Topic: Mixed Bivariate Data and Time Series Data – The Investigation Process

Time: 45 mins
Marks: /45 marks

Calculator Assumed

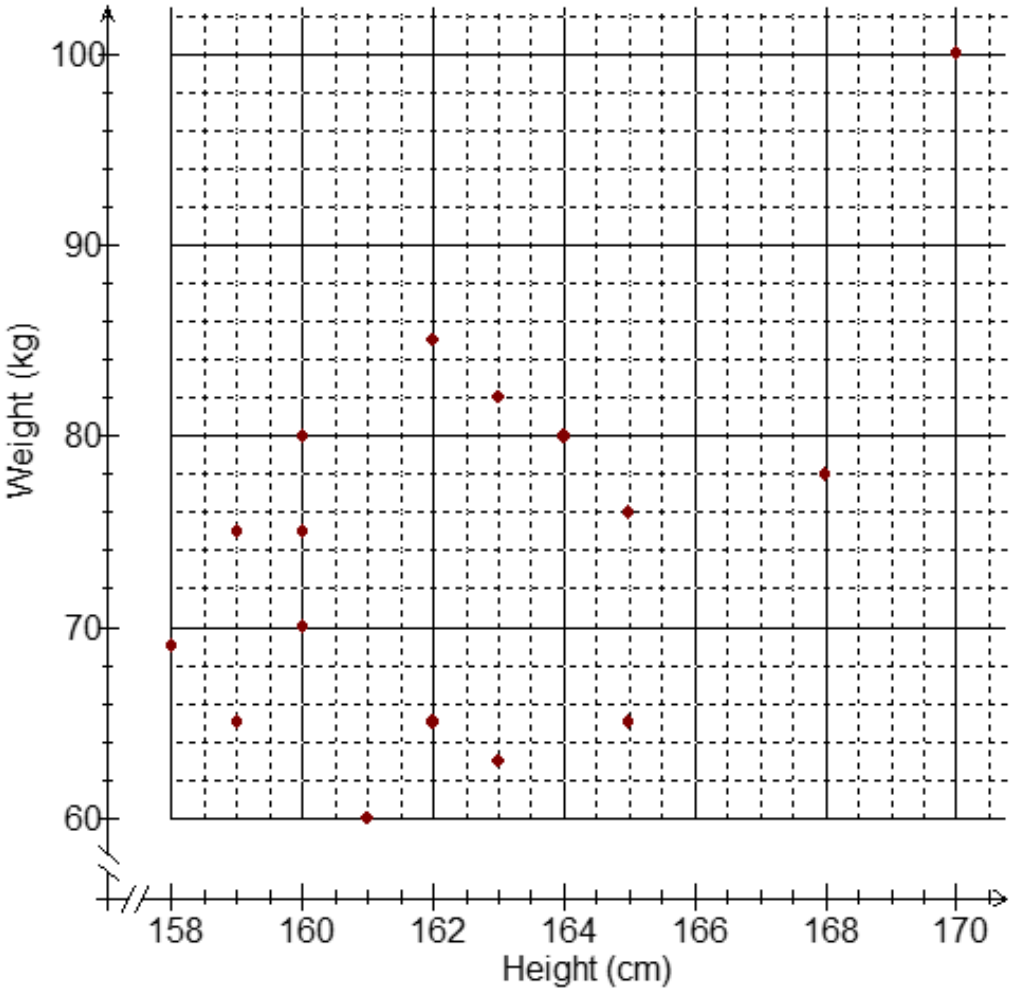
Question One: [2, 3, 1, 1, 3, 3, 2, 2, 3: 20 marks]

a) If we want to use a 20 year old female person's height to predict their weight, explain a process in which this may be possible.

Justin decided to collect and analyse the heights and weights of his fellow classmates and produced the following table of data.

Individual	Height (cm)	Weight (kg)
1	160	80
2	159	65
3	163	82
4	162	85
5	160	75
6	165	76
7	170	100
8	168	78
9	162	65
10	158	69
11	159	75
12	168	78
13	164	80
14	161	60
15	160	70
16	163	63
17	165	65
18	167	71
19	159	54
20	158	54
21	165	80
22	165	86

b) Complete the scatterplot for the data collected by Justin by adding height and weight for the last 5 individuals to the graph.



c) Identify the response variable.

d) Calculate the correlation coefficient, r_{HW}.

e) Calculate the coefficient of determination and describe what this value means.

f) Without calculating these predictions, comment just on the reliability of predicting the weight of an individual who is:

i) 140 cm tall

ii) 160 cm tall

iii) 190 cm tall

Justin realizes he forgot to add the data for four individuals into his calculations.

Individual	Height (cm)	Weight (kg)
23	166	85
24	168	90
25	158	60
26	160	68

g) How do these four data points effect the reliability of prediction the weights for the heights mentioned in part f) ?

h) State the least squares regression line for predicting weight, based on height from all the data collected by Justin.

Recently it has been claimed that the Body Mass Index (BMI) calculator is hugely inaccurate.

BMI is calculated by dividing weight, in kg, by height, **in metres**, and then dividing this answer by height again.

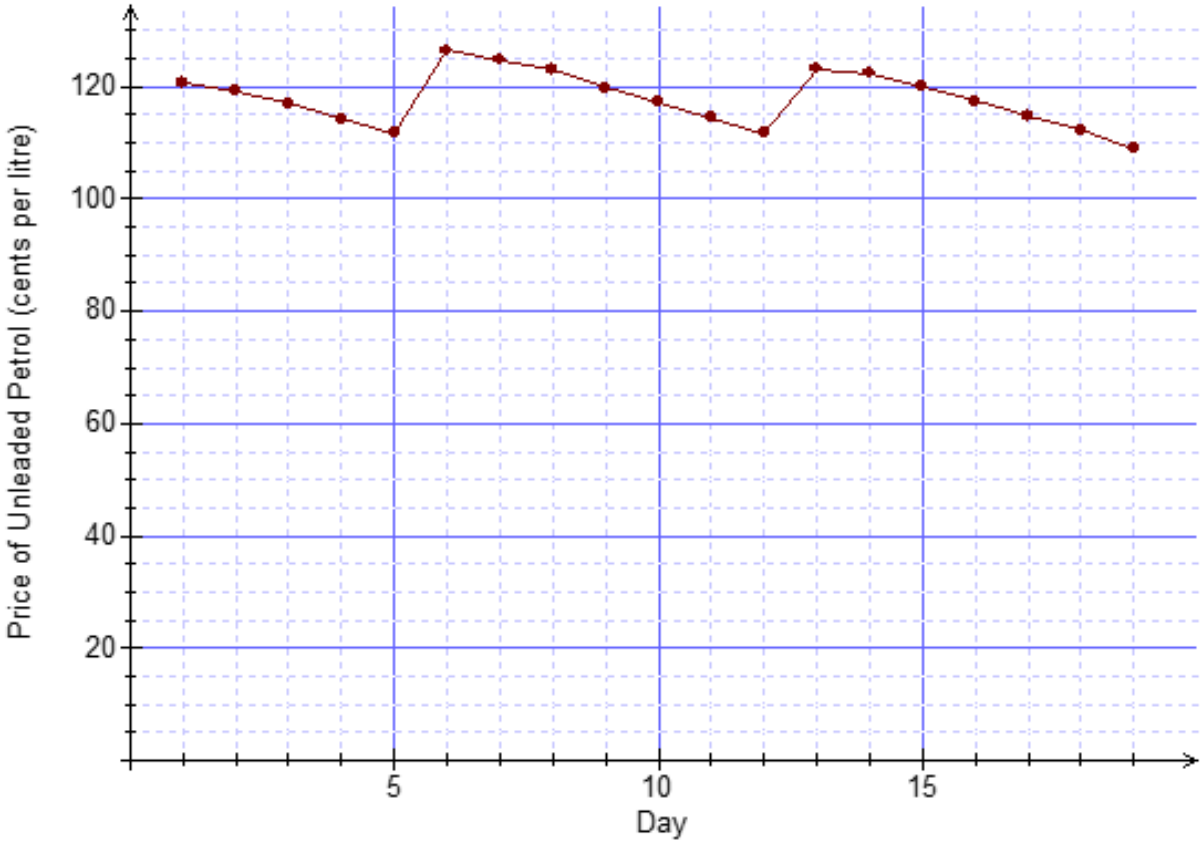
i) If it is known that someone has a BMI of 31 and a height of 162 cm compare the weight found by using the BMI calculator to the weight found using the data Justin collected.

Question Two: [1, 1, 1, 4, 4, 2, 2, 5: 20 marks]

Emma and Lachy were watching the news when they heard a lot of discussion about the "Petrol Price Cycle". They decided to investigate whether or not their local petrol station had a similar cycle.

They began recording the price of unleaded petrol at their local petrol station each day, until they were sure of that the cycle was.

Their graph of the petrol prices is shown below.



a) What is the length of the cycle?

b) If they began recording the price of petrol on a Thursday, what appears to be the best day to fill up with petrol according to this data?

The prices of petrol they collected are shown in the table below.

Emma and Lachy want to investigate the seasonal index for each day of the week.

c) Explain why Lachy and Emma have only used the shaded part of the table in their calculations for "weekly average" and "price as a % of average".

b) Calculate values **A, B, C and D** from the table below.

Date	Price	Weekly Average	Price as % of average
Thursday Jan 7	120.7		
Friday Jan 8	119.2		
Saturday Jan 9	117		
Sunday Jan 10	114.3		
Monday Jan 11	111.8		93.4
Tuesday Jan 12	126.4		105.6
Wednesday Jan 13	124.7		104.2
Thursday Jan 14	123	**A**	**C**
Friday Jan 15	119.8		100.1
Saturday Jan 16	117.3		98.0
Sunday Jan 17	114.5		95.7
Monday Jan 18	111.8		95.2
Tuesday Jan 19	123.2		105.0
Wednesday Jan 20	122.4		**D**
Thursday Jan 21	120	**B**	102.2
Friday Jan 22	117.4		100.0
Saturday Jan 23	114.7		97.7
Sunday Jan 24	112.2		95.6
Monday Jan 25	109		

c) Calculate the seasonal indices for Wednesday and Saturday and explain what this figures mean in terms of which day of the week it is best to fill up one's car with petrol.

Seasonal Index Monday	Seasonal Index Tuesday	Seasonal Index Wednesday	Seasonal Index Thursday	Seasonal Index Friday	Seasonal Index Saturday	Seasonal Index Sunday
94.3	105.3		102.5	100.1		95.7

On data analyzed in the past, Tuesday's used to have the lowest seasonal index.

d) Suggest a reason why this might have changed in recent weeks.

Lachy and Emma decide to compare what a predicted price of petrol for Monday 25th Janurary would be compared to the actual price which they recorded.

e) Complete the table they began for the deseasonalised data.

Day	Deseasonalised Data
Monday Jan 11	100.0
Tuesday Jan 12	101.3
Wednesday Jan 13	100.9
Thursday Jan 14	101.2
Friday Jan 15	101.0
Saturday Jan 16	101.1
Sunday Jan 17	101.0
Monday Jan 18	100.0
Tuesday Jan 19	98.7
Wednesday Jan 20	
Thursday Jan 21	
Friday Jan 22	
Saturday Jan 23	

f) Predict the actual petrol price for Monday January 25th (justify your answer with sufficient working) and compare this prediction with that of the price they recorded.

Question Three: [3, 2: 5 marks]

A company's profits show a quarterly seasonal pattern.

The least squares regression line for predicting the deseasonalised profits in thousands of dollars is: $\hat{p} = 0.0995n + 9.8323$ where p is the profit (in \$1000) and n is the quarter since profit was first recorded.

The least squares regression line and the seasonal index are used to predict that the actual sales for the 29th quarter will be \$14 449.

a) Calculate the seasonal index for the first quarter.

The deseasonalised profit for the very first quarter is 10.0343 and the average sales for the first year is 10.15.

b) What is the profit as a percentage of the average for the very first quarter?

Topic: Mixed Bivariate Data and Time Series Data SOLUTIONS

Time: 45 mins Marks: /45 marks

Calculator Assumed

Question One: [2, 3, 1, 1, 3, 3, 2, 2, 3: 20 marks]

a) If we want to use a 20 year old female person's height to predict their weight, explain a process in which this may be possible.

Collect the heights and weights of a random sample of many females aged 20 years old.

Investigate the correlation coefficient of the height vs weight to determine the strength of the relationship between these two variables. Consider the coefficient of determination.

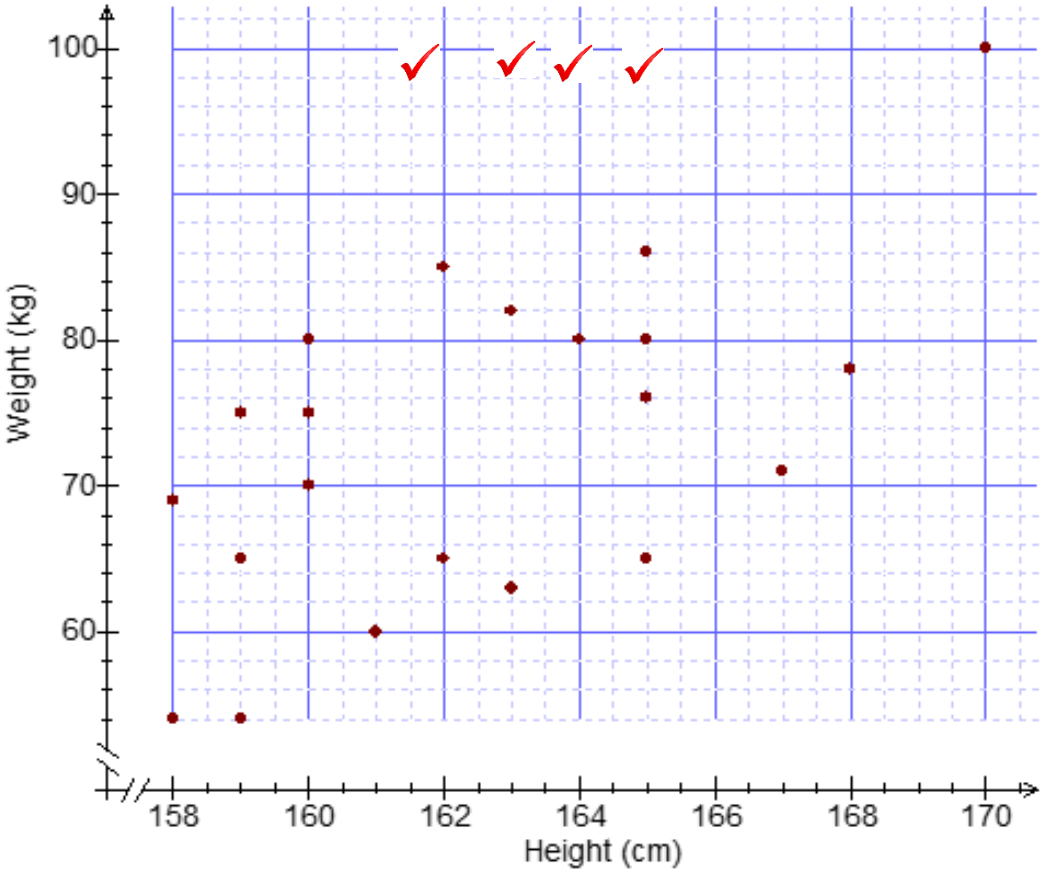
Calculate the Line of Best Fit and use this line to make predictions.

✓ ✓

Justin decided to collect and analyse the heights and weights of his fellow classmates and produced the following table of data.

Individual	Height (cm)	Weight (kg)
1	160	80
2	159	65
3	163	82
4	162	85
5	160	75
6	165	76
7	170	100
8	168	78
9	162	65
10	158	69
11	159	75
12	168	78
13	164	80
14	161	60
15	160	70
16	163	63
17	165	65
18	167	71
19	159	54
20	158	54
21	165	80
22	165	86

b) Complete the scatterplot for the data collected by Justin by adding height and weight for the last 5 individuals to the graph.



c) Identify the response variable.

Weight (kg)

d) Calculate the correlation coefficient, r_{HW}.

$r = 0.5976$ (4 dp) ✓

A weak, positive linear relationship. A positive relationship exists but it is not strong.

e) Calculate the coefficient of determination and describe what this value means.

✓ ✓

$r^2 = 0.3571$ (4 dp) Approximately 36% of the variation in weight between these individuals can be explained by the variation in height. The other 64% is unexplained.

✓

f) Without calculating these predictions, comment just on the reliability of predicting the weight of an individual who is:

i) 140 cm tall Unreliable due to extrapolating ✓

ii) 160 cm tall Unreliable due to weak correlation coefficient ✓

iii) 190 cm tall Unreliable due to extrapolating ✓

Justin realizes he forgot to add the data for four individuals into his calculations.

Individual	Height (cm)	Weight (kg)
23	166	85
24	168	90
25	158	60
26	160	68

g) How do these four data points effect the reliability of prediction the weights for the heights mentioned in part f) ? ✓

Adding these four points in increases r to $r = 0.676\ (3dp)$ thus allowing a prediction for 160 cm to be more reliable but the other predictions would remain unreliable as they would still be extrapolations of the data. ✓

h) State the least squares regression line for predicting weight, based on height from the data collected by Justin.

$$w = 2.0939h - 267.2923$$ ✓ ✓

Recently it has been claimed that the Body Mass Index (BMI) calculator is hugely inaccurate.

BMI is calculated by dividing weight, in kg, by height, **in metres**, and then dividing this answer by height again.

i) If it is known that someone has a BMI of 31 and a height of 162 cm compare the weight found by using the BMI calculator to the weight found using the data Justin collected.

$$31 = \frac{weight}{1.62 \div 1.62}$$

$$weight = 81.36\ kg$$

$$\hat{w} = 71.9\ kg\ (1dp)$$

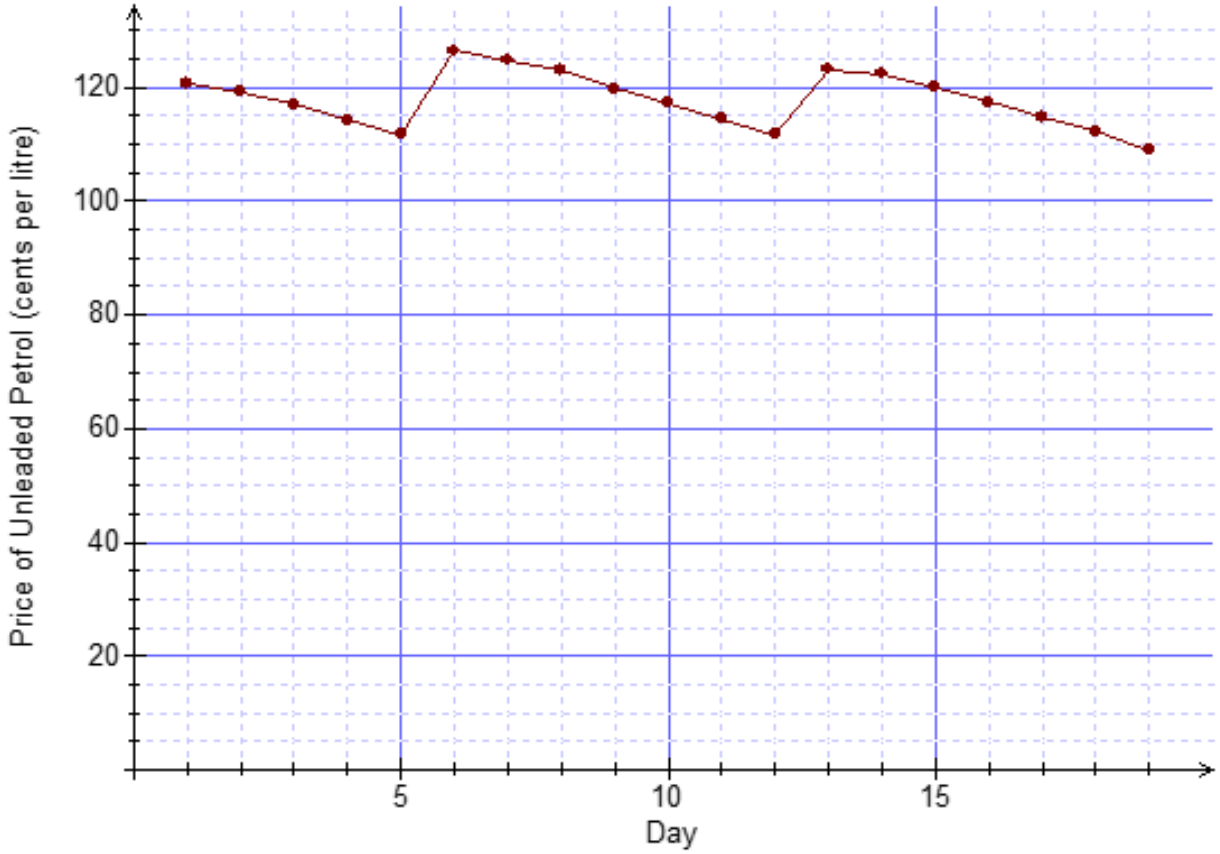
These two answers do not match but Justin's predictions are not very reliable due to a correlation coefficient which is not strong and we have been told that there are questions around the reliability of the BMI calculator.

Question Two: [1, 1, 1, 4, 4, 2, 2, 5: 20 marks]

Emma and Lachy were watching the news when they heard a lot of discussion about the "Petrol Price Cycle". They decided to investigate whether or not their local petrol station had a similar cycle.

They began recording the price of unleaded petrol at their local petrol station each day, until they were sure of that the cycle was.

Their graph of the petrol prices is shown below.



a) What is the length of the cycle?

✓ It appears to be a 7 day cycle.

b) If they began recording the price of petrol on a Thursday, what appears to be the best day to fill up with petrol according to this data?

✓ Monday's appear to have the lowest petrol price.

The prices of petrol they collected are shown in the table below.

Emma and Lachy want to investigate the seasonal index for each day of the week.

c) Explain why Lachy and Emma have only used the shaded part of the table in their calculations for "weekly average" and "price as a % of average".

The shaded part of the table is two full cycles of data. They cannot calculate the weekly average for the first four days because they do not have the data for the full week/cycle.

b) Calculate values **A, B, C and D** from the table below.

Date	Price	Weekly Average	Price as % of average
Thursday Jan 7	120.7		
Friday Jan 8	119.2		
Saturday Jan 9	117		
Sunday Jan 10	114.3		
Monday Jan 11	111.8		93.4
Tuesday Jan 12	126.4		105.6
Wednesday Jan 13	124.7		104.2
Thursday Jan 14	123	**A** 119.6	102.8 **C**
Friday Jan 15	119.8		100.1
Saturday Jan 16	117.3		98.0
Sunday Jan 17	114.5		95.7
Monday Jan 18	111.8		95.2
Tuesday Jan 19	123.2		105.0
Wednesday Jan 20	122.4		104.3 **D**
Thursday Jan 21	120	**B** 117.4	102.2
Friday Jan 22	117.4		100.0
Saturday Jan 23	114.7		97.7
Sunday Jan 24	112.2		95.6
Monday Jan 25	109		

c) Calculate the seasonal indices for Wednesday and Saturday and explain what this figures mean in terms of which day of the week it is best to fill up one's car with petrol.

Seasonal Index Monday	Seasonal Index Tuesday	Seasonal Index Wednesday	Seasonal Index Thursday	Seasonal Index Friday	Seasonal Index Saturday	Seasonal Index Sunday
94.3	105.3	104.3 ✓	102.5	100.1	97.9 ✓	95.7

The seasonal index shows the average percentage of increased or decrease each day of the cycle has from the mean. Monday has the lowest petrol price because it has the lowest seasonal index. Monday's petrol price is on average the lowest below the weekly average for the week.

✓ ✓

On data analyzed in the past, Tuesday's used to have the lowest seasonal index.

d) Suggest a reason why this might have changed in recent weeks.

Once people are in the habit of filling up their cars with petrol on a Tuesday and then the cycle changes and Tuesday becomes the highest price in the cycle, some people may end up still filling up their cars with petrol on a Tuesday and the petrol companies will make more money. ✓ ✓

Lachy and Emma decide to compare what a predicted price of petrol for Monday 25th Janurary would be compared to the actual price which they recorded.

e) Complete the table they began for the deseasonalised data.

Day	Deseasonalised Data
Monday Jan 11	118.6
Tuesday Jan 12	120.04
Wednesday Jan 13	119.61
Thursday Jan 14	120.0
Friday Jan 15	119.68
Saturday Jan 16	119.82
Sunday Jan 17	119.71
Monday Jan 18	118.56
Tuesday Jan 19	116.999
Wednesday Jan 20	117.35
Thursday Jan 21	117.07
Friday Jan 22	117.28
Saturday Jan 23	117.16
Sunday Jan 24	107.38

f) Predict the actual petrol price for Monday January 25th (justify your answer with sufficient working) and compare this prediction with that of the price they recorded.

$\hat{p} = -0.528d + 121.765$ ✓ ✓

$\hat{p} = -0.528(15) + 121.765 = 113.84$ ✓

predicting for Monday

$\therefore$ *prediction of actual price*: $113.84 \times 0.943 = \$107.35$ ✓

The actual price for this day was recorded and was \$109.

This prediction is very close. ✓

Question Three: [3, 2: 5 marks]

A company's profits show a quarterly seasonal pattern.

The least squares regression line for predicting the deseasonalised profits in thousands of dollars is: $\hat{p} = 0.0995n + 9.8323$ where p is the profit (in \$1000) and n is the quarter since profit was first recorded.

The least squares regression line and the seasonal index are used to predict that the actual sales for the 29th quarter will be \$14 449.

a) Calculate the seasonal index for the first quarter.

$\hat{p} = 0.0995(29) + 9.8323$ ✓

$= 12.7178$

$12.7178 \times x = 14.449$

$x = 1.136$ ✓

$\therefore 113.6\%$ is the seasonal index for first quarter. ✓

The deseasonalised profit for the very first quarter is 10.0343 and the average sales for the first year is 10.15.

b) What is the profit as a percentage of the average for the very first quarter?

$Actual\ sales\ figure$: $10.0343 \times 1.136 = 11.40\ (2dp)$ ✓

$Profit\ as\ percentage\ of\ average$: $11.40 \div 10.15 \times 100 = 112.3\%\ (1dp)$ ✓