The University of Queensland
Australia
CREATE CHANGE

ECON1310
Introductory Statistics for Social Sciences

Tutorial 11: SIMPLE LINEAR REGRESSION I

Tutor: Francisco Tavares Garcia

Happy Chinese New Year !!

Source: https://www.chinahighlights.com/travelguide/festivals/chinese-new-year-greetings.htm

Week 8 Timetable – Australian Day Public Holiday

Timetable Changes for Week 8 due to Australia Day Public Holiday

Posted on: Friday, 20 January 2023 15:53:21 o'clock AEST

Hi All

Next Thursday (Week 8) is a public holiday for Australia Day. To accommodate this, all Thursday tutorials will take place on **Wednesday** instead (not Friday as originally indicated in the course profile). Some of the consultation sessions have also changed as a result.

Please check the new Week 8 timetables under the Course Help tab for all the details. Let me know if you have any queries about this.

Kind Regards

Dominic

TIME	MON (23/01/2023)	TUE (24/01/2023)	WED (25/01/2023)	THU (26/01/2023)	FRI (27/01/2023)
10:00-10:30		TUT1/01 – PETER Online https://uqz.zoom.us/j/84419335972	TUT2/01 – PETER Online https://uqz.zoom.us/j/84419335972	AUSTRALIA DAY PUBLIC HOLIDAY 26 JANUARY 2023 NO CLASSES	
10:30-11:00					
11:00-11:30					
11:30-12:00		TUT1/02 - BEN Online https://uqz.zoom.us/j/7884658078	TUT2/02 - BEN Online https://uqz.zoom.us/j/7884658078		
12:00-12:30					
12:30-13:00					
13:00-13:30					
13:30-14:00					
14:00-14:30					
14:30-15:00					
15:00-15:30					
15:30-16:00					
16:00-16:30		TUT1/04 - FRANCISCO Online https://uqz.zoom.us/j/3181814065	TUT2/04 - FRANCISCO Online https://uqz.zoom.us/j/3181814065		
16:30-17:00					
17:00-17:30					

TIME	MON (23/01/2023)	TUE (24/01/2023)	WED (25/01/2023)	THU (26/01/2023)	FRI (27/01/2023)
10:00-10:30				AUSTRALIA DAY PUBLIC HOLIDAY 26 JANUARY 2023 NO CONSULTATION	
10:30-11:00					
11:00-11:30					
11:30-12:00					
12:00-12:30					DOMINIC (12pm – 1pm) https://uqz.zoom.us/j/5207526654
12:30-13:00					
13:00-13:30		BEN (1pm – 3pm) https://uqz.zoom.us/j/7884658078	PETER (1pm – 3pm) https://uqz.zoom.us/j/84419335972		
13:30-14:00					
14:00-14:30					
14:30-15:00					
15:00-15:30			FRANCISCO (3pm – 4pm) https://uqz.zoom.us/j/3181814065		
15:30-16:00					
16:00-16:30	FRANCISCO (4pm – 5pm) https://uqz.zoom.us/j/3181814065				
16:30-17:00					

CML 05 (2^{nd}) and CML 06 (only)

CML 5 and CML6 Reminder

Item is not available.

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

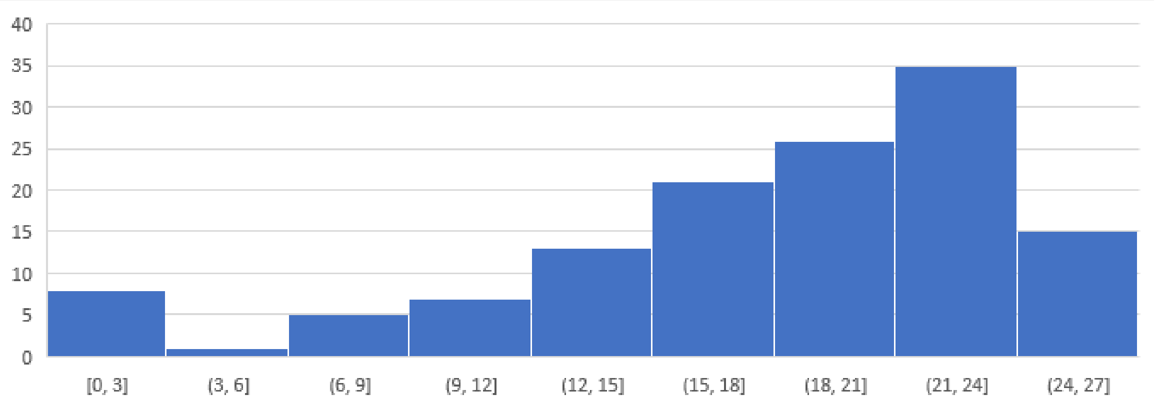
adm T11 Q1 Q2 Q3

The University of Queensland Australia

LBRT #1

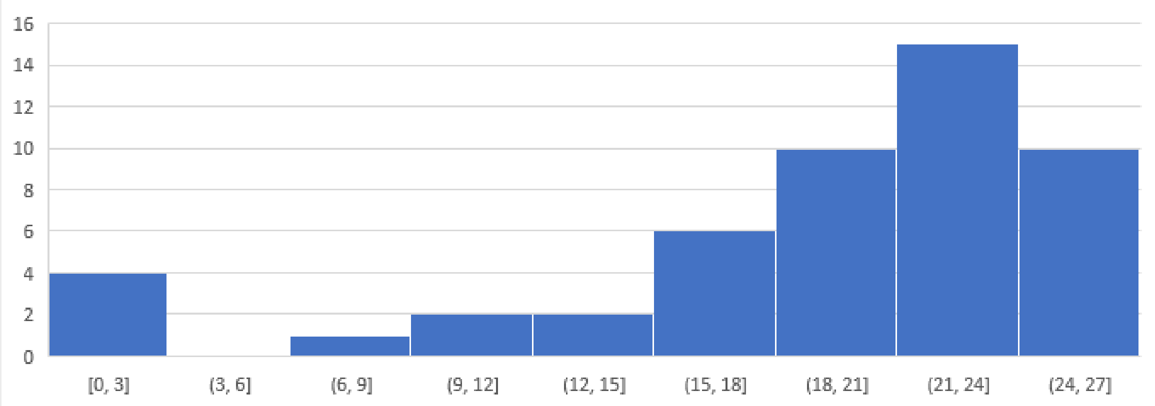
(excludes scores of 0)

All students
(n = 131)



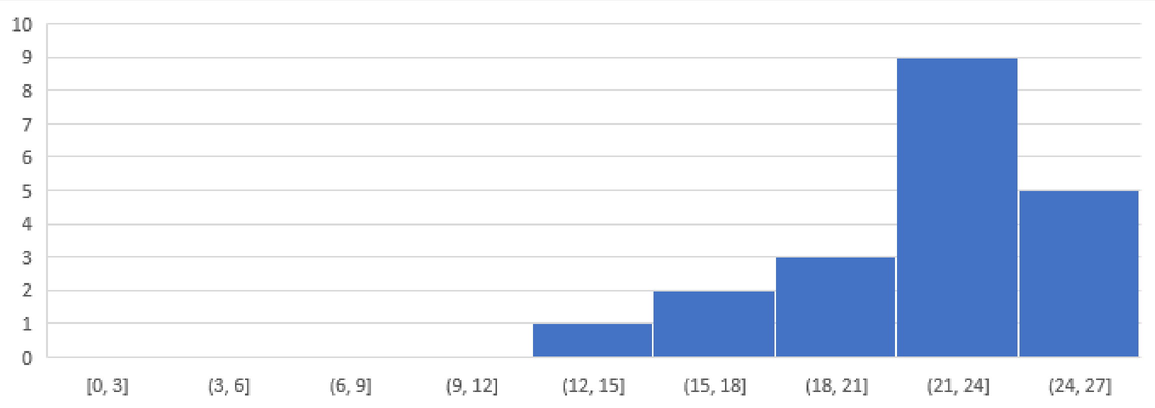
Mean = 18.99
Median = 20

Our tutorial
(n = 50)



Mean = 20.70
Median = 22

Students who attended at least 70% by Tutorial 10
(n = 20)

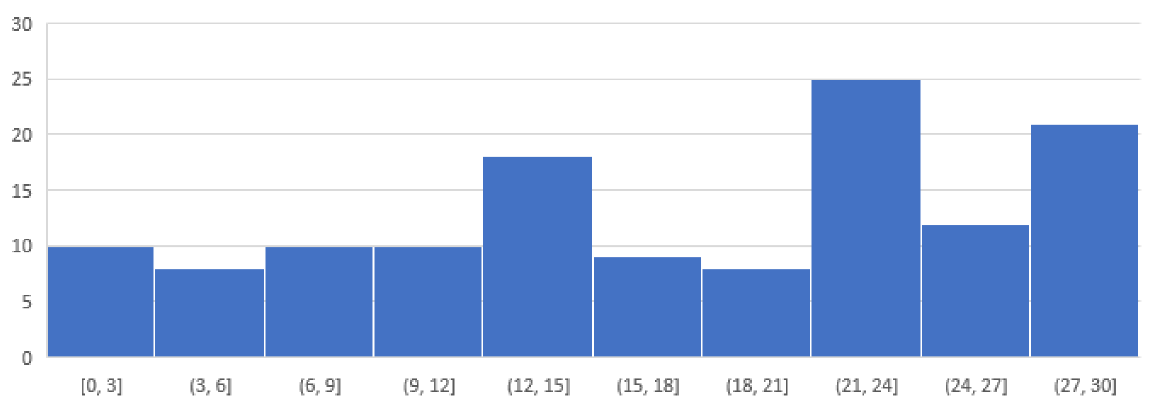


Mean = 22.48
Median = 24

LBRT #2

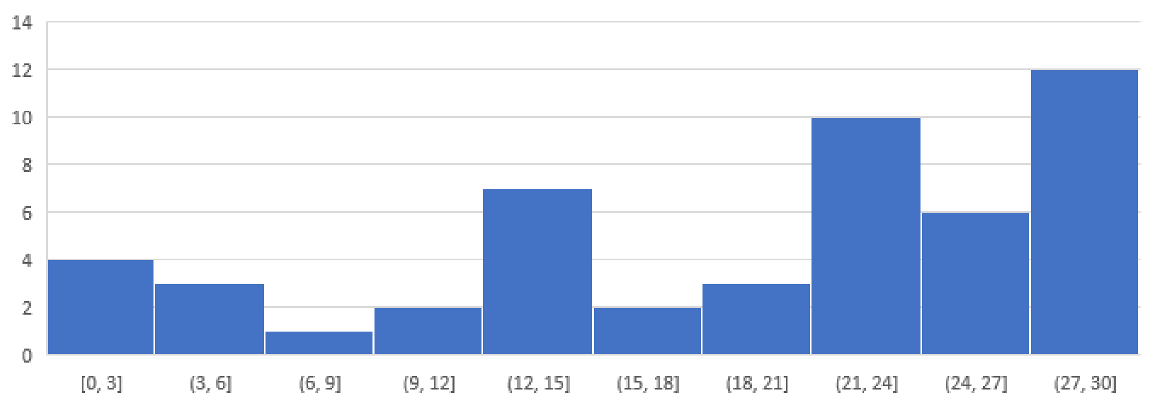
(excludes scores of 0)

All students
(n = 131)



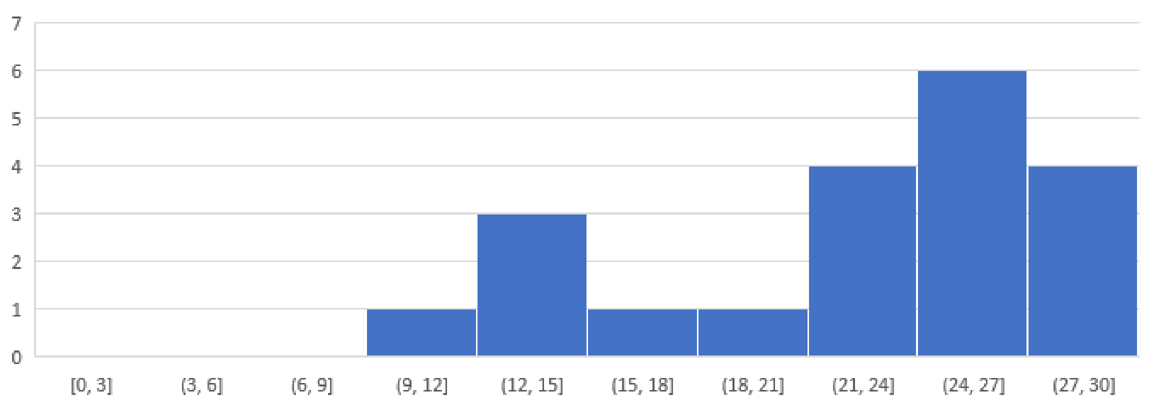
Mean = 18.70
Median = 20

Our tutorial
(n = 50)



Mean = 20.60
Median = 24

Students who attended at least 70% by Tutorial 10
(n = 20)



Mean = 22.85
Median = 25

ECON1310

Tutorial 11 – Week 12

SIMPLE LINEAR REGRESSION I

At the end of this tutorial you should be able to

- Formulate a SLR model and interpret the coefficients.
- Estimate a SLR equation using Excel.
- Interpret the coefficient of determination and standard error of the regression, given Excel output.
- Construct a confidence interval for the slope coefficient.

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis?

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used?
State the estimated simple linear regression equation.
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

d) Assuming no assumptions have been violated, predict ACDE when ANI is $15 000 and $50 000.
What reservations might you have about these predictions?

e) Calculate the 95% confidence interval for β_1.

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis?

(Answers in chat)

Why use Simple Linear Regression?

- used to **predict** the value of one variable (dependent variable) based on a given value of another variable (independent variable).
- used to explain the **impact** of a change in the independent variable on the dependent variable.
- SLR is an **inferential statistics technique** allowing conclusions to be made about a population parameter based on a sample statistic.

4

adm T11 Q1 Q2 Q3

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

It makes more sense to use income to explain expenditure than the other way around

→ expenditure depends on income ✓

→ income depends on expenditure ✗

Why use Simple Linear Regression?

- used to **predict** the value of one variable (dependent variable) based on a given value of another variable (independent variable).
- used to explain the **impact** of a change in the independent variable on the dependent variable.
- SLR is an **inferential statistics technique** allowing conclusions to be made about a population parameter based on a sample statistic.

4

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used?
State the estimated simple linear regression equation.
Interpret the slope coefficient of ANI and the constant value.

Example 1.
Analysing Excel Regression Output

n = sample size
(can find from df. residual = n-2, or from df total = n-1)

SS = Sum of Squares
SSR = Sum of Squares Regression
SSE = Sum of Squares Residuals (Errors)
SST = Sum of Squares Total = SSR + SSE

MSE = Mean Square Error

38

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used? n = 10
State the estimated simple linear regression equation.
Interpret the slope coefficient of ANI and the constant value.

Example 1.
Analysing Excel Regression Output

n = sample size
(can find from df. residual = n-2, or from df total = n-1)

SS = Sum of Squares
SSR = Sum of Squares Regression
SSE = Sum of Squares Residuals (Errors)
SST = Sum of Squares Total = SSR + SSE

MSE = Mean Square Error

38

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used? $n = 10$
State the estimated simple linear regression equation.
Interpret the slope coefficient of ANI and the constant value.

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used? n = 10
State the estimated simple linear regression equation.
Interpret the slope coefficient of ANI and the constant value.

$$\hat{Y}_i = 14.783 + 2.217 * X_i$$

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

$$\hat{Y}_i = 14.783 + 2.217 * X_i$$

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

adm T11 Q1 Q2 Q3

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

$$\hat{Y}_i = 14.783 + 2.217 * X_i$$

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * \$100.

adm T11 Q1 Q2 Q3

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

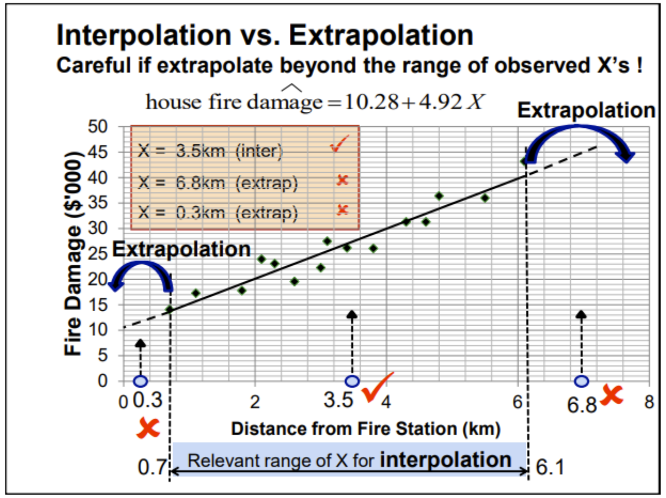
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

$$\hat{Y}_i = 14.783 + 2.217 * X_i$$

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

$r^2 = ?$

Coefficient of Determination, r^2

The **coefficient of determination** is the portion of the total variation in the dependent variable **that is explained** by variation in the independent variable.

$$r^2 = \frac{SSR}{SST} = \frac{\text{regression } \textit{sum} \text{ of squares}}{\textit{total} \text{ sum of squares}}$$
$$= \frac{SSR}{SSR + SSE}$$

Note that r^2 can only take values between **0 and 1.** It gives a measure of how useful is the SLR model.

13

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

$$r^2 = \frac{SSR}{SST} = \frac{3618.783}{4352} = 0.8315$$

Coefficient of Determination, r^2

The **coefficient of determination** is the portion of the total variation in the dependent variable **that is explained** by variation in the independent variable.

$$r^2 = \frac{SSR}{SST} = \frac{\text{regression } \textit{sum} \text{ of squares}}{\textit{total} \text{ sum of squares}} = \frac{SSR}{SSR + SSE}$$

Note that r^2 can only take values between **0 and 1.** It gives a measure of how useful is the SLR model.

13

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

$$r^2 = \frac{SSR}{SST} = \frac{3618.783}{4352} = 0.8315$$

83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

Coefficient of Determination, r^2

The **coefficient of determination** is the portion of the total variation in the dependent variable **that is explained** by variation in the independent variable.

$$r^2 = \frac{SSR}{SST} = \frac{\text{regression } \textit{sum} \text{ of squares}}{\textit{total} \text{ sum of squares}} = \frac{SSR}{SSR + SSE}$$

Note that r^2 can only take values between **0 and 1.** It gives a measure of how useful is the SLR model.

13

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

r^2 = 0.8315, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is $15 000 and $50 000.
What reservations might you have about these predictions?

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

$\hat{Y}_i = 14.783 + 2.217 * X_i$

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * \$100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

r^2 = 0.8315, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is \$15 000 and \$50 000.
What reservations might you have about these predictions?

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

$$\hat{Y}_i = 14.783 + 2.217 * X_i$$

$$\hat{Y}_i = 14.783 + 2.217 * 15 = 48.038$$

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	SS	MS	F
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	Standard Error	t Stat	P-value
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

$r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is $15 000 and $50 000.
What reservations might you have about these predictions?

adm T11 Q1 Q2 Q3

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

$\hat{Y}_i = 14.783 + 2.217 * X_i$

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

$\hat{Y}_i = 14.783 + 2.217 * 15 = 48.038$

When ANI = \$15,000,
ACDE $= 48.038 * 100 =$ \$4,803.80

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * \$100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

$r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is \$15 000 and \$50 000.
What reservations might you have about these predictions?

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * \$100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

$r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is \$15 000 and \$50 000.
What reservations might you have about these predictions?

$\hat{Y}_i = 14.783 + 2.217 * X_i$

$\hat{Y}_i = 14.783 + 2.217 * 15 = 48.038$

When ANI = \$15,000,
ACDE $= 48.038 * 100 =$ \$4,803.80

$\hat{Y}_i = 14.783 + 2.217 * 50 = 125.633$

When ANI = \$50,000,
ACD $= 125.633 * 100 =$ \$12,563.30

adm T11 Q1 Q2 Q3

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

$\hat{Y}_i = 14.783 + 2.217 * X_i$

$\hat{Y}_i = 14.783 + 2.217 * 15 = 48.038$

When ANI = \$15,000,

ACDE $= 48.038 * 100 =$ \$4,803.80

$\hat{Y}_i = 14.783 + 2.217 * 50 = 125.633$

When ANI = \$50,000,

ACD $= 125.633 * 100 =$ \$12,563.30

Extrapolation?

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * \$100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10

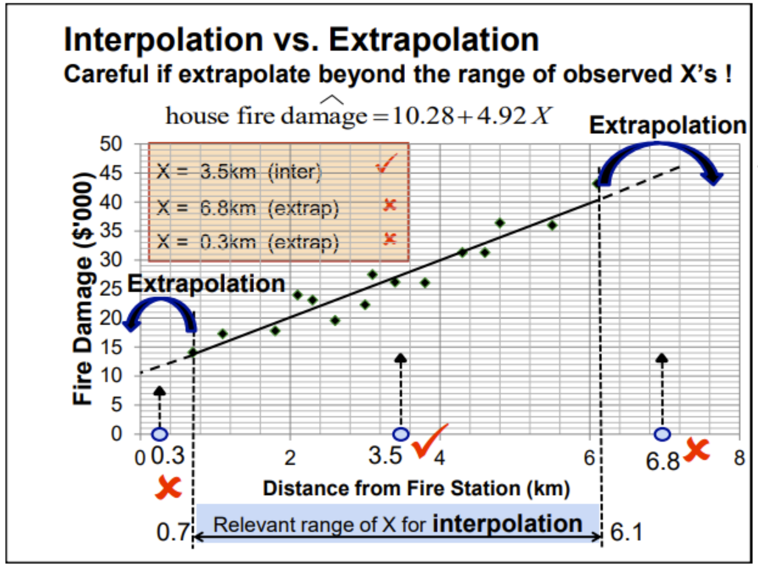
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$

Interpret the slope coefficient of ANI and the constant value.

Calculate and interpret the coefficient of determination. $r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

Assuming no assumptions have been violated, predict ACDE when ANI is \$15 000 and \$50 000.

What reservations might you have about these predictions?



adm T11 Q1 Q2 Q3

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * \$100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.

r² = 0.8315, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is \$15 000 and \$50 000.
What reservations might you have about these predictions?

$\hat{Y}_i = 14.783 + 2.217 * X_i$

$\hat{Y}_i = 14.783 + 2.217 * 15 = 48.038$

When ANI = \$15,000,
ACDE $= 48.038 * 100 = \$4{,}803.80$

$\hat{Y}_i = 14.783 + 2.217 * 50 = 125.633$

When ANI = \$50,000,
ACD $= 125.633 * 100 = \$12{,}563.30$

Predictions for ANI = \$50,000 is an extrapolation which assumes the same linear relationship holds beyond slope of sample. This may not be true.

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination.
$r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is $15 000 and $50 000.
What reservations might you have about these predictions?
When ANI = $15,000, ACDE $= 48.038 * 100 = \$4{,}803.80$
When ANI = $50,000, ACD $= 125.633 * 100 = \$12{,}563.30$

e) Calculate the 95% confidence interval for β_1.

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).

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in \$100's, and Annual Net Income (ANI) in \$1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from \$8 000 to \$40 000.

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

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by \$1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * \$100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * \$100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination. $r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is \$15 000 and \$50 000.
What reservations might you have about these predictions?

When ANI = \$15,000, ACDE $= 48.038 * 100 = \$4{,}803.80$
When ANI = \$50,000, ACD $= 125.633 * 100 = \$12{,}563.30$

e) Calculate the 95% confidence interval for β_1.

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).

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

$$\beta_1 = b_1 \pm t_{\alpha/2,\, df} * s_{b_1}$$
$$\beta_1 = 2.217 \pm t_{0.025,\, 10-2} * 0.353$$

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination. $r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is $15 000 and $50 000.
What reservations might you have about these predictions?

When ANI = $15,000, ACDE $= 48.038 * 100 = \$4{,}803.80$
When ANI = $50,000, ACD $= 125.633 * 100 = \$12{,}563.30$

e) Calculate the 95% confidence interval for β_1.

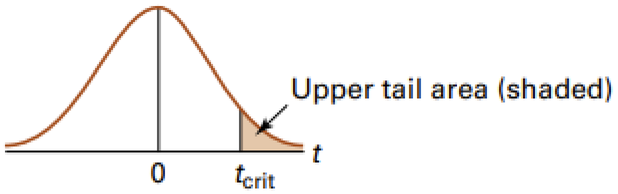
Confidence interval for β_1
(the slope coefficient for the population)

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).

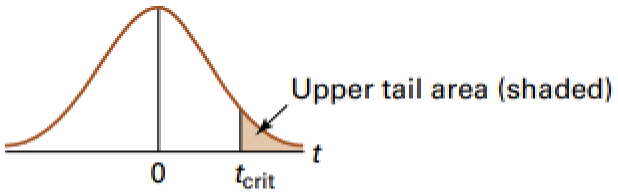
adm T11 Q1 Q2 Q3



$t_{0.025,\,8}$

	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



$t_{0.025,\,8}$

	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

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination. $r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is $15 000 and $50 000.
What reservations might you have about these predictions?
When ANI = $15,000, ACDE $= 48.038 * 100 = \$4{,}803.80$
When ANI = $50,000, ACD $= 125.633 * 100 = \$12{,}563.30$

e) Calculate the 95% confidence interval for β_1.

$$\beta_1 = b_1 \pm t_{\alpha/2,\, df} * s_{b_1}$$
$$\beta_1 = 2.217 \pm 2.306 * 0.353$$
$$1.403 < \beta_1 < 3.031$$

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).

The University Of Queensland Australia

Q1. It is commonly believed by economists that there is a close relationship between Annual Consumer Durables Expenditure (ACDE) in $100's, and Annual Net Income (ANI) in $1000's.

a) Which of ACDE or ANI would be the dependent variable in a regression analysis? ACDE

The following output was obtained with ANI ranging from $8 000 to $40 000.

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	3618.783	3618.783	39.484
Residual	8	733.217	91.652	
Total	9	4352		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	14.783	8.994	1.644	0.139
X	2.217	0.353	6.284	0.000

When ANI (X_i) increases by $1000, your expected ACDE ($\hat{Y}_i$) increases by 2.217 * $100.

When ANI (X_i) is zero, ACDE ($\hat{Y}_i$) is expected to be 14.783 * $100. But it involves extrapolation, so it may not be accurate.

b) How large a sample was used? n = 10
State the estimated simple linear regression equation. $\hat{Y}_i = 14.783 + 2.217 * X_i$
Interpret the slope coefficient of ANI and the constant value.

c) Calculate and interpret the coefficient of determination. $r^2 = 0.8315$, 83.15% of the variability in ACDE (Y) can be explained by the regression equation with ANI (X).

d) Assuming no assumptions have been violated, predict ACDE when ANI is $15 000 and $50 000.
What reservations might you have about these predictions?

When ANI = $15,000, ACDE $= 48.038 * 100 = \$4{,}803.80$
When ANI = $50,000, ACD $= 125.633 * 100 = \$12{,}563.30$

e) Calculate the 95% confidence interval for β_1.

$$\beta_1 = b_1 \pm t_{\alpha/2,\,df} * s_{b_1}$$
$$\beta_1 = 2.217 \pm 2.306 * 0.353$$
$$1.403 < \beta_1 < 3.031$$

The slope of the regression line between ANI and ACDE is estimated, with 95% confidence, to be between 1.403 and 3.031.

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).

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$*. Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation.
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

(Poll)

a) State the units for the variables and sample statistics.
b) State the estimated regression equation.
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

1. What unit would you give to Y (electricity consumption)? (Single Choice) *

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

2. What unit would you give to X (outdoor temperature)? (Single Choice) *

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

3. What unit would you give to b0 (Intercept)? (Single Choice) *

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

4. What unit would you give to b1 (Slope coefficient)? (Single Choice) *

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\,\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

(Poll)

a) State the units for the variables and sample statistics.
b) State the estimated regression equation.
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

1. What unit would you give to Y (electricity consumption)? (Single Choice) *

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

2. What unit would you give to X (outdoor temperature)? (Single Choice) *

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

3. What unit would you give to b0 (Intercept)? (Single Choice) *

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

4. What unit would you give to b1 (Slope coefficient)?

- Kw
- °F
- Kw/°F
- °C
- Kw/°C

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

kw kw $\frac{kw}{°F}$ °F

36

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$*. Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation.
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

$$\hat{Y}_i = 169.45 - 1.86 * X_i$$

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation.
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

The consumption of electricity (Y) falls by 1.86 Kw for each increase in outdoor temperature (X) of 1 °F.

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$\hat{Y}_i = 169.45 - 1.86 * X_i$

$X_1 = 50°\text{F}$?

(Answers in chat)

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

The University of Queensland
AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$$\hat{Y}_i = 169.45 - 1.86 * X_i$$

$X_1 = 50°F$?

$$\hat{Y}_1 = 169.45 - 1.86 * 50$$

$$\hat{Y}_1 = 76.45\ Kw$$

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\,\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$$\hat{Y}_i = 169.45 - 1.86 * X_i$$

$X_1 = 50°\text{F}$?

$$\hat{Y}_1 = 169.45 - 1.86 * 50$$

$$\hat{Y}_1 = 76.45\ Kw$$

$X_1 = 90°\text{F}$?

(Answers in chat)

Example 1.
Simple Linear Regression

b. Using the output from Excel, write the estimated linear regression equation.

$$\hat{Y}_i = b_0 + b_1 X_i$$

36

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

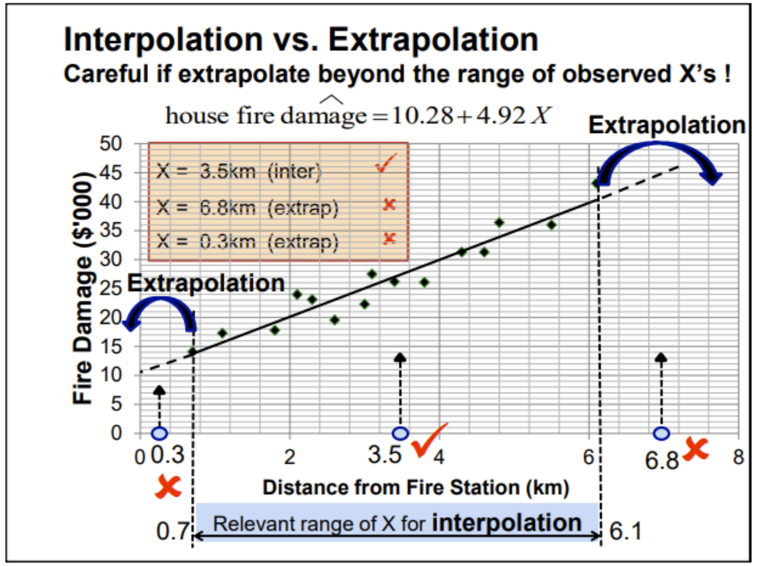
	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$$\hat{Y}_i = 169.45 - 1.86 * X_i$$

$$X_1 = 50°F\ ?$$
$$\hat{Y}_1 = 169.45 - 1.86 * 50$$
$$\hat{Y}_1 = 76.45\ Kw$$

$$X_2 = 90°F\ ?$$
$$\hat{Y}_2 = 169.45 - 1.86 * 90$$
$$\hat{Y}_2 = 2.05\ Kw$$



Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment.
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$$\hat{Y}_i = 169.45 - 1.86 * X_i$$

$X_1 = 50°F$? Within 25 - 78°F, so OK

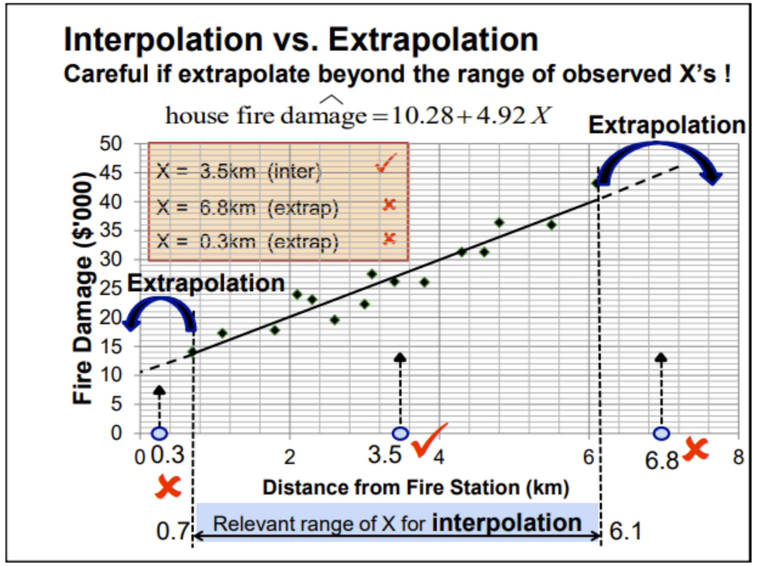
$$\hat{Y}_1 = 169.45 - 1.86 * 50$$
$$\hat{Y}_1 = 76.45\ Kw$$

$X_2 = 90°F$? Not within 25 - 78°F, so this is an extrapolation that may give a biased prediction.

$$\hat{Y}_2 = 169.45 - 1.86 * 90$$
$$\hat{Y}_2 = 2.05\ Kw$$

The same linear relationship will not hold past 78°F because when temperature increases, more electricity usage is expected due to fans and air conditioning being used.

Problem with extrapolation.



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

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

(Answers in chat)

Coefficient of Determination, r^2

The **coefficient of determination** is the portion of the total variation in the dependent variable **that is explained** by variation in the independent variable.

$$r^2 = \frac{SSR}{SST} = \frac{\text{regression } sum \text{ of squares}}{total \text{ sum of squares}} = \frac{SSR}{SSR+SSE}$$

Note that **r^2** can only take values between **0 and 1.** It gives a measure of how useful is the SLR model.

13

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning.
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$$r^2 = \frac{SSR}{SST} = \frac{25094.12}{28071.34} = 0.8939$$

89.39% of the total variation in electricity consumption (Y) is explained by the relationship with temperature (X).

Coefficient of Determination, r^2

The **coefficient of determination** is the portion of the total variation in the dependent variable **that is explained** by variation in the independent variable.

$$r^2 = \frac{SSR}{SST} = \frac{\text{regression } \mathit{sum} \text{ of squares}}{\mathit{total} \text{ sum of squares}} = \frac{SSR}{SSR+SSE}$$

Note that r^2 can only take values between **0 and 1.** It gives a measure of how useful is the SLR model.

13

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$*. Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning. $r^2 = 0.8939$
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

r = ?

(Answers in chat)

Compare r^2 and r.

r^2 = Coefficient of Determination

r = Correlation Coefficient $r = \pm\sqrt{r^2}$

From Example 1, $r^2 = 0.923478$, so

$$r = \pm\sqrt{r^2}$$
$$= \sqrt{0.923478}$$
$$= 0.961$$

Since the slope coefficient b_1 is positive, need to have the **same sign** for the correlation coefficient, so r = +0.961

18

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

***Note:** the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning. $r^2 = 0.8939$
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$$r = \pm\sqrt{r^2} = -\sqrt{0.8939} = -0.9455$$

Because b_1 is negative

Compare r^2 and r.

r^2 = Coefficient of Determination

r = Correlation Coefficient $r = \pm\sqrt{r^2}$

From Example 1, r^2 = 0.923478, so

$$r = \pm\sqrt{r^2}$$
$$= \sqrt{0.923478}$$
$$= 0.961$$

Since the slope coefficient b_1 is positive, need to have the same sign for the correlation coefficient, so r = +0.961

18

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning. $r^2 = 0.8939$
f) Compute the correlation coefficient.
g) Determine the standard error of the estimate. What does this represent?

$$r = \pm\sqrt{r^2} = -\sqrt{0.8939} = -0.9455$$

There is a high negative correlation between electricity consumption and outdoor temperature.

Compare r^2 and r.

r^2 = Coefficient of Determination

r = Correlation Coefficient $r = \pm\sqrt{r^2}$

From Example 1, $r^2 = 0.923478$, so

$$r = \pm\sqrt{r^2} = \sqrt{0.923478} = 0.961$$

Since the slope coefficient b_1 is positive, need to have the **same sign** for the correlation coefficient, so r = +0.961

18

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning. $r^2 = 0.8939$
f) Compute the correlation coefficient. $r = -0.9455$
g) Determine the standard error of the estimate. What does this represent?

$$s_e = ?$$

(Answers in chat)

Standard Error of Estimate, s_e

The standard deviation of all the (observed) points around the estimated regression line = **standard deviation of the error of the model**

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

where

SSE = sum of squares of error
MSE = mean of squares error
n = sample size

19

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning. $r^2 = 0.8939$
f) Compute the correlation coefficient. $r = -0.9455$
g) Determine the standard error of the estimate. What does this represent?

$$S_e = \sqrt{\frac{SSE}{df}} = \sqrt{\frac{SSE}{n-2}} =$$

$$\sqrt{\frac{2977.22}{22}} = 11.63$$

Standard Error of Estimate, s_e

The standard deviation of all the (observed) points around the estimated regression line = **standard deviation of the error of the model**

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

where

SSE = sum of squares of error
MSE = mean of squares error
n = sample size

19

The University Of Queensland
AUSTRALIA

Q2. (Based on Levine, Stephan, Krehbiel, and Berenson, p537, Q12.31).

A linear regression model has been proposed for predicting electricity consumption (in Kilowatts, Kw) based on outdoor temperature (in Fahrenheit, °F) in northern USA. The monthly consumption data, collected to develop the model, ranged from 39 to 138 Kw and the outdoor temperature data ranged from 25 to 78°F.

Note: *the formula to convert from Fahrenheit degrees to Centigrade is* $C\frac{5}{9}(F-32)$. *Calculate the range of outdoor temperature in Centigrade degrees for better understanding.*

Regression analysis produced the following output:

ANOVA

	df	*SS*	*MS*
Regression	1	25094.12	25094.12
Residual	22	2977.22	135.33
Total	23	28071.34	

	Coefficients	*Standard Error*	*t Stat*
Intercept	169.45	7.17	23.63
Temperature	-1.86	0.14	-13.28

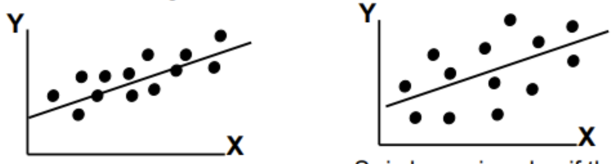
a) State the units for the variables and sample statistics.
b) State the estimated regression equation. $\hat{Y}_i = 169.45 - 1.86 * X_i$
c) Interpret the meaning of the slope coefficient.
d) Predict electricity consumption when temperature is 50 °F and 90°F. Comment. $\hat{Y}_1 = 76.45\ Kw,\ \hat{Y}_2 = 2.05\ Kw$
e) Compute the coefficient of determination AND interpret its meaning. $r^2 = 0.8939$
f) Compute the correlation coefficient. $r = -0.9455$
g) Determine the standard error of the estimate. What does this represent? $s_e = 11.63$

$$s_e = \sqrt{\frac{SSE}{df}} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{2977.22}{22}} = 11.63$$

This is the standard deviation of all the 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.

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

(Poll)

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

2. **What symbol would you give to level of profits?** (Single Choice) *
   - Y
   - b0 (b zero)
   - b1 (b one)
   - X
   - SSE
   - SSX
   - n

3. **What symbol would you give to the value -16.3?** (Single Choice) *
   - Y
   - b0 (b zero)
   - b1 (b one)
   - X
   - SSE
   - SSX
   - n

4. **What symbol would you give to the value 32.7?** (Single Choice) *
   - Y
   - b0 (b zero)
   - b1 (b one)
   - X
   - SSE
   - SSX
   - n

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43   Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

(Poll)

1. What symbol would you give to value of loan portifolio? (Single Choice) *

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

2. What symbol would you give to level of profits? (Single Choice) *

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

3. What symbol would you give to the value -16.3? (Single Choice) *

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

4. What symbol would you give to the value 32.7? (Single Choice) *

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

The University Of Queensland Australia

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
$\alpha = 1\% = 0.01$

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

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

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
α = 1% = 0.01
n = 4*6 = 24

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

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

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
$\alpha = 1\% = 0.01$
$n = 4*6 = 24$

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

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

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

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
α = 1% = 0.01
n = 4*6 = 24

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

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

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

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

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
$\alpha = 1\% = 0.01$
$n = 4*6 = 24$

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43   Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

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

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

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

(Answers in chat)

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

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
α = 1% = 0.01
n = 4*6 = 24

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

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

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

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{937.43}{22}} = 6.527668$$

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

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
$\alpha = 1\% = 0.01$
$n = 4*6 = 24$

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

Y = -16.3 + 32.7X

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

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

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}} = \frac{6.5277}{\sqrt{287.21}} = 0.385175$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{937.43}{22}} = 6.527668$$

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

$b_0 = -16.3$
$b_1 = 32.7$
$SSE = 937.43$
$SSX = 287.21$
α = 1% = 0.01
n = 4*6 = 24

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

Y = -16.3 + 32.7X

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

$$\beta_1 = b_1 \pm t_{\alpha/2,\, n-2} * s_{b_1} = 32.7 \pm t_{0.005,\, 22} * 0.385175 = ?$$

$$s_{b_1} = \frac{s_e}{\sqrt{SS_{XX}}} = \frac{6.5277}{\sqrt{287.21}} = 0.385175$$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{937.43}{22}} = 6.527668$$

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

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

$$SS_{YY} = SST$$

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

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

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

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

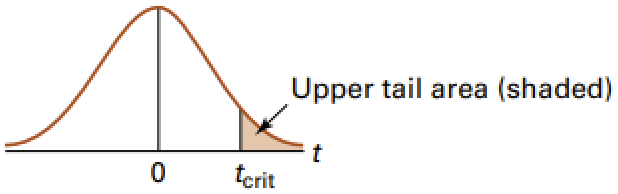
$$s_e = \sqrt{\frac{SSE}{n\text{-}2}} = \sqrt{MSE}$$

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

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

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

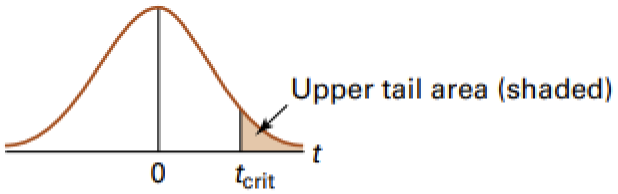
29



$t_{0.005,\,22}$

	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

adm T11 Q1 Q2 Q3



$t_{0.005, 22}$

	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

adm T11 Q1 Q2 Q3

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

$b_0 = -16.3$

$b_1 = 32.7$

$SSE = 937.43$

$SSX = 287.21$

α = 1% = 0.01

n = 4*6 = 24

Q3. A large corporation was interested in determining the relationship between the value of its loan portfolio (billions of dollars), and its level of profits (millions of dollars). Using quarterly data from its previous six years, it estimated the following:

$$Y = -16.3 + 32.7X$$

Error sum of squares: 937.43 Sum of squares of X: 287.21

Determine the 99% confidence interval for the slope coefficient.

$\beta_1 = b_1 \pm t_{\alpha/2,\, n-2} * s_{b_1}$ = 32.7 $\pm$ $t_{0.05,\,22}$ * 0.385175

= 32.7 $\pm$ 2.819 * 0.385175

$$31.614 < \beta_1 < 33.785$$

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

ECON1310

Tutorial 11 – Week 12

SIMPLE LINEAR REGRESSION I

At the end of this tutorial you should be able to

- Formulate a SLR model and interpret the coefficients.
- Estimate a SLR equation using Excel.
- Interpret the coefficient of determination and standard error of the regression, given Excel output.
- Construct a confidence interval for 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.