A clothes shop manager records the weekly sales figures, £ *s*, and the average weekly   
temperature, *t* °C, for 6 weeks during the summer. The sales figures were coded so that  
*w* =**

The data are summarised as follows

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S*ww* = 50 | = 784 | = 2435 | = 119 | = 42 |

(*a*)Find S*wt* and S*tt*

**(3)**

(*b*)Write down the value of S*ss* and the value of S*st*

**(2)**

(*c*)Find the product moment correlation coefficient between *s* and *t*.

**(2)**

The manager of the clothes shop believes that a linear regression model may be appropriate

to describe these data.

(*d*)State, giving a reason, whether or not your value of the correlation coefficient supports

the manager’s belief.

**(1)**

(*e*)Find the equation of the regression line of *w* on *t*, giving your answer

in the form *w* = *a* + *bt*

**(3)**

(*f*)Hence find the equation of the regression line of *s* on *t*, giving your answer

in the form *s* = *c* + *dt*, where *c* and *d* are correct to 3 significant figures.

**(2)**

(*g*)Using your equation in part (*f*), interpret the effect of a 1°C increase in average

weekly temperature on weekly sales during the summer.

**(1)**

**Total 14 marks**

**S1 June 2017 qu.1**

A biologist is studying the behaviour of bees in a hive. Once a bee has located a source of food, it returns to the hive and performs a dance to indicate to the other bees how far away the source of the food is. The dance consists of a series of wiggles. The biologist records the distance, *d* metres, of the food source from the hive and the average number of wiggles, *w*, in the dance.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Distance, *d* m** | 30 | 50 | 80 | 100 | 150 | 400 | 500 | 650 |
| **Average number of wiggles, *w*** | 0.725 | 1.210 | 1.775 | 2.250 | 3.518 | 6.382 | 8.185 | 9.555 |

[You may use  = 33.6 = 13833 S*dd* = 394600 S*ww* = 80.481 (to 3 decimal places)]

(*a*)Show that S*dw* = 5601.

**(2)**

(*b*)State, giving a reason, which is the response variable.

**(1)**

(*c*)Calculate the product moment correlation coefficient for these data.

**(2)**

(*d*)Calculate the equation of the regression line of *w* on *d*, giving your answer in the form  
 *w*= *a* + *bd*.

**(4)**

A new source of food is located 350 m from the hive.

(*e*)(i) Use your regression equation to estimate the average number of wiggles in the corresponding dance.

(ii) Comment, giving a reason, on the reliability of your estimate.

**(2)**

**Total 11 marks**

**S1 June 2016 qu.1**

Before going on holiday to *Seapron*, Tania records the weekly rainfall (*x* mm) at *Seapron* for   
8 weeks during the summer. Her results are summarised as

= 86.8  = 985.88

(*a*)Find the standard deviation, *σx*, for these data.

**(3)**

Tania also records the number of hours of sunshine (*y* hours) per week at *Seapron* for these 8 weeks and obtains the following

** = 58 *σy =* 9.461 (correct to 4 significant figures)  = 4900.5

(*b*)Show that S*yy* = 716 (correct to 3 significant figures).

**(1)**

(*c*)Find S*xy*.

**(2)**

(*d*)Calculate the product moment correlation coefficient, *r*, for these data.

**(2)**

During Tania’s week-long holiday at *Seapron* there are 14 mm of rain and 70 hours of sunshine.

(*e*)State, giving a reason, what the effect of adding this information to the above data would be on the value of the product moment correlation coefficient.

**(2)**

**Total 10 marks**

**S1 June 2016 qu.3**

The table shows data on the number of visitors to the UK in a month, *v* (1000s), and the amount of money they spent, *m* (£ millions), for each of 8 months.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of visitors  *v* (1000s) | 2450 | 2480 | 2540 | 2420 | 2350 | 2290 | 2400 | 2460 |
| Amount of money spent  *m* (£ millions) | 1370 | 1350 | 1400 | 1330 | 1270 | 1210 | 1330 | 1350 |

You may use

*Svv* = 42587.5 *Svm* = 31512.5 *Smm* = 25187.5 Σ*v =* 19390 Σ*m* = 10610

(*a*) Find the product moment correlation coefficient between *m* and *v*.

**(2)**

(*b*) Give a reason to support fitting a regression model of the form *m* = *a* + *bv* to these data.

**(1)**

(*c*) Find the value of *b* correct to 3 decimal places.

**(2)**

(*d*) Find the equation of the regression line of *m* on *v*.

**(2)**

(*e*) Interpret your value of *b*.

**(2)**

(*f*) Use your answer to part (*d*) to estimate the amount of money spent when the number of visitors to the UK in a month is 2 500 000.

**(2)**

(*g*) Comment on the reliability of your estimate in part (*f*). Give a reason for your answer.

**(2)**

**Total 13 marks**

**S1 June 2014 qu.3**

A large company is analysing how much money it spends on paper in its offices every year.   
The number of employees, *x*, and the amount of money spent on paper, *p* (£ hundreds), in   
8 randomly selected offices are given in the table below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 8 | 9 | 12 | 14 | 7 | 3 | 16 | 19 |
| *p* (£ hundreds) | 40.5 | 36.1 | 30.4 | 39.4 | 32.6 | 31.1 | 43.4 | 45.7 |

(You may use Σ*x*2 = 1160 Σ*p* = 299.2 Σ*p*2 = 11 422 Σ*xp* = 3449.5)

(*a*) Show that *Spp* = 231.92 and find the value of *Sxx* and the value of *Sxp*.

**(5)**

(*b*) Calculate the product moment correlation coefficient between *x* and *p*.

**(2)**

The equation of the regression line of *p* on *x* is given in the form *p* = *a + bx*.

(*c*) Show that, to 3 significant figures, *b* = 0.824 and find the value of *a*.

**(4)**

(*d*) Estimate the amount of money spent on paper in an office with 10 employees.

**(2)**

(*e*) Explain the effect each additional employee has on the amount of money spent on paper.

**(1)**

Later the company realised it had made a mistake in adding up its costs, *p*. The true costs were actually half of the values recorded. The product moment correlation coefficient and the equation   
of the linear regression line are recalculated using this information.

(*f*) Write down the new value of

(i) the product moment correlation coefficient,

(ii) the gradient of the regression line.

**(2)**

**Total 16 marks**

**S1 June 2014R qu.3**

Statistical models can provide a cheap and quick way to describe a real world situation.

(*a*) Give two other reasons why statistical models are used.

**(2)**

A scientist wants to develop a model to describe the relationship between the average daily temperature, *x* °C, and her household’s daily energy consumption, *y* kWh, in winter.

A random sample of the average daily temperature and her household’s daily energy consumption are taken from 10 winter days and shown in the table.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | –0.4 | –0.2 | 0.3 | 0.8 | 1.1 | 1.4 | 1.8 | 2.1 | 2.5 | 2.6 |
| *y* | 28 | 30 | 26 | 25 | 26 | 27 | 26 | 24 | 22 | 21 |

[You may use ∑ *x*2 = 24.76 ∑ *y* = 255 ∑∑ *xy* = 283.8 S*xx* = 10.36]

(*b*) Find S*xy* for these data.

**(3)**

(*c*) Find the equation of the regression line of *y* on *x* in the form *y* = *a* + *bx*.

Give the value of *a* and the value of *b* to 3 significant figures.

**(4)**

(*d*) Give an interpretation of the value of *a*.

**(1)**

(*e*) Estimate her household’s daily energy consumption when the average daily temperature is 2°C.

**(2)**

The scientist wants to use the linear regression model to predict her household’s energy consumption in the summer.

(*f*) Discuss the reliability of using this model to predict her household’s energy consumption in the summer.

**(2)**

**Total 14 marks**

**S1 June 2015 qu.4**

The age, *t* years, and weight, *w* grams, of each of 10 coins were recorded. These data are summarised below.

∑ *t*2 = 2688 ∑ *tw* = 1760.62 ∑ *t* = 158 ∑*w* = 111.75 *Sww* = 0.16

(*a*) Find *Stt* and *Stw* for these data.

**(3)**

(*b*) Calculate, to 3 significant figures, the product moment correlation coefficient between *t* and *w*.

**(2)**

(*c*) Find the equation of the regression line of *w* on *t* in the form *w* = *a* + *bt*.

**(4)**

(*d*) State, with a reason, which variable is the explanatory variable.

**(2)**

(*e*) Using this model, estimate

(i) the weight of a coin which is 5 years old,

(ii) the effect of an increase of 4 years in age on the weight of a coin.

**(2)**

It was discovered that a coin in the original sample, which was 5 years old and weighed 20 grams, was a fake.

(*f*) State, without any further calculations, whether the exclusion of this coin would increase or decrease the value of the product moment correlation coefficient. Give a reason for your answer.

**(2)**

**Total 15 marks**

**S1 January 2012 qu.5**

A teacher took a random sample of 8 children from a class. For each child the teacher recorded the length of their left foot, *f* cm, and their height, *h* cm. The results are given in the table below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *f* | 23 | 26 | 23 | 22 | 27 | 24 | 20 | 21 |
| *h* | 135 | 144 | 134 | 136 | 140 | 134 | 130 | 132 |

(You may use ∑ *f* =186 ∑*h* =1085 S*ff* = 39.5 S*hh* =139.875 ∑ *fh* = 25 291)

(*a*) Calculate S*fh*.

**(2)**

(*b*) Find the equation of the regression line of *h* on f in the form *h* = *a* + *bf*.

Give the value of *a* and the value of *b* correct to 3 significant figures.

**(5)**

(*c*) Use your equation to estimate the height of a child with a left foot length of 25 cm.

**(2)**

(*d*) Comment on the reliability of your estimate in part (*c*), giving a reason for your answer.

**(2)**

The left foot length of the teacher is 25 cm.

(*e*) Give a reason why the equation in part (*b*) should not be used to estimate the teacher’s height.

**(1)**

**Total 12 marks**

**S1 May 2011 qu.7**

A junior judge is being trained by a senior judge to learn how to assess ice skaters. After

the training, the judges each assess 6 ice skaters *A*, *B*, *C*, *D*, *E* and *F*. They each list them

in order of preference with the best ice skater first. The results are shown in the table

below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Rank | 1 | 2 | 3 | 4 | 5 | 6 |
| Senior Judge | *A* | *B* | *D* | *C* | *F* | *E* |
| Junior Judge | *B* | *D* | *A* | *F* | *C* | *E* |

(*a*)Calculate Spearman’s rank correlation coefficient for these data.

**(5)**

(*b*)Test, at the 5% level of significance, whether or not there is evidence of a positive

correlation between the rankings of the junior judge and the senior judge. State your

hypotheses clearly.

**(4)**

(*c*)Comment on the effectiveness of the training delivered by the senior judge.

**(1)**

**Total 10 marks**

**S3 May 2017 qu.3**

**Mark scheme**

|  |  |  |  |
| --- | --- | --- | --- |
| **Question Number** | **Scheme** | **Marks** | |
| **1. (a)** |  | M1 | A1 |
|  | or  (accept awrt 74.8) | A1 |
|  |  | (3) | |
| **(b)** | or **50 000 000** (o.e.) | B1 | |
|  | **– 49 000** | B1ft | |
|  |  | (2) | |
| **(c)** | or  =,  = awrt **– 0.801** | M1, A1 | |
|  |  | (2) | |
| **(d)\*** | *r* is close to – 1 or |*r*| is close to 1 or “**strong**” (o.e.) [negative] **correlation** | B1ft | |
|  | … so “**yes**”or  **does** support the belief |
|  |  | (1) | |
| **(e)** | or *a* = 7 – *b* | M1, M1 | |
|  | So ***w* = 20.0 – 0.655*t*** | A1 | |
|  |  | (3) | |
| **(f)** | ***s*  = 20 000 – 655*t*** or ***c* = 20 000**  and ***d =*  – 655** | B1ft B1ft | |
|  |  | (2) | |
| **(g)** | **Decrease**  in sales of [£] **655** ( ignore any minus sign) | B1ft | |
|  |  | (1) | |
|  |  | **[14]** | |

Whole question assesses spec point 7 Correlation and linear regression.

\*Part (d) is AO3

|  |  |  |
| --- | --- | --- |
| **Question Number** | **Scheme** | **Marks** |
| **1. (a)** | or  (But 13833 – 8232 is M0) | M1 |
|  | = **5601** (\*) | A1 cso |
|  |  | (2) |
| **(b)** | *w*, since the number of wiggles depends on the distance or *w* depends on *d* | B1 |
|  |  | (1) |
| **(c)** | , = 0.99389… awrt **0.994** | M1, A1 |
|  |  | (2) |
| **(d)\*** | , = 0.014194… (awrt 0.014) | M1, A1 |
|  | [=0.72244..] | M1 |
|  | ***w* = 0.722 + 0.0142*d*** | A1 |
|  |  | (4) |
| **(e)\*(i)** | [ 0.722 + 0.0142 × 350 = ] **awrt: 5.7** or **5.6** | B1 |
| **(ii)** | Reliable since 350 m is in the range of the data | B1 |
|  |  | (2) |
|  |  | **[11 marks]** |

Whole question assesses spec point 7 Correlation and linear regression.

\*Part (e) is AO3

|  |  |  |
| --- | --- | --- |
| **Question Number** | **Scheme** | **Marks** |
| **3. (a)** |  | M1 |
|  | or | A1 |
|  | = 2.3478… = awrt **2.35** | A1 |
|  |  | (3) |
| **(b)** | = 716 (3 sf) but may see |  |
|  | or  or 716.08… (= 716 to 3 sf)(\*) | B1 cso |
|  |  | (1) |
| **(c)** | or | M1 |
|  | = **–133.9** (Allow –134) | A1 |
|  |  | (2) |
| **(d)** | or | M1 |
|  | = awrt **–0.753** or **–0.754** | A1 |
|  |  | (2) |
| **(e)\*** | *r* < 0 means high sunshine and low rain; this is high sunshine high rain | B1 |
|  | [this is not in keeping with the trend so] *r* is closer to 0 or |*r|* decreases | B1 |
|  |  | (2) |
|  |  | **[10 marks]** |

Parts (b), (c), (d) and (e) assess spec point 7 Correlation and linear regression.

\*Part (e) is AO3

| **Question Number** | **Scheme** | **Marks** |
| --- | --- | --- |
| **3** (a) | *r* = awrt **0.962** | M1 A1  **(2)** |
| (b) | *r* is close to 1 or a **strong** **correlation**.  [“points are close to a straight line” isB0] | B1 |
|  | [Just “positive” correlation is B0]  [Use of “relationship” or “skew” not “correlation” is B0] | **(1)** |
| (c) | *b* =  =0.739947... = 0.740 (3 dp) **0.740** (only) | M1  A1cao |
|  |  | **(2)** |
| (d) | *a* = 1326.25 – (0.7399… 2423.75) [ = – 467.2 or awrt – 467] | M1 |
|  | So ***m* = – 467+ 0.74*ν*** | A1 **(2)** |
| (e) | *b* is the money (spent) per visitor. (i.e. definition of a rate in words.)  [ignore values]  So each 1000 visitors generates an extra £0.74 million  or each visitor spends £740 oe | B1  B1ft  **(2)** |
| (f) | *m* = – 467+ 0.74  *m* = 1383 (£ million) awrt **1380** | M1  A1 |
|  |  | **(2)** |
| (g)\* | As 2500 is within the range of the data set or it involves interpolation.  The value of money spent is reliable | B1  dB1 **(2)** |
|  |  | **Total 13** |

Whole question assesses spec point 7 Correlation and linear regression.

\*Part (g) is AO3

|  |  |  |
| --- | --- | --- |
| **Question Number** | **Scheme** | **Marks** |
| **3. (a)** |  | B1 |
|  | (**\*)** | B1cso |
|  |  | M1 A1 |
|  |  |  |
|  | **awrt** 158 | A1 |
|  |  | **(5)** |
| **(b)** |  | M1 |
|  | 0.7501726031…. **awrt** 0.750 | A1 |
|  |  | **(2)** |
| **(c)** | **(\*)** | M1 A1cso |
|  | **awrt** 28.3 | M1 A1 |
|  |  | **(4)** |
| **(d)** | = 36.57552…. **awrt** £3700 | M1 A1 **(2)** |
| **(e)** | Goes up £82.40 | B1 |
|  |  | **(1)** |
| **(f)\*** | (i) *r* = 0.750 | B1ft |
|  | (ii)*b* = 0.412 | B1 **(2)** |
|  |  | **[Total 16]** |

Whole question assesses spec point 7 Correlation and linear regression.

\*Part (f) is AO3

|  |  |  |
| --- | --- | --- |
| **Question** | **Scheme** | **Marks** |
| **4. (a)** | To simplify (or represent) a real world problem (o.e.)  To improve understanding (o.e.)  To analyse a real world problem or can change variables/replicate easily (oe)  To make predictions or find estimates (o.e.) | B1g  B1h |
|  |  | (2) |
| **(b)** | S*xy* = = **22.2** | B1  M1,A1cao |
|  |  | (3) |
| **(c)** | (A1 for awrt 2.1) | M1A1  M1  A1 |
|  |  | (4) |
| **(d)** | (28.1 kWh) of energy are used when the temperature is 0[°C] | B1 |
|  |  | (1) |
| **(e)** | awrt **23.8** | M1  A1 |
|  |  | (2) |
| **(f)\*** | The regression model is based on temperatures from the winter,  so not reliable in the summer. | B1  dB1 |
|  | Stating it **is** reliable (whatever the reason) is B0B0 | (2) |
|  |  | (14 marks) |

Whole question assesses spec point 7 Correlation and linear regression.

\*Part (f) is AO3

|  |  |  |
| --- | --- | --- |
| Question Number | Scheme | Marks |
| **5 (a)** | *Stt*= = 191.6 awrt 191.6 | M1  A1 |
|  | *S*tw = = -5.03 awrt -5.03 | A1 |
|  |  | (3) |
| **(b)** | r =  = - 0.908469… awrt -0.908(5) | M1A1 |
|  |  | (2) |
| **(c)** | *b* =  awrt -0.026 | M1 A1 |
|  |  |  |
|  | *a* = 11.175 + 0.0263 | M1 |
|  | = 11.59 |  |
|  | *w* = 11.6 – 0.0263*t* | A1 |
|  |  | (4) |
| **(d)** | The explanatory variable is the age of each coin. This is because the age is set and the weight varies. | B1 B1 |
|  |  | (2) |
| **(e) (i)** | awrt 11.5 | B1 |
| **(ii)** | Decrease(in weight of coin of 0.1052 g) = 0.1 or –0.1 or increase of –0.1  awrt (–0.1) | B1 |
|  |  | (2) |
| **(f)\*** | Decrease; removing the fake will result in a better linear fit so *r* will be closer to -1 | B1;B1 |
|  |  | (2) |
|  |  | **(15 marks)** |

Whole question assesses spec point 7 Correlation and linear regression.

\*Part (f) is AO3

|  |  |  |
| --- | --- | --- |
| Question Number | Scheme | Marks |
| **7.** |  |  |
| **(a)** |  | M1 |
|  | = 64.75 (accept 64.8) | A1 |
|  |  | (2) |
| **(b)** | *b* =, = 1.6392…. (awrt 1.6) | M1, A1 |
|  | *a* = , = 97.512… (awrt 97.5) | M1, A1 |
|  | *h* = 97.5 + 1.64*f* | A1ft (dep on M1M1) |
|  |  | (5) |
| **(c)** | *h* = 97.5+ 1.6425 , = 138~139 (final answer in [138, 139]) | M1, A1 |
|  |  | (2) |
| **(d)\*** | Should be reliable, since 25 cm(or *f* or footlength) is within the range of the data | B1, B1 |
|  |  | (2) |
| **(e)\*** | Line is for children – a different equation would apply to adults  or  Children are still growing, height will increase more than foot length | B1 |
|  |  | (1) |
|  |  | **12** |

Whole question assesses spec point 7 Correlation and linear regression.

\*Parts (d) and (e) are AO3

|  |  |  |
| --- | --- | --- |
| Question  Number | Scheme | Marks |
| **3.** |  |  |
| **(a)** | |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Senior Judge | Junior Judge |  |  | | *A* | 1 | 3 | -2 | 4 | | *B* | 2 | 1 | 1 | 1 | | *C* | 4 | 5 | -1 | 1 | | *D* | 3 | 2 | 1 | 1 | | *E* | 6 | 6 | 0 | 0 | | *F* | 5 | 4 | 1 | 1 | |  |  |  |  | 8 | | M1A1 |
|  | or 62 | A1 |
|  |  | dM1A1 |
|  |  | (5) |
| **(b)** |  | B1 |
|  | Critical value | B1 |
|  | (0.771<0.8286) so insufficient evidence to reject | M1 |
|  | There is insufficient evidence to suggest a **positive** correlation between the judges. | A1 ft |
|  |  | (4) |
| **(c)\*** | (For positive correlation c.v.is 0.8286>0.771) |  |
|  | Training of junior judge was ineffective. | B1 ft |
|  |  | (1) |
|  |  | **Total 10** |

Whole question assesses spec point 7 Correlation and linear regression.

\*Part (c) is AO3