|  |  |  |
| --- | --- | --- |
|  | Gbwehkag;wej | |
| Due: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Total marks: /25  (10% of total assessment mark) | | Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

**Extended investigation Part 1:** **Preparation activities**

The purpose of this investigation is to provide students with experience in the statistical investigation cycle. There will be opportunities to:

* clarify the problem and pose questions that can be answered with data
* design and implement a plan to collect data
* select and apply graphical or numerical techniques to analyse the data
* interpret and communicate the results of the analysis

(See glossary for Topic 3.1)

Activities associated with the statistical investigation are outlined below with examples provided and questions you could consider to address each part of the process.

**Activity 1: The problem**

For example:

*Teachers in Agerla College are concerned about the amount of time that students have available to study on the weekends. In fact, the teachers believe that the students are undertaking too many hours of paid work and this is affecting the time spent studying.*

Examples of questions … Consider this problem with respect to:

* Are all words or phrases clearly understood?
* Do any words need to be defined?
* What is meant by the “weekend”? Does it include Friday night?
* Which students are being considered? Should this investigation consider only one year group? Explain.
* What is meant by paid work? Should work around the home when pocket money is given for chores be included?
* What is meant by study? Does this include doing homework? Painting a piece of artwork for school?
* Does it matter who the teachers are? Or how many of them think this?

**Activity 2: Pose a question**

Example: One question that could be answered by collecting data is:

*Do students do enough study on the weekend?*

This question is one that relates to the topic but there are better questions that can be related to the topic.

**Activity 3: Design the data collection**

Plan and design the collection of relevant data to answer the question. Use the following ideas to guide the planning.

* What data is needed? How many variables are there? What are they?
* What type of data will be collected? Numeric? Continuous or discrete?
* From where will data be collected? This school, which class, which year groups?
* How will the data be collected? Written or oral survey?
* How will the data be kept secure?
* Where will the data be stored? What backup of data is needed?
* How many students will provide data? How do we know that there is enough data to determine any relationships?
* Do the students need to be randomly selected for the survey?
* Is the sample biased? Will it represent the population of students which is the focus for this investigation?
* When will the data be collected?
* What procedures might be necessary to ensure the data is valid and reliable?

*Note: The original problem described is suitable for the collection of bivariate data. Check that this fits with your design.*

**Activity 4: Data collection**

Implement your plan to collect the data.

**Activity 5: Analyse data**

For bivariate data consider the following actions for the analysis:

* Identify the response and the explanatory variables.
* Construct a scatterplot.
* Can you identify any outliers?
* Does the scatterplot show a pattern? Describe any apparent pattern.
* Determine the correlation coefficient and the coefficient of determination.
* Determine the equation for the least-squares line.
* Construct a residual plot.

**Activity 6: Interpret results**

* What type of relationship exists between the variables?
* Is a linear model appropriate? How do you know?
* What does the correlation coefficient infer?
* How much variation in the response variable is explained by the explanatory variable?
* What other factors may have influenced your results?
* What does the gradient of the regression line infer?
* What does the vertical intercept of the regression line infer?
* Is the regression line reliable for predictions?
* How do your results answer the original question posed in Activity 2?
* Are there further questions to answer as a result of your findings?

**Activity 7: Communicate findings**

Present your findings as a report, (can be emailed), or poster in a *systematic and concise manner.*

[Glossary]

**YOUR TASK**

You are required to demonstrate the Statistical Investigation Process for bivariate data by using the data source below.

Data will be sourced from the following website <https://new.censusatschool.org.nz/explore/>

* Use CHROME as Excel may not work on Safari

🡪 Download or explore a sample (orange button)

🡪 Read and tick the box for I agree.

🡪 Complete criteria

database: CensusAtSchool NZ 2019

select subpopulation: year 13

select variables: all

select sample type: random sample

enter sample size: 40-50

🡪 Generate sample (orange button)

🡪 Download sample (blue button)

* Data is randomly generated. No two samples will be the same.

Complete Activities 2 to 7 using the data source.

Work is to be submitted on the day of the assessment, before assessment commences.

Ensure you include a print out of the raw data you use.