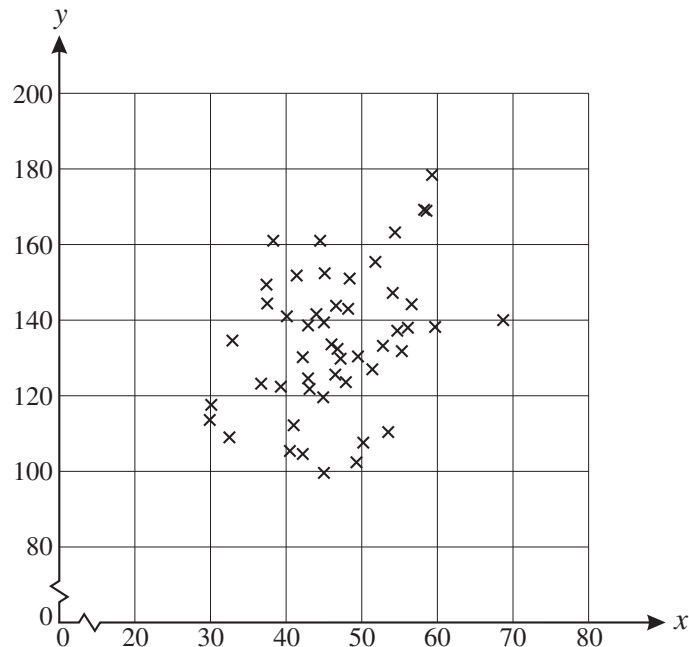


- 1 An investment analyst thinks that there may be correlation between the cost of oil, x dollars per barrel, and the price of a particular share, y pence. The analyst selects 50 days at random and records the values of x and y . Summary statistics for these data are shown below, together with a scatter diagram.

$$\Sigma x = 2331.3 \quad \Sigma y = 6724.3 \quad \Sigma x^2 = 111\,984 \quad \Sigma y^2 = 921\,361 \quad \Sigma xy = 316\,345 \quad n = 50$$



- (i) Calculate the sample product moment correlation coefficient. [5]
- (ii) Carry out a hypothesis test at the 5% significance level to investigate the analyst's belief. State your hypotheses clearly, defining any symbols which you use. [6]
- (iii) An assumption that there is a bivariate Normal distribution is required for this test to be valid. State whether it is the sample or the population which is required to have such a distribution. State, with a reason, whether in this case the assumption appears to be justified. [3]
- (iv) Explain why a 2-tail test is appropriate even though it is clear from the scatter diagram that the sample has a positive correlation coefficient. [2]

- 2 Jess is watching a shower of meteors (shooting stars). During the shower, she sees meteors at an average rate of 1.3 per minute.

(i) State conditions required for a Poisson distribution to be a suitable model for the number of meteors which Jess sees during a randomly selected minute. [2]

You may assume that these conditions are satisfied.

(ii) Find the probability that, during one minute, Jess sees

(A) exactly one meteor,

(B) at least 4 meteors. [4]

(iii) Find the probability that, in a period of 10 minutes, Jess sees exactly 10 meteors. [3]

(iv) Use a suitable approximating distribution to find the probability that Jess sees a total of at least 100 meteors during a period of one hour. [5]

(v) Jess watches the shower for t minutes. She wishes to be at least 99% certain that she will see one or more meteors. Find the smallest possible integer value of t . [5]

- 3 Intensity of light is measured in lumens. The random variable X represents the intensity of the light from a standard 100 watt light bulb. X is Normally distributed with mean 1720 and standard deviation 90. You may assume that the intensities for different bulbs are independent.

(i) Show that $P(X < 1700) = 0.4121$. [4]

(ii) These bulbs are sold in packs of 4. Find the probability that the intensities of exactly 2 of the 4 bulbs in a randomly chosen pack are below 1700 lumens. [3]

(iii) Use a suitable approximating distribution to find the probability that the intensities of at least 20 out of 40 randomly selected bulbs are below 1700 lumens. [5]

A manufacturer claims that the average intensity of its 25 watt low energy light bulbs is 1720 lumens. A consumer organisation suspects that the true figure may be lower than this. The intensities of a random sample of 20 of these bulbs are measured. A hypothesis test is then carried out to check the claim.

(iv) Write down a suitable null hypothesis and explain briefly why the alternative hypothesis should be $H_1 : \mu < 1720$. State the meaning of μ . [3]

(v) Given that the standard deviation of the intensity of such bulbs is 90 lumens and that the mean intensity of the sample of 20 bulbs is 1703 lumens, carry out the test at the 5% significance level. [5]

- 4 In a traffic survey a random sample of 400 cars passing a particular location during the rush hour is selected. The type of car and the sex of the driver are classified as follows.

		Sex		Row totals
		Male	Female	
Type of car	Hatchback	96	36	132
	Saloon	77	35	112
	People carrier	38	44	82
	4WD	19	8	27
	Sports car	22	25	47
Column totals		252	148	400

- (i) Carry out a test at the 5% significance level to examine whether there is any association between type of car and sex of driver. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. [12]
- (ii) For each type of car, comment briefly on how the number of drivers of each sex compares with what would be expected if there were no association. [5]

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations, is given to all schools that receive assessment material and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1PB.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

4767 Statistics 2

Question 1

(i)	<p>EITHER:</p> $S_{xy} = \sum xy - \frac{1}{n} \sum x \sum y = 316345 - \frac{1}{50} \times 2331.3 \times 6724.3$ $= 2817.8$ $S_{xx} = \sum x^2 - \frac{1}{n} (\sum x)^2 = 111984 - \frac{1}{50} \times 2331.3^2 = 3284.8$ $S_{yy} = \sum y^2 - \frac{1}{n} (\sum y)^2 = 921361 - \frac{1}{50} \times 6724.3^2 = 17036.8$ $r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{2817.8}{\sqrt{3284.8 \times 17036.8}} = 0.377$ <p>OR:</p> $\text{cov}(x, y) = \frac{\sum xy}{n} - \bar{x} \bar{y} = 316345/50 - 46.626 \times 134.486$ $= 56.356$ $\text{rmsd}(x) = \sqrt{\frac{S_{xx}}{n}} = \sqrt{(3284.8/50)} = \sqrt{65.696} = 8.105$ $\text{rmsd}(y) = \sqrt{\frac{S_{yy}}{n}} = \sqrt{(17036.8/50)} = \sqrt{340.736} = 18.459$ $r = \frac{\text{cov}(x, y)}{\text{rmsd}(x) \text{rmsd}(y)} = \frac{56.356}{8.105 \times 18.459} = 0.377$	<p>M1 for method for S_{xy}</p> <p>M1 for method for at least one of S_{xx} or S_{yy}</p> <p>A1 for at least one of S_{xy}, S_{xx} or S_{yy} correct</p> <p>M1 for structure of r A1 (AWRT 0.38)</p> <p>M1 for method for cov (x,y)</p> <p>M1 for method for at least one msd A1 for at least one of cov(x,y), rmsd(x) or rmsd(y) correct</p> <p>M1 for structure of r A1 (AWRT 0.38)</p>	5
(ii)	<p>$H_0: \rho = 0$ $H_1: \rho \neq 0$ (two-tailed test)</p> <p>where ρ is the population correlation coefficient</p> <p>For $n = 50$, 5% critical value = 0.2787</p> <p>Since $0.377 > 0.2787$ we can reject H_0:</p> <p>There is sufficient evidence at the 5% level to suggest that there is correlation between oil price and share cost</p>	<p>B1 for H_0, H_1 in symbols B1 for defining ρ B1FT for critical value</p> <p>M1 for sensible comparison leading to a conclusion A1 for result B1 FT for conclusion in context</p>	6
(iii)	<p>Population</p> <p>The scatter diagram has a roughly elliptical shape, hence the assumption is justified.</p>	<p>B1 B1 elliptical shape E1 conclusion</p>	3
(iv)	<p>Because the alternative hypothesis should be decided without referring to the sample data and there is no suggestion that the correlation should be positive rather than negative.</p>	<p>E1 E1</p>	2
		TOTAL	16

Question 2

(i)	Meteors are seen randomly and independently There is a uniform (mean) rate of occurrence of meteor sightings	B1 B1	2
(ii)	(A) <i>Either</i> $P(X = 1) = 0.6268 - 0.2725 = 0.3543$ <i>Or</i> $P(X = 1) = e^{-1.3} \frac{1.3^1}{1!} = 0.3543$ (B) Using tables: $P(X \geq 4) = 1 - P(X \leq 3)$ $= 1 - 0.9569$ $= 0.0431$	M1 for appropriate use of tables or calculation A1 M1 for appropriate probability calculation A1	4
(iii)	$\lambda = 10 \times 1.3 = 13$ $P(X = 10) = e^{-13} \frac{13^{10}}{10!} = 0.0859$	B1 for mean M1 for calculation A1 CAO	3
(iv)	Mean no. per hour = $60 \times 1.3 = 78$ Normal approx. to the Poisson, $X \sim N(78, 78)$ $P(X \geq 100) = P\left(Z > \frac{99.5 - 78}{\sqrt{78}}\right)$ $= P(Z > 2.434) = 1 - \Phi(2.434)$ $= 1 - 0.9926 = 0.0074$	B1 for Normal approx. B1 for correct parameters (SOI) B1 for continuity corr. M1 for correct Normal probability calculation using correct tail A1 CAO, (but FT wrong or omitted CC)	5
(v)	<i>Either</i> $P(\text{At least one}) = 1 - e^{-\lambda} \frac{\lambda^0}{0!} = 1 - e^{-\lambda} \geq 0.99$ $e^{-\lambda} \leq 0.01$ $-\lambda \leq \ln 0.01$, so $\lambda \geq 4.605$ $1.3 t \geq 4.605$, so $t \geq 3.54$ Answer $t = 4$ <i>Or</i> $t = 1, \lambda = 1.3, P(\text{At least one}) = 1 - e^{-1.3} = 0.7275$ $t = 2, \lambda = 2.6, P(\text{At least one}) = 1 - e^{-2.6} = 0.9257$ $t = 3, \lambda = 3.9, P(\text{At least one}) = 1 - e^{-3.9} = 0.9798$ $t = 4, \lambda = 5.2, P(\text{At least one}) = 1 - e^{-5.2} = 0.9944$ Answer $t = 4$	M1 formation of equation/inequality using $P(X \geq 1) = 1 - P(X = 0)$ with Poisson distribution. A1 for correct equation/inequality M1 for logs A1 for 3.54 A1 for t (correctly justified) M1 at least one trial with any value of t A1 correct probability. M1 trial with either $t = 3$ or $t = 4$ A1 correct probability of $t = 3$ and $t = 4$ A1 for t	5
		TOTAL	19

Question 3

(i)	$X \sim N(1720, 90^2)$ $P(X < 1700) = P\left(Z < \frac{1700 - 1720}{90}\right)$ $= P(Z < -0.2222)$ $= \Phi(-0.2222) = 1 - \Phi(0.2222)$ $= 1 - 0.5879$ $= 0.4121$	M1 for standardising A1 M1 use of tables (correct tail) A1 CAO NB ANSWER GIVEN	4
(ii)	$P(2 \text{ of } 4 \text{ below } 1700)$ $= \binom{4}{2} \times 0.4121^2 \times 0.5879^2 = 0.3522$	M1 for coefficient M1 for $0.4121^2 \times 0.5879^2$ A1 FT (min 2sf)	3
(iii)	Normal approx with $\mu = np = 40 \times 0.4121 = 16.48$ $\sigma^2 = npq = 40 \times 0.4121 \times 0.5879 = 9.691$ $P(X \geq 20) = P\left(Z \geq \frac{19.5 - 16.48}{\sqrt{9.691}}\right)$ $= P(Z \geq 0.9701) = 1 - \Phi(0.9701)$ $= 1 - 0.8340 = 0.1660$	B1 B1 B1 for correct continuity corr. M1 for correct Normal probability calculation using correct tail A1 CAO, (but FT wrong or omitted CC)	5
(iv)	$H_0: \mu = 1720;$ H_1 is of this form since the consumer organisation suspects that the mean is below 1720 μ denotes the mean intensity of 25 Watt low energy bulbs made by this manufacturer.	B1 E1 B1 for definition of μ	3
(v)	Test statistic $= \frac{1703 - 1720}{90/\sqrt{20}} = \frac{-17}{20.12}$ $= -0.8447$ Lower 5% level 1 tailed critical value of $z = -1.645$ $-0.8447 > -1.645$ so not significant. There is not sufficient evidence to reject H_0 There is insufficient evidence to conclude that the mean intensity of bulbs made by this manufacturer is less than 1720	M1 must include $\sqrt{20}$ A1FT B1 for -1.645 No FT from here if wrong. Must be -1.645 unless it is clear that absolute values are being used. M1 for sensible comparison leading to a conclusion. FT only candidate's test statistic A1 for conclusion in words in context	5
		TOTAL	20

Question 4

(i)	<p>H_0: no association between type of car and sex; H_1: some association between type of car and sex;</p> <table><tr><th>EXPECTED</th><th>Male</th><th>Female</th></tr><tr><td>Hatchback</td><td>83.16</td><td>48.84</td></tr><tr><td>Saloon</td><td>70.56</td><td>41.44</td></tr><tr><td>People carrier</td><td>51.66</td><td>30.34</td></tr><tr><td>4WD</td><td>17.01</td><td>9.99</td></tr><tr><td>Sports car</td><td>29.61</td><td>17.39</td></tr></table> <table><tr><th>CONTRIBUTION</th><th>Male</th><th>Female</th></tr><tr><td>Hatchback</td><td>1.98</td><td>3.38</td></tr><tr><td>Saloon</td><td>0.59</td><td>1.00</td></tr><tr><td>People carrier</td><td>3.61</td><td>6.15</td></tr><tr><td>4WD</td><td>0.23</td><td>0.40</td></tr><tr><td>Sports car</td><td>1.96</td><td>3.33</td></tr></table> <p>$\chi^2 = 22.62$</p> <p>Refer to χ^2_4 Critical value at 5% level = 9.488</p> <p>$22.62 > 9.488$ Result is significant There is evidence to suggest that there is some association between sex and type of car.</p> <p>NB if H_0 H_1 reversed, or 'correlation' mentioned, do not award first B1 or final A1</p>	EXPECTED	Male	Female	Hatchback	83.16	48.84	Saloon	70.56	41.44	People carrier	51.66	30.34	4WD	17.01	9.99	Sports car	29.61	17.39	CONTRIBUTION	Male	Female	Hatchback	1.98	3.38	Saloon	0.59	1.00	People carrier	3.61	6.15	4WD	0.23	0.40	Sports car	1.96	3.33	<p>B1</p> <p>M1 A2 for expected values (to 2 dp)</p> <p>(allow A1 for at least one row or column correct)</p> <p>M1 for valid attempt at $(O-E)^2/E$ A1 for all correct NB These M1A1 marks cannot be implied by a correct final value of χ^2</p> <p>M1 for summation A1 for χ^2 CAO</p> <p>B1 for 4 deg of f B1 CAO for cv</p> <p>M1 sensible comparison leading to a conclusion A1</p>	12
EXPECTED	Male	Female																																					
Hatchback	83.16	48.84																																					
Saloon	70.56	41.44																																					
People carrier	51.66	30.34																																					
4WD	17.01	9.99																																					
Sports car	29.61	17.39																																					
CONTRIBUTION	Male	Female																																					
Hatchback	1.98	3.38																																					
Saloon	0.59	1.00																																					
People carrier	3.61	6.15																																					
4WD	0.23	0.40																																					
Sports car	1.96	3.33																																					
(ii)	<ul style="list-style-type: none">In hatchbacks, male drivers are more frequent than expected.In saloons, male drivers are slightly more frequent than expected.In people carriers, female drivers are much more frequent than expected.In 4WDs the numbers are roughly as expectedIn sports cars, female drivers are more frequent than expected.	<p>E1</p> <p>E1</p> <p>E1</p> <p>E1</p> <p>E1</p>	5																																				
		TOTAL	17																																				

Deleted: ¶