OCR Spec Statistics Practice Paper 1

1. A set of bivariate data is summarised as follow.

$$n = 20, \Sigma x = 111.1, \Sigma y = 91.1, \Sigma y^2 = 584.69, \Sigma xy = 533.09$$

The product moment correlation coefficient for these data is 0.1736.

a) Find the value of Σx^2, giving your answer to two decimal places.

[3]

The least squares regression line of y on x for these data is calculated.

b) (i) Explain the meaning of "least squares" in this context

[2]

(ii) Is the least squares regression line likely to be a good model for this data? Explain your answer.

[2]

2. A restaurant has customers for both lunch and dinner. The owner of the restaurant believes that whether a customer attends for lunch or dinner affects the probability that the customer will order dessert.

The owner collects the following data on the customers in the restaurant.

	Lunch	Dinner
Dessert	7	35
No Dessert	14	24

Perform a suitable statistical test, at the 5% significance level, to test the owner's belief.

[8]

3. A post office has two types of customers. One type of customer wants to post a letter, and the other type of customer wants to post a parcel.

Customers who want to post a letter arrive at a constant rate of 20 per hour.

Customers who want to post a parcel arrive at a constant rate of 5 per hour.

a) Calculate the probability that, in any given 15 minute period, exactly 5 customers arrive to post a letter.

[2]

Customers arriving to post a letter and customers arriving to post a parcel arrive independently from each other.

b) Calculate the probability that more than 20 customers arrive in a one hour period.

[2]

For every customer who enters the post office, there is a 0.8 probability that the customer wants to post a letter, and a 0.2 probability that the customer wants to post a parcel.

The random variable X represents the number of customers up to and including the first customer who wants to post a parcel from when the post office opens in the morning.

c) Work out $\mathrm{Var}(X)$.

[2]

4. A game is played using a set of 24 cards. Of these 24 cards, 16 are *number* cards and 8 are *action* cards. The deck of 24 cards are shuffled randomly.

a) At the start of the game, a player draws a hand of 5 cards. Work out the probability that the player's hand has exactly one *action* card. Give your answer as an exact fraction.

[2]

These 5 cards are returned to the deck and the full deck of 24 cards is again shuffled randomly.

b) Work out the probability that all 8 of the *action* cards are next to each other in the deck. Give your answer as an exact fraction.

[3]

c) Work out the probability that none of the 8 *action* cards are next to each other in the deck. Give your answer as an exact fraction.

[3]

5. James can travel to school by bus or by train. He must arrive at school by 9am.

If he catches the bus, he leaves home at 8am, and the length of the journey in minutes can be modelled by a $N(55, 10^2)$ distribution.

If he catches the train, he leaves home at 8:15am, and the length of the journey in minutes can be modelled by a $N(42, 2^2)$ distribution.

a) Show that James is more likely to be late to school if he travels by bus. Do **not** use statistical functions of your calculator or statistical tables.

[4]

In a particular week, James travelled to school once by bus and four times by train.

b) Calculate the probability that the total time of his journeys by train was more than three times the time of his journey on the bus.

[6]

6. Two judges ranked the performance of 8 dancers, labelled A to H, performing a breakdancing routine. Their ranks are shown in the table.

	1	2	3	4	5	6	7	8
Judge 1	F	B	A	D	E	C	H	G
Judge 2	F	B	C	A	H	G	D	E

a) Calculate Spearman's rank correlation coefficient for these data.

[4]

b) Perform a hypothesis test at the 5% level for this data using your answer to a) to determine if there is correlation between the judges' rankings. State your hypotheses clearly.

[4]

7. It is known that the median length of time people take to complete a particular crossword puzzle is 22 minutes.

a) A group of people from a puzzle club complete the puzzle. Their times, in minutes, are as follows.

7	10	13	18	20	23	39

Perform a sign test to determine if the median time taken by members this club is less than 22 minutes.

[5]

b) Members of a second puzzle club also complete the puzzle. Their times, in minutes, are as follows.

11	14	19	21	22

Perform a suitable hypothesis test to determine if there is a difference between the median times to complete the puzzle between the two puzzle clubs.

[8]

8. A continuous random variable X has probability density function given by

$$f(x) = \frac{1}{k(1 + x^2)} \qquad x \in \mathbb{R}$$

where k is a constant.

a) Show that $k = \pi$.

[4]

b) State the curve $y = f(x)$, showing the coordinates of any points of intersection with the coordinate axes.

[2]

c) State the median of X.

[1]

d) State the mode of X.

[1]

e) Show that $E(X)$ does not exist.

[3]

f) Find the cumulative distribution function of X

[4]

1.

a)

$$r = \frac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma y}{\sqrt{\left(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right)\left(\Sigma y^2 - \frac{1}{n}(\Sigma y)^2\right)}}$$

$$r^2\left(\Sigma y^2 - \frac{1}{n}(\Sigma y)^2\right) = \frac{\left(\Sigma xy - \frac{1}{n}\Sigma x \Sigma y\right)^2}{\left(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2\right)}$$

$$\Sigma x^2 = \frac{1}{n}(\Sigma x)^2 + \frac{\left(\Sigma xy - \frac{1}{n}\Sigma x \Sigma y\right)^2}{r^2\left(\Sigma y^2 - \frac{1}{n}(\Sigma y)^2\right)}$$

M1M1

$$= 759.99$$

A1

b) (i)

The **straight line** that **minimises**...

B1

The **sum** of the **squared vertical distances** (from the points to the line)

B1

c)

No

B1

Because the PMCC is small/close to zero/the correlation is not strong/the points lie far from a straight line

B1

2.

H_0: Lunch/Dinner is independent of dessert

H_1: Lunch/Dinner is not independent of dessert

B2

Expected values

	Lunch	Dinner
Dessert	11.025	30.975
No Dessert	9.975	28.025

M1

$$X^2_{\text{Yates}} = \frac{(|7-11.025|-0.5)^2}{11.025} + \frac{(|35-30.975|-0.5)^2}{30.975} + \frac{(|14-9.975|-0.5)^2}{9.975} + \frac{(|24-28.025|-0.5)^2}{28.025}$$

M1 (no Yates = M0)

$= 3.217 \ldots$

A1ft (no Yates = A0)

$\nu = 1$, CV=3.41

A1

Do not reject the null hypothesis,

M1

Insufficient evidence to conclude that lunch/dinner are not independent of dessert.

A1ft

3.

a) $X \sim \text{Po}(5)$

M1

$P(X = 5) = 0.175$

A1

b) $X \sim \text{Po}(25)$

M1

$P(X > 20) = 1 - P(X < 20)$

$= 1 - 0.18549$

$= 0.815$

A1

c) $X \sim \text{Geo}(0.2)$

M1

$$\text{Var}(X) = \frac{1 - 0.2}{0.2^2} = 20$$

A1

4.

a) $\frac{16}{24} \times \frac{15}{23} \times \frac{14}{22} \times \frac{13}{21} \times \frac{8}{20} \times 5$

M1 (with or without 5)

$$= \frac{260}{759}$$

A1

b) The 8 action cards can begin at any location from 1 to 17, and can be in any order, as can the number cards. So probability is

$$\frac{17 \times 8! \times 16!}{24!} = \frac{1}{43263}$$

M1 for 17

M1 for $\frac{8! \times 16!}{24!}$

A1

c) With the 16 number cards laid out, there are 17 locations the action cards can fit in. So the probability is

$$\frac{\binom{17}{8} \times 8! \times 16!}{24!} = \frac{130}{3933}$$

M1 for $\binom{17}{8}$

M1 for $\frac{8! \times 16!}{24!}$

A1

5.

a) P(late by bus) = $P(N(55, 10^2) > 60) = P(Z > \frac{5}{10})$

M1A1

P(late by train) = $P(N(42,2^2) > 45) = P(Z > \frac{3}{2})$

M1

$\frac{3}{2} > \frac{1}{2}$ so James is more likely to be late by bus.

A1

b) Train total $\sim N(168, 16)$

M1A1

Bus total $\sim N(165, 900)$

M1A1

Train – Bus $\sim N(3, 916)$

M1

P(Train > Bus) = P(Train – Bus > 0) = 0.539

A1

6.

a)

	A	B	C	D	E	F	G	H
Judge 1	3	2	6	4	5	1	8	7
Judge 2	4	2	3	7	8	1	6	5
d	-1	0	3	-3	-3	0	2	2
d^2	1	0	9	9	9	0	4	4

M1

$$1-\frac{6\Sigma d^2}{n(n^2-1)}=1-\frac{6\times 36}{8\times 63}$$

M1A1

$$=\frac{4}{7}\approx 0.571$$

A1

b)

H_0: No association between judges' rankings

H_1: Association between judges' rankings

B1

CV at $n=8, \alpha=0.05$ is 0.7391

B1

0.571 < 0.7391 Do not reject H_0

M1 (ft CV and correlation)

Insufficient evidence to conclude there is association between the judges' rankings.

A1

7.

a) Single-sample sign test

H_0: Median of this club is 22

H_1: Median of this club is less than 22

B1

If X is the number of values above 22 then $X \sim B(7, 0.5)$

M1

Observed $X = 2, P(X \leq 2) = 0.227$

M1

$0.227 > 0.05$, do not reject H_0

M1

Insufficient evidence to conclude that the median time for this club is less than 22 minutes.

A1

b) Wilcoxon rank-sum test

Club 1	Rank	Club 2	Rank
7	1		
10	2		
		11	3
13	4		
		14	5
18	6		
		19	7
20	8		
		21	9
		22	10
23	11		
39	12		

$m = 5, n = 7$

$R_m = 34$

M1M1A1

$m(m + n + 1) - R_m = 31$

M1A1

$W = 31$

CV=13

B1

$31 > 13$, do not reject H_0

M1

Insufficient evidence to conclude that the **median** time of the two clubs to complete the puzzle is different.

A1

8.

a) $\int \frac{1}{k(1+x^2)}\,\mathrm{d}x = \frac{1}{k}\arctan x\,(+c)$

B1

$$\lim_{t\to-\infty}\left[\frac{1}{k}\arctan x\right]_t^0 + \lim_{u\to\infty}\left[\frac{1}{k}\arctan x\right]_0^u = 1$$

M1

$$\frac{1}{k}\left(\lim_{u\to\infty}\arctan u - \lim_{t\to-\infty}\arctan t\right) = 1$$

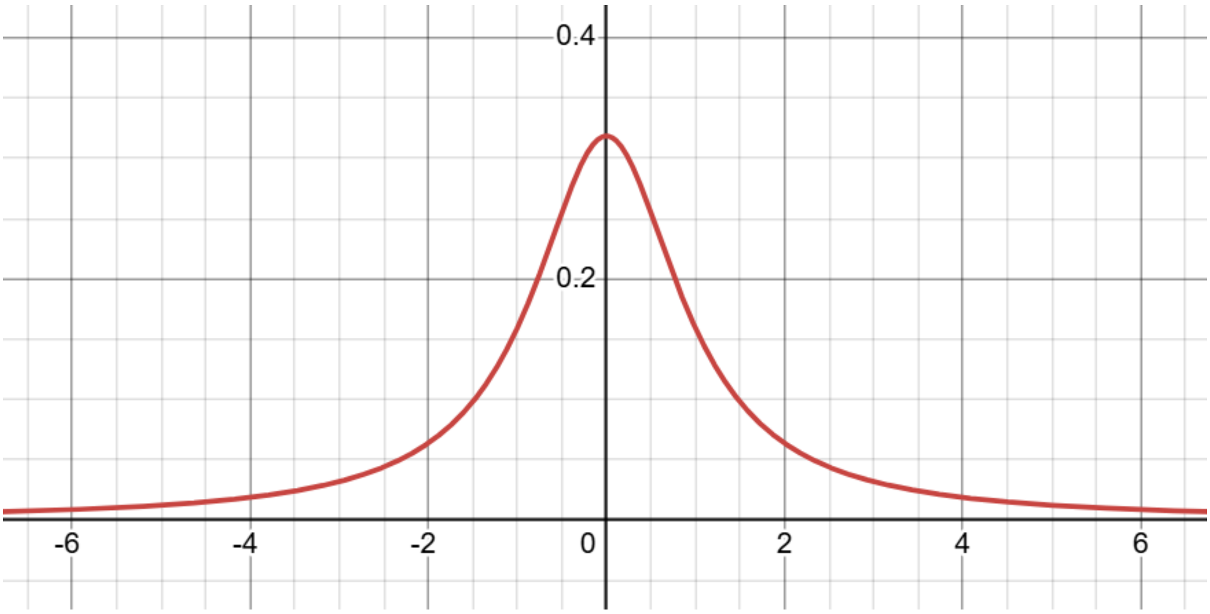
$$\frac{1}{k}\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = 1$$

M1

$$k = \pi$$

A1

b)



B1 correct shape

B1 intersection at $\left(0, \frac{1}{\pi}\right)$

c) 0

B1

d) 0

B1

e) $E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} \mathrm{d}x$

M1

= $\lim_{t \to -\infty} \left[\frac{1}{2\pi} \ln(1 + x^2)\right]_t^0 + \lim_{u \to \infty} \left[\frac{1}{2\pi} \ln(1 + x^2)\right]_0^u$

M1

Which does not exist as $\lim_{x \to \infty} \ln x$ does not exist.

A1

f) $P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\pi(1+y^2)} \mathrm{d}y$

M1

$$= \lim_{t \to -\infty} \left[\frac{1}{\pi} \arctan y\right]_t^x$$

M1

$$= \frac{1}{\pi} \arctan x - \lim_{t \to -\infty} \frac{1}{\pi} \arctan t$$

$$= \frac{1}{\pi} \arctan x - \frac{1}{\pi} \times -\frac{\pi}{2}$$

M1

$$= \frac{1}{\pi} \arctan x + \frac{1}{2}$$

A1