with the call. The probability of a caller, chosen at random, being connected to the wrong agent is p.

The probability of at least 1 call in 5 consecutive calls being connected to the wrong agent is 0.049

The call centre receives 1000 calls each day.

(a) Find the mean and variance of the number of wrongly connected calls a day.
(7)

(b) Use a Poisson approximation to find, to 3 decimal places, the probability that more than 6 calls each day are connected to the wrong agent.
(2)

(c) Explain why the approximation used in part (b) is valid.
(2)

The probability that more than 6 calls each day are connected to the wrong agent using the binomial distribution is 0.8711 to 4 decimal places.

(d) Comment on the accuracy of your answer in part (b).
(1)

(a) Let the random variable X = the number of wrongly connected calls a day, when there are 1000 calls a day.

$X \sim B(1000, p)$

Given $P(X \geq 1) = p = 0.049$

$1 - P(X=0) = 0.049$

$P(X=0) = 0.951$

$x^5 = 0.951$

$x = 0.99$

$\therefore \underline{p = 0.01}$

So $X \sim B(1000, 0.01)$

Question 2 continued

Mean $= np = 1000 \times 0.01$

$\boxed{= 10}$

Variance $= np(1-p)$

$= 1000 \times 0.01\,(0.99)$

$= 10 \times 0.99$

$\boxed{= 9.9}$

(b) Under a Poisson approximation,
$X \sim Po(np) \Rightarrow X \sim Po(10)$

$P(X > 6) = 1 - P(X \leq 6)$

$= 1 - 0.1301$

$= \boxed{0.870 \text{ (to 3d.p.)}}$

(c) The approximation used in part (b) is valid because the number of calls, n is large, and the probability of connecting to the wrong agent, p is small.

(d) The answer in part (b) under a binomial distribution would $= 0.871\ldots$ so the answer is correct to 2 d.p.

3. Bags of £1 coins are paid into a bank. Each bag contains 20 coins.

The bank manager believes that 5% of the £1 coins paid into the bank are fakes. He decides to use the distribution $X \sim B(20, 0.05)$ to model the random variable X, the number of fake £1 coins in each bag.

The bank manager checks a random sample of 150 bags of £1 coins and records the number of fake coins found in each bag. His results are summarised in Table 1. He then calculates some of the expected frequencies, correct to 1 decimal place.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	53.8	56.6	28.3	8.9	2.4

Table 1

(a) Carry out a hypothesis test, at the 5% significance level, to see if the data supports the bank manager's statistical model. State your hypotheses clearly.

(10)

The assistant manager thinks that a binomial distribution is a good model but suggests that the proportion of fake coins is higher than 5%. She calculates the actual proportion of fake coins in the sample and uses this value to carry out a new hypothesis test on the data. Her expected frequencies are shown in Table 2.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	44.5	55.7	33.2	12.5	4.1

Table 2

(b) Explain why there are 2 degrees of freedom in this case.

(2)

(c) Given that she obtains a χ^2 test statistic of 2.67, test the assistant manager's hypothesis that the binomial distribution is a good model for the number of fake coins in each bag. Use a 5% level of significance and state your hypotheses clearly.

(2)

(a) Fill out the table to find the Expected frequencies for $x=2$ and $x \geq 4$.

For $x=2$, $P(X=2) = 0.1887$

So Expected freq. $= 0.1887 \times 150$
$= 28.3$ (to 1d.p.)

Question 3 continued

So for $x \geq 4$, Expected frequency,

$= 150 - (53.8 + 56.6 + 28.3 + 8.9)$

$= 150 - 147.6$

$= 2.4$ (to 1d.p.)

H_0: The use of a binomial distribution $B(20, 0.05)$ is supported by the data.

H_1: The use of a binomial distribution $B(20, 0.05)$ is not supported by the data.

$\alpha = 5\%$.

Since the expected frequency for $x \geq 4$ is less than 5, combine the last two columns:

No. of fake coins in each bag	0	1	2	≥ 3
Observed frequency	43	62	26	19
Expected frequency	53.8	56.6	28.3	11.3

Number of degrees of freedom, $\nu = 4 - 1 = 3$

Critical value, $\chi^2_3(0.05) = 7.815$

To find the test statistic, use $\frac{(O_i - E_i)^2}{E_i}$:

Question 3 continued

	O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
0	43	53.8	2.1680
1	62	56.6	0.5152
2	26	28.3	0.1869
≥ 3	19	11.3	5.2469
	$\Sigma O_i = 150$	$\Sigma E_i = 150$	$\Sigma \frac{(O_i-E_i)^2}{E_i} = 8.117$

Test statistic, $\chi = 8.117$

Since $8.117 > 7.815$, there is sufficient evidence to reject H_0 at the 5% level of significance. There is evidence to reject the manager's model.

(b) If the assistant manager calculates the actual proportions, then an extra constraint is created. So there are now 2 constraints with 4 columns (Since the last 2 need to be combined to make an expected frequency of greater than 5.)

So, number of degrees of freedom, $\nu = 4-2 = 2$

(c) H_0: The assistant manager's model is good.
H_1: The assistant manager's model is not good.

Critical value, $\chi^2_2(0.05) = 5.991$

Since $2.67 < 5.991$, there is insufficient evidence to reject H_0 at the 5% significance level: the assistant manager's model is good.

4. A random sample of 100 observations is taken from a Poisson distribution with mean 2.3

Estimate the probability that the mean of the sample is greater than 2.5

(4)

$X \sim Po(2.3)$

$n = 100, \mu = 2.3$

Using the central limit theorem,

$\bar{X} \sim N\left(\mu, \frac{\sigma}{n}\right)$

$\bar{X} \sim N\left(2.3, \frac{2.3}{100}\right)$

$\boxed{P(\bar{X} > 2.5) = 0.0936 \text{ (to 4d.p.)}}$

Use σ as $\sqrt{\frac{2.3}{100}}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

5. The probability of Richard winning a prize in a game at the fair is 0.15

Richard plays a number of games.

(a) Find the probability of Richard winning his second prize on his 8th game,
(2)

(b) State two assumptions that have to be made, for the model used in part (a) to be valid.
(2)

Mary plays the same game, but has a different probability of winning a prize. She plays until she has won r prizes. The random variable G represents the total number of games Mary plays.

(c) Given that the mean and standard deviation of G are 18 and 6 respectively, determine whether Richard or Mary has the greater probability of winning a prize in a game.
(4)

(a) Let the random variable X = the number of wins in 8 games, then

$X \sim$ Negative $B(2, 0.15)$

$$P(X=8) = \binom{8-1}{2-1} \cdot 0.15^2 \cdot (1-0.15)^6$$

$$= \binom{7}{1} \cdot 0.15^2 \cdot 0.85^6$$

$$= 7 \times 0.15^2 \times 0.85^6$$

$$\boxed{= 0.0594 \quad (\text{to } 4\text{d.p.})}$$

(b) It is assumed that the games are independent of one another, and the probability of winning a prize is constant, i.e. 0.15

(c) Mean $= \frac{r}{p}$

So $18 = \frac{r}{p}$

DO NOT WRITE IN THIS AREA

Question 5 continued

$$\text{Variance} = \frac{r(1-p)}{p^2}$$

$$6^2 = \frac{r(1-p)}{p^2}$$

$$36 = \frac{r(1-p)}{p^2}$$

$$36 = \frac{r}{p}\,\frac{(1-p)}{p}$$

So, $36 = 18\,\frac{(1-p)}{p}$, since $18 = \frac{r}{p}$

$$2 = \frac{1-p}{p}$$

$$2p = 1-p$$

$$3p = 1$$

$$\underline{p = \frac{1}{3}} = 0.\dot{3}$$

Since $\frac{1}{3} > 0.15$, Mary has the greater probability of winning a prize in a game.

(Total for Question 5 is 8 marks)

6. The probability generating function of the discrete random variable X is given by

$$G_X(t) = k(3 + t + 2t^2)^2$$

(a) Show that $k = \frac{1}{36}$ **(2)**

(b) Find $P(X = 3)$ **(2)**

(c) Show that $\text{Var}(X) = \frac{29}{18}$ **(8)**

(d) Find the probability generating function of $2X + 1$ **(2)**

(a) $G_x(1) = 1 \Rightarrow k(3+1+2(1)^2)^2 = 1$

$k(6)^2 = 1$

$k \times 6^2 = 1$

$36k = 1$

$\therefore \boxed{k = \frac{1}{36}}$

(b) $G_x(t) = \frac{1}{36}(3+t+2t^2)^2$

$= \frac{1}{36}(2t^2+t+3)(2t^2+t+3)$

$= \frac{1}{36}(4t^4+4t^3+13t^2+6t+9)$

$= \frac{1}{9}t^4 + \underline{\frac{1}{9}t^3} + \frac{13}{36}t^2 + \frac{1}{6}t + \frac{1}{4}$

$\therefore \boxed{P(X=3) = \frac{1}{9}}$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Question 6 continued

(c) $\text{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$ From formula book

$$G_X(t) = \frac{1}{9}t^4 + \frac{1}{9}t^3 + \frac{13}{36}t^2 + \frac{1}{6}t + \frac{1}{4}$$

$$G'_X(t) = \frac{4}{9}t^3 + \frac{1}{3}t^2 + \frac{26}{36}t + \frac{1}{6}$$

$$= \frac{4}{9}t^3 + \frac{1}{3}t^2 + \frac{13}{18}t + \frac{1}{6}$$

$$G'_X(1) = \frac{4}{9} + \frac{1}{3} + \frac{13}{18} + \frac{1}{6}$$

$$= \frac{8+6+13+3}{18}$$

$$= \frac{30}{18}$$

$$= \frac{5}{3}$$

$$(G'_X(1))^2 = \left(\frac{5}{3}\right)^2 = \frac{25}{9}$$

$$G''_X(t) = \frac{12}{9}t^2 + \frac{2}{3}t + \frac{13}{18}$$

$$= \frac{4}{3}t^2 + \frac{2}{3}t + \frac{13}{18}$$

Question 6 continued

$$G''_X(1) = \frac{4}{3} + \frac{2}{3} + \frac{13}{18}$$

$$= \frac{24+12+13}{18}$$

$$= \frac{49}{18}$$

$$\therefore \operatorname{Var}(X) = \frac{49}{18} + \frac{5}{3} - \frac{25}{9}$$

$$\operatorname{Var}(X) = \frac{49+30-50}{18}$$

$$\therefore \boxed{\operatorname{Var}(X) = \frac{29}{18}}$$

(d) Let $Y = 2X+1$

Then $P(X=4) = P(Y=9)$
$P(X=3) = P(Y=7)$
$P(X=2) = P(Y=5)$
$P(X=1) = P(Y=3)$
$P(X=0) = P(Y=1)$

So $G_Y(t) = \frac{1}{9}t^9 + \frac{1}{9}t^7 + \frac{13}{36}t^5 + \frac{1}{6}t^3 + \frac{1}{4}t$

$$G_{2X+1}(t) = \frac{t}{36}\left(9 + 6t^2 + 13t^4 + 4t^6 + 4t^8\right)$$

$$\boxed{G_{2X+1}(t) = \frac{t}{36}\left(3 + t^2 + 2t^4\right)^2}$$

7. Sam and Tessa are testing a spinner to see if the probability, p, of it landing on red is less than $\frac{1}{5}$. They both use a 10% significance level.

Sam decides to spin the spinner 20 times and record the number of times it lands on red.

(a) Find the critical region for Sam's test.
(2)

(b) Write down the size of Sam's test.
(1)

Tessa decides to spin the spinner until it lands on red and she records the number of spins.

(c) Find the critical region for Tessa's test.
(6)

(d) Find the size of Tessa's test.
(1)

(e) (i) Show that the power function for Sam's test is given by

$$(1-p)^{19}(1+19p)$$

(ii) Find the power function for Tessa's test.
(4)

(f) With reference to parts (b), (d) and (e), state, giving your reasons, whether you would recommend Sam's test or Tessa's test when $p = 0.15$
(4)

(a) Let the random variable X = the number of times the spinner lands on red from 20 spins.

$X \sim B(20, 0.2)$ $\quad \alpha = 0.1$

So, check c so that $P(X \leq c) < 0.1$

$P(X \leq 2) = 0.2061 \quad > 0.1$

$P(X \leq 1) = 0.0692 \quad < 0.1$

$\therefore$ the $\boxed{\text{critical region is } X \leq 1}$

(b) Size of sam's test $\boxed{= 0.0692}$

Question 7 continued

Let the random variable Y = the number of spins until red is spun.

$$Y \sim \text{Geo}(0.2)$$

Since the mean, $\mu = \frac{1}{p}$, if $p < 0.2$ then μ would be greater.

So check for d so that $P(Y \geq d) < 0.1$

$$P(Y \geq d) = (1-p)^{d-1}$$
$$= 0.8^{d-1}$$

So, $0.8^{d-1} < 0.1$

$$\log (0.8)^{d-1} < \log (0.1)$$
$$(d-1)(\log(0.8)) < \log(0.1)$$
$$d-1 > \frac{\log(0.1)}{\log(0.8)}$$

the sign switch since $\log(0.8) < 0$

$$d > \frac{\log(0.1)}{\log(0.8)} + 1$$
$$d > 11.3189\ldots$$

$\therefore$ the $\boxed{\text{critical region is } Y \geq 12}$

[redacted]

DO NOT WRITE IN THIS AREA

Question 7 continued

(d) $P(Y \geq 12) = (1-0.2)^{12-1}$

$= 0.8^{11}$

$\boxed{= 0.0859}$

$\therefore$ the size of Tessa's test is 0.0859

(e) (i) Power = P(reject H_0 when false)

$= P(X \leq 1 \mid X \sim B(20, p))$

$= P(X=0) + P(X=1)$

$= \binom{20}{0} p^0 (1-p)^{20} + \binom{20}{1} p^1 (1-p)^{19}$

$= (1-p)^{20} + 20p(1-p)^{19}$

$\boxed{= (1-p)^{19}(1+19p)}$

(ii) For Tessa's test:

Power $= (Y \geq 12 \mid Y \sim \text{Geo}(0.2))$

$\boxed{= (1-p)^{11}}$

(f) The power of Sam's test, according to part (e) (i) $= (1-p)^{19}(1+19p)$.

So when $p = 0.15$, Power $= (1-0.15)^{19}(1+19(0.15))$

$= 0.85^{19} \times 3.85$

$= 0.1756$ (to 4d.p.)

Question 7 continued

The power of Tessa's test, according to part (e) (ii) $= (1-p)^{11}$

So, when $p=0.15$, Power $= (1-0.15)^{11}$

$$= 0.85^{11}$$

$$\underline{= 0.1673 \text{ (to 4d.p)}}$$

Also, for Sam, $P(\text{Type I error}) = P(X \leq 1 \mid p=0.20)$

$$\underline{= 0.0692}$$

For Tessa, $P(\text{Type I error}) = P(Y \geq 12 \mid p=0.20)$

$$\underline{= 0.0859}$$

Since Sam's test has a smaller P(Type I error) there is less probability for a Type I error, so is better.

The power of Sam's test is also greater than Tessa's.

So $\boxed{\text{for } p=0.15\text{, Sam's test is recommended.}}$