

GY472

Coursework 2

Bivariate Analysis and Tests of Difference – 1000 Words

Aim

By the end of this exercise, which involves you writing a report, you should have become confident in utilising selected parametric and non-parametric statistical tests to determine if data sets display patterns/relationships/ differences. You should become competent in the steps utilised in hypothesis testing and in the use of Minitab.

You are expected to be able to select a sample from a large data set, to describe the data (graphs and text) and then to explore patterns/relationships/differences by selecting and using appropriate statistical tests. You have the freedom to select the data set you wish to study and, to some extent, the relationships you wish to explore. The requirement is that you conduct 3 bivariate tests and two tests of difference, in addition to required normality testing.

You will also be expected to undertake some additional reading to try and explain any patterns/relationships/differences you identify and to examine theories presented in the literature that can be tested over the course of this report.

The two data sets provided are:-

- 1. Gross Domestic Product and Death Rate**
- 2. Coastal Erosion at Holderness**

The further details associated with the two data sets are outlined below together with an outline of the statistical tests that you need to undertake on your selected data sets.

The following must be attempted following appropriate normality testing to decide appropriate test:

- i) For your selected data set you must test the relationship between three sets of two variables (e.g. rate of erosion v cliff height, rate of erosion v northing, rate of erosion v easting). It is ok if no relationship is present in your tests but if a significant correlation is identified then linear regression should be undertaken in each instance to describe the nature of the relationship. Essentially you must conduct minimum of 3 bivariate tests (regression and/or correlation). If no significant correlations are identified then feel free to still show your skills in regression analysis but comment upon why the results may not be valid.
- ii) For your selected data set undertake two tests of difference to determine if particular attributes impact on the data (e.g. rates of erosion north and south of a coastal defence).

The work is to be written as a report so ensure that you adopt a logical report structure. Word savings can be made via the Methodology.

Don't forget to use references throughout. Only choose one of the datasets to work on.

Work to be submitted electronically via Turnitin by **17.00 Friday 23rd March, 2018.**

For this assignment you will need to undertake a series of statistical tests and report them from within a report structure. For each test you will need to:-

- a) State the rationale for the study including research questions.
- b) Select a sample from the data set chosen with careful consideration regarding sampling.
- c) Graphically display the sampled data using appropriate graphs
- d) Briefly describe the data set and highlight any potential patterns/relationships
- e) State the Significance Level and therefore confidence being adopted
- f) Undertake appropriate normality tests to determine which statistical test is appropriate (e.g. Parametric or non-parametric)
- g) Clearly state which statistical test is being utilised.
- h) Define the null and alternate hypotheses.
- i) Utilise Minitab to calculate the test statistic, and present these results.
- j) Clearly state if a significant result has been determined and note if the Null hypothesis is to be rejected.
- k) Relate the p value and test statistic to the critical value from the table (if appropriate) and then to the real world scenario.
- l) Try to explain the pattern/relationship you have identified by making reference to appropriate literature.

Specifics for Data Set 1 - Gross Domestic Product and Death Rates for Each Country

The data set (available on Student Central) consists of the quinquennial (every 5 years) GDP and Death rates of every country in the world (235). The data set is arranged by geographical region (e.g. Central Africa, Eastern Africa, Europe etc.) and lists every country in the world. The data includes

- a) Country
- b) Broad Geographical Region
- c) Sub Region
- d) Average Gross Domestic Product (US\$) for every 5 years from 1960.
- e) Average Death Rate (per 1,000 population) for every 5 years from 1950

You are expected to describe/explore:-

1. Relationship between GDP and death rate (Bivariate Test 1)
2. Relationship between death rate and time for a developing region (Bivariate Test 2)
3. Relationship between death rate and time for a developed region (Bivariate Test 3)
4. Differences in death rates between different sub regions (at least three sub-regions to be utilised) (Two tests of difference)

Specifics for Data Set 2 – Coastal Erosion along the Holderness Coastline

The data set (available on Student Central) consists of 123 sites where the amount of coastal erosion in a particular year (1950-2004) has been recorded. The cliffs are composed of unconsolidated silts, clays and sands deposited by the Last British Ice Sheet. The area is recognised as one of the fastest eroding sites in the European Union. The data set (available on Student Central) consists of the following:

Worksheet 1

- a) Post Number – Numbered from North to South = Sample Number
- b) Date Set Up – When measurements of erosion started (Date)
- c) Location of Post – Description of the site the post is located at
- d) Height of Cliff – The height of the cliff at the post location (m)
- e) Easting – The easting grid reference of the post (m)
- f) Northing – The northing grid reference of the post (m)
- g) Total Erosion – Total erosion recorded at the site during period of measurement (m)
- h) Number of Readings – The number of readings taken at each site (in years)
- i) Average Erosion Rate – The average rate of erosion at the site during the period of measurement.

Coastal defences are located between the following post numbers:-

Bridlington – Between posts 4 and 5

Hornsea – Between posts 43 and 44

Withernsea – Between posts 86 and 87

Easington – Between posts 106 and 109

You are expected to explore, following normality testing:

- 1. Relationship between cliff height and rate of erosion (Bivariate Test 1)
- 2. Relationship between rate of erosion and northing (Bivariate Test 2)
- 3. Relationship between rate of erosion and easting (Bivariate Test 3)
- 4. Differences in rates of erosion north and south of coastal defences (at least two sites)
(Two tests of difference)

The following additional information is also provided

Figure 2 – The location of the sample sites.

The following web-site may be of use <http://urbanrim.org.uk/Holderness.htm>

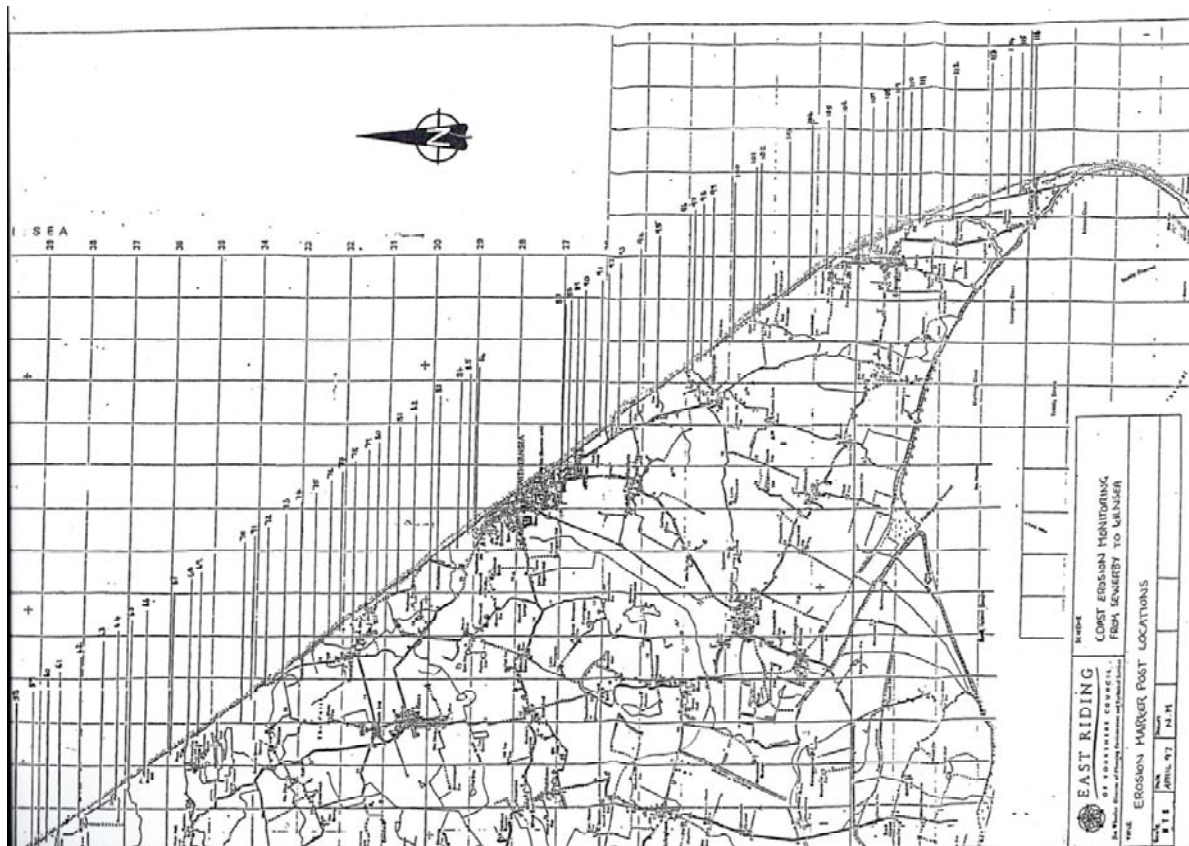
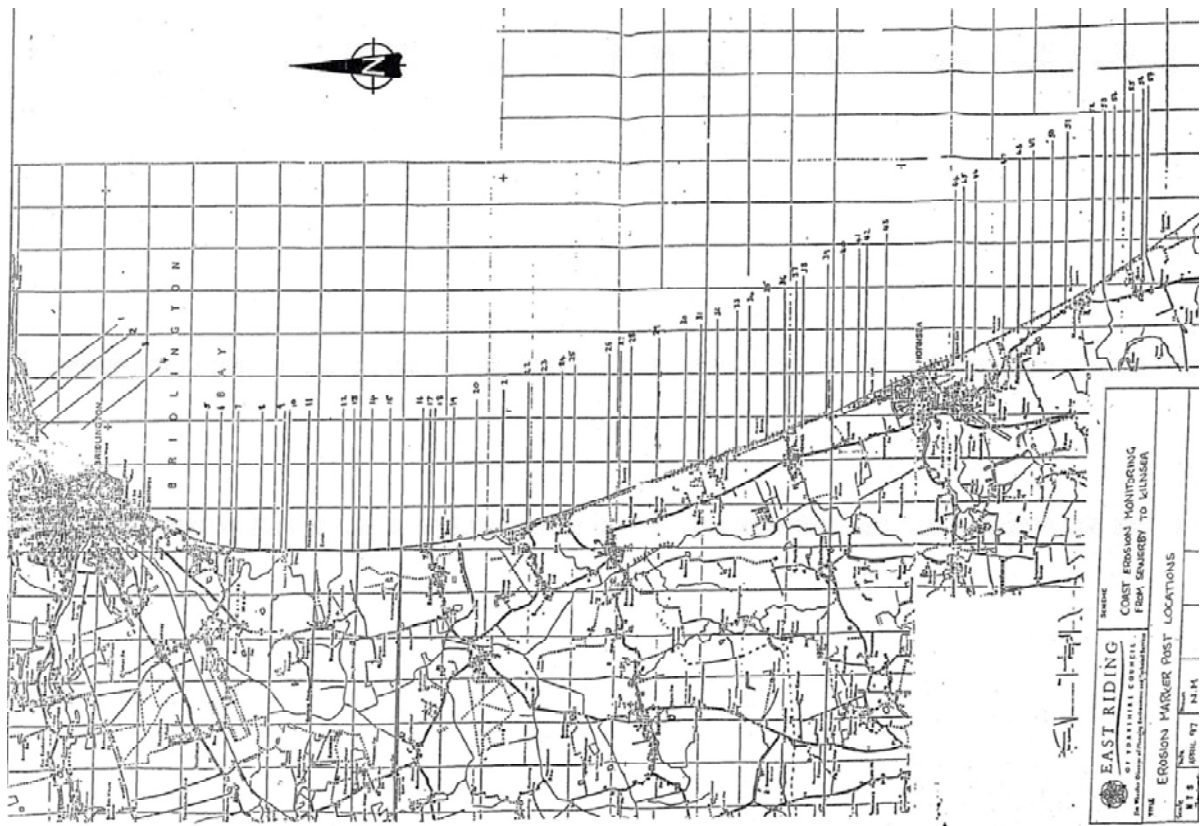


Figure 2: The location of the Holderness sites where cliff erosion has been measured.