THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

ECON1310
Introductory Statistics for Social Sciences

Tutorial 12: SIMPLE LINEAR REGRESSION II

Tutor: Francisco Tavares Garcia

LBRT #3

LBRT #3

Type: Online Quiz
Learning Objectives Assessed: 1, 2, 3, 4, 5
Due Date: 07 Feb 23 9:00 - 08 Feb 23 16:00 2nd attempt: 9-10 Feb 2023, 09:00-16:00
Weight: 20%
Reading: 0 minutes
Duration: 90 minutes
Format: Multiple-choice, Problem solving
Task Description:
LBRT #3 will involve solving problems based on the learning materials covered in Lectures 9 to 12 inclusively. This includes all learning materials presented in Lectures 9 to 12 and the associated tutorials, as well as CML5 and CML6. All answers must be entered into Blackboard by the due date and time.
Criteria & Marking:
UQ Students: Please access the profile from Learn.UQ or mySI-net to access marking criteria held in this profile.

CML 5(2nd) and 6 – first and only attempt

CML 5 and CML6 Reminder

Posted on: Wednesday, 25 January 2023 09:00:00 o'clock AEST

Dear Students,

A reminder that:

1. **CML 5 (2nd Attempt)** is now open and will **close at 4pm this Friday** (27 January).
2. **CML 6** is now open and will **close at 4pm Monday 6 February**. Note that there is **NO second attempt** for CML 6.
3. Please ensure you **check, save and submit** your CMLs, as CMLs do not auto-submit.

Best of luck!

Dominic

ECON1310

Tutorial 12 – Week 13

SIMPLE LINEAR REGRESSION II

At the end of this tutorial you should be able to

- Describe the assumptions that underpin the SLR model.
- Carry out analysis of the regression residuals to test whether the assumptions hold.
- Carry out hypothesis tests on the slope coefficient.

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.
b) State the units for the constant and coefficient.
c) State the assumptions on which the calculations are based.
d) Test if the linear relationship is **downward sloping** using 5% level of significance.

(Answers in chat)

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.
b) State the units for the constant and coefficient.
c) State the assumptions on which the calculations are based.
d) Test if the linear relationship is **downward sloping** using 5% level of significance.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient.

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

(Answers in chat)

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient.

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Constant: b_0 = $thousands

Slope coefficient: $b_1 = \frac{\$thousands}{km}$

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient. b_0 = \$thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\ 6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

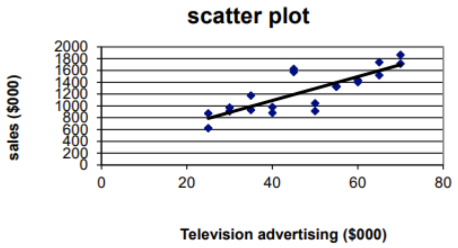
b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

- Linearity.

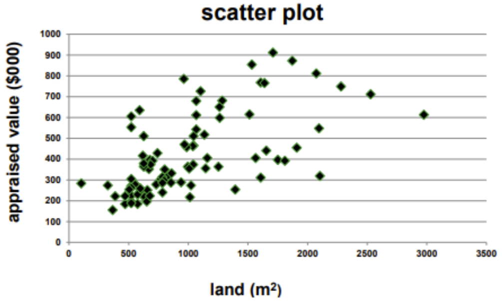
Check Assumption 1 – is the model linear?



Plot looks linear, so a **linear model** can be used. The "linear assumption" is satisfied.

7

Check Assumption 1 – is the model linear?



Does **NOT** look linear, so a linear model should **NOT** be used. The "linear assumption" is **violated.**

8

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\ 6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

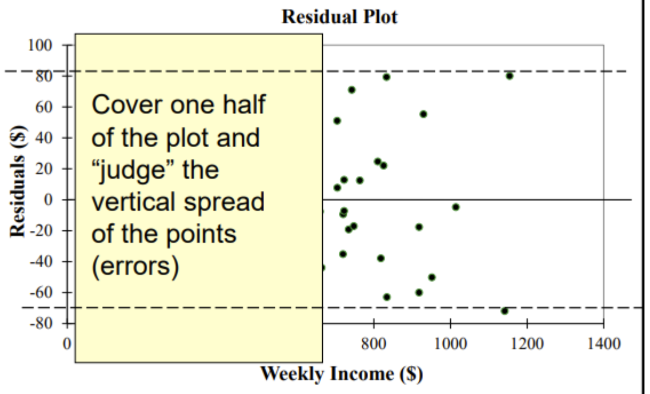
b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

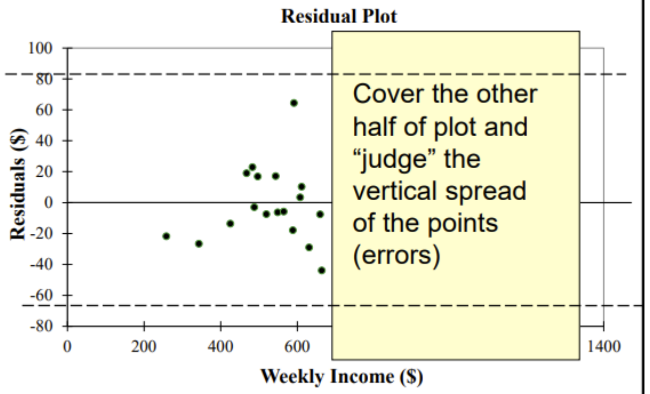
d) Test if the linear relationship is **downward sloping** using 5% level of significance.

- Linearity.
- The errors have constant variance around the regression line for all values of X.

Example 2
Weekly Income and Food Expenditure



Example 2
Weekly Income and Food Expenditure



Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

adm T12 Q1 Q2 Q3

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 486.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

- Linearity.
- The errors have constant variance around the regression line for all values of X.
- Errors are independent of each value of X as well as each other. When data is gathered over time, errors in adjacent time periods should not be correlated (auto correlation).

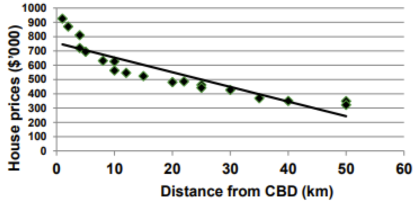
Residual plot to check Assumptions 3

Independent and random errors (= good).

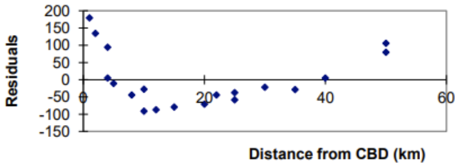
- the residual plot should show **no pattern in the residuals.**
- several consecutive positive errors followed by several consecutive negative errors (a pattern) as X increases can indicate a violation of the **independence** of errors assumption.
- If **time** is on the horizontal axis (or observations are ordered as measured), and a pattern in the residuals exists, this violation is called **autocorrelation.**

23

Residual plot examples.



(X,Y) scatter plot looks **non-linear, so** assumption 1 about being linear is **violated.**



Residual plot has a pattern as X increases, and errors are not random. **Violates** assumption 3 and the independence of errors.

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

- Linearity.
- The errors have constant variance around the regression line for all values of X.
- Errors are independent of each value of X as well as each other. When data is gathered over time, errors in adjacent time periods should not be correlated (auto correlation).
- Errors around the regression line are normally distributed at each value of X with mean 0.

Assumptions 4 – Normality of Errors

The error terms are assumed to be normally distributed with an average, or expected value, equal to zero ie: $E(e_i) = 0$

The residual plot is **NOT** used to check the assumption of normality of errors.

A normality plot (or histogram of errors showing the distribution) is needed and this will NOT be covered in ECON1310.

25

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is downward sloping using 5% level of significance.

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient. b_0 = \$thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

One tail test

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $|t_{calc}| > t_{crit}$

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\ 6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Step 1: State H_0 and H_1

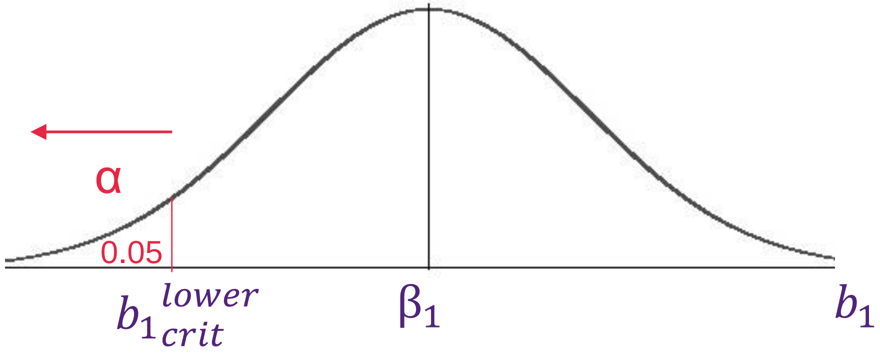
H_0: $\beta_1 \geq 0$

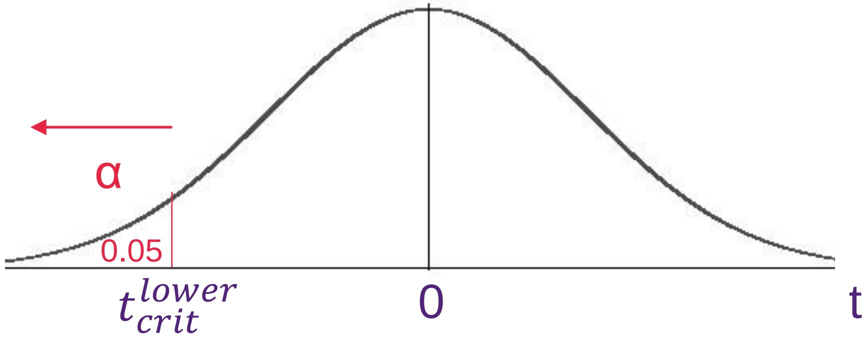
H_1: $\beta_1 < 0$ (downward sloping)

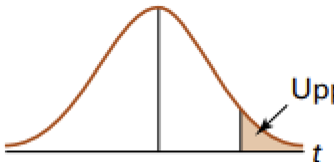
Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36}$ = ?

Rejection regions

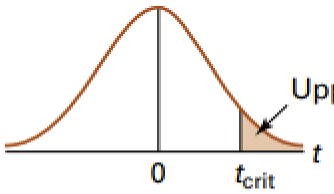






$t_{0.05,\,36}$

df	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.001}$
			Upper tail areas			
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
31	1.309	1.696	2.040	2.453	2.744	3.375
32	1.309	1.694	2.037	2.449	2.738	3.365
33	1.308	1.692	2.035	2.445	2.733	3.356
34	1.307	1.691	2.032	2.441	2.728	3.348
35	1.306	1.690	2.030	2.438	2.724	3.340
36	1.306	1.688	2.028	2.434	2.719	3.333
37	1.305	1.687	2.026	2.431	2.715	3.326
38	1.304	1.686	2.024	2.429	2.712	3.319
39	1.304	1.685	2.023	2.426	2.708	3.313
40	1.303	1.684	2.021	2.423	2.704	3.307
41	1.303	1.683	2.020	2.421	2.701	3.301
42	1.302	1.682	2.018	2.418	2.698	3.296
43	1.302	1.681	2.017	2.416	2.695	3.291
44	1.301	1.680	2.015	2.414	2.692	3.286
45	1.301	1.679	2.014	2.412	2.690	3.281
46	1.300	1.679	2.013	2.410	2.687	3.277
47	1.300	1.678	2.012	2.408	2.685	3.273
48	1.299	1.677	2.011	2.407	2.682	3.269
49	1.299	1.677	2.010	2.405	2.680	3.265
50	1.299	1.676	2.009	2.403	2.678	3.261



$t_{0.05,\,36}$

	Upper tail areas					
df	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.001}$
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
31	1.309	1.696	2.040	2.453	2.744	3.375
32	1.309	1.694	2.037	2.449	2.738	3.365
33	1.308	1.692	2.035	2.445	2.733	3.356
34	1.307	1.691	2.032	2.441	2.728	3.348
35	1.306	1.690	2.030	2.438	2.724	3.340
36	1.306	1.688	2.028	2.434	2.719	3.333
37	1.305	1.687	2.026	2.431	2.715	3.326
38	1.304	1.686	2.024	2.429	2.712	3.319
39	1.304	1.685	2.023	2.426	2.708	3.313
40	1.303	1.684	2.021	2.423	2.704	3.307
41	1.303	1.683	2.020	2.421	2.701	3.301
42	1.302	1.682	2.018	2.418	2.698	3.296
43	1.302	1.681	2.017	2.416	2.695	3.291
44	1.301	1.680	2.015	2.414	2.692	3.286
45	1.301	1.679	2.014	2.412	2.690	3.281
46	1.300	1.679	2.013	2.410	2.687	3.277
47	1.300	1.678	2.012	2.408	2.685	3.273
48	1.299	1.677	2.011	2.407	2.682	3.269
49	1.299	1.677	2.010	2.405	2.680	3.265
50	1.299	1.676	2.009	2.403	2.678	3.261

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\ 6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

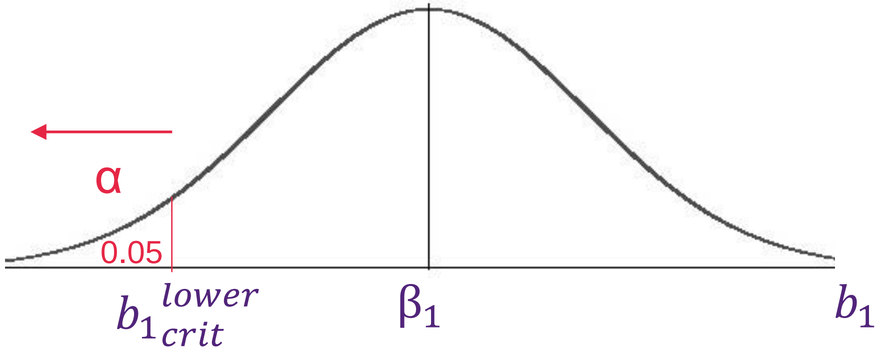
a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

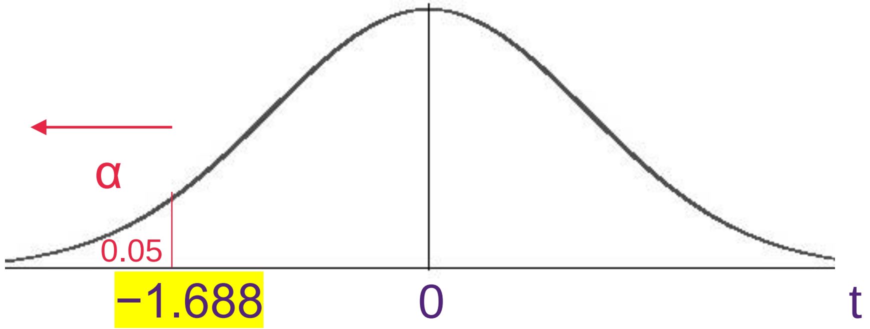
b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Rejection regions





Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36}$ = -1.688

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\ 6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

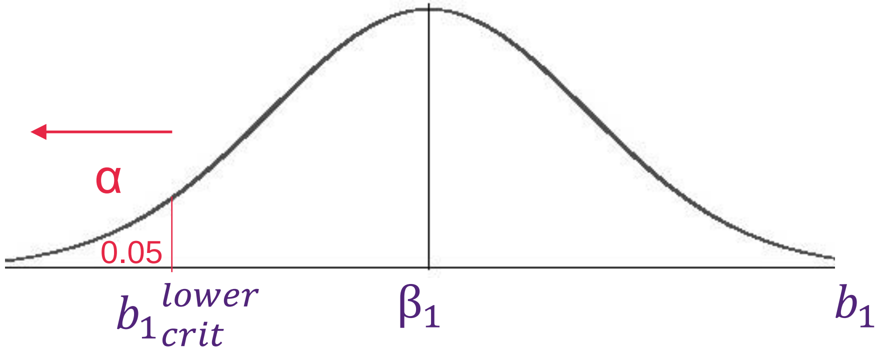
When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

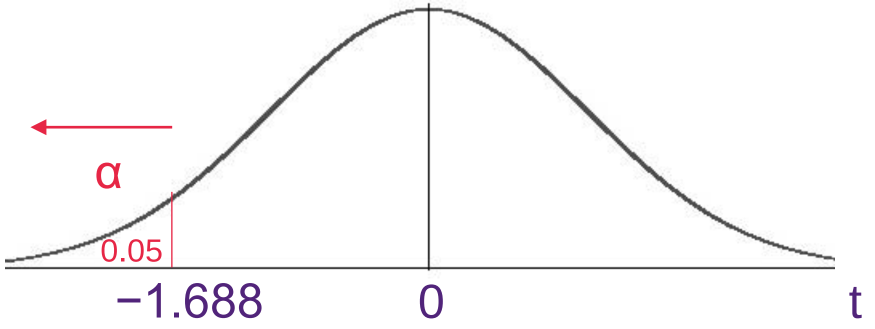
b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Rejection regions





Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36}$ = -1.688

Step 3: Calculate t_{calc}

$t_{calc} = \frac{b_1 - \beta_1}{s_{b_1}}$ = ?

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

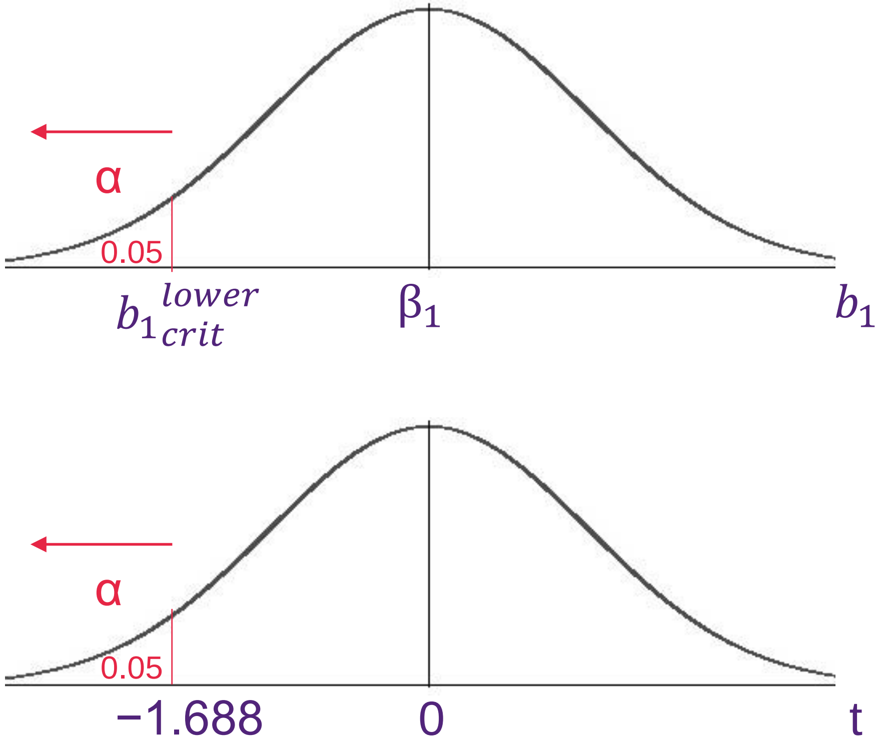
The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.
b) State the units for the constant and coefficient.
c) State the assumptions on which the calculations are based.
d) Test if the linear relationship is **downward sloping** using 5% level of significance.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

b_0 = \$thousands, $b_1 = \frac{\$thousands}{km}$

Rejection regions



Step 1: State H_0 and H_1
H_0: $\beta_1 \geq 0$
H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule
Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36}$ = -1.688

Step 3: Calculate t_{calc}
$t_{calc} = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-1.577 - 0}{s_{b_1}}$

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 486.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

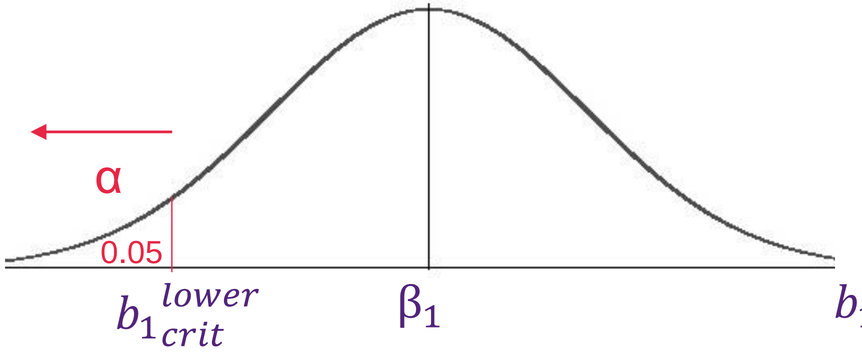
a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

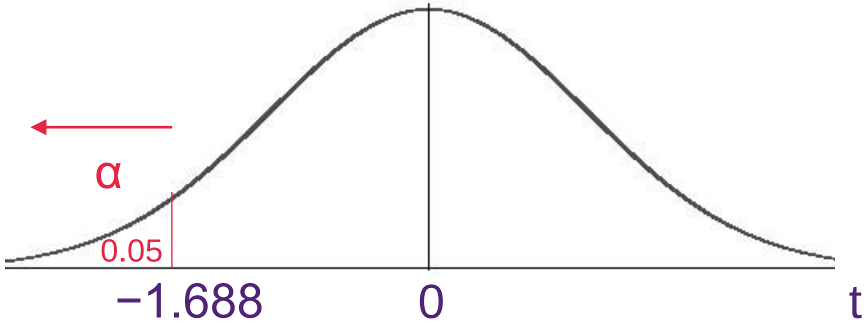
b) State the units for the constant and coefficient. b_0 = \$thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Rejection regions





Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36} = -1.688$

Step 3: Calculate t_{calc}

$$t_{calc} = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-1.577 - 0}{s_{b_1}}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}} =$$

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = **sum of squares of X (sometimes written SSX)**

29

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

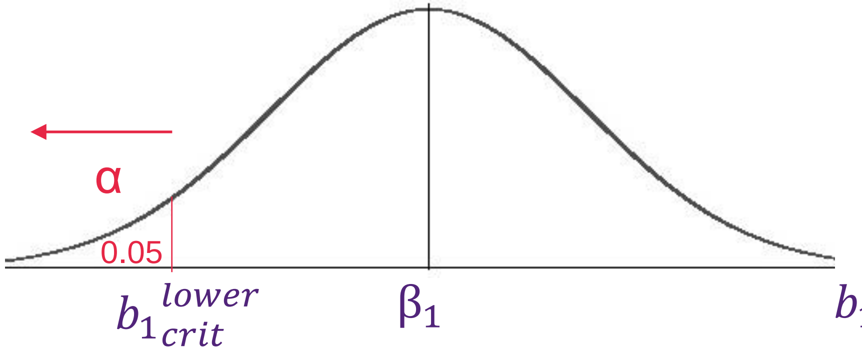
The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

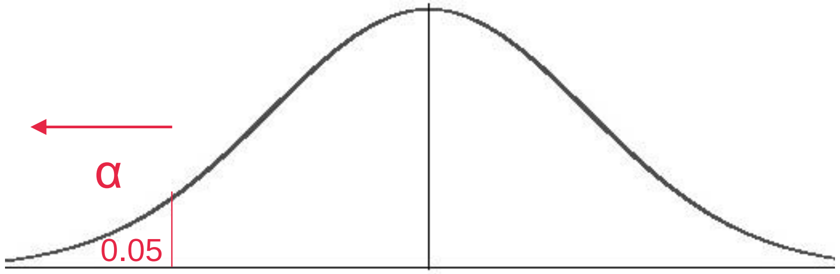
a) Interpret the value of the coefficient.
b) State the units for the constant and coefficient.
c) State the assumptions on which the calculations are based.
d) Test if the linear relationship is **downward sloping** using 5% level of significance.

When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

Rejection regions





Step 1: State H_0 and H_1
H_0: $\beta_1 \geq 0$
H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule
Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha,\,n-2} = t_{0.05,36}$ = -1.688

Step 3: Calculate t_{calc}

$$t_{calc} = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-1.577 - 0}{s_{b_1}}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}} = \frac{113.7}{\sqrt{44229.1}} = 0.5406$$

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 486.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

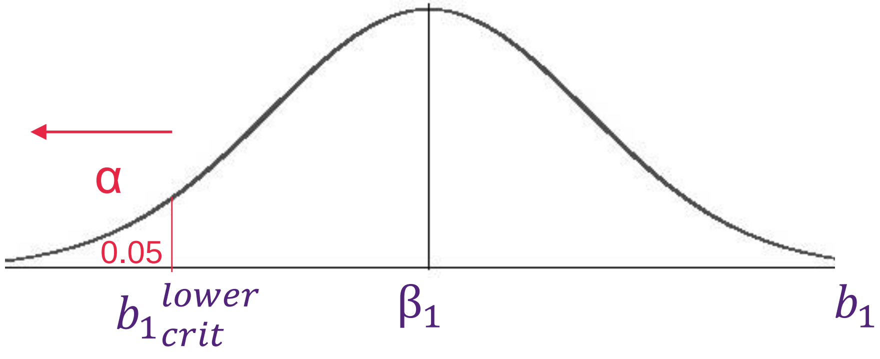
a) Interpret the value of the coefficient. When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

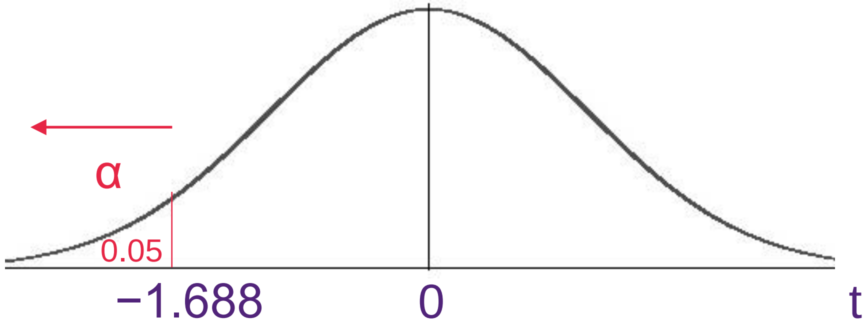
b) State the units for the constant and coefficient. b_0 = \$thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Rejection regions





Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha,\,n-2} = t_{0.05,36} = -1.688$

Step 3: Calculate t_{calc}

$$t_{calc} = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-1.577 - 0}{0.5406} = -2.917$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}} = \frac{113.7}{\sqrt{44229.1}} = 0.5406$$

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

Q1. The price (in \$thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\,6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

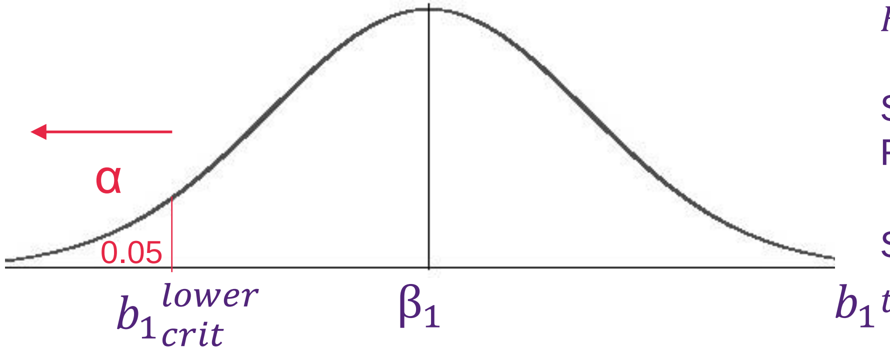
When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by \$1,577 (-1.577 * 1000).

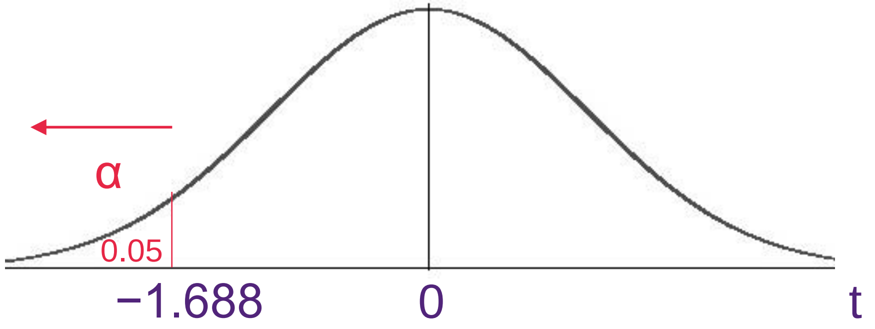
b) State the units for the constant and coefficient. b_0 = \$thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Rejection regions





Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36}$ = -1.688

Step 3: Calculate t_{calc}

$t_{calc} = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-1.577 - 0}{0.5406}$ = -2.917

Step 4: Make a decision

$t_{calc} < t_{crit}$ → -2.917 < -1.688 → ?

adm T12 Q1 Q2 Q3

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\ 6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

a) Interpret the value of the coefficient.

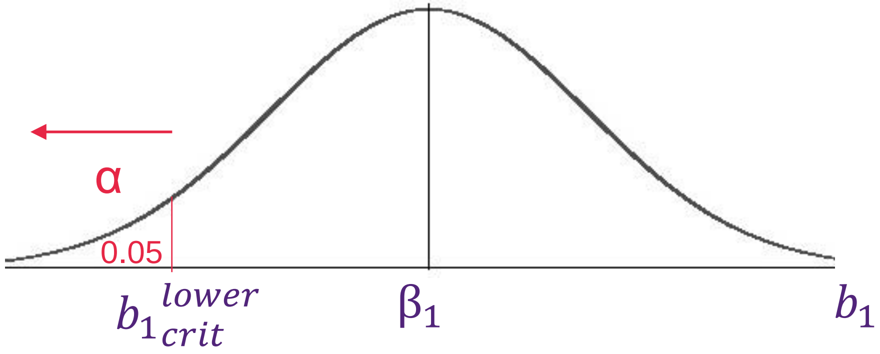
When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

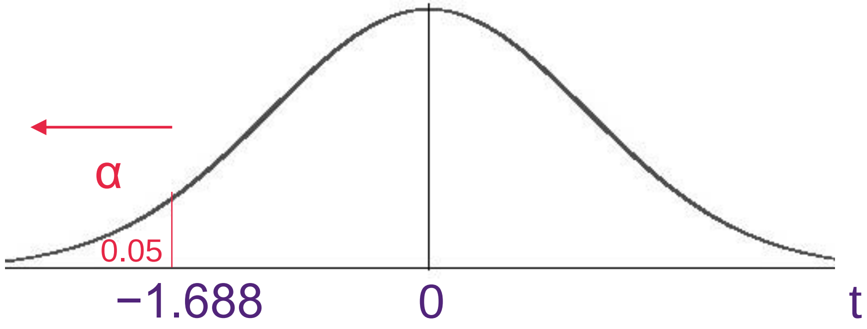
b) State the units for the constant and coefficient. b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Rejection regions





Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36}$ = -1.688

Step 3: Calculate t_{calc}

$$t_{calc} = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{-1.577 - 0}{0.5406} = -2.917$$

Step 4: Make a decision

$t_{calc} < t_{crit}$ → -2.917 < -1.688 → Reject H_0.

adm T12 Q1 Q2 Q3

Q1. The price (in $thousands) of 38 houses was regressed on the distance (km) from the Central Business District (CBD) and the following equation was estimated:

$$\hat{P} = 48\ 6.33 - 1.577D$$

The standard error of the estimate was found to be 113.7 and the sum of squares for distance was 44 229.1

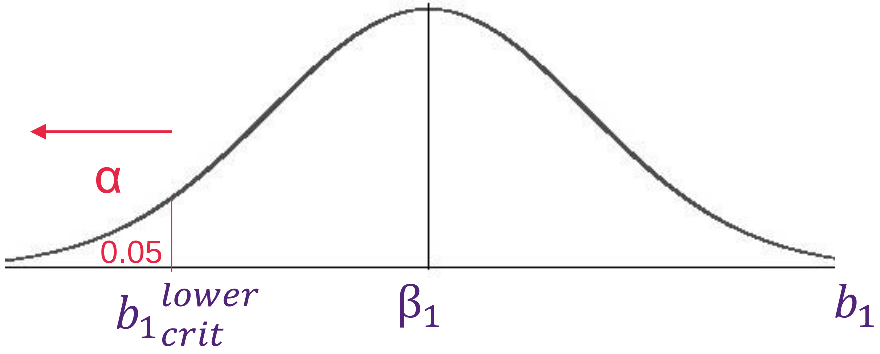
a) Interpret the value of the coefficient.
When the distance from CBD (D) increases by 1 km, the estimated price the of house ($\hat{p}$) decreases by $1,577 (-1.577 * 1000).

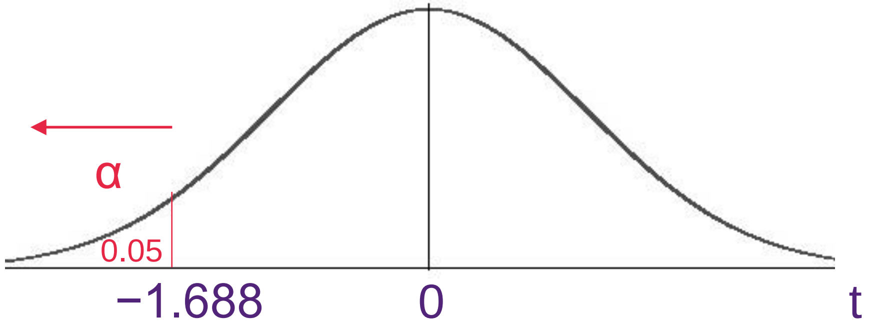
b) State the units for the constant and coefficient.
b_0 = $thousands, $b_1 = \frac{\$thousands}{km}$

c) State the assumptions on which the calculations are based.

d) Test if the linear relationship is **downward sloping** using 5% level of significance.

Rejection regions





Step 1: State H_0 and H_1

H_0: $\beta_1 \geq 0$

H_1: $\beta_1 < 0$ (downward sloping)

Step 2: Decision rule

Reject H_0 if $t_{calc} < t_{crit} = t_{\alpha, n-2} = t_{0.05,36}$ = -1.688

Step 3: Calculate t_{calc}

$t_{calc} = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{-1.577 - 0}{0.5406}$ = -2.917

Step 4: Make a decision

$t_{calc} < t_{crit}$ → -2.917 < -1.688 → Reject H_0.

Step 5: Conclusion

There is sufficient evidence at the 5% level of significance to suggest that there is a negative relationship (downward sloping).

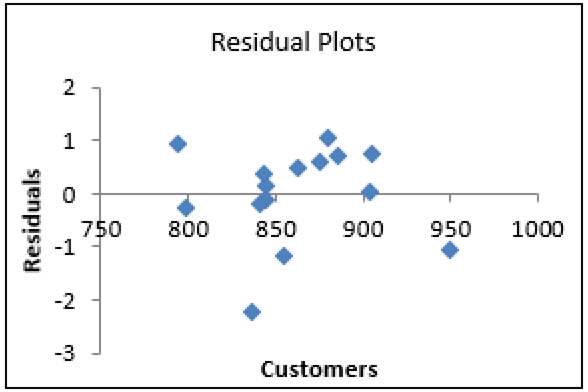
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

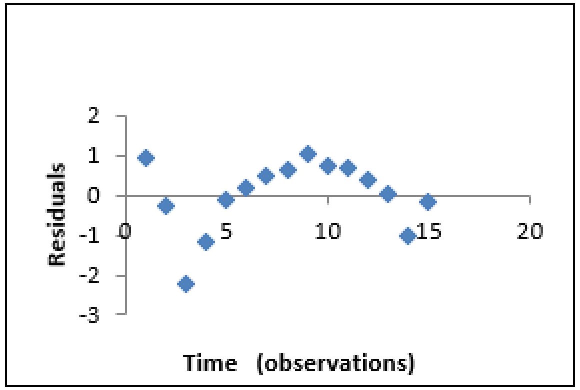
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables.
b) How well does the model fit the data?
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





adm T12 Q1 Q2 Q3

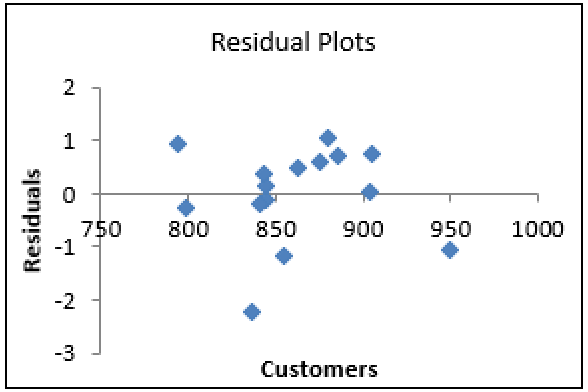
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

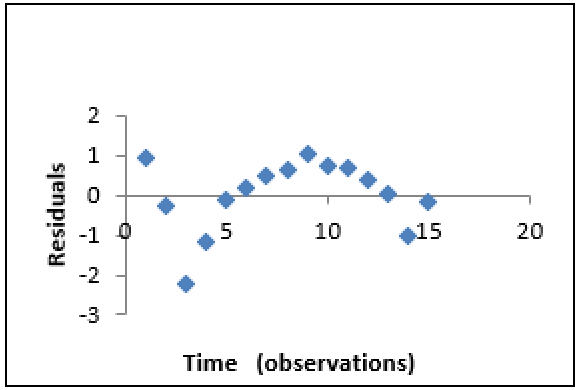
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables.
b) How well does the model fit the data?
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





(Poll)

1. What symbol would you give to the number of customers? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

2. What symbol would you give to sales (in $thous)? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

3. What symbol would you give to the value -16.032? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

4. What symbol would you give to the value 0.031? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

The University Of Queensland
AUSTRALIA

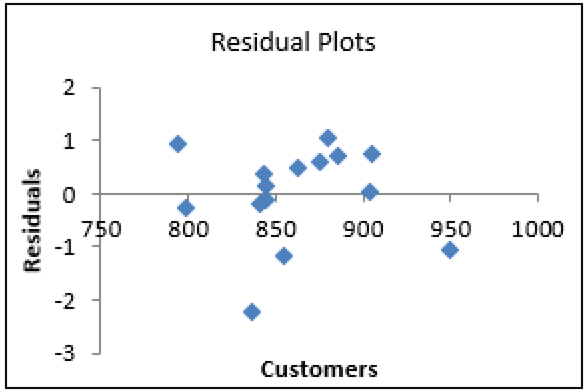
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

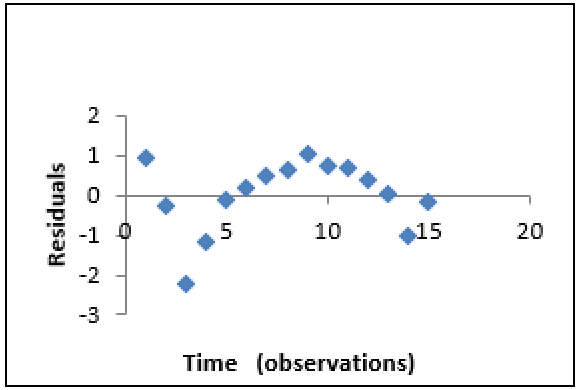
ANOVA

	df	SS	MS
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	Standard Error	t Stat
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables.
b) How well does the model fit the data?
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





$$\hat{Y}_i = -16.032 + 0.031 * X_i$$

1. What symbol would you give to the number of customers? (Single Choice) *
- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

2. What symbol would you give to sales (in $thous)? (Single Choice) *
- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

3. What symbol would you give to the value -16.032? (Single Choice) *
- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

4. What symbol would you give to the value 0.031? (Single Choice) *
- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

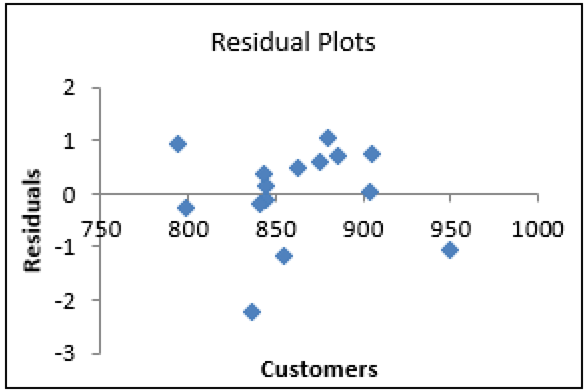
b) How well does the model fit the data?

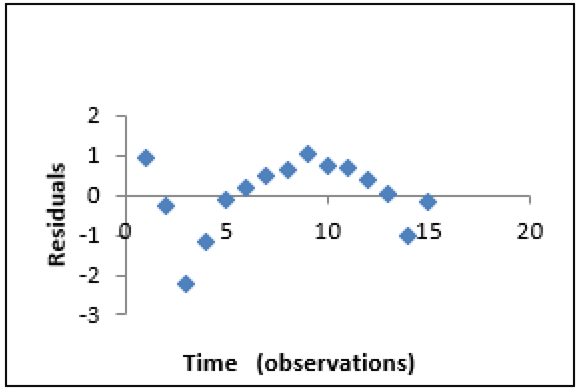
c) Test at the 1% level whether there is a significant relationship.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n\text{-}2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = **sum of squares of X (sometimes written SSX)**

29

adm T12 Q1 Q2 Q3

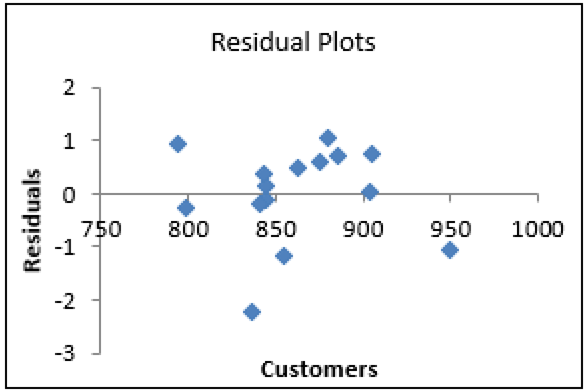
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

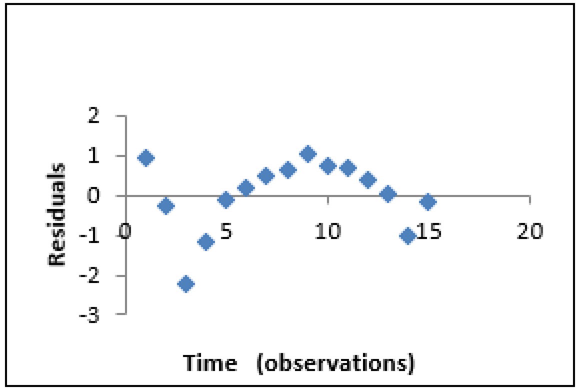
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data?
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





$$r^2 = \frac{SSR}{SST} = ?$$

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n\text{-}2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

adm T12 Q1 Q2 Q3

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

b) How well does the model fit the data?

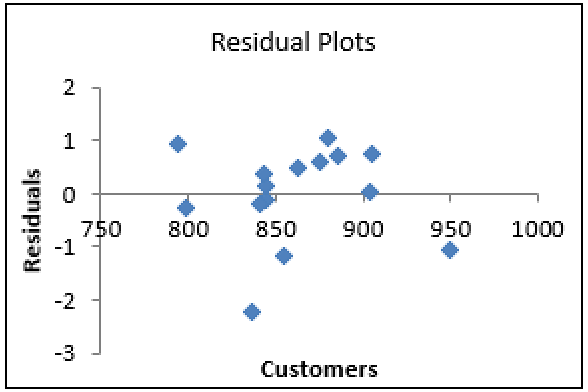
c) Test at the 1% level whether there is a significant relationship.

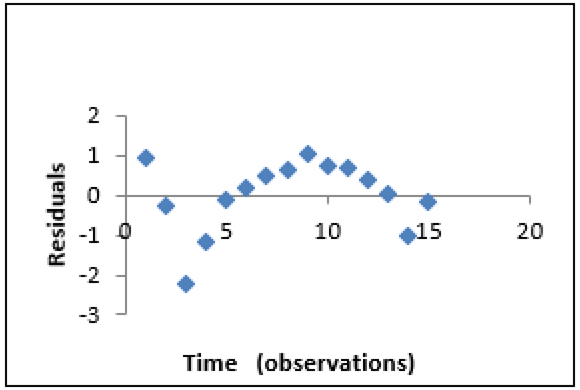
d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.

$$r^2 = \frac{SSR}{SST} = \frac{21.8604}{33.2506} = 0.65744$$





Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

adm T12 Q1 Q2 Q3

The University Of Queensland
AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

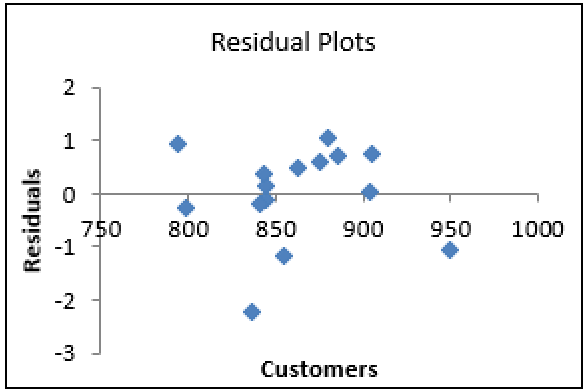
b) How well does the model fit the data?

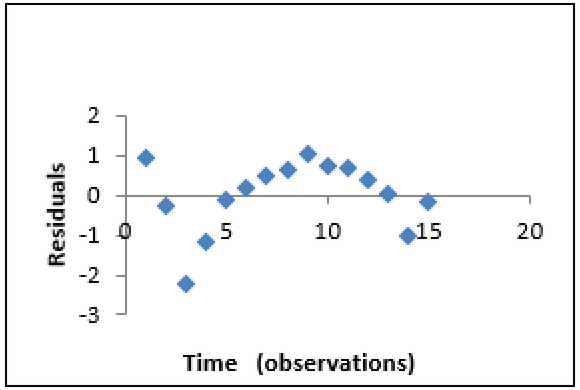
c) Test at the 1% level whether there is a significant relationship.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.

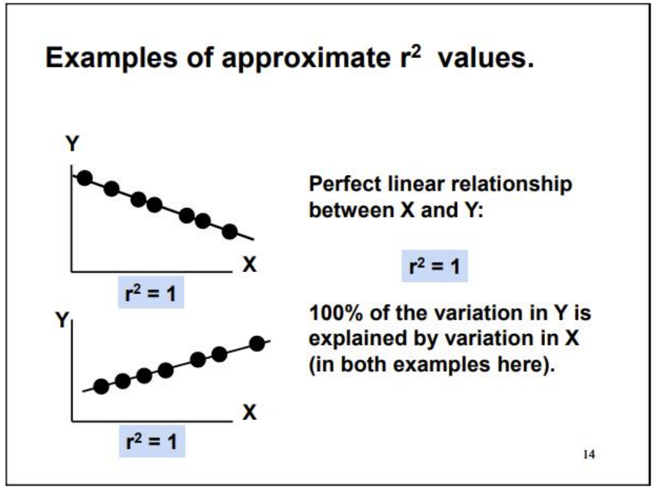


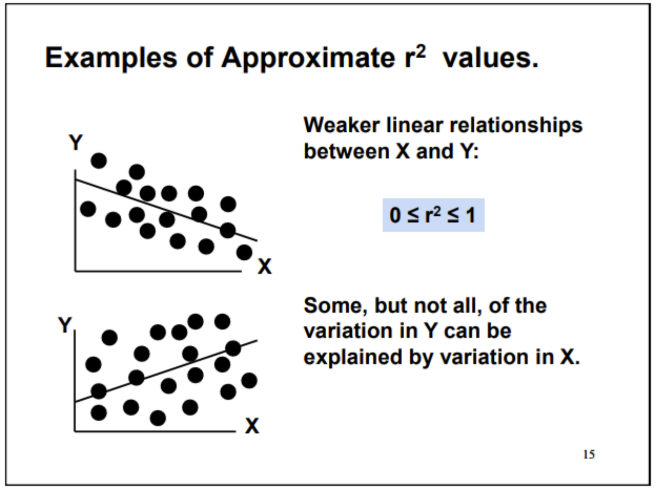


$$r^2 = \frac{SSR}{SST} = \frac{21.8604}{33.2506} = 0.65744$$

65.7% of the variability in sales (Y) can be explained by the variation in the number of customers (X).

Based on r² = moderate fit





Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

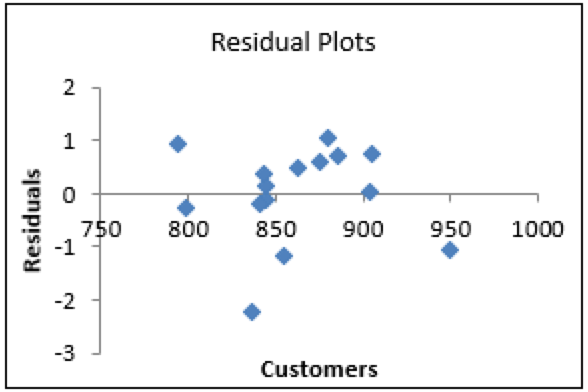
b) How well does the model fit the data? $r^2 = 0.65744$

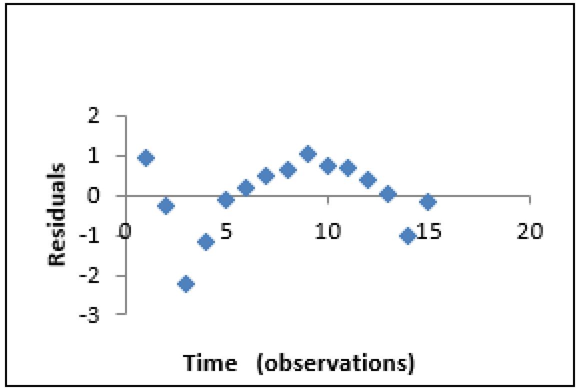
c) Test at the 1% level whether there is a significant relationship.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

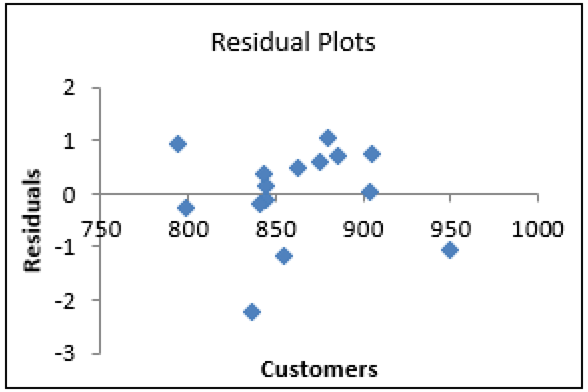
b) How well does the model fit the data? r² = 0.65744

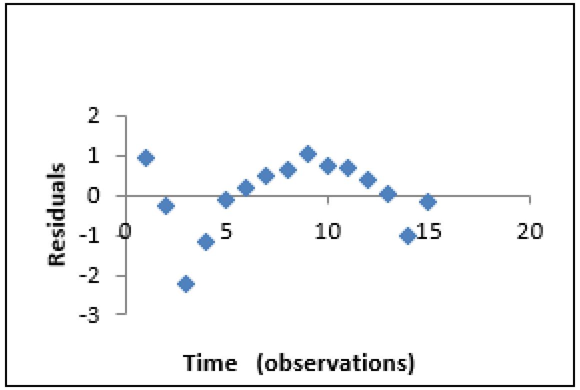
c) Test at the 1% level whether there is a significant relationship.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.



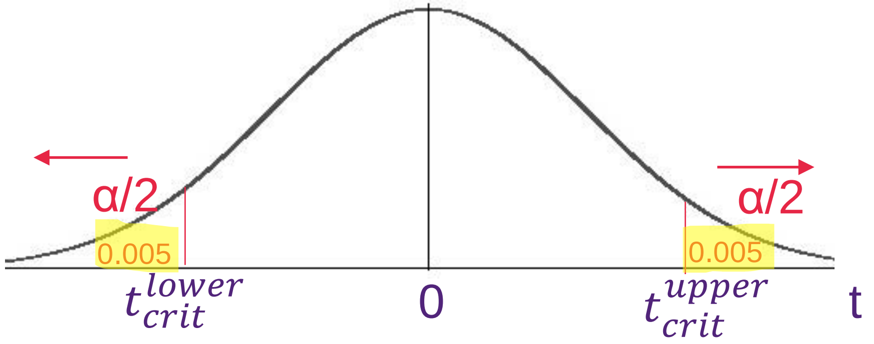


Step 1: State H_0 and H_1

H_0: $\beta_1 = 0$ (no significant relationship)

H_1: $\beta_1 \neq 0$ (significant relationship)

Two tail test



Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

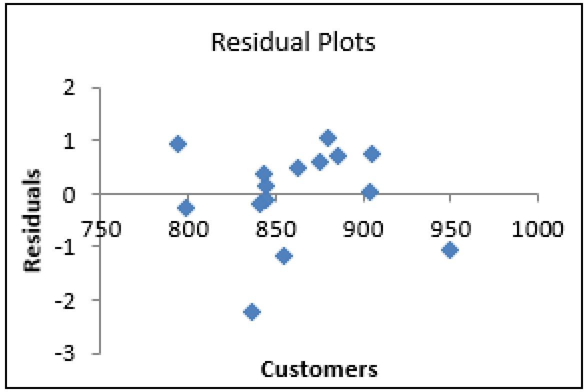
b) How well does the model fit the data? $r^2 = 0.65744$

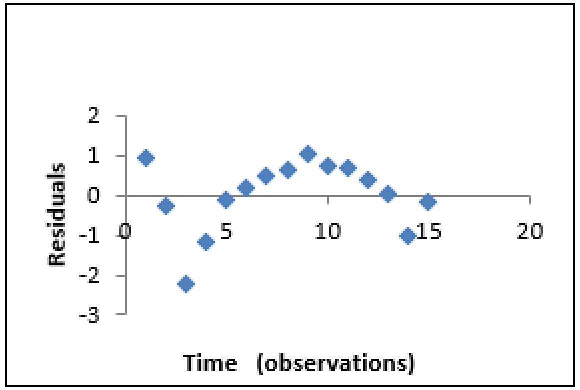
c) Test at the 1% level whether there is a significant relationship.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





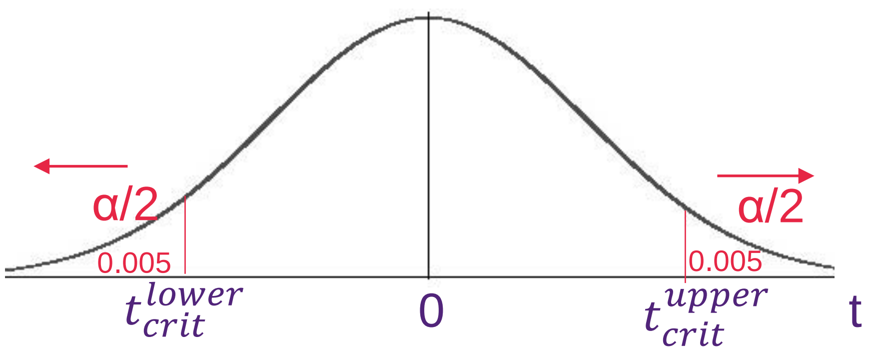
Step 1: State H_0 and H_1

H_0: $\beta_1 = 0$ (no significant relationship)

H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule

Reject H_0 if $|t_{calc}| > t_{crit}$ = ?



Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

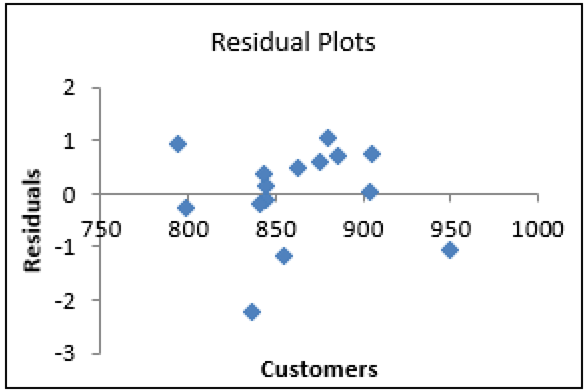
b) How well does the model fit the data? $r^2 = 0.65744$

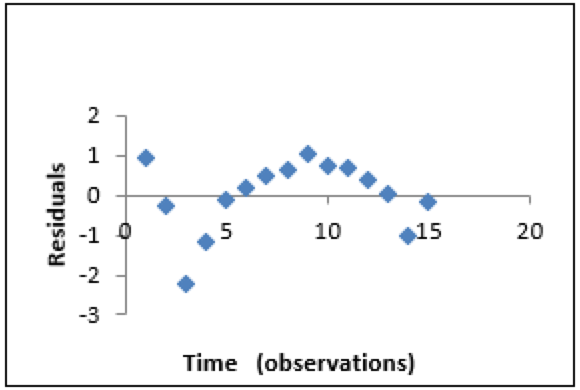
c) Test at the 1% level whether there is a significant relationship.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





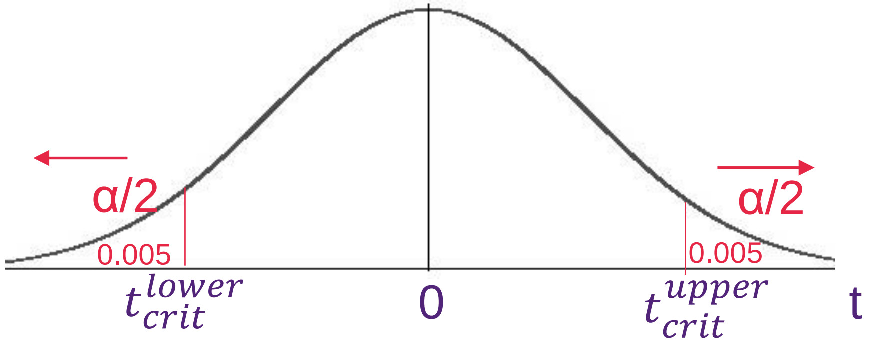
Step 1: State H_0 and H_1

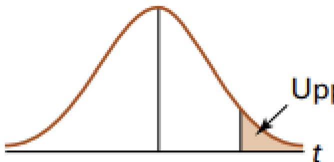
H_0: $\beta_1 = 0$ (no significant relationship)

H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule

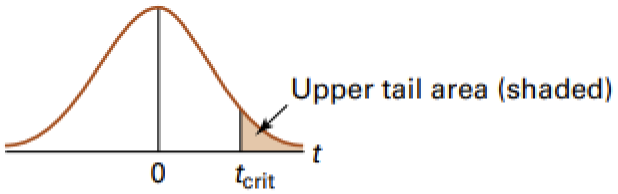
Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2 \;\; n-2} = t_{0.005,13}$ = ?





$t_{0.005,\, 13}$

	Upper tail areas					
df	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450



$t_{0.005,\,13}$

	Upper tail areas					
df	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450

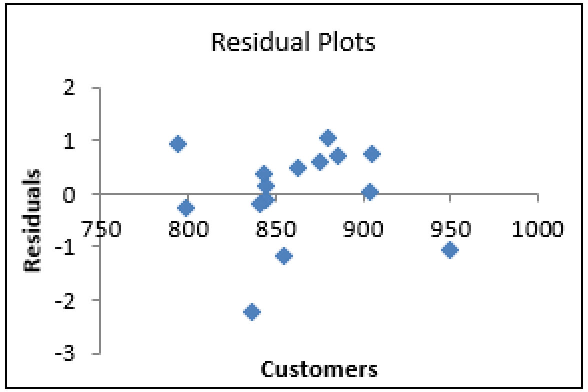
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

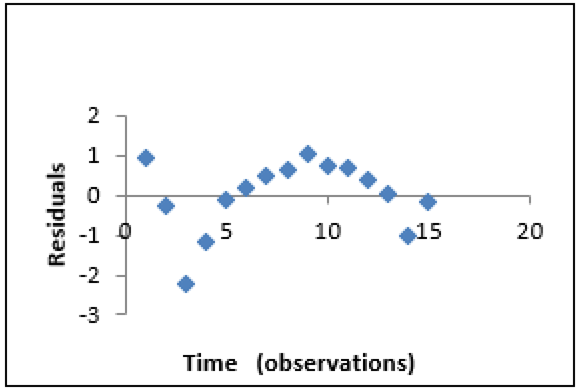
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data? $r^2 = 0.65744$
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





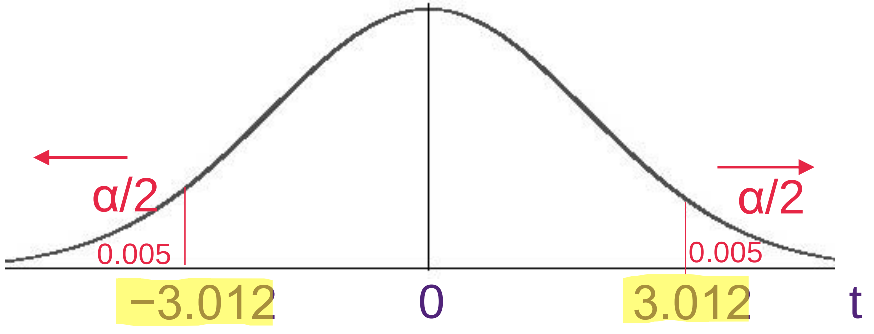
Step 1: State H_0 and H_1

H_0: $\beta_1 = 0$ (no significant relationship)

H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule

Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2\ \ n-2} = t_{0.005,13} = 3.012$



The University Of Queensland
AUSTRALIA

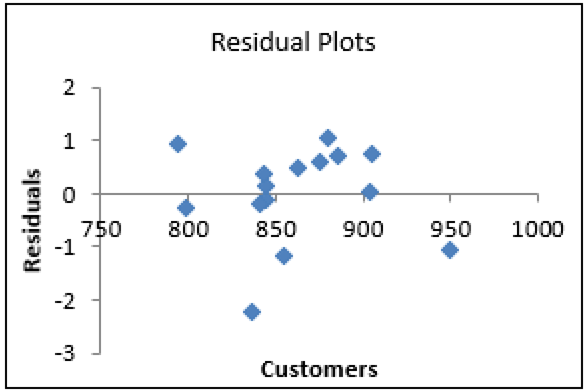
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

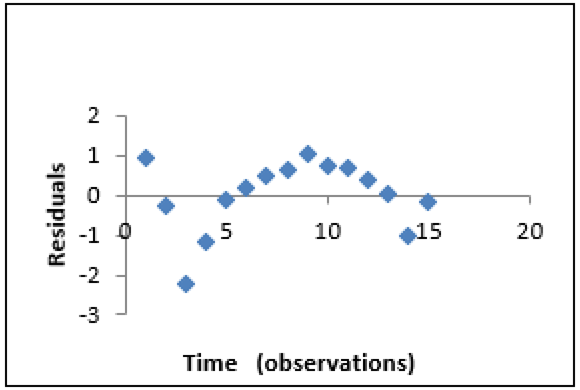
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data? $r^2 = 0.65744$
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.

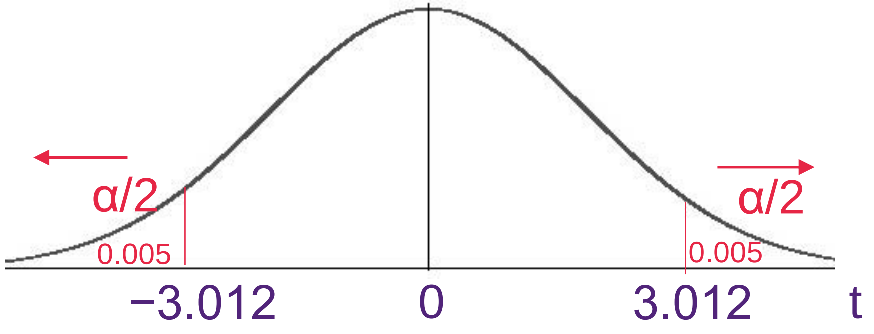




Step 1: State H_0 and H_1
H_0: $\beta_1 = 0$ (no significant relationship)
H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule
Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2 \;\; n-2} = t_{0.005,13} = 3.012$

Step 3: Calculate t_{calc}
t_{calc} = ?



adm T12 Q1 Q2 Q3

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

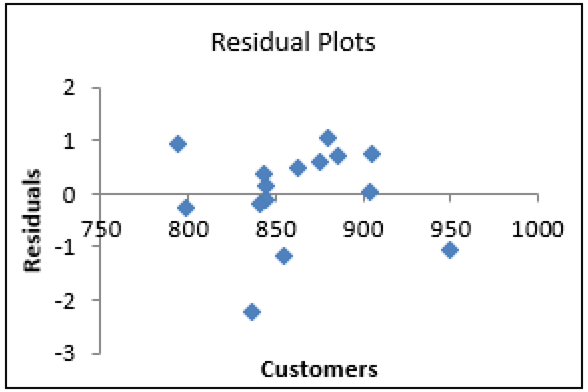
b) How well does the model fit the data? $r^2 = 0.65744$

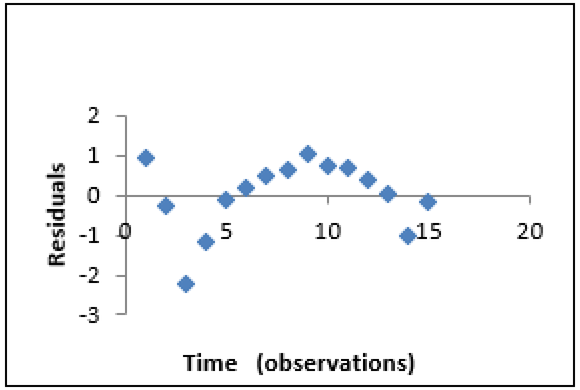
c) Test at the 1% level whether there is a significant relationship.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Step 1: State H_0 and H_1

H_0: $\beta_1 = 0$ (no significant relationship)

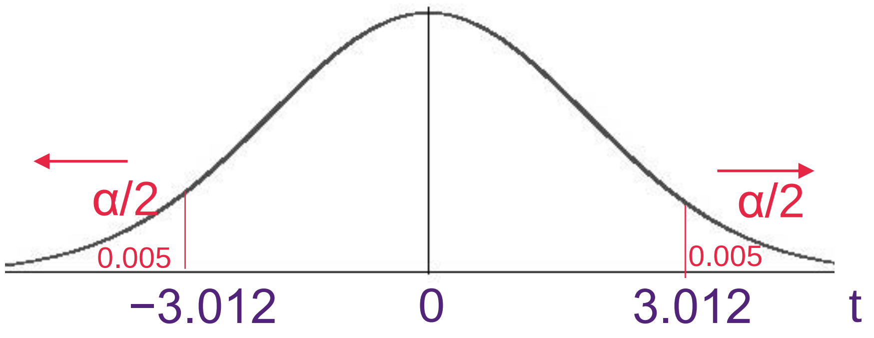
H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule

Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2 \;\; n-2} = t_{0.005,13} = 3.012$

Step 3: Calculate t_{calc}

t_{calc} = 4.995



The University Of Queensland AUSTRALIA

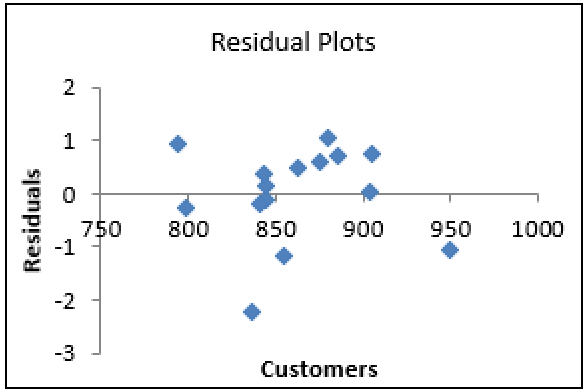
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

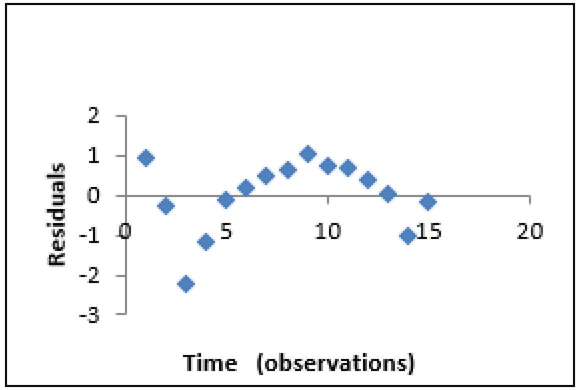
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data? r² = 0.65744
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.



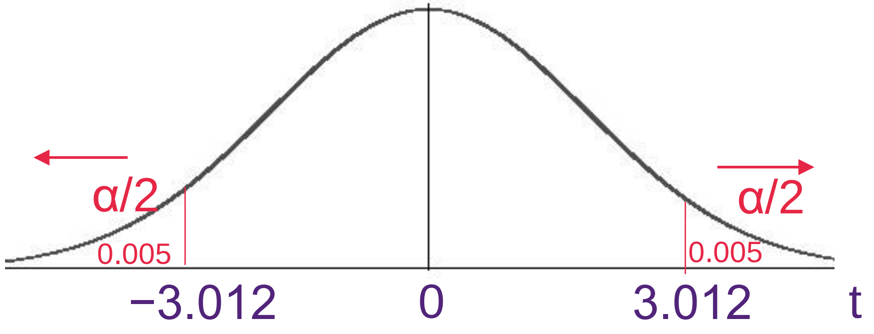


Step 1: State H_0 and H_1
H_0: $\beta_1 = 0$ (no significant relationship)
H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule
Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2\ \ n-2} = t_{0.005,13} = 3.012$

Step 3: Calculate t_{calc}
$t_{calc} = 4.995$

Step 4: Make a decision
$|t_{calc}| > t_{crit} \rightarrow |4.995| > 3.012 \rightarrow$?



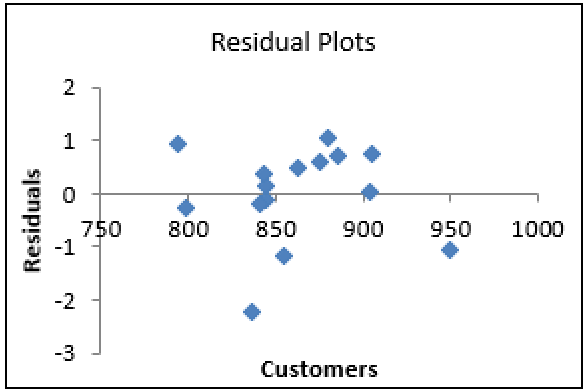
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

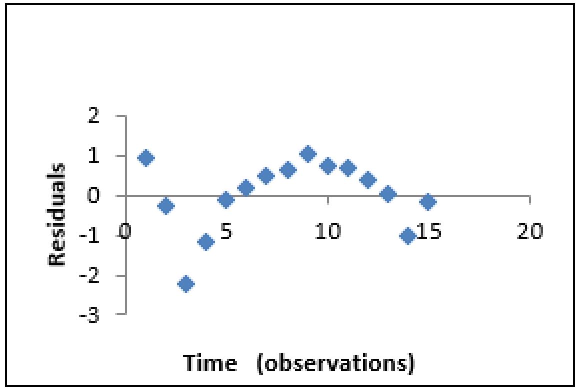
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data? $r^2 = 0.65744$
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Step 1: State H_0 and H_1

H_0: $\beta_1 = 0$ (no significant relationship)

H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule

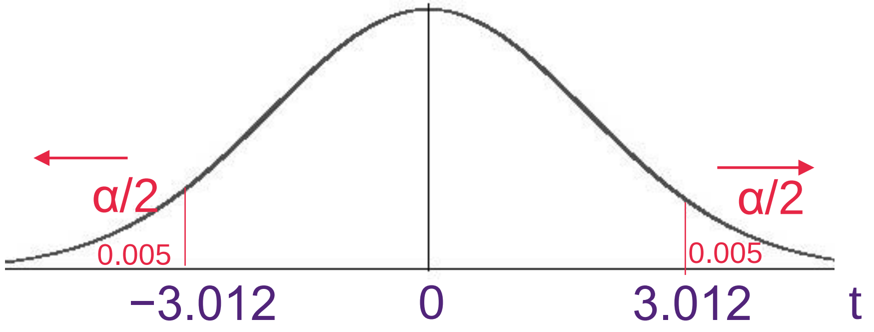
Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2\ \ n-2} = t_{0.005,13} = 3.012$

Step 3: Calculate t_{calc}

$t_{calc} = 4.995$

Step 4: Make a decision

$|t_{calc}| > t_{crit} \rightarrow |4.995| > 3.012 \rightarrow$ Reject H_0.



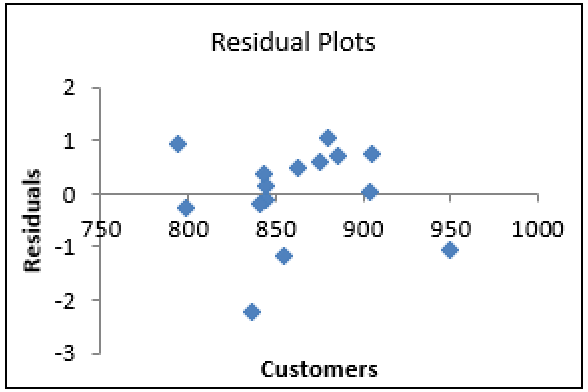
Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

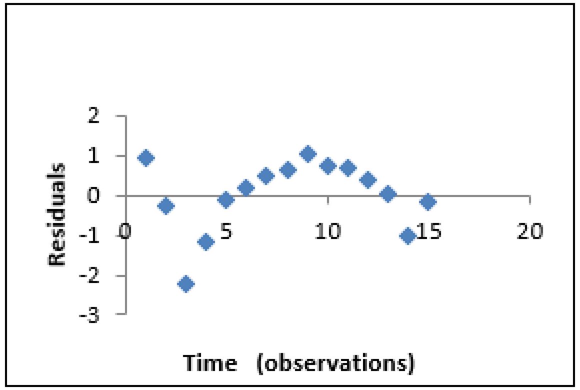
ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data? r² = 0.65744
c) Test at the 1% level whether there is a significant relationship.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





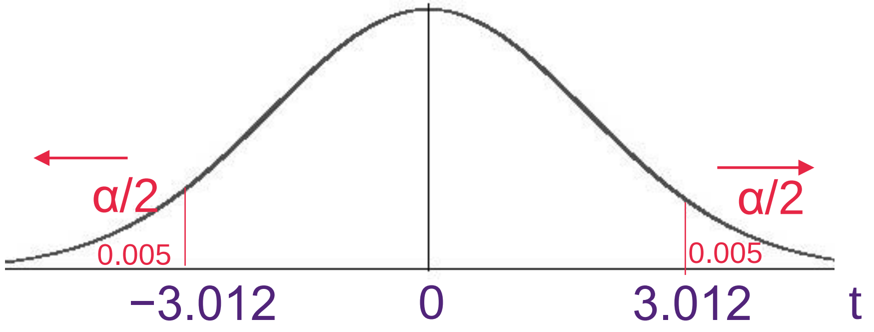
Step 1: State H_0 and H_1
H_0: $\beta_1 = 0$ (no significant relationship)
H_1: $\beta_1 \neq 0$ (significant relationship)

Step 2: Decision rule
Reject H_0 if $|t_{calc}| > t_{crit} = t_{\alpha/2\ \ n-2} = t_{0.005,13} = 3.012$

Step 3: Calculate t_{calc}
$t_{calc} = 4.995$

Step 4: Make a decision
$|t_{calc}| > t_{crit} \rightarrow |4.995| > 3.012 \rightarrow$ Reject H_0.

Step 5: Conclusion
There is sufficient evidence at the 1% level of significance to conclude that there is a relationship between sales and the number of customers.



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	SS	MS
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	Standard Error	t Stat
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

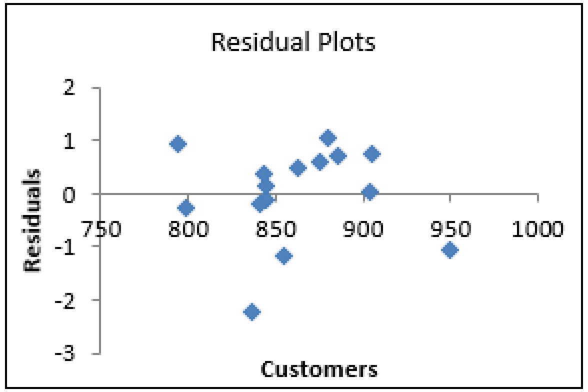
b) How well does the model fit the data? $r^2 = 0.65744$

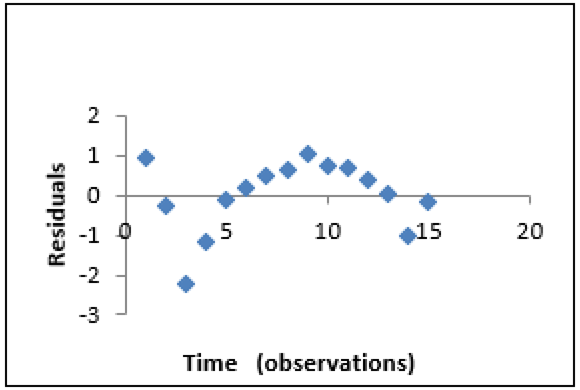
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Formulae for Simple Linear Regression

$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$ $SS_{YY} = SST$ $b_1 = \frac{SS_{XY}}{SS_{XX}}$

$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$ $SSR = b_1^2 * SS_{XX}$

$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$ $s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$

$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$ $s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

$$s_e = \sqrt{\frac{SSE}{n-2}} = ?$$

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

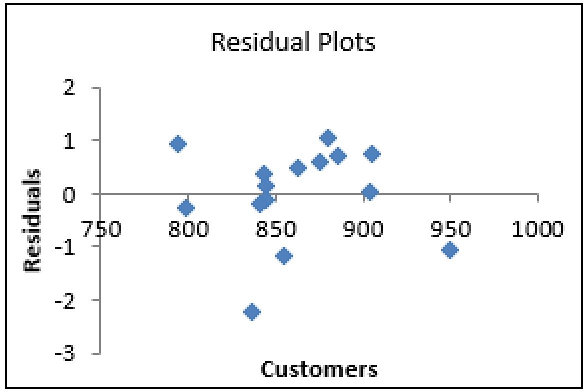
b) How well does the model fit the data? $r^2 = 0.65744$

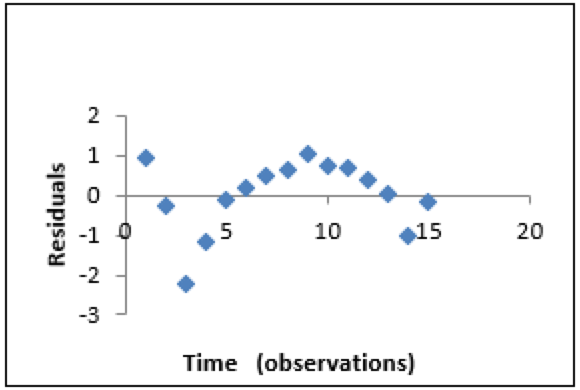
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

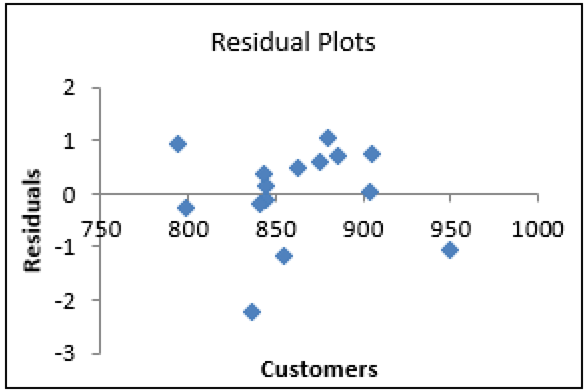
$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE} = ?$$

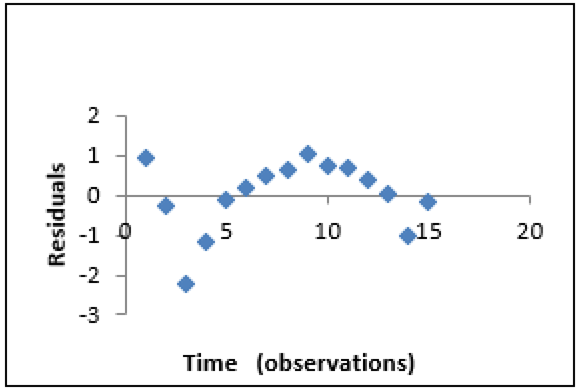
ANOVA

	df	SS	MS
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	Standard Error	t Stat
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data? $r^2 = 0.65744$
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.
d) Calculate the standard error of the estimate and explain what it represents.
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE} = \sqrt{0.8762} = 0.936$$

ANOVA

	df	SS	MS
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	Standard Error	t Stat
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

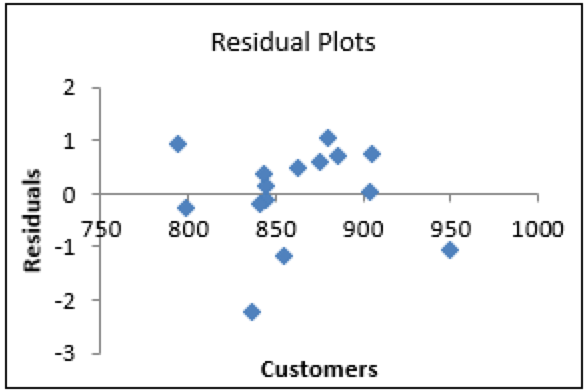
b) How well does the model fit the data? $r^2 = 0.65744$

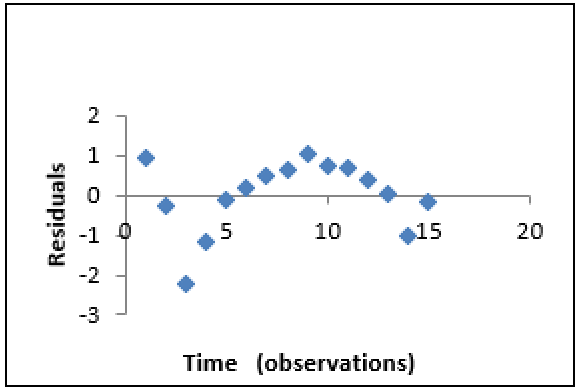
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}} \qquad SS_{YY} = SST \qquad b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n} \qquad SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n} \qquad s_e = \sqrt{\frac{SSE}{n\text{-}2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n} \qquad s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = sum of squares of X (sometimes written SSX)

29

adm T12 Q1 Q2 Q3

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

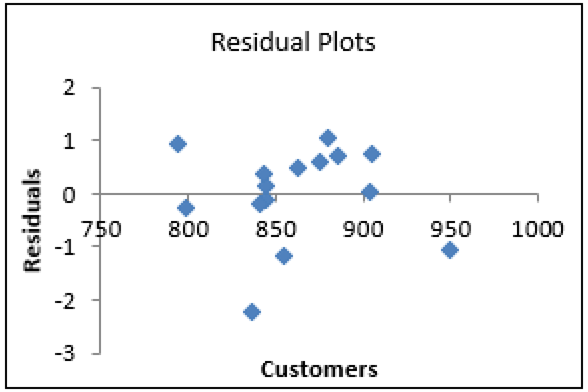
b) How well does the model fit the data? $r^2 = 0.65744$

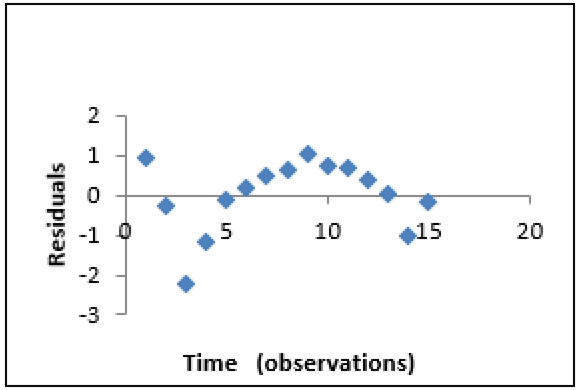
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents.

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.



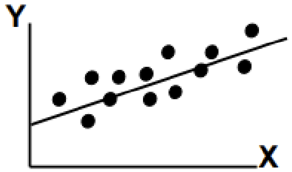


$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE} = \sqrt{0.8762} = 0.936$$

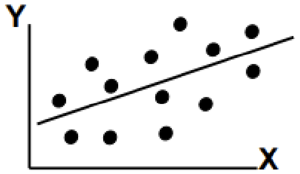
Standard deviation of the error of all points around the estimated regression line.

Comparing Standard Errors

S_e is a measure of the variation of observed Y values from the regression line



S_e is small in value if there is a strong linear relationship



S_e is larger in value if there is a weak linear relationship

The magnitude of s_e should always be judged relative to the size of the Y values in the sample data.

For example, a value of s_e = 2.3 ($'000) = $2,300 is small when compared to the fire damage values in the range of $14,000 to $43,000.

adm T12 Q1 Q2 Q3

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

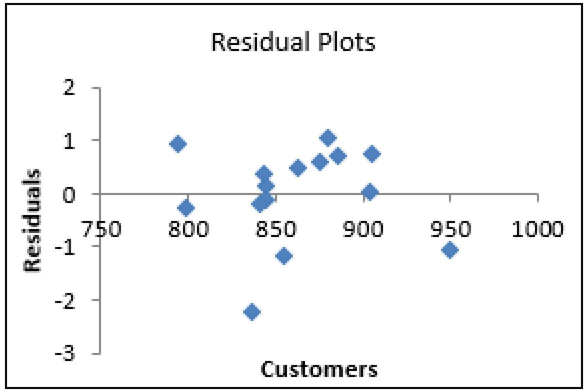
b) How well does the model fit the data? $r^2 = 0.65744$

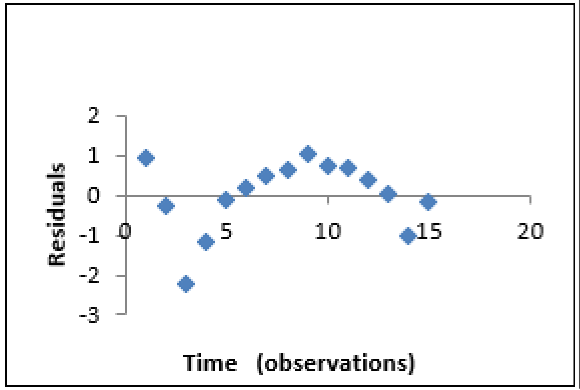
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Confidence interval for β_1
(the slope coefficient for the population)

The confidence interval estimate for β_1

$$\beta_1 = b_1 \pm t_{(n-2),\, \alpha/2} * s_{b1}$$

- this gives the **upper and lower limits of the slope** for the population linear regression equation.
- The **standard error for the slope coefficient (s_{b1})** is printed in the ANOVA table output (under the column Standard Error, next to coefficients).

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} **= sum of squares of X (sometimes written SSX)**

29

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

$$\beta_1 = b_1 \pm t_{\alpha/2,\, n-2} * s_{b_1}$$

ANOVA

	df	SS	MS
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	Standard Error	t Stat
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

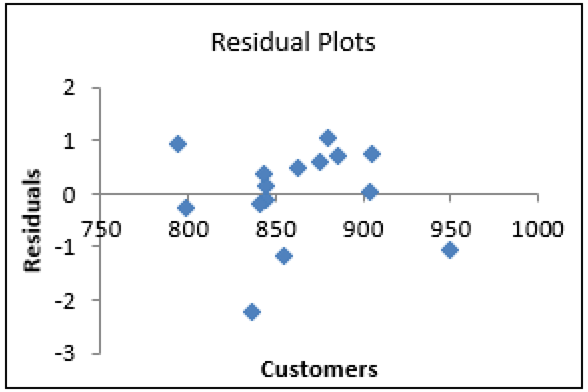
b) How well does the model fit the data? $r^2 = 0.65744$

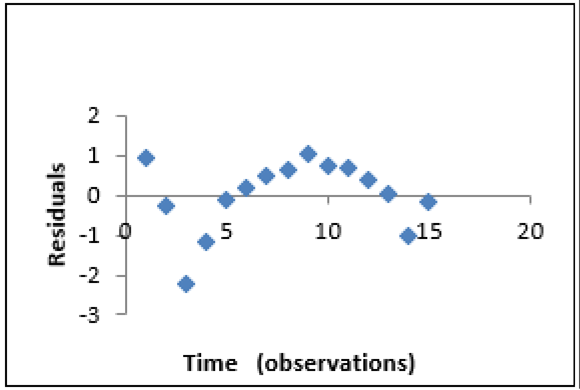
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Confidence interval for β_1
(the slope coefficient for the population)

The confidence interval estimate for β_1

$$\beta_1 = b_1 \pm t_{(n-2),\, \alpha/2} * s_{b1}$$

- this gives the **upper and lower limits of the slope** for the population linear regression equation.
- The **standard error for the slope coefficient (s_{b1})** is printed in the ANOVA table output (under the column Standard Error, next to coefficients).

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = **sum of squares of X (sometimes written SSX)**

29

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in \$thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	SS	MS
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	Standard Error	t Stat
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

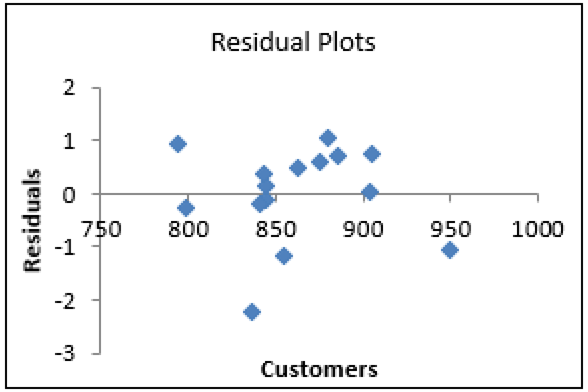
b) How well does the model fit the data? $r^2 = 0.65744$

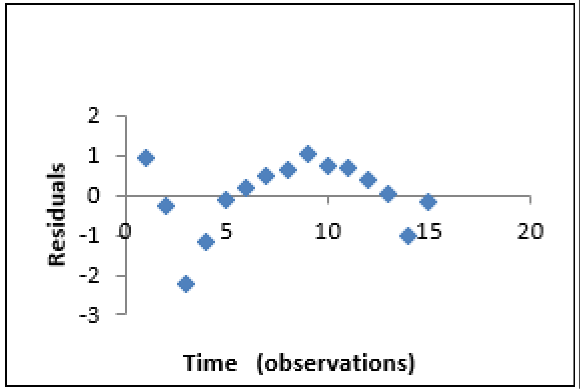
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





$$\beta_1 = b_1 \pm t_{\alpha/2,\, n-2} * s_{b_1}$$

$$s_{b_1} = 0.006$$

Confidence interval for β_1
(the slope coefficient for the population)

The confidence interval estimate for β_1

$$\beta_1 = b_1 \pm t_{(n-2),\, \alpha/2} * s_{b1}$$

- this gives the **upper and lower limits of the slope** for the population linear regression equation.
- The **standard error for the slope coefficient (s_{b1})** is printed in the ANOVA table output (under the column Standard Error, next to coefficients).

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = **sum of squares of X (sometimes written SSX)**

29

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

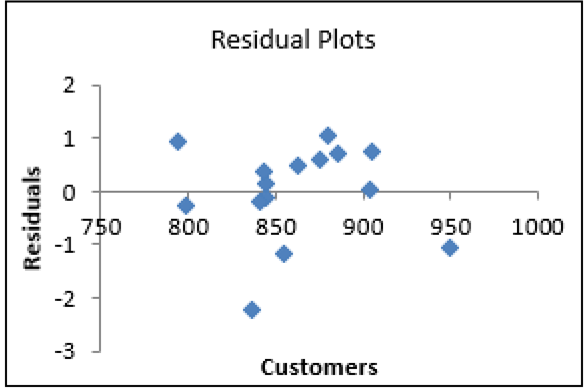
	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

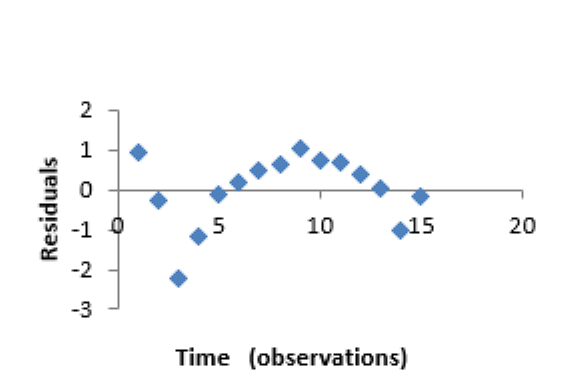
	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$
b) How well does the model fit the data? $r^2 = 0.65744$
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.
d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.
f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.

$$\beta_1 = b_1 \pm t_{\alpha/2,\, n-2} * s_{b_1}$$

$$= 0.031 \pm t_{0.025,\, 13} * 0.006$$





Confidence interval for β_1
(the slope coefficient for the population)

The confidence interval estimate for β_1

$$\beta_1 = b_1 \pm t_{(n-2),\, \alpha/2} * s_{b1}$$

- this gives the **upper and lower limits of the slope** for the population linear regression equation.
- The **standard error for the slope coefficient (s_{b1})** is printed in the ANOVA table output (under the column Standard Error, next to coefficients).

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum(Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum(X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

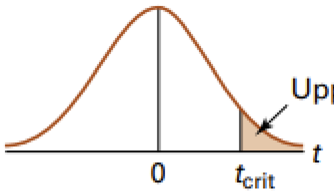
$$SS_{XY} = \sum(X_i * Y_i) - \frac{\sum(X_i) * \sum(Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = **sum of squares of X (sometimes written SSX)**

29

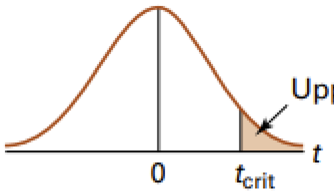
THE UNIVERSITY OF QUEENSLAND AUSTRALIA



$t_{0.025,\,13}$

	Upper tail areas					
df	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450

THE UNIVERSITY OF QUEENSLAND AUSTRALIA



$t_{0.025,\,13}$

	Upper tail areas					
df	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

b) How well does the model fit the data? $r^2 = 0.65744$

c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

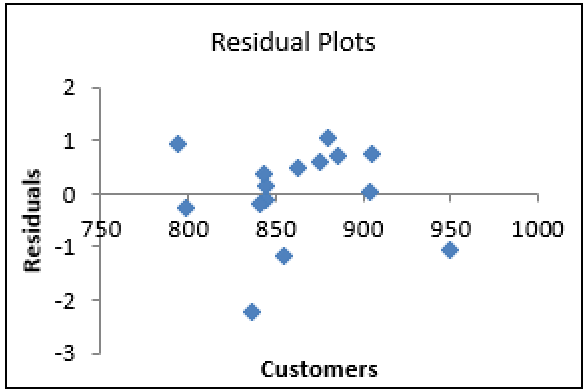
d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

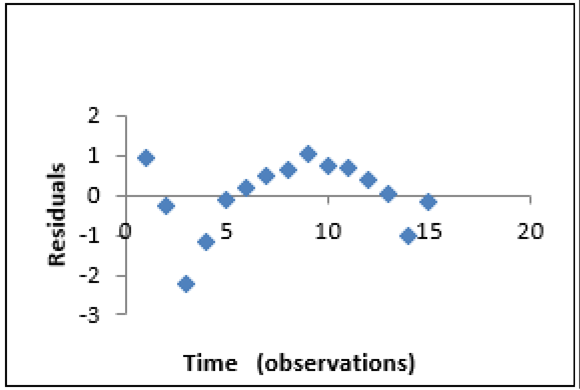
e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.

$$\beta_1 = b_1 \pm t_{\alpha/2,\, n-2} * s_{b_1}$$

$$= 0.031 \pm 2.16 * 0.006$$





Confidence interval for β_1
(the slope coefficient for the population)

The confidence interval estimate for β_1

$$\beta_1 = b_1 \pm t_{(n-2),\, \alpha/2} * s_{b1}$$

- this gives the **upper and lower limits of the slope** for the population linear regression equation.
- The **standard error for the slope coefficient (s_{b1})** is printed in the ANOVA table output (under the column Standard Error, next to coefficients).

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = **sum of squares of X (sometimes written SSX)**

29

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

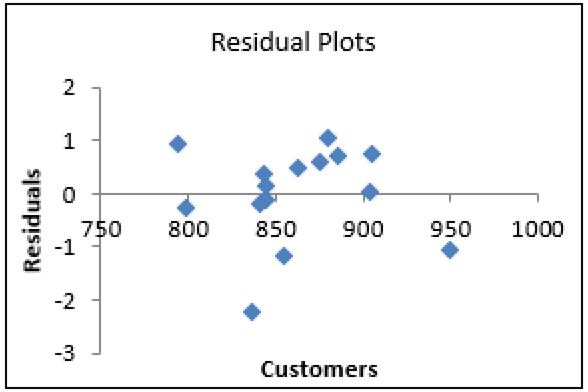
b) How well does the model fit the data? $r^2 = 0.65744$

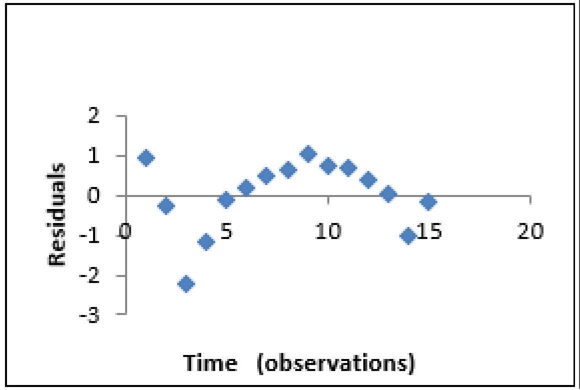
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents.

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





$$\beta_1 = b_1 \pm t_{\alpha/2,\, n-2} * s_{b_1}$$

$$= 0.031 \pm 2.16 * 0.006$$

$$0.018 < \beta_1 < 0.044$$

The slope of the regression relationship in the population is estimated with 95% confidence to be between 0.018 and 0.044.

Confidence interval for β_1
(the slope coefficient for the population)

The confidence interval estimate for β_1

$$\beta_1 = b_1 \pm t_{(n-2),\, \alpha/2} * s_{b1}$$

- this gives the **upper and lower limits of the slope** for the population linear regression equation.
- The **standard error for the slope coefficient (s_{b1})** is printed in the ANOVA table output (under the column Standard Error, next to coefficients).

Formulae for Simple Linear Regression

$$r^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{SSR}{SS_{YY}}$$

$$SS_{YY} = SST$$

$$b_1 = \frac{SS_{XY}}{SS_{XX}}$$

$$SS_{YY} = \sum Y_i^2 - \frac{\sum (Y_i)^2}{n}$$

$$SSR = b_1^2 * SS_{XX}$$

$$SS_{XX} = \sum X_i^2 - \frac{\sum (X_i)^2}{n}$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$$

$$SS_{XY} = \sum (X_i * Y_i) - \frac{\sum (X_i) * \sum (Y_i)}{n}$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}}$$

SS_{XX} = **sum of squares of X (sometimes written SSX)**

29

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

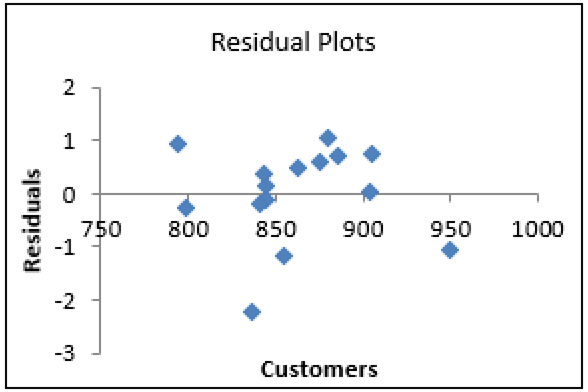
b) How well does the model fit the data? $r^2 = 0.65744$

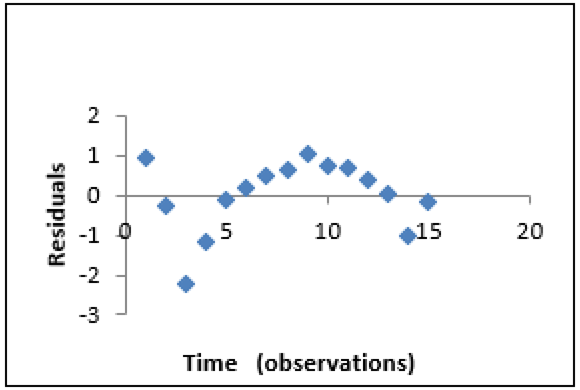
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents. $0.018 < \beta_1 < 0.044$

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	SS	MS
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	Standard Error	t Stat
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

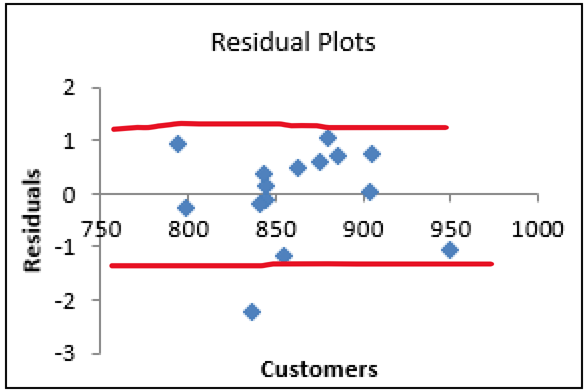
b) How well does the model fit the data? r² = 0.65744

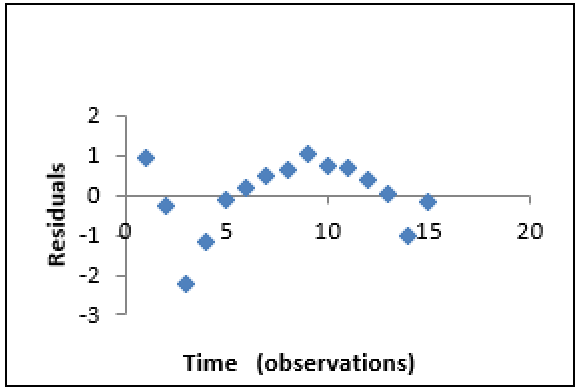
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents. $0.018 < \beta_1 < 0.044$

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





- The errors have constant variance around the regression line for all values of X.

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

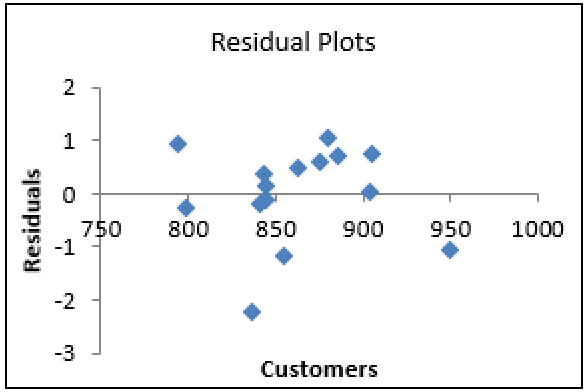
b) How well does the model fit the data? $r^2 = 0.65744$

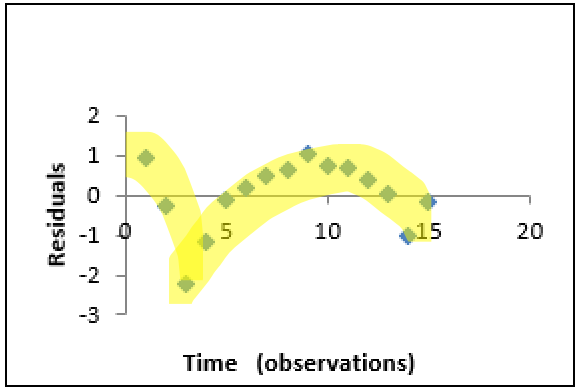
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β1. Explain in your own words what this confidence interval represents. $0.018 < \beta_1 < 0.044$

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





- The errors have constant variance around the regression line for all values of X.
- Errors are not independent of time or not random.

Least Squares Method Assumptions.

1. The model is linear.

<u>Error term assumptions.</u>

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

Q2. The manager of a discount electrical store wants to predict weekly sales based on the number of customers making purchases for a period of 15 weeks. The regression relationship between the number of customers and the sales (in $thous) is presented below, together with two residual plots – one with time and one with X on the horizontal axis.

ANOVA

	df	*SS*	*MS*
Regression	1	21.8604	
Residual	13		0.8762
Total		33.2506	

	Coefficients	*Standard Error*	*t Stat*
Intercept	-16.032	5.310	-3.019
Customers	0.031	0.006	4.995

a) State the estimated regression equation, explaining the variables. $\hat{Y}_i = -16.032 + 0.031 * X_i$

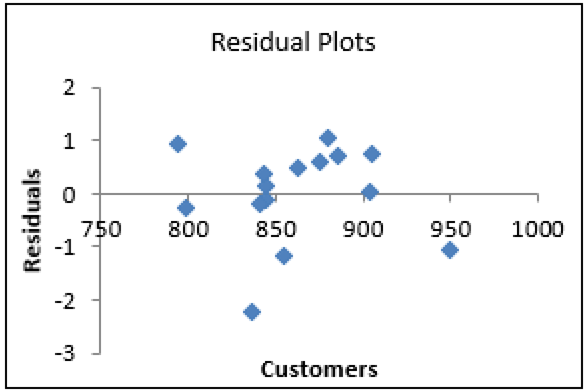
b) How well does the model fit the data? $r^2 = 0.65744$

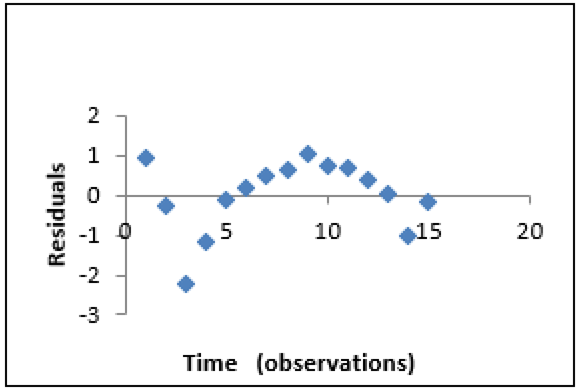
c) Test at the 1% level whether there is a significant relationship. Yes → Reject H_0.

d) Calculate the standard error of the estimate and explain what it represents. $s_e = 0.936$

e) Compute a 95% confidence interval for β_1. Explain in your own words what this confidence interval represents. $0.018 < \beta_1 < 0.044$

f) It is important that the assumptions about the error term are not violated in order to obtain a valid and reliable model which can be used for prediction. Observe the two residual plots below (from Excel and Kaddstat). Is there evidence that any of the assumptions have been violated? Discuss.





- The errors have constant variance around the regression line for all values of X.
- Errors are not independent of time or not random.
- Errors around the regression line are normally distributed at each value of X with mean 0. (We cannot determine using residual plots).

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

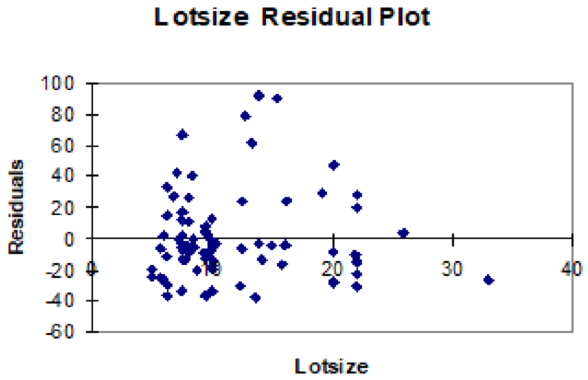
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables.

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

(Poll)

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

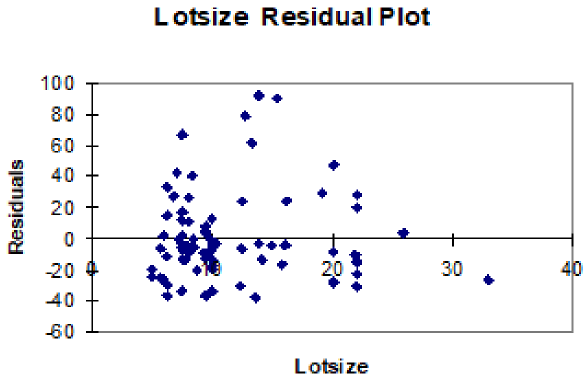
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables.

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



1. What symbol would you give to the value 137.35? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

2. What symbol would you give to the value 1.49? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

3. What symbol would you give to appraised value of houses? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

4. What symbol would you give to lot size? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

$$\hat{Y}_i = 137.35 + 1.49 * X_i$$

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

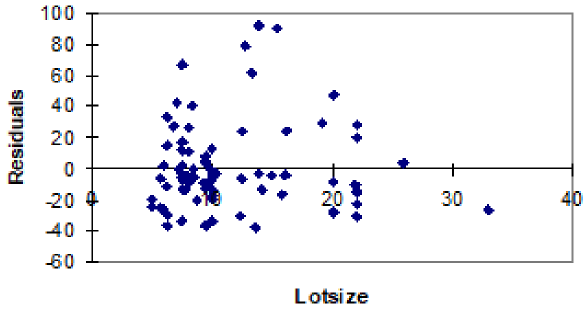
a) State the estimated linear relationship, explaining the variables.

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?

Lotsize Residual Plot



1. What symbol would you give to the value 137.35? (Single Choice)

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

*3. What symbol would you give to appraised value of houses? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

2. What symbol would you give to the value 1.49? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

4. What symbol would you give to lot size? (Single Choice) *

- Y
- b0 (b zero)
- b1 (b one)
- X
- SSE
- SSX
- n

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

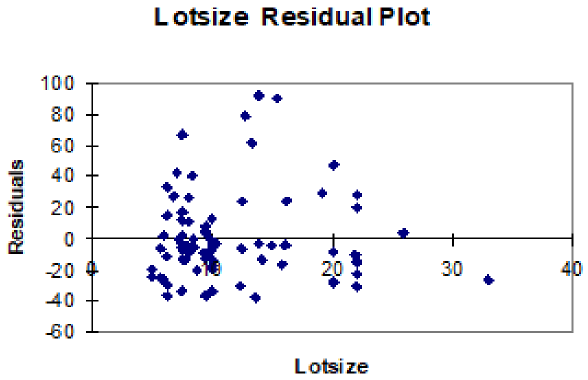
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables.

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



$$\hat{Y}_i = 137.35 + 1.49 * X_i$$

Y: (in $ thous) appraised value of houses.
X: ($100m^2$) lot size.

adm T12 Q1 Q2 Q3

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

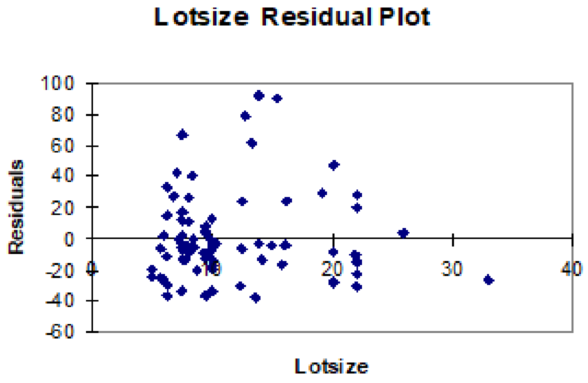
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



(Answers in chat)

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

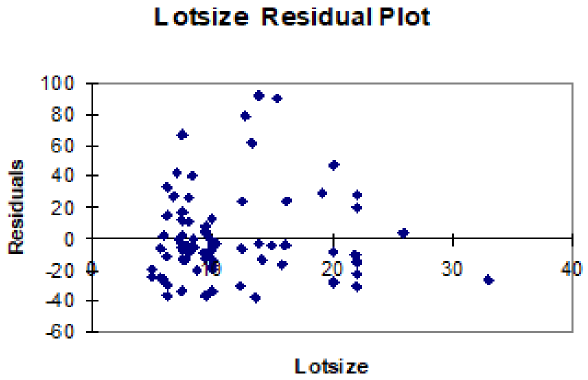
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



$b_1 = 1.49$, for every additional 100m² of lot size, the appraised value is expected to increase by 1.49*$1,000 = $1,490.

So every additional square meter of area on the lot changes the appraised value by $14,90.

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

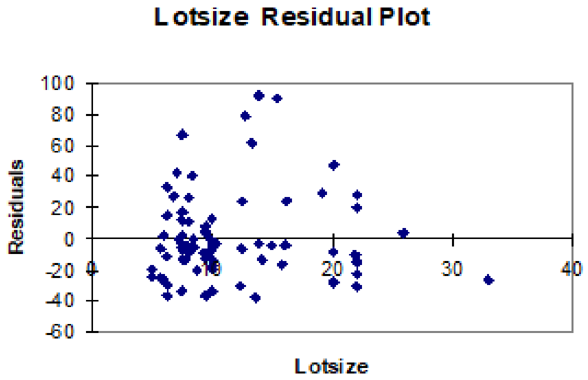
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change? $14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Step 1: State H_0 and H_1

H_0:

H_1:

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*	
Regression	1	6350.05	6350.05	
Residual			769.00	
Total		69408.42		

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

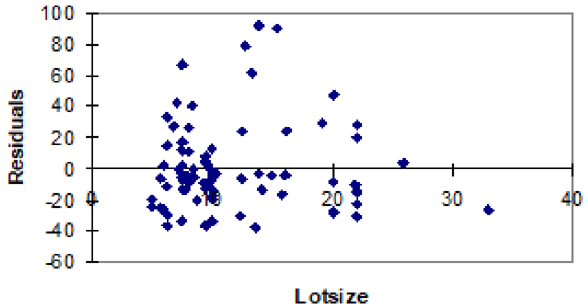
b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?

Lotsize Residual Plot

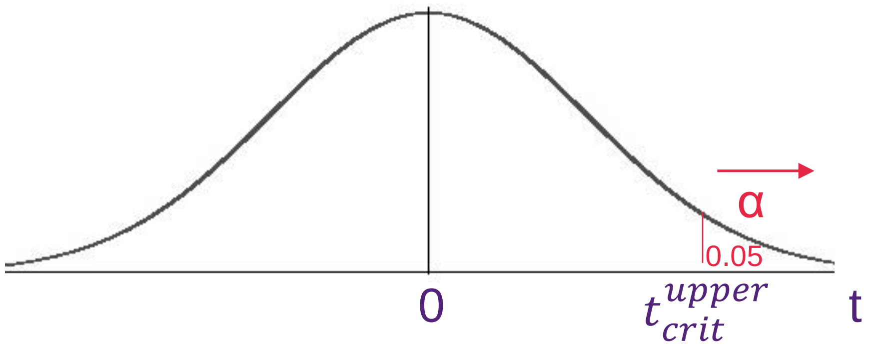


Step 1: State H_0 and H_1

H_0: $\beta_1 \leq 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

One tail test



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

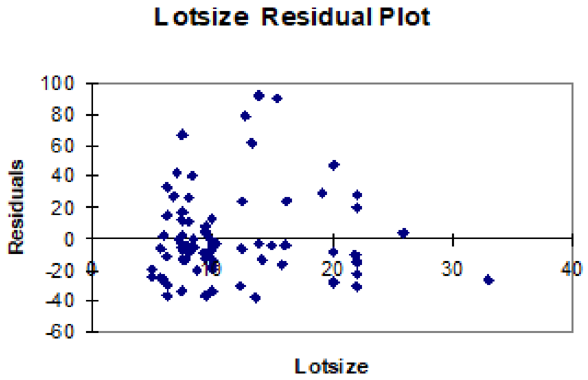
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



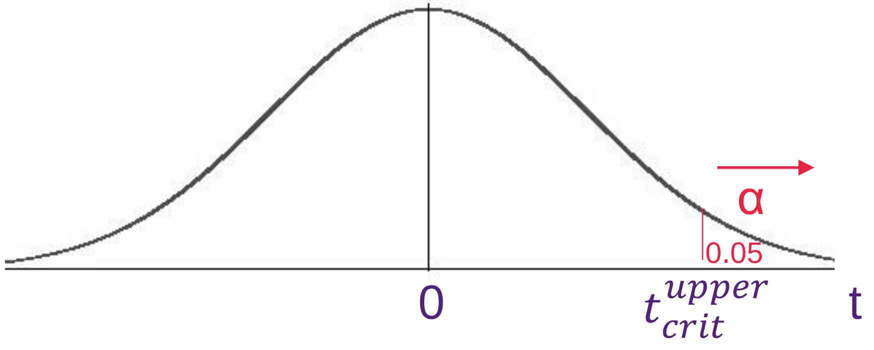
Step 1: State H_0 and H_1

H_0: $\beta_1 \leq 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

Reject H_0 if p-value < α



Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

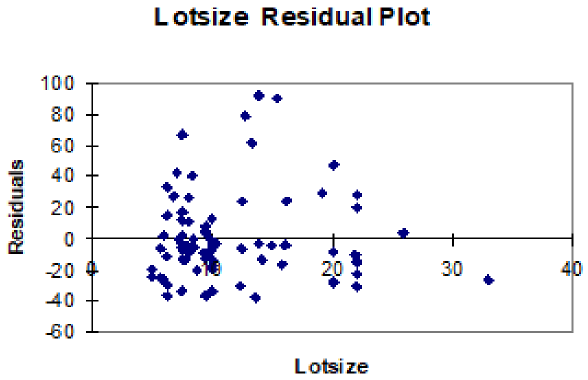
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



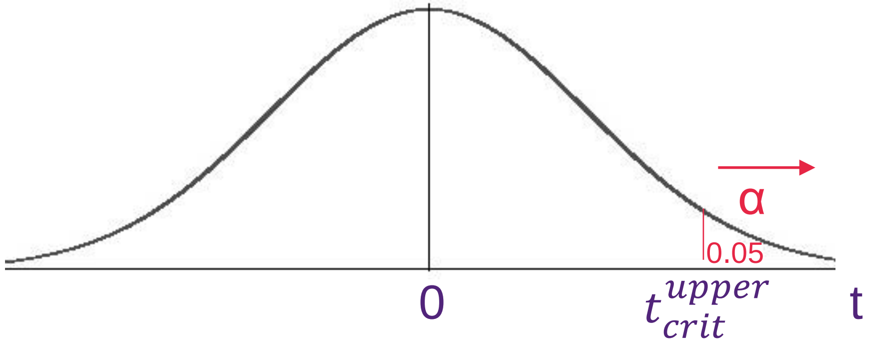
Step 1: State H_0 and H_1

H_0: $\beta_1 \le 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

Reject H_0 if p-value < α = 0.05



Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

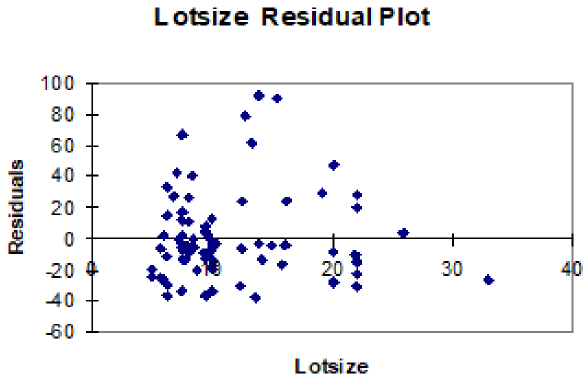
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Step 1: State H_0 and H_1

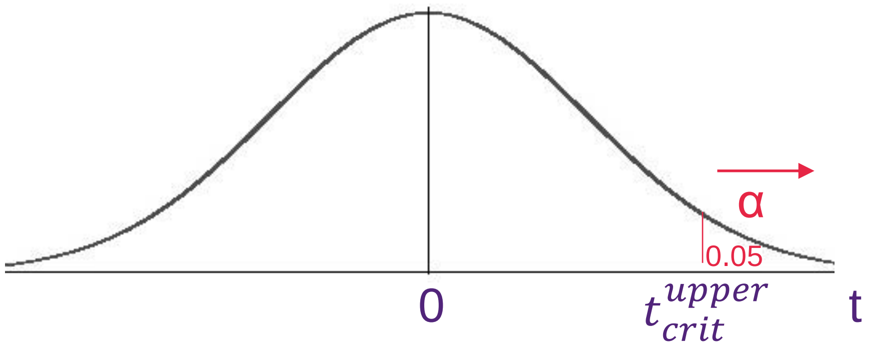
H_0: $\beta_1 \leq 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

Reject H_0 if p-value < α = 0.05

Step 3: Calculate p-value



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

Two tail test

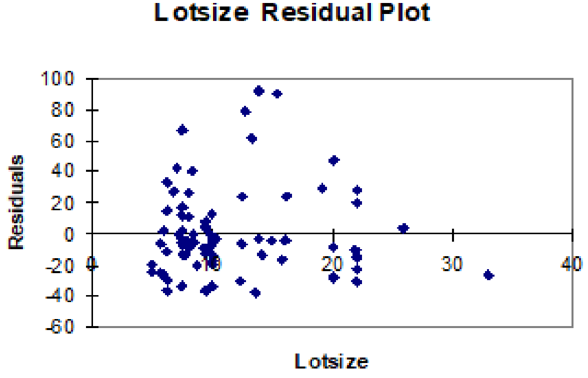
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Step 1: State H_0 and H_1

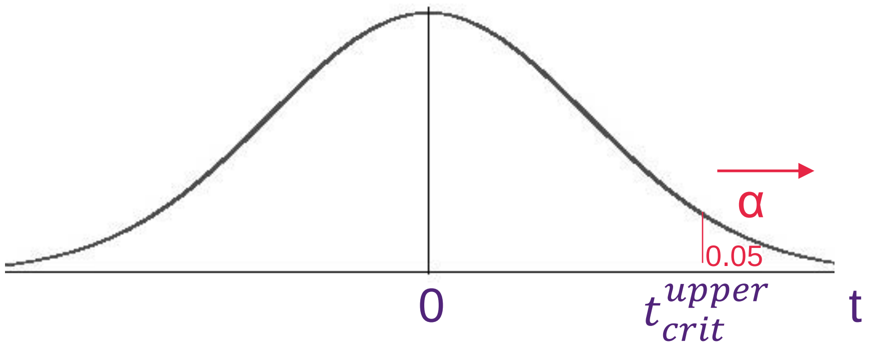
H_0: $\beta_1 \leq 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

Reject H_0 if p-value < α = 0.05

Step 3: Calculate p-value



Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

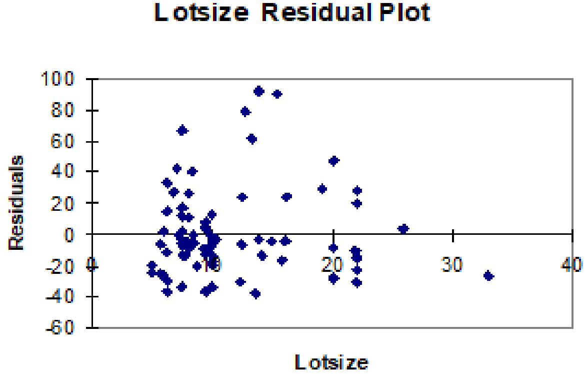
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Step 1: State H_0 and H_1

H_0: $\beta_1 \leq 0$ (none or negative relationship)

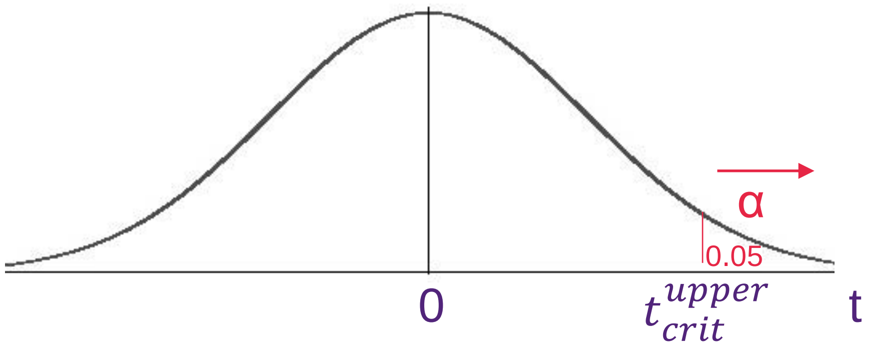
H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

Reject H_0 if p-value < α = 0.05

Step 3: Calculate p-value

p-value = 0.0516/2 = 0.0258



adm T12 Q1 Q2 Q3

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

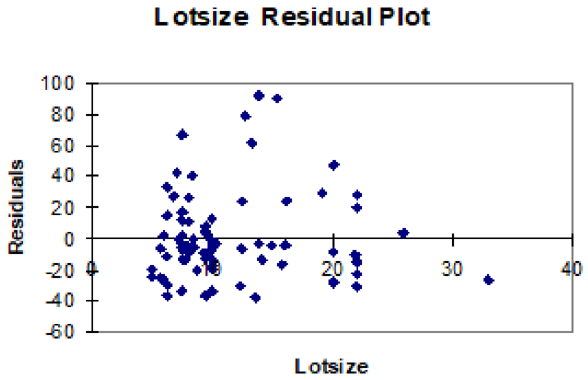
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Step 1: State H_0 and H_1

H_0: $\beta_1 \leq 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

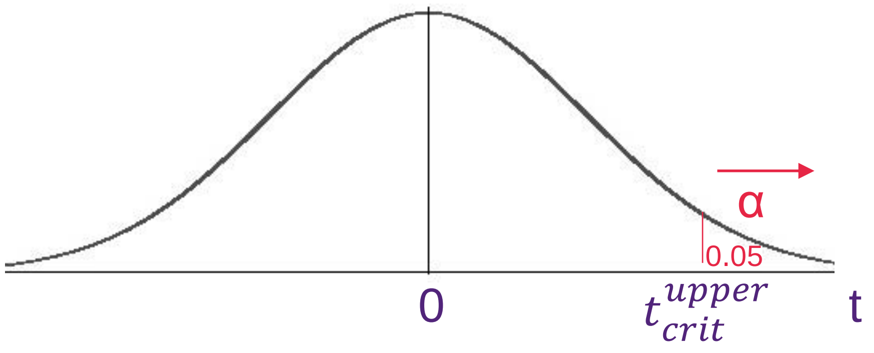
Reject H_0 if p-value < α = 0.05

Step 3: Calculate p-value

p-value = 0.0516/2 = 0.0258

Step 4: Make a decision

p-value < α → 0.0258 < 0.05 → ?



Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

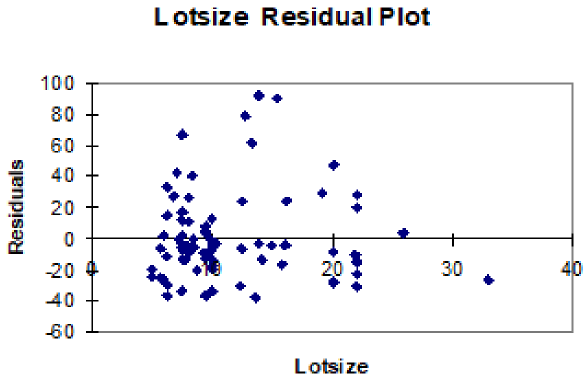
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Step 1: State H_0 and H_1

H_0: $\beta_1 \leq 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

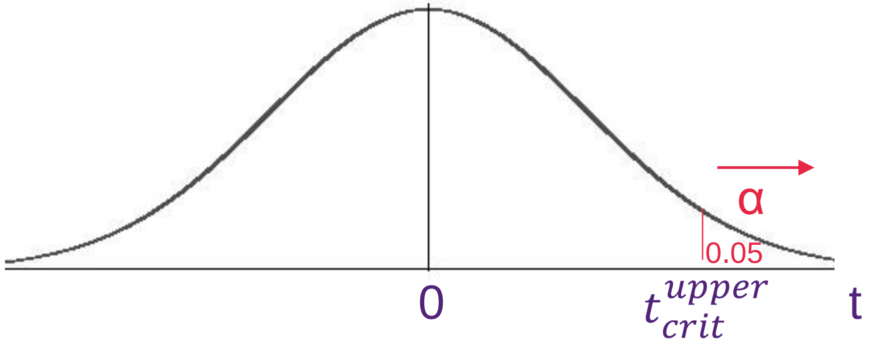
Reject H_0 if p-value < α = 0.05

Step 3: Calculate p-value

p-value = 0.0516/2 = 0.0258

Step 4: Make a decision

p-value < α → 0.0258 < 0.05 → Reject H_0.



Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

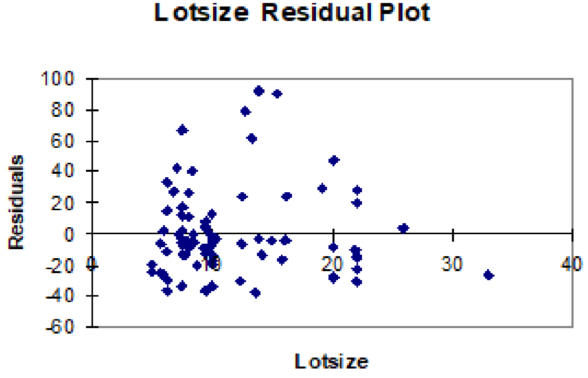
a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change?

$14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house?

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Step 1: State H_0 and H_1

H_0: $\beta_1 \le 0$ (none or negative relationship)

H_1: $\beta_1 > 0$ (positive relationship)

Step 2: Decision rule

Reject H_0 if p-value < α = 0.05

Step 3: Calculate p-value

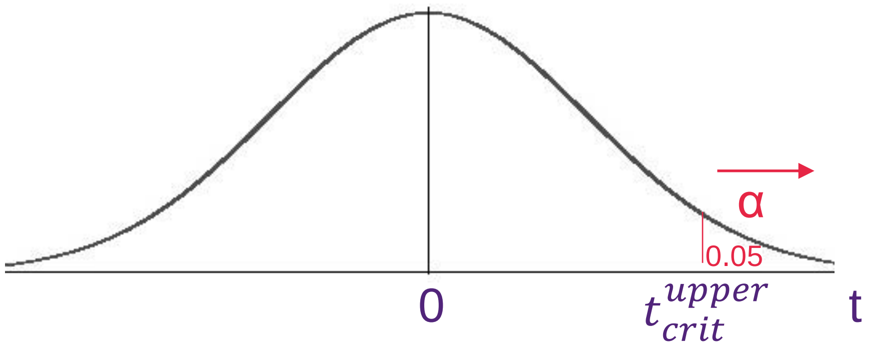
p-value = 0.0516/2 = 0.0258

Step 4: Make a decision

p-value < α → 0.0258 < 0.05 → Reject H_0.

Step 5: Conclusion

There is sufficient evidence at the 5% level of significance to conclude that there is a positive relationship between lot size and appraised value.



Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

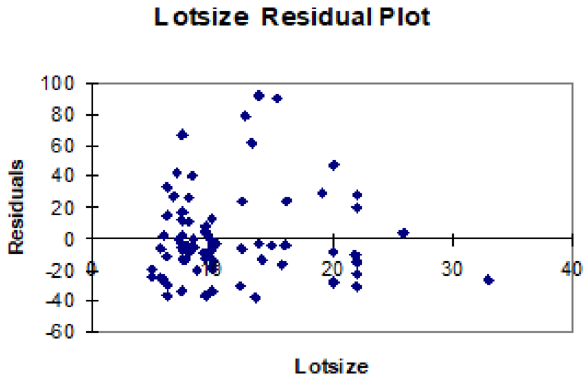
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables. $\widehat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change? $14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house? Yes $\rightarrow$ Reject H_0.

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



Least Squares Method Assumptions.

1. The model is linear.

<u>Error term assumptions.</u>

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

adm T12 Q1 Q2 Q3

The University Of Queensland
AUSTRALIA

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

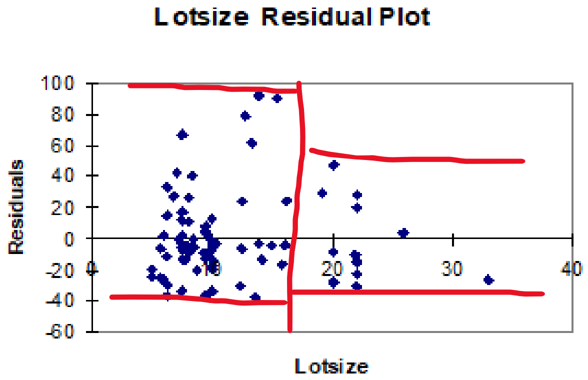
	df	*SS*	*MS*	
Regression	1	6350.05	6350.05	
Residual			769.00	
Total		69408.42		
	Coefficients	***St Error***	***t Stat***	***p-value***
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change? $14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house? Yes → Reject H_0.

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



- Constant variance: There is a greater variance when the lot size is smaller than 17 compared to greater than 17, so the variance is not constant.

Problem: heteroskedasticity

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

Q3. A market researcher wishes to determine the relationship between the appraised value of houses, measured in thousands of dollars, and their lot size, measured in hundreds of square metres, for a semi-rural community. A sample of 84 was used.

ANOVA

	df	*SS*	*MS*
Regression	1	6350.05	6350.05
Residual			769.00
Total		69408.42	

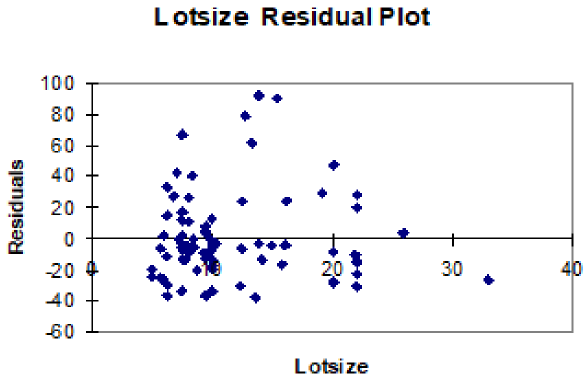
	Coefficients	*St Error*	*t Stat*	*p-value*
Intercept	137.35	6.80	20.20	1.39E-33
Lot size	1.49			0.0516

a) State the estimated linear relationship, explaining the variables. $\hat{Y}_i = 137.35 + 1.49 * X_i$

b) For every additional square metre of area on the lot, by how many dollars would you expect the appraised value to change? $14,90

d) The researcher believed that a **larger lot size would mean a higher appraised value**. Test this belief at the 5% level of significance using the *p*-value approach. Using this test, does a larger lot size lead to a higher appraised value for the house? Yes $\rightarrow$ Reject H_0.

e) Is there any evidence of a violation of assumptions which would lead us to question the validity of the model?



- Constant variance: There is a greater variance when the lot size is smaller than 17 compared to greater than 17, so the variance is not constant.

Problem: heteroskedasticity

- There are no patterns in the residuals so errors are independent of each value of X as well as each other.

Least Squares Method Assumptions.

1. The model is linear.

Error term assumptions.

2. The error terms have constant variance.
3. The error terms are independent (ie: they are not correlated) and occur randomly.
4. The error terms are normally distributed with an expected value (=mean) of zero.

ie: $E(e_i)=0$.

5

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

ECON1310

Tutorial 12 – Week 13

SIMPLE LINEAR REGRESSION II

At the end of this tutorial you should be able to

- Describe the assumptions that underpin the SLR model.
- Carry out analysis of the regression residuals to test whether the assumptions hold.
- Carry out hypothesis tests on the slope coefficient.

Thank you

Francisco Tavares Garcia
Academic Tutor | School of Economics
tavaresgarcia.github.io

Reference

Black et al. (2016), Australasian Business Statistics, 4th Edition, Wiley Australia.