MODEL ANSWERS

Write your name here

| Surname | Other names |
|---|---|
| | |

**Pearson Edexcel**
**International**
**Advanced Level**

Centre Number | | | | | |

Candidate Number | | | | |

# Statistics S3

## Advanced/Advanced Subsidiary

Wednesday 24 May 2017 – Morning
**Time: 1 hour 30 minutes**

Paper Reference
**WST03/01**

**You must have:**
Mathematical Formulae and Statistical Tables (Blue)

Total Marks

**Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra 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). Coloured pencils and highlighter pens must not be used.
- **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.
- Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

## Information

- 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* ▶

P48248A



1/1/

P 4 8 2 4 8 A 0 1 2 8

Pearson

Leave blank

**1.** The ages, in years, of a random sample of 8 parrots are shown in the table below.

| Parrot | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
|---|---|---|---|---|---|---|---|---|
| Age | 10 | 4 | 13 | 15 | 2 | 1 | 8 | 6 |

A parrot breeder does not know the ages of these 8 parrots. She examines each of these 8 parrots and is asked to put them in order of decreasing age. She puts them in the order

*D* *G* *H* *C* *A* *B* *F* *E*

(a) Find, to 3 decimal places, Spearman's rank correlation coefficient between the breeder's order and the actual order.

**(5)**

(b) Use your value of Spearman's rank correlation coefficient to test for evidence of the breeder's ability to order parrots correctly, by their age, after examining them. Use a 1% significance level and state your hypotheses clearly.

**(4)**

a) $r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$

| Parrot | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Age rank | 3 | 6 | 2 | 1 | 7 | 8 | 4 | 5 |
| Breeder's rank | 5 | 6 | 4 | 1 | 8 | 7 | 2 | 3 |
| $d^2$ | 4 | 0 | 4 | 0 | 1 | 1 | 4 | 4 |

$\Sigma d^2 = 18$

$r_s = 1 - \frac{6(18)}{8(8^2-1)}$

$= 1 - \frac{108}{504}$

$= 0.785714$

$\simeq 0.786$

b) $H_0 : \rho = 0$

$H_1 : \rho > 0$

$n = 8$

At 1% significance level, cv = 0.8333

$0.786 < 0.8333$

∴ Do not reject $H_0$. There is insufficient evidence to show that breeder can order the parrots correctly by age.

Leave blank

**Question 1 continued**

**Q1**

**(Total 9 marks)**

Leave blank

**2.** A school uses online report cards to promote both hard work and good behaviour of its pupils. Each card details a pupil's recent achievement and contains exactly one of three inspirational messages *A*, *B* or *C*, chosen by the pupil's teacher.

The headteacher believes that there is an association between the pupil's gender and the inspirational message chosen. He takes a random sample of 225 pupils and examines the card for each pupil. His results are shown in Table 1.

| | | Inspirational message | | | Total |
|---|---|---|---|---|---|
| | | ***A*** | ***B*** | ***C*** | |
| **Pupil's gender** | **Male** | 25 | 37 | 45 | 107 |
| | **Female** | 32 | 50 | 36 | 118 |
| **Total** | | 57 | 87 | 81 | 225 |

**Table 1**

Stating your hypotheses clearly, test, at the 10% level of significance, whether or not there is evidence to support the headteacher's belief. Show your working clearly.
You should state your expected frequencies correct to 2 decimal places.

**(10)**

$H_0$ : There is an association between pupil's gender and inspirational message chosen.

$H_1$ : There is no association between pupil's gender and inspirational message chosen.

Expected outcomes :

| | A | B | C |
|---|---|---|---|
| Male | $\frac{107(57)}{225} = 27.1067$ | $\frac{107(87)}{225} = 41.3733$ | $\frac{107(81)}{225} = 38.52$ |
| Female | $\frac{118(57)}{225} = 29.8933$ | $\frac{118(87)}{225} = 45.6266$ | $\frac{118(81)}{225} = 42.48$ |

| O | E | $\frac{(O-E)^2}{E}$ |
|---|---|---|
| 25 | 27.1067 | 0.1637 |
| 37 | 41.3733 | 0.4623 |
| 45 | 38.52 | 1.0901 |
| 32 | 29.8933 | 0.1485 |
| 50 | 45.6266 | 0.4192 |
| 36 | 42.48 | 0.9885 |

$\Sigma \frac{(O-E)^2}{E} = 3.2721$

$\simeq 3.27$

Leave blank

**Question 2 continued**

$v = (2-1)(3-1)$

$= 2$

At 10% significance level, CV = 4.605

$3.27 < 4.605$

$\therefore$ Do not reject $H_0$. There is insufficient evidence to support the headteacher's belief.

Leave blank

**Question 2 continued**

Leave blank

**Question 2 continued**

**(Total 10 marks)**

**Q2**

Leave blank

**3.** The manager of a gym claimed that the mean age of its customers is 30 years. A random sample of 75 customers is taken and their ages have a mean of 28.2 years and a standard deviation, $s$, of 8.5 years.

(a) Stating your hypotheses clearly and using a 10% level of significance, test whether or not the manager's claim is supported by the data.

**(5)**

(b) Explain the relevance of the Central Limit Theorem to your calculation in part (a).

**(1)**

(c) State an additional assumption needed to carry out the test in part (a).

**(1)**

a) $H_0: \mu = 30$

$H_1: \mu \neq 30$

$\bar{X} \sim N\left(28.2, \frac{8.5^2}{75}\right)$

$Z = \frac{28.2 - 30}{\frac{8.5}{\sqrt{75}}}$

$= -1.8339$

two-tailed test so at 10% significance level,

critical region: $z \geq 1.6449$ or $z \leq -1.6449$

$-1.8339 < -1.6449$

$\therefore$ Reject $H_0$. The manager's claim is not supported.

b) The sample mean is approximately normally distributed

c) The variance of the sample is the same as the population variance.

Leave blank

**Question 3 continued**

Leave blank

**Question 3 continued**

Leave blank

**Question 3 continued**

**Q3**

**(Total 7 marks)**

Leave blank

**4.** The number of emergency plumbing calls received per day by a local council was recorded over a period of 80 days. The results are summarised in the table below.

| **Number of calls, $x$** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| **Frequency** | 3 | 13 | 14 | 15 | 10 | 8 | 8 | 6 | 3 |

(a) Show that the mean number of emergency plumbing calls received per day is 3.5

**(1)**

A council officer suggests that a Poisson distribution can be used to model the number of emergency plumbing calls received per day. He uses the mean from the sample above and calculates the expected frequencies shown in the table below.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 or more |
|---|---|---|---|---|---|---|---|---|---|
| **Expected frequency** | 2.42 | 8.46 | 14.80 | $r$ | 15.10 | 10.57 | 6.17 | 3.08 | $s$ |

(b) Calculate the value of $r$ and the value of $s$, giving your answers correct to 2 decimal places.

**(3)**

(c) Test, at the 5% level of significance, whether or not the Poisson distribution is a suitable model for the number of emergency plumbing calls received per day. State your hypotheses clearly.

**(7)**

a) $\frac{13 + 2(14) + 3(15) + 4(10) + 5(8) + 6(8) + 7(6) + 8(3)}{80} = \frac{280}{80}$

$= 3.5$

b) $X \sim Po(3.5)$

$r = e^{-3.5}\left(\frac{3.5^3}{3!}\right)(80)$

$= 0.2158\,(80)$

$= 17.2628$

$\simeq 17.26$

$2.42 + 8.46 + 14.80 + 15.10 + 10.57 + 6.17 + 3.08 + r + s = 80$

$s = 2.14$

Leave blank

**Question 4 continued**

c) $H_0$ : The Poisson distribution is a suitable model.

$H_1$ : The Poisson distribution is not a suitable model.

| $x$ | O | E | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|
| $\leq 1$ | 16 | 10.88 | 2.4094 |
| 2 | 14 | 14.80 | 0.0432 |
| 3 | 15 | 17.26 | 0.2959 |
| 4 | 10 | 15.10 | 1.7225 |
| 5 | 8 | 10.57 | 0.6249 |
| 6 | 8 | 6.17 | 0.5428 |
| $\geq 7$ | 9 | 5.22 | 2.7372 |
| | | $\Sigma \frac{(O-E^2)}{E} =$ | 8.3759 |

$v = 7 - 1 - 1$

$= 5$

At 5% significance level, C.V. $= 11.070$

$8.3759 < 11.070$

$\therefore$ Do not reject $H_0$. The Poisson distribution is a suitable model.

Leave blank

**Question 4 continued**

Leave blank

**Question 4 continued**

**(Total 11 marks)**

**Q4**

Leave blank

**5.** A dance studio has 800 dancers of which

452 are beginners
251 are intermediates
97 are professionals

(a) Explain in detail how a stratified sample of size 50 could be taken.

**(3)**

(b) State an advantage of stratified sampling rather than simple random sampling in this situation.

**(1)**

Independent random samples of 80 beginners and 60 intermediates are chosen. Each of these dancers is given an assessment score, $x$, based on the quality of their dancing. The results are summarised in the table below.

| | $\bar{x}$ | $s^2$ | $n$ |
|---|---|---|---|
| **Beginners** | 31.7 | 57.3 | 80 |
| **Intermediates** | 36.9 | 38.1 | 60 |

The studio manager believes that the mean score of intermediates is more than 3 points greater than the mean score of beginners.

(c) Stating your hypotheses clearly and using a 5% level of significance, test whether or not these data support the studio manager's belief.

**(7)**

a) $\frac{452}{800} \times 50 = 28.25$

$\simeq 28$

$\frac{251}{800} \times 50 = 15.6875$

$\simeq 16$

$\frac{97}{800} \times 50 = 6.0625$

$\simeq 6$

Number the beginners from 1 to 452, intermediates from 1 to 251 and professionals from 1 to 97. Using random numbers, choose 28 beginners, 16 intermediates and 6 professionals.

b) The sample reflects the population structure

Leave blank

**Question 5 continued**

c) $H_0 : \mu_I - \mu_B = 3$

$H_1 : \mu_I - \mu_B > 3$

$$z = \frac{36.9 - 31.7 - 3}{\sqrt{\frac{57.3}{80} + \frac{38.1}{60}}}$$

$$= \frac{2.2}{\sqrt{\frac{1081}{800}}}$$

$= 1.89258$

$\simeq 1.89$

$1.89 > 1.6449$

$\therefore$ Reject $H_0$. Manager's belief is correct.

Leave blank

**Question 5 continued**

Leave blank

**Question 5 continued**

**(Total 11 marks)**

**Q5**

Leave blank

**6.** A company produces a certain type of mug. The masses of these mugs are normally distributed with mean $\mu$ and standard deviation 1.2 grams. A random sample of 5 mugs is taken and the mass, in grams, of each mug is measured. The results are given below.

229.1 229.6 230.9 231.2 231.7

(a) Find a 95% confidence interval for $\mu$, giving your limits correct to 1 decimal place.

**(4)**

Sonia plans to take 20 random samples, each of 5 mugs. A 95% confidence interval for $\mu$ is to be determined for each sample.

(b) Find the probability that more than 3 of these intervals will not contain $\mu$.

**(3)**

a) $\bar{x} = \frac{229.1 + 229.6 + 230.9 + 231.2 + 231.7}{5}$

$= \frac{1152.5}{5}$

$= 230.5$

$230.5 + 1.96\left(\frac{1.2}{\sqrt{5}}\right) = 231.552$

$\simeq 231.6$

$230.5 - 1.96\left(\frac{1.2}{\sqrt{5}}\right) = 229.448$

$\simeq 229.4$

$(229.4, 231.6)$

b) $X$ = no. of CI which does not contain $\mu$

$X \sim B(20, 0.05)$

$P(X > 3) = 1 - P(X \leq 3)$

$= 1 - 0.9841$

$= 0.0159$

Leave blank

**Question 6 continued**

**Q6**

**(Total 7 marks)**

Leave blank

**7.** The independent random variables $X$ and $Y$ are such that

$$X \sim \mathrm{N}(30, 4.5^2) \text{ and } Y \sim \mathrm{N}(20, 3.5^2)$$

The random variables $X_1$, $X_2$ and $X_3$ are independent and each has the same distribution as $X$. The random variables $Y_1$ and $Y_2$ are independent and each has the same distribution as $Y$.

Given that the random variable $A$ is defined as

$$A = \frac{X_1 + X_2 + X_3 + Y_1 + Y_2}{5}$$

(a) find $\mathrm{P}(A < 24)$

**(6)**

The random variable $W$ is such that $W \sim \mathrm{N}(\mu, 2.8^2)$

Given that $\mathrm{P}(W - X < 4) = 0.1$ and that $W$ and $X$ are independent,

(b) find the value of $\mu$, giving your answer to 3 significant figures.

**(6)**

a) $E(A) = \frac{3(30) + 2(20)}{5}$

$= 26$

$\mathrm{Var}(A) = \frac{3(4.5)^2 + 2(3.5)^2}{5^2}$

$= 3.41$

$A \sim N(26, 3.41)$

$P(A < 24) = P\left(z < \frac{24-26}{\sqrt{3.41}}\right)$

$= P(z < -1.083)$

$= 1 - 0.8599$

$= 0.1401$

$\simeq 0.140$

b) $\mathrm{Var}(W - X) = 2.8^2 + 4.5^2$

$= 28.09$

$P(W - X < 4) = 0.1$

$P\left(z < \frac{4 - (\mu - 30)}{\sqrt{28.09}}\right) = 0.1$

$\frac{4 - (\mu - 30)}{\sqrt{28.09}} = -1.2816$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

**Question 7 continued**

$-\mu = -1.2816\left(\sqrt{28.09}\right) - 34$

$\mu = 40.792$

$\simeq 40.8$

Leave blank

**Question 7 continued**

Leave blank

**Question 7 continued**

**(Total 12 marks)**

**Q7**

Leave blank

**8.** The random variable $X$ has a continuous uniform distribution over the interval $[\alpha + 3, 2\alpha + 9]$ where $\alpha$ is a constant.

The mean of a random sample of size $n$, taken from this distribution, is denoted by $\overline{X}$

(a) Show that $\overline{X}$ is a biased estimator of $\alpha$

**(2)**

(b) Hence find the bias, in terms of $\alpha$, when $\overline{X}$ is used as an estimator of $\alpha$

**(1)**

Given that $Y = \frac{2\overline{X}}{3} + k$ is an unbiased estimator of $\alpha$

(c) find the value of the constant $k$

**(2)**

A random sample of 8 values of $X$ is taken and the results are as follows

4.8 5.8 6.5 7.1 8.2 9.5 9.9 10.6

(d) Use the sample to estimate the maximum value that $X$ can take.

**(3)**

a) $E(\bar{X}) = \frac{\alpha+3+2\alpha+9}{2}$

$= \frac{3\alpha+12}{2}$

$= \frac{3\alpha}{2} + 6 \neq \alpha$

$\therefore \bar{x}$ is a biased estimator

b) $\frac{3\alpha}{2} + 6 - \alpha = \frac{\alpha+12}{2}$

c) $\frac{2(\frac{3\alpha}{2}+6)}{3} + k = \alpha$

$\alpha + 4 + k = \alpha$

$k = -4$

d) $\bar{x} = (4.8 + 5.8 + 6.5 + 7.1 + 8.2 + 9.5 + 9.9 + 10.6)\frac{1}{8}$

$= (62.4)\frac{1}{8}$

$= 7.8$

$\alpha: \frac{2}{3}(7.8) - 4 = 1.2$

Max: $2(1.2) + 9 = 11.4$

Leave blank

**Question 8 continued**

Leave blank

**Question 8 continued**

Q8

**(Total 8 marks)**

**TOTAL FOR PAPER: 75 MARKS**

**END**