



Unit 6: One and Two Variable data analysis

Lesson 6.2: Dynamic analysis of Two-variable data

Learning Goal: analyze two-variable data for & account for impact of outliers and hidden or extraneous variables

Once you know there is a correlation between variables in a correlational study, it is important to consider what outside influences, if any, may be present. _____ and _____ or other _____ can have a dramatic impact on a linear regression model which may skew analysis and interpretation of the data.

Outlier: _____.

Hidden variable: _____.

Detecting outliers & hidden variables

It is easy to spot an outlier in a scatter plot when it is very far from the trend of the data, but what about outliers that may not be so obvious. Creating a residual plot of the data can help to identify outliers.

Residual:

_____.

Residual Plot:

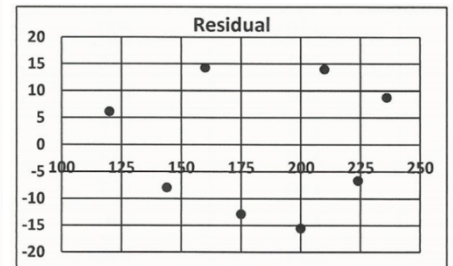
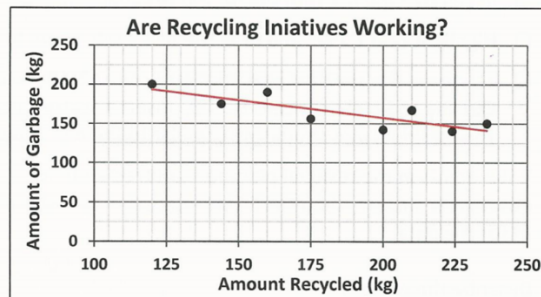
_____.

Example:

Will an increase in recycling result in a reduction of landfill? The table below compares the mass of garbage and recycling for a town during a recycling campaign.

Create a scatter plot and perform a linear regression.

Amount Recycled (kg)	Amount of Garbage (kg)
120	200
144	175
160	190
175	156
200	142
210	167
224	140
236	150



a) Create a scatter plot and perform a linear regression.

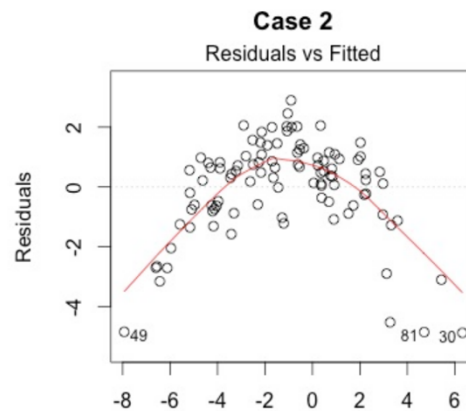
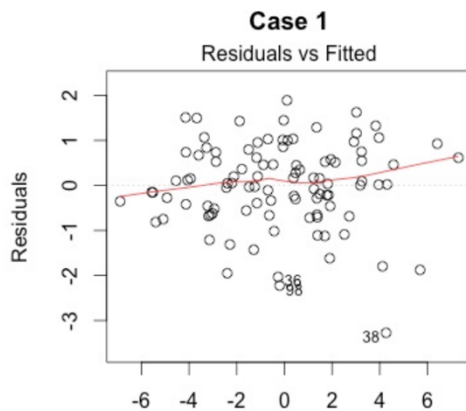


THE ERINDALE ACADEMY

1576 Dundas Street West, Mississauga ON L5C 1E5
www.teacademy.ca Tel: 905-232-1576
Email: info@teacademy.ca

Teacher: Ella

- b) Describe the trend. Is a linear model reasonable in this case?
- c) Interpret the residuals for (120, 200) and (200, 142).
- d) Construct a residual plot of the data. And describe the pattern.





THE ERINDALE ACADEMY

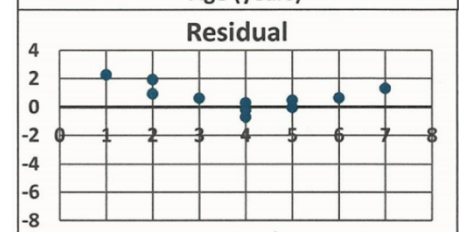
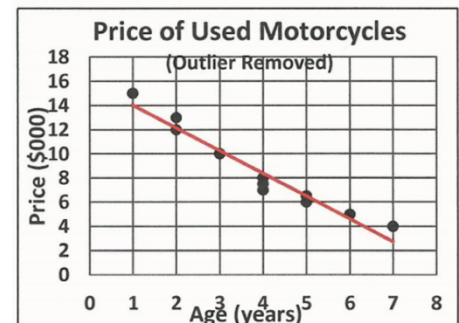
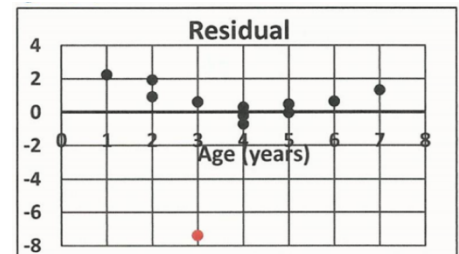
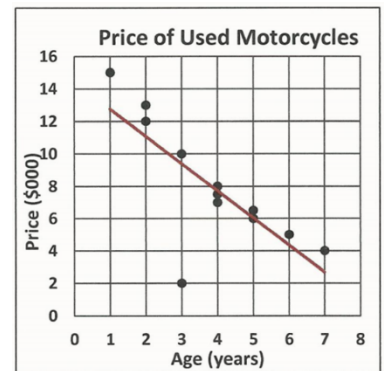
1576 Dundas Street West, Mississauga ON L5C 1E5
 www.teacademy.ca Tel: 905-232-1576
 Email: info@teacademy.ca

Teacher: Ella

Practice: The table shows the sale price for several used motorcycles of the same model and their age.

Age (years)	Price (\$000)
1	15
2	13
2	12
3	2
3	10
4	8
4	7.5
4	7
5	6.5
5	6
6	5
7	4

- Create a scatter plot of sale price versus age. It should look similar to the one shown here.
- Perform a linear regression and correlation analysis and discuss the goodness of fit.
- Follow a similar process as in Example 1 to construct a residual lot and identify any outliers. It should look similar to the one shown here. Suggest reasons why outliers may exist.
- Repeat the regression and the residual plot with the outlier remove. It should look similar to those shown here. Compare the two linear models.





Unit 6: One and Two Variable data analysis

Lesson 6.3: Cause & Effect, uses & misuses of data

*Learning Goal: distinguish between **correlation & causality** and identify types of relationships between variables*

See how data can be distorted to impact the way it is interpreted and thus influence perception

Data analysis involves much more than fitting a line or a curve to a set of data points. Once you know there is a correlation, it is important to consider how and why such a correlation exists – to distinguish between **correlation** and **causality**.

Causal Relationships

Direct cause & effect:

Common cause:

Reverse cause & effect:



THE ERINDALE ACADEMY

1576 Dundas Street West, Mississauga ON L5C 1E5
www.teacademy.ca Tel: 905-232-1576
Email: info@teacademy.ca

Teacher: Ella

Accidental relationship:

Presumed relationship:

Part II: Uses & Misuses of data

You can display data in multiple ways. Sometimes data are deliberately distorted to make an argument more convincing. The media often sensationalizes data to generate public interest. How you choose to display data can have a significant impact on how a reader is likely to interpret them.

Data can become easily distorted and perceptions easily influenced in several ways

- Bias in the wording
- Misrepresentation of the graphical data
- Inadequate data collection
- Poor analysis