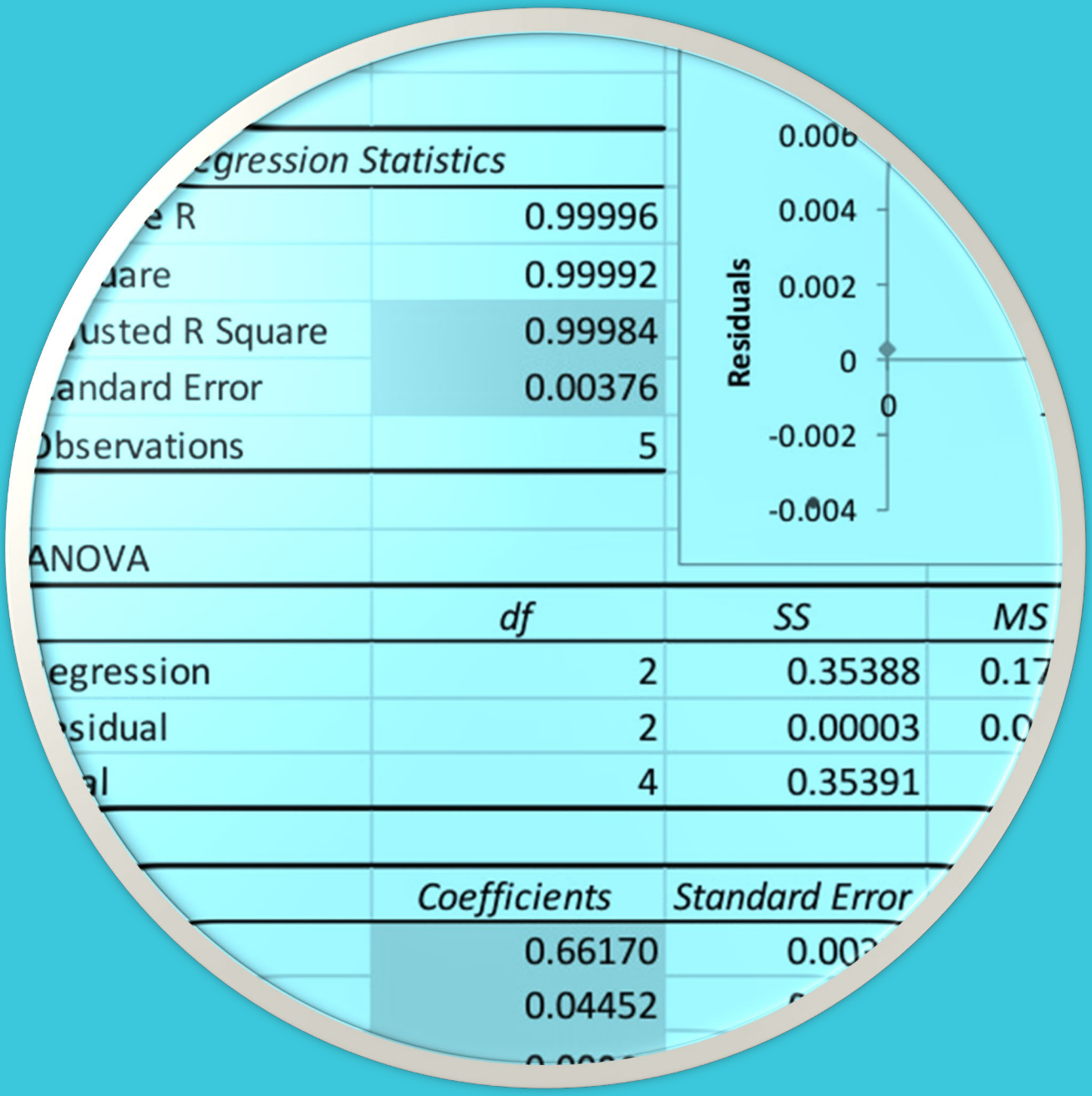
Researchers in Schools



egression Statistics	
e R	0.99996
uare	0.99992
usted R Square	0.99984
andard Error	0.00376
bservations	5

ANOVA

	df	SS	MS
egression	2	0.35388	0.17
sidual	2	0.00003	0.0
al	4	0.35391	

	Coefficients	Standard Error
	0.66170	0.00
	0.04452	

Hey Maths, Science needs your help!

Key Stage 4 Programme

Pupil Name

Coursebook Designed by Dr Josephine Lawal

Contents

Timetable and Assignment Submission 1
Timetable – Tutorials 1
Timetable – Homework Assignments 1
Assignment Submission – Lateness and Plagiarism 1
Course Rationale 2
Uni Pathways Mark Scheme 2020 – maths-based assignments 3
Baseline Assignment (problem-set): Pupil Feedback Report 5
Final Assignment (problem-set): Pupil Feedback Report 6
Subject Vocabulary 7
Launch Event 8
Tutorial 1 – All in the Name of Data 9
What is the Purpose of Tutorial 1? 9
Tutorial 2 – What are the Scores 16
What is the Purpose of Tutorial 2? 16
Tutorial 3 – Visualising Numbers 19
What is the Purpose of Tutorial 3? 19
Tutorial 4 – Any Sense in this Data? 22
What is the Purpose of Tutorial 4? 22
Tutorial 5 – How Do the Numbers Add Up? 25
What is the Purpose of Tutorial 5? 25
Tutorial 6 – Feedback tutorial 37
What is the Purpose of Tutorial 6? 37
Final assignment feedback from your Uni Pathways Teacher 37
Tutorial 7 – Final tutorial 39
Appendix 1 – Referencing correctly 41
How do I reference? 42
Appendix 2 – Resource for Baseline and Final Assignments 43
Bibliography 48
Notes 49

Timetable and Assignment Submission

Timetable – Tutorials

Tutorial	Date	Time	Location
1			
2			
3			
4			
5			
6 (Feedback)			
7 (Feedback)			

Timetable – Homework Assignments

Homework Assignment	Description	Due Date
Tutorial 1	Baseline assignment	
Tutorial 2		
Tutorial 3		
Tutorial 4		
Tutorial 5	Final assignment	

Assignment Submission – Lateness and Plagiarism

Lateness	
Submission after midnight on 9th August	10 marks deducted
Plagiarism	
Some plagiarism	10 marks deducted
Moderate plagiarism	20 marks deducted
Extreme plagiarism	Automatic fail

Course Rationale

My PhD was a science-based research aimed at designing and producing multi-functional metallic coatings by a process known as magnetron sputtering. This method of coating production is an environmentally friendly alternative to traditional methods of producing wear- and corrosion-resistant coatings.

The main objectives of the research were:

a) Design experiments to produce aluminium-rich metallic nanostructured/amorphous coatings by reactive magnetron sputtering.

b) Produce coatings.

c) Investigate structural, mechanical, tribological and electrochemical properties of the coatings produced.

d) Establish relationships between the properties of the coatings and the parameters used e.g. nitrogen content.

Few of the coating characteristics aimed for and achieved are:

- Thickness in excess of 10 microns
- High hardness up to 10GPa
- Internal stress <4GPa
- High thermal stability
- Sacrificial corrosion properties
- Excellent abrasive wear resistance

Even though my PhD was science-based, none of my results would have made any meaning without the use mathematical methods to make inferences; the use of statistical tools played a huge part in interpreting my results.

One of the main reasons why we study maths is because "it is critical to science, technology and engineering". Therefore, the connection between maths and science, especially with experimental data interpretation, is the focus of this course.

In this course, you will experience an appreciation for data and the difference between raw and derived data (scores). You will have an opportunity to compute some derived scores based on real experimental data, as well as be able to visualise raw data by graphical presentation. This will give a hint into the correlation between the dependent and independent variables. However, a numerical correlation will be computed using the derived scores. This is what is known simply as correlation analysis. The Regression analysis will follow by generating the regression model. This model is an equation that can be used to make academic inference.

The process of correlation and regression analyses is how scientific data can be presented to make sense and be used to make useful decisions. This is the essence of this course.

Uni Pathways Mark Scheme 2020 – maths-based assignments

	Subject Knowledge	Critical Thinking	Written Communication
1st	The work shows a depth of knowledge and understanding of key concepts and mathematical methods, drawing on sources of evidence where appropriate. Knowledge is used to build and support highly effective mathematical arguments and explanations.	Analyses the available data/evidence and identifies all the relevant mathematical information. Identifies and critically evaluates available mathematical approaches and evidence, deciding on their credibility, strength and relative significance. Makes connections between different pieces of information and puts them together to form accurate solutions and verify answers and assumptions.	Mathematical arguments are complete, and the reasoning behind each step has been clearly communicated to the reader. The work has a coherent flow and is well structured. The writing style is appropriate; mathematical language and key mathematical terms are used accurately and effectively to support the arguments and explanations made. There are no, or very few, errors in spelling or grammar. Consistent referencing, appropriate paragraphing and use of correctly labelled tables and figures; matching the style taught in the course.
2:1	The work shows an understanding of key concepts and mathematical methods, drawing on sources of evidence where appropriate. Knowledge is used to build and support effective mathematical arguments and explanations.	Analyses the available data/evidence and identifies all the relevant mathematical information. Engages with multiple approaches and critically evaluates their utility, forming judgements and using reasoning and evidence to inform problem solving strategy. Shows some understanding of the relative value of evidence and data.	Mathematical arguments are clearly made, with few gaps in logical reasoning. The work is well-structured. The writing style is appropriate; mathematical language and key terms are used correctly. There are few errors in spelling or grammar. Mostly consistent referencing and use of tables and figures; matching the style taught in the course.
2:2	The work shows an understanding of key concepts and mathematical methods, with no major misconceptions. Beginning to apply this knowledge to build and support effective mathematical arguments and explanations.	Identifies and uses basic mathematical evidence and reasoning. Evaluates the quality of mathematical reasoning and evidence, forming judgements. Not yet showing understanding of the relative value of evidence and reasoning.	Mathematical arguments have been constructed, but there is little explanation from step to step. The work has some structure. The writing style can sometimes be informal; occasionally key terms are not used when it would be appropriate to do so.

			There are some errors in grammar and spelling, but these do not get in the way of communicating the content. Referencing has some consistency; matching the style taught in the course. Limited use of tables and figures.
3rd	Shows a developing understanding of key concepts and mathematical methods, with some misconceptions. Not yet applying this knowledge to build and support mathematical arguments and explanations.	Beginning to analysis mathematical evidence and reasoning. Describes or attempts to solve the problem without any/or little evaluation of whether the approach was reasonable or accurate.	Mathematical arguments to solve problems are poorly sequenced, with little or no explanation. The grammar, spelling, style, and structure of the work need improving in order to communicate ideas to the reader. Key terms are not always used correctly. Limited, or no use of referencing, tables and figures.

Baseline Assignment (problem-set): Pupil Feedback Report

Name of Pupil	
Name of School	
Name of RIS teacher	
Title of Assignment	

How your assignment is graded:

Grade	Marks	What this means
1st	70+	Performing to an excellent standard at A-level
2:1	60-69	Performing to a good standard at A-level
2:2	50-59	Performing to an excellent standard at GCSE
3rd	40-49	Performing to a good standard at GCSE
Working towards a pass	0-39	Performing below a good standard at GCSE
Did not submit	DNS	No assignment received by The Brilliant Club

Lateness	
Any lateness	10 marks deducted
Plagiarism	
Some plagiarism	10 marks deducted
Moderate plagiarism	20 marks deducted
Extreme plagiarism	Automatic fail

Marks

OVERALL MARK / 100		FINAL MARK / 100 (including any deductions)	
DEDUCTED MARKS		FINAL GRADE	

If marks have been deducted (e.g. late submission, plagiarism) the teacher should give an explanation in this section:

Mark Breakdown and Feedback

Subject knowledge	
	mark
Critical thinking	
	mark
Written communication	
	mark

Problem Set	
Comment	*mark*

Final Assignment (problem-set): Pupil Feedback Report

Name of Pupil	
Name of School	
Name of RIS teacher	
Title of Assignment	

How your assignment is graded:

Grade	Marks	What this means
1st	70+	Performing to an excellent standard at A-level
2:1	60-69	Performing to a good standard at A-level
2:2	50-59	Performing to an excellent standard at GCSE
3rd	40-49	Performing to a good standard at GCSE
Working towards a pass	0-39	Performing below a good standard at GCSE
Did not submit	DNS	No assignment received by The Brilliant Club

Lateness	
Any lateness	10 marks deducted
Plagiarism	
Some plagiarism	10 marks deducted
Moderate plagiarism	20 marks deducted
Extreme plagiarism	Automatic fail

Marks

OVERALL MARK / 100		FINAL MARK / 100 (including any deductions)	
DEDUCTED MARKS		FINAL GRADE	

If marks have been deducted (e.g. late submission, plagiarism) the teacher should give an explanation in this section:

Mark Breakdown and Feedback

Subject knowledge	
	mark
Critical thinking	
	mark
Written communication	
	mark
Problem Set	
Comment	*mark*

Subject Vocabulary

Word	Definition	In a sentence
Compute	To calculate or work out something	Compute the difference between the -3.025 and 3.025
Data	A series of observations, measurements, or facts that can be used to gain knowledge or make decision	The data from the market survey were unreliable
variables	A quantity that can have any of set of values	The variables in these experiments are nitrogen flow rate and hardness
Dependent Variables	Anything that is observed in an experiment or study whose value is determined by the value of other things within the experiment or study	Dependent variables are plotted on the y-axis
Independent Variables	Anything that is manipulated in an experiment or study whose value determines the value of observed quantity	Independent variables are plotted on the x-axis
Raw Data	Any unprocessed data i.e. data that has not altered in any way	Age of school children is an example of raw data
Derived Scores	Any data that has undergone mathematic transformation	The mean, median, mode and range of any data are examples of derived scores
Correlation	The observed relationship between two things	There is a strong correlation between energy bills and room temperature
Correlation Coefficient	The value that describes the strength and direction of the relationship between two variables	The correlation coefficient can be a negative or positive value
Regression	The technique used for predicting the dependent variable from the independent variable with a consideration for error	The data analysist carried out regression analysis using the data from the two experiments
Regression Model	The equation that is formulated by which predictions can be made for the dependent variable using the independent variable	The was significant error in the regression model
Coefficient of Determination	The quantity that assesses the quality of the regression model. It describes the proportion of the variation in the dependent variable that can be predicted from the independent variable	Sometimes, coefficient of determination can be the same as the correlation coefficient

Launch Event

The launch event is your opportunity to learn more about this course and Uni Pathways. You will watch some videos, listen to your teacher and have some discussion. Below is some room to make notes for some of the videos you will watch during the event.

Virtual Campus Tours

Notes…

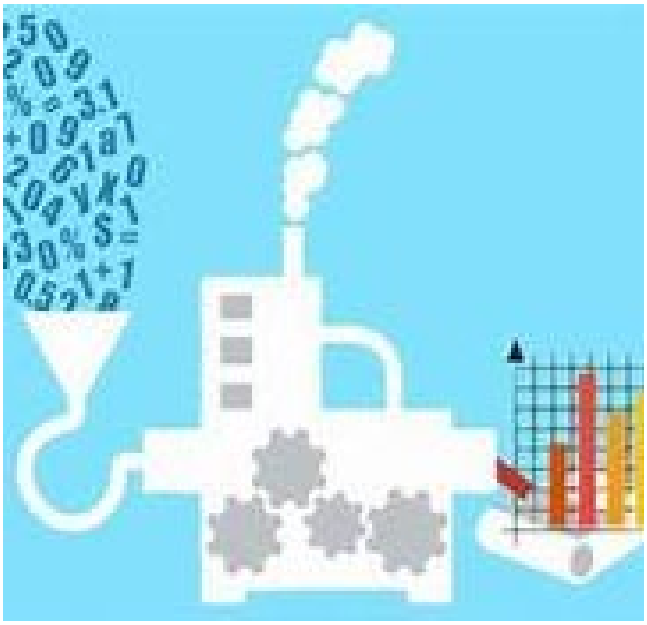
The Scholars Programme Alumni

Notes…

Student Ambassadors

Notes…

Tutorial 1 – All in the Name of Data



What is the Purpose of Tutorial 1?

- Objective 1. Pupils can explain the difference between raw data and derived scores using an example

- Objective 2. Pupils can categorise different notations used in basic correlation and regression analysis computations into raw data or derived scores

- Objective 3. Pupils can identify and define the summation notation (∑) called sigma

Definitions:

- **Data - A series of observations, measurements, or facts that can be used to gain knowledge or make decision**

- **Raw Data - Any unprocessed data i.e. data that has not been altered in any way. E.g. age of school children is an example of raw data**

10B2_000.	133.5179	8025.914	102.1934	608785.6	193.0768	192.4202	116.0448	13.18348	5.771841	73.47432	1.514531
10B2_001.	141.3345	8021.982	101.3261	668495.1	201.0607	200.712	109.8011	12.00006	7.270744	84.14649	1.472344
10B2_002.	132.1307	8023.89	98.48918	598467.2	193.4795	193.233	112.7985	13.4074	9.220971	77.22815	1.4239
10B2_003.	139.1242	8023.916	100.7303	651341.2	199.1488	198.8673	110.5836	12.31907	8.301361	83.54027	1.447787
10B2_004.	133.8134	8025.003	102.8792	610994.7	192.847	192.3165	116.6121	13.13433	7.592751	77.75438	1.468667
10B2_005.	141.0627	8022.985	99.2289	666374.2	202.1261	201.7027	107.6996	12.03976	9.618291	87.01877	1.418419
10B2_006.	134.9581	8022.874	99.88761	619586.9	195.6856	195.1973	112.4334	12.94875	9.499189	80.96537	1.422228
10B2_007.	134.6082	8020.003	103.3422	616954.9	193.1933	192.8129	116.5697	12.99933	9.149557	81.26998	1.437293
10B2_008.	147.9753	8023.969	104.0697	721314.5	206.0396	205.8017	108.5668	11.12409	7.432452	92.04958	1.475349
10B2_009.	135.3739	8022.954	98.29835	622722.2	196.7206	196.5877	110.3656	12.88368	10.99094	82.90456	1.392861
10B2_010.	137.122	8020.9	100.6199	635985.9	197.1915	196.9081	111.7879	12.61176	8.534741	81.90226	1.442716
10B2_011.	146.3341	8021.911	99.92605	708082	207.2769	206.543	105.2136	11.32907	7.586758	89.22249	1.461419
10B2_012.	142.9368	8023.817	101.678	681063.5	202.544	202.1223	109.1611	11.78131	7.222921	85.77473	1.474359
10B2_013.	133.3477	8024.818	100.2795	607515.5	193.8536	193.3661	113.9904	13.20924	8.981092	78.69461	1.432954
10B2_014.	135.2603	8022.864	98.35753	621864.7	196.8408	196.4365	110.5082	12.9013	10.0991	81.60816	1.407756
10B2_015.	134.2534	8024.87	101.6461	614290.7	194.1521	193.4653	114.9049	13.06364	8.490805	79.29461	1.446129

- **Derived Scores - Any data that has undergone mathematical transformation. E.g. mean, range of any data**

Activity 1:

In a few words, explain the main difference between raw data and derived scores? Give one example for each.

Introduction to basic Correlation and Regression Analysis

Mathematical Notations

Raw Data	Derived Score
X – independent variable	$\sum X$
Y – dependent variable	$\sum Y$
	$\bar{X} = \sum X/n$
	$\bar{Y} = \sum Y/n$

Activity 2:

True or false questions

Are the following items derived scores or not?

Item	Derived Scores
Shoe size	
Height	
X, Y	
Mode	
Median	
Test score	
$\bar{X}$	
$\sum x^2$	

Activity 3:

Here are some numbers assigned to $\mathbf{X}$: 5.23, 3.08, 4.5, 7.88, 6 Calculate $\sum \mathbf{X}$ and $\overline{\mathbf{X}}$

Baseline assignment questions:

Write a short description of basic correlation and regression analysis and explain their importance in understanding scientific data

Details of the assignment:

Essay: write a 150 to 200 words essay on the question above.

Problem set:

1) use your calculator to work out

$$\frac{4.94 \times \sqrt{99.78}}{2.03^3}$$

a) Write out all the figures displayed on your calculator.

b) Covert your answer to 3 significant figures

2)

a) $x = y + 4$

b) $x = \frac{y+4}{5}$

c) $x = \frac{y+4}{y+5}$

ii. Rearrange each of the equations and make y the subject

iii. For each of the new equation, find the value of y when $x = 2.056$ (use your calculator).

Success criteria – how to structure your assignment and what to include

Your essay (use resources from tutorial 1 and the introduction chapter in your workbook) must include:

- **An introduction**
- **What correlation analysis is**
- **What regression analysis is**
- **Examples of mathematical notations used in correlation and regression analysis**
- **Conclude by explaining the importance of correlation and regression in understanding scientific data**

For your essay:

****Use the notes in appendix 2****

For the problem set:

****Show all your working out clearly****

In this assignment, show me that you **understand**:	Show me that you **are able to**:
How to structure an essay The importance of correlation and regression analysis as discussed in tutorial 1	Remember some of the mathematical notations you learnt in tutorial 1 Use the calculator, competently Rearrange equations and substitute to find the value the unknown in an equation

---

Tutorial 2 – What are the Scores

What is the Purpose of Tutorial 2?

- Objective 1: Pupils can compute simple summations (i.e. $\sum X$ and $\sum Y$) using raw data from actual scientific experiment.
- Objective 2: Pupils can calculate the mean (i.e., $\bar{X}$, $\bar{Y}$) of the raw data.
- Objective 3: Pupils can compute complex summations (i.e. $\sum x^2$, $\sum y^2$ and $\sum xy$) using raw data from actual scientific experiment.

Do Now:

Use your calculator to work out the following. Give your answer to 2 decimal places

$$\frac{2.579 + 3.8 \times 4.3^2}{\sqrt{3.45 - 1.5}}$$

More Mathematical Notations

Raw Data	Derived Scores
X – independent variable	$\sum X$
Y – dependent variable	$\sum Y$
n – number of data items	$\bar{X} = \sum X/n$
	$\bar{Y} = \sum Y/n$
	$x = X - \bar{X}$
	$y = Y - \bar{Y}$
	x^2
	y^2
	xy
	$\sum x^2$
	$\sum y^2$
	$\sum (xy)$

Activity 1 and 2: Computing derived scores – fill in the table when prompted

Table 1: for Activity 1 and 2 – answers to 1 decimal place

X	Y	x	y	x^2	y^2	xy
8.5	**4.2**					
12.6	**2.0**					
14.5	**1.8**					
15.3	**1.3**					
15.2	**1.1**					
14.7	**0.9**					

$\sum X$		$\sum Y$				$\sum x^2$		$\sum y^2$		$\sum(xy)$	
$\bar{X}$		$\bar{Y}$									

Tutorial 3 – Visualising Numbers

What is the Purpose of Tutorial 3?

- Objective 1: Pupils can draw a scatter graph
- Objective 2: Pupils can draw the line of best fit and compute the slope of the line of best fit using the derived scores.
- Objective 3: Pupils can describe the correlation between dependent and independent variables.

Do Now:

Use your calculator to work out the following. Give your answer to 2 significant figures

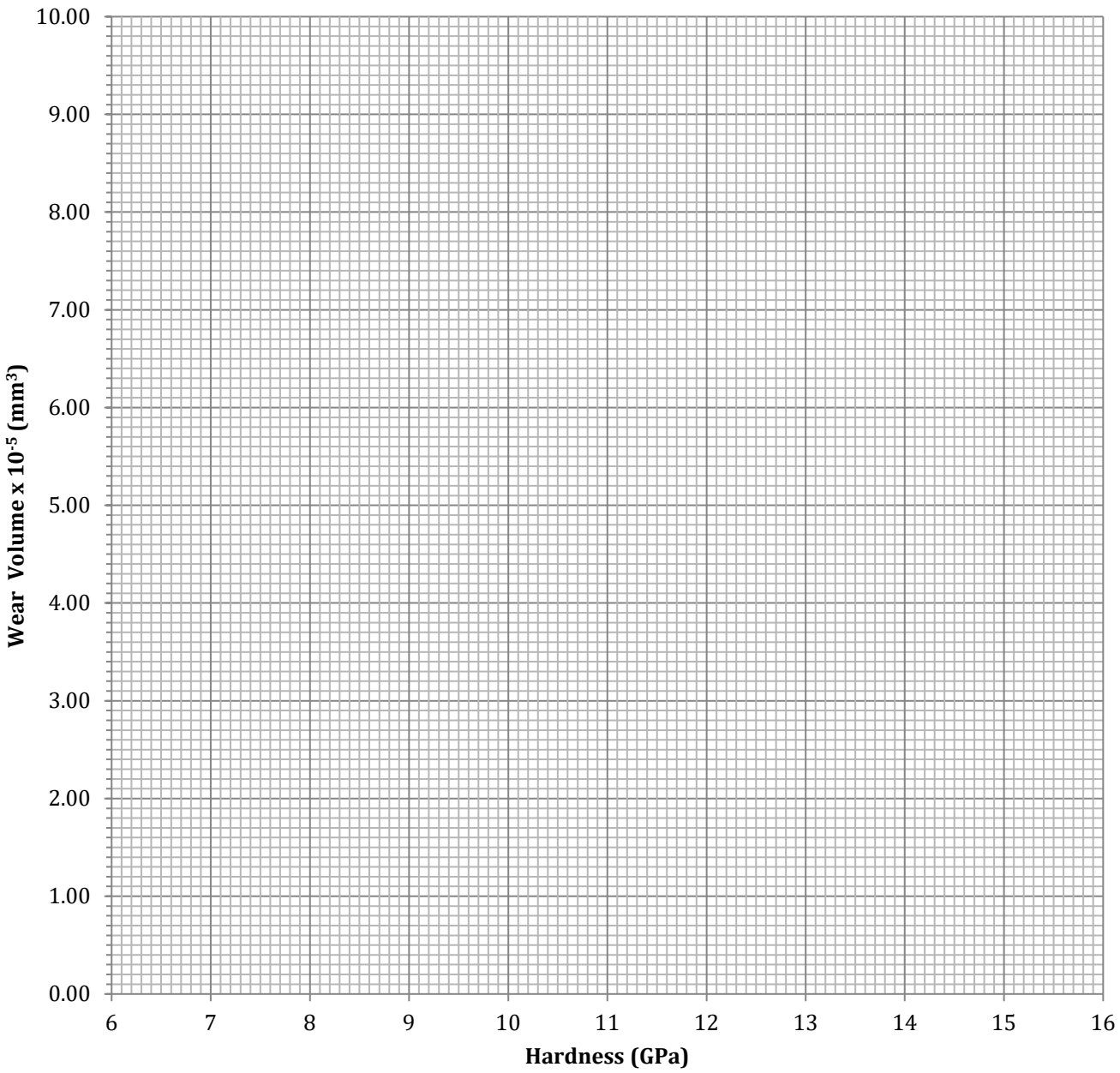
$$\frac{2.579^3 + 0.8 \times 4.3^2}{\sqrt{3.45 - 1.5}}$$

Activity 1: Practicing plotting a scatter graph

Using the table and graph template below, plot a scatter graph to present the data visually.

Table 2:

Coating	Hardness (GPa)	Wear Volume $x10^{-5}$ (mm^3)
0L	8.5	4.2
10L	12.6	2.0
15L	14.5	1.8
10M	15.3	1.3
15M	15.2	1.1
15T	14.7	0.9



Drawing the Line of Best Fit

The line of best fit is the line (drawn on a scatter graph) which describes the direction of the data.

Activity 2: Drawing line of best fit

a) **On the scatter graph above, draw a line of best fit. Follow the example on the board**

b) **Compute the slope of the line of best fit**

Pick two suitable points on the line and record the x and y coordinates for the points.

e.g. (x_1, y_1) , (x_2, y_2) **substitute the coordinates in the equation below to calculate the slope.**

Slope = $\frac{y_2 - y_1}{x_2 - x_1}$

Activity 3: Describing correlation

Describe fully the correlation of the data in table 2

Tutorial 4 – Any Sense in this Data?

What is the Purpose of Tutorial 4?

- Objective 1: Pupils can calculate the correlation coefficient using derived scores.
- Objective 2: Pupils can generate equation of the line of best fit.

Do Now:

Use your calculator to work out the following. Give your answer to 2 decimal places

$$\frac{\sqrt{0.57} + 3.8 \times 14.3^2}{\sqrt{0.45 + 1.15}}$$

Activity 1: Calculate the correlation coefficient using the computed scores in table 1

Correlation coefficient (r_{xy}) can be calculated using:

$$\frac{\Sigma(xy)}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

The rule of thumb when using correlation coefficient in decision making is to determine if $|r_{xy}| > \frac{2}{\sqrt{n}}$ where n is the number of samples

Activity 2: Can the correlation coefficient calculated above be used in decision making? Give reason for your answer.

Activity 3: Generate the equation of the line of best fit for the data in Table 1

Where does the line of best fit intercept the y-axis?

The equation of the line of best fit can be generated using:

$$y = mx + c$$

Where y **is the dependent variable**
m **is the slope of the line**
x **is the independent variable**
c **is the y-intercept**

Tutorial 5 – How Do the Numbers Add Up?

What is the Purpose of Tutorial 5?

- Objective 1: Pupils can generate the regression model.
- Objective 2: Pupils can use the regression model to make prediction and academic inference within and outside the data set.

Do Now:

Use your calculator to work out the following. Give your answer to 3 significant figures

$$\frac{\sqrt{0.57} + 3.8 \times 14.3^2}{\sqrt[3]{0.389 + 0.568}}$$

The regression model is expressed as:

$$\boldsymbol{y_i = b_1 x_i + b_0}$$

Where: y_i **is the dependent variable for observation *i***
x_i **is the independent variable for observation *i***
ε **is the error term for observation *i***

b_1 **is slope of the regression line:**

$$\boldsymbol{b_1} = \frac{\sum[(x_i - \overline{X})(y_i - \overline{Y})]}{\sum[(x_i - \overline{X})^2]}$$

b_0 **is the y-intercept of the regression line and it is constant expressed as:**

$$\boldsymbol{b_0 = \overline{Y} - b_1 \overline{X}}$$

Activity 1: Generate the regression model

Activity 2: Use your model to predict the wear resistance of a coating with hardness of 13.5 GPa and 18 GPa.

Activity 3: Think/pair/share: How reliable are your predictions?

- Write down your thoughts
- Discuss with your partner
- Share your convictions

Final assignment question:

What is the role of statistical methods in presenting and interpreting results of scientific experiments?

Details of the assignment: This assignment consists of two part – the essay and the problem set.

Essay: write a 500 to 750 words essay on the question above.

Problem set: You will be provided with a set of data; you are required to:

1. **Decide on the best way to present this data.**
2. **Calculate the correlation coefficient.**
3. **Determine the regression model.**
4. **Use your regression model to make a reasonable prediction – in the confines of your data (table 3) for the hardness of a coating.**
5. **Provide a justification for your prediction**

Success criteria – how to structure your assignment and what to include:

In your essay, focus on scatter graphs, correlation and regression analysis

- **An introduction;**
- **What scatter graphs are;**
- **A brief explanation of the importance of scatter graphs in data presentation and analysis.**
- **A brief description of correlation and regression analyses.**
    - **Definitions or formulas of key notations such as – independent variables, dependent variables, correlation coefficient, error term, coefficient of determination, y-intercept, and regression model.**
- **The purpose/importance and limitations of correlation and regression analysis in data analysis and interpretation;**
- **A conclusion.**

****Use notes in appendix 2 and other source as you deem fit****

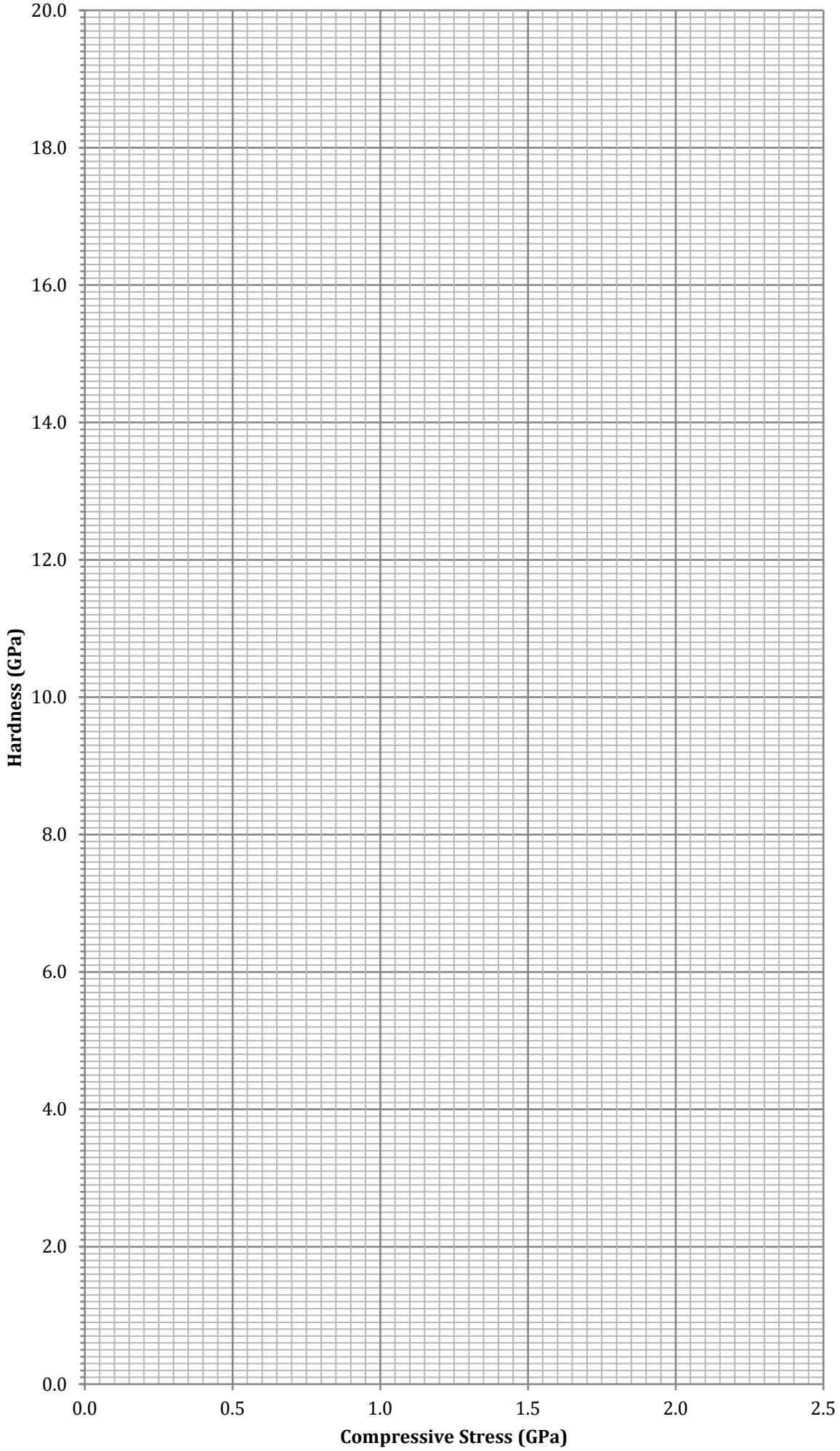
In this assignment, show me that you **understand**:	Show me that you **are able to**:
1. The key concepts such as raw data and derived score; correlation and regression analysis. 2. How to present scientific data. 3. How to interpret scientific data	1. Correctly use mathematical notations. 2. Apply your knowledge and understanding in making academic inferences and predictions.

Below is the data for the final assignment:

Table 3:

Coating	Compressive Stress (GPa)	Hardness (GPa)
0L	0.5	8.5
0M	0.6	10.1
5L	1.0	11.1
10L	1.3	12.6
15L	1.7	14.5
15M	1.8	15.2
20L	2.0	17.0

X		Y		x	y	x^2		y^2		xy	
$\sum X$		$\sum Y$				$\sum x^2$		$\sum y^2$		$\sum(xy)$	
$\bar{X}$		$\bar{Y}$									


20.0
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
Hardness (GPa)
0.0
0.5
1.0
1.5
2.0
2.5
Compressive Stress (GPa)

Tutorial 6 – Feedback tutorial

What is the Purpose of Tutorial 6?

- To receive feedback on your final assignment
- To respond to the feedback from your Uni Pathways teacher
- To write targets for improvement on your final assignment

Final assignment feedback from your Uni Pathways Teacher

(Remember to look at the mark scheme to help you understand what you have done well so far, and how you can do even better in your final assignment)

Here are three things that my Uni Pathways Teacher thought I did well in my draft assignment
•   •   •

Here are three things that my Uni Pathways Teacher thinks that I could do to get a higher mark in my final assignment
•   •   •

Tasks from my Uni Pathways Teacher to do during the feedback tutorial to help me improve
•   •   •
My response:

Actions I will take to improve my final assignment after this tutorial…
•   •   •

Hand in date for my final assignment:	

Tutorial 7 – Final tutorial

What is the Purpose of Tutorial 7?

- To receive feedback and a grade on your final assignment.
- To reflect on the programme including what you enjoyed and what was challenging.
- To ask any questions you may have about university.

Final assignment feedback from my Uni Pathways Teacher

Final mark:	University style grade:
Feedback: Here are three things that my Uni Pathways teacher thought I did well in my final assignment • • •	

Here are three things that my Uni Pathways teacher thinks I should remember for when I am doing this kind of study in the future
• • •

University

What questions do you still have about University after taking part in Uni Pathways?
• • •

Reflecting on Uni Pathways

What did you most enjoy about Uni Pathways?
• • •

What did you find challenging about the programme?	How did you overcome these challenges?
• • •	• • •

Appendix 1 – Referencing correctly

When you get to university, you will need to include references in the assignments that you write, so we would like you to start getting into the habit of referencing in your Brilliant Club assignment. This is really important, because it will help you to avoid plagiarism. Plagiarism is when you take someone else's work or ideas and pass them off as your own. Whether plagiarism is deliberate or accidental, the consequences can be severe. In order to avoid losing marks in your final assignment, or even failing, you must be careful to reference your sources correctly.

What is a reference?

A reference is just a note in your assignment which says if you have referred to or been influenced by another source such as book, website or article. For example, if you use the internet to research a particular subject, and you want to include a specific piece of information from this website, you will need to reference it.

Why should I reference?

Referencing is important in your work for the following reasons:

- It gives credit to the authors of any sources you have referred to or been influenced by.
- It supports the arguments you make in your assignments.
- It demonstrates the variety of sources you have used.
- It helps to prevent you losing marks, or failing, due to plagiarism.

When should you use a reference?

You should use a reference when you:

- Quote directly from another source.
- Summarise or rephrase another piece of work.
- Include a specific statistic or fact from a source.

How do I reference?

There are a number of different ways of referencing, and these often vary depending on what subject you are studying. The most important to thing is to be consistent. This means that you need to stick to the same system throughout your whole assignment. Here is a basic system of referencing that you can use, which consists of the following two parts:

1. **A marker in your assignment**: After you have used a reference in your assignment (you have read something and included it in your work as a quote, or re-written it your own words) you should mark this is in your text with a number, e.g. [1]. The next time you use a reference you should use the next number, e.g. [2].
2. **Bibliography**: This is just a list of the references you have used in your assignment. In the bibliography, you list your references by the numbers you have used, and include as much information as you have about the reference. The list below gives what should be included for different sources.

    a. **Websites** – Author (if possible), title of the web page, website address, [date you accessed it, in square brackets].

    E.g. Dan Snow, 'How did so many soldiers survive the trenches?', http://www.bbc.co.uk/guides/z3kgjxs#zg2dtfr [11 July 2014].

    b. **Books** – Author, date published, title of book (in italics), pages where the information came from.

    E.g. S. Dubner and S. Levitt, (2006) *Freakonomics*, 7-9.

    c. **Articles** – Author, 'title of the article' (with quotation marks), *where the article comes from* (newspaper, journal etc.), date of the article.

    E.g. Maev Kennedy, 'The lights to go out across the UK to mark First World War's centenary', Guardian, 10 July 2014.

Appendix 2 – Resource for Baseline and Final Assignments

When to Use Correlation and What It Tells Us

Researchers compute correlation coefficients when they want to know how two variables are related to each other. For a Pearson product-moment correlation, both of the variables must be measured on an interval or ratio scale and are known as **continuous variables**. For example, suppose I want to know whether there is a relationship between the amount of time students spend studying for an exam and their scores on the exam. I suspect that the more hours students spend studying, the higher their scores will be on the exam. But I also suspect that there is not a perfect correspondence between time spent studying and test scores. Some students will probably get low scores on the exam even if they study for a long time, simply because they may have a hard time understanding the material. Indeed, there will probably be a number of students who spend an inordinately long period of time studying for the test precisely *because* they are having trouble understanding the material. On the other hand, there will probably be some students who do very well on the test without spending very much time studying. Despite these "exceptions" to my rule, I still hypothesize that, *on average*, as the amount of time spent studying increases, so do students' scores on the exam.

There are two fundamental characteristics of correlation coefficients researchers care about. The first of these is the **direction** of the correlation coefficient. Correlation coefficients can be either positive or negative. A **positive correlation** indicates that the values on the two variables being analyzed move in the same direction. That is, as scores on one variable go up, scores on the other variable go up as well (on average). Similarly, on average, as scores on one variable go down, scores on the other variable go down. Returning to my earlier example, if there is a positive correlation between the amount of time students spend studying and their test scores, I can tell that, on average, the more time students spend studying, the higher their scores are on the test. This is *equivalent* to saying that, on average, the *less* time they spend studying, the *lower* their scores are on the test. Both of these represent a *positive* correlation between time spent studying and test scores. (*Note:* I keep saying "on average" because it is important to note that the presence of a correlation between two variables does *not* mean that this relationship holds true

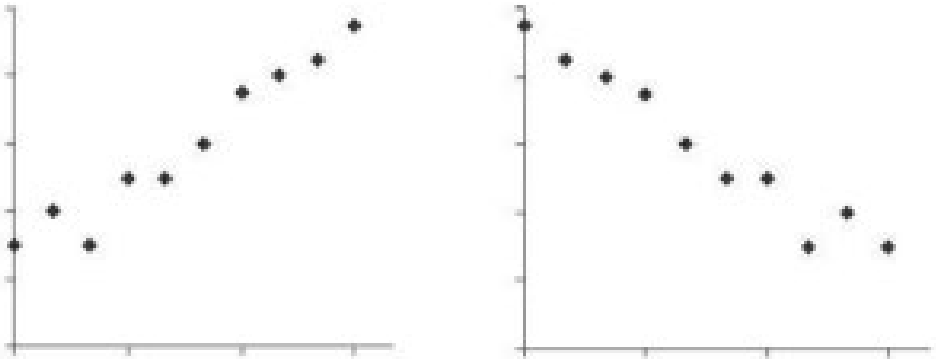


FIGURE 8.1 Examples of positive and negative correlations.

for each member of the sample or population. Rather, it means that, in general, there is a relationship of a given direction and strength between two variables in the sample or population.)

A **negative correlation** indicates that the values on the two variables being analyzed move in *opposite* directions. That is, as scores on one variable go up, scores on the other variable go down, and vice versa (on average). If there were a negative correlation between the amount of time spent studying and test scores, I would know that, on average, the more time students spend studying for the exam, the *lower* they actually score on the exam. Similarly, with a negative correlation I would also conclude that, on average, the *less* time students spend studying, the *higher* their scores are on the exam. These positive and negative correlations are represented by **scattergrams** in Figure 8.1. Scattergrams are simply graphs that indicate the scores of each case in a sample simultaneously on two variables. For example, in the positive correlation scattergram in Figure 8.1, the first case in the sample studied for 1 hour and got a score of 30 on the exam. The second case studied for 2 hours and scored a 40 on the exam.

The second fundamental characteristic of correlation coefficients is the **strength** or **magnitude** of the relationship. Correlation coefficients range in strength from −1.00 to +1.00. A correlation coefficient of .00 indicates that there is no relationship between the two variables being examined. That is, scores on one of the variables are not related in any meaningful way to scores on the second variable. The closer the correlation coefficient is to either −1.00 or +1.00, the stronger the relationship is between the two variables. A **perfect negative correlation** of −1.00 indicates that for every member of the sample or population, a higher score on one variable is related to a lower score on the other variable. A **perfect positive correlation** of +1.00 reveals that for every member of the sample or population, a higher score on one variable is related to a higher score on the other variable.

Perfect correlations are never found in actual social science research. Generally, correlation coefficients stay between −.70 and +.70. Some textbook authors suggest that correlation coefficients between −.20 and +.20 indicate a weak relation between two variables, those between .20 and .50 (either positive or negative) represent a moderate relationship, and those larger than .50 (either positive or negative) represent a strong relationship. These general rules of thumb for judging the relevance of correlation coefficients must be taken with a grain of salt. For example, even a "small" correlation between alcohol consumption and liver disease (e.g., +.15) is important, whereas a strong correlation between how much children like vanilla and chocolate ice cream (e.g., +.70) may not be so important.

The scattergrams presented in Figure 8.1 represent very strong positive and negative correlations ($r = .97$ and $r = -.97$ for the positive and negative correlations, respectively; r is the symbol for the sample Pearson correlation coefficient). In Figure 8.2, a scattergram representing virtually no correlation between the number of hours spent studying and the scores on the exam is presented. Notice that there is no discernible pattern between the scores on the two variables. In other words, the data presented in Figure 8.2 reveal that it would be virtually impossible to

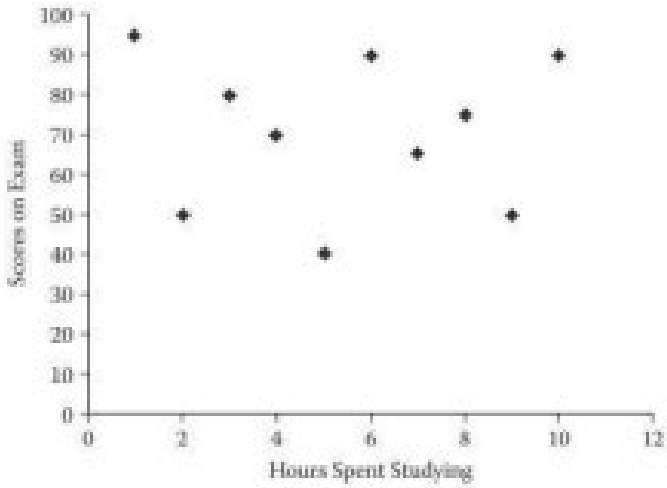


FIGURE 8.2 No correlation between hours spent studying and exam scores.

predict an individual's test score simply by knowing how many hours the person studied for the exam.

Pearson Correlation Coefficients in Depth

The first step in understanding how Pearson correlation coefficients are calculated is to notice that we are concerned with a sample's scores on *two* variables at the same time. Returning to our previous example of study time and test scores, suppose that we randomly select a sample of five students and measure the time they spent studying for the exam and their exam scores. The data are presented in Table 8.1 (with a scattergram in Figure 8.3).

TABLE 8.1 Data for Correlation Coefficient

	Hours Spent Studying (*X* Variable)	Exam Score (*Y* Variable)
Student 1	5	80
Student 2	6	85
Student 3	7	70
Student 4	8	90
Student 5	9	85

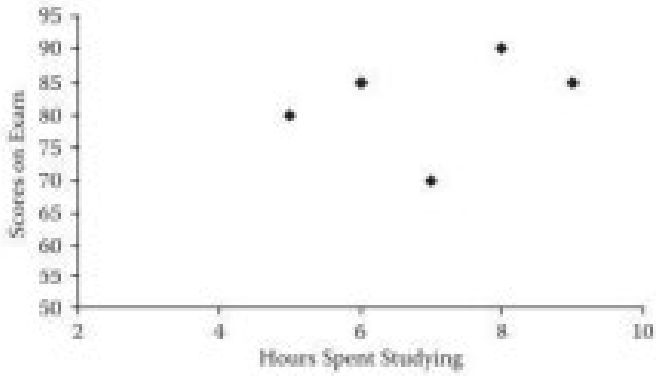


FIGURE 8.3 Scattergram of data from Table 8.1.

CHAPTER 13

Regression

In Chapter 8, the concept of correlation was introduced. Correlation involves a measure of the degree to which two variables are related to each other. A closely related concept, coefficient of determination, was also introduced in that chapter. This statistic provides a measure of the strength of the association between two variables in terms of percentage of variance explained. Both of these concepts are present in regression. In this chapter, the concepts of **simple linear regression** and **multiple regression** are introduced.

Regression is a very common statistic in the social sciences. One of the reasons it is such a popular technique is that it is so versatile. Regression, particularly multiple regression, allows researchers to examine the nature and strength of the relations between the variables, the relative predictive power of several independent variables on a dependent variable, and the unique contribution of one or more independent variables when controlling for one or more covariates. It is also possible to test for interactions in multiple regression. With all of the possible applications of multiple regression, it is clear that it is impossible to describe all of the functions of regression in this brief chapter. Therefore, the focus of this chapter is to provide an introduction to the concept and uses of regression, and to refer the reader to resources providing additional information.

Simple versus Multiple Regression

The difference between simple and multiple regression is similar to the difference between one-way and factorial ANOVA. Like one-way ANOVA, simple regression analysis involves a single **independent,** or **predictor variable** and a single **dependent,** or **outcome variable**. This is the same number of variables used in a simple correlation analysis. The difference between a Pearson correlation coefficient and a simple regression analysis is that whereas the correlation does not distinguish between independent and dependent variables, in a regression analysis there is always a designated predictor variable and a designated dependent variable. That is because the purpose of regression analysis is to make *predictions* about the value of the dependent variable given certain values of the predictor variable. This is a simple extension of a correlation analysis. If I am interested in the relationship between height and weight, for example, I could use simple regression analysis to answer this question: If I know a man's height, what would I *predict* his weight to be? Of course, the accuracy of my prediction will only be as good as my correlation will allow, with stronger correlations leading to more accurate predictions. Therefore, simple linear regression is not really a more powerful tool than simple correlation analysis. But it does give me another way of conceptualizing the relation between two variables, a point I elaborate on shortly.

The real power of regression analysis can be found in multiple regression. Like factorial ANOVA, multiple regression involves models that have two or more predictor variables and a single dependent variable. For example, suppose that, again, I am interested in predicting how much a person weighs (i.e., weight is the dependent variable). Now, suppose that in addition

to height, I know how many minutes of exercise the person gets per day, and how many calories a day he consumes. Now I've got three predictor variables (height, exercise, and calories consumed) to help me make an educated guess about the person's weight. Multiple regression analysis allows me to see, among other things, (1) how much these three predictor variables, as a group, are related to weight, (2) the strength of the relationship between each predictor variable and the dependent variable *while controlling for the other predictor variables in the model*, (3) the *relative* strength of each predictor variable, and (4) whether there are interaction effects between the predictor variables. As you can see, multiple regression is a particularly versatile and powerful statistical technique.

Variables Used in Regression

As with correlation analysis, in regression the dependent and independent variables need to be measured on an interval or ratio scale. **Dichotomous** (i.e., categorical variables with two categories) predictor variables can also be used. There is a special form of regression analysis, logit regression, that allows us to examine dichotomous dependent variables, but this type of regression is beyond the scope of this book. In this chapter, we limit our consideration of regression to those types that involve a continuous dependent variable and either continuous or dichotomous predictor variables.

Regression in Depth

Regression, particularly simple linear regression, is a statistical technique that is very closely related to correlations (discussed in Chapter 8). In fact, when examining the relationship between two continuous (i.e., measured on an interval or ratio scale) variables, either a correlation coefficient or a regression equation can be used. Indeed, the Pearson correlation coefficient is nothing more than a simple linear regression coefficient that has been standardized. The benefits of conducting a regression analysis rather than a correlation analysis are (1) regression analysis yields more information, particularly when conducted with one of the common statistical software packages, and (2) the regression equation allows us to think about the relation between the two variables of interest in a more intuitive way. Whereas the correlation coefficient provides us with a single number (e.g., $r = .40$), which we can then try to interpret, the regression analysis yields a formula for calculating the **predicted value** of one variable when we know the actual value of the second variable. Here's how it works.

The key to understanding regression is to understand the formula for the regression equation. So I begin by presenting the most simple form of the regression equation, describe how it works, and then move on to more complicated forms of the equation. In Table 13.1, the regression equation used to find the predicted value of Y is presented along with definitions of the components.

In simple linear regression, we begin with the assumption that the two variables are **linearly** related. In other words, if the two variables are actually related to each other, we assume that every time there is an increase of a given size in value on the X variable (called the **predictor** or **independent** variable), there is a corresponding increase (if there is a positive correlation) or decrease (if there is a negative correlation) *of a given size* in the Y variable (called the **dependent,**

TABLE 13.1 The Regression Equation

$$\hat{Y} = bX + a$$

where $\hat{Y}$ is the predicted value of the Y variable
b is the unstandardized regression coefficient, or the slope
a is intercept (i.e., the point where the regression line intercepts the Y axis)

Bibliography

T. C. Urdan. (2010). *Statistics in Plain English,* 79-81; 145-146

Notes

UNI PATHWAYS

UNI PATHWAYS

thebrilliantclub.org