PC Lab 6 – Pearson’s Correlation, Simple Linear Regression and Multiple Linear Regression

**SPSS Exercises**

Your first task this week is to examine any association between academic achievement by ACU students in their first year, first semester studies and their self-reported stress level. You are asked to do so in relation to the entire ACU student population, and for male and female students separately.

**While this tutorial can be completed quickly using SPSS, try to do the “ACU student population” calculation by hand (Excel – PLEASE DO NOT USE PEN AND PAPER!). This will give you a greater understanding of the concepts. See the Weeks 5 and 6 lecture notes for examples of how to do this. Only use SPSS to check your Excel answers.**

You are provided with data on:

* Average unit mark: an average of all marks obtained by first year students in their first semester of study at ACU. The range of possible marks is zero to 100.
* Stress levels: in the mid-semester break of semester 1, students were required to complete a survey that measures their stress levels. The range of final “stress scores” is zero (complete absence of stress) to 100 (completely debilitated by stress). A stress level of 50 on this scale is considered to be the population norm.

Further details on these variables can be found in the tables below.

To complete this task, please provide:

* A scatterplot of the data for the population, using stress as the independent (x) variable and average unit marks as the dependent (y) variable
* Correlation results for the overall ACU student population
* Correlation results for just male ACU students
* Correlation results for just female ACU students
* Your interpretation of all of the above results

Your next task is to expand upon the above by generating a simple linear equation that describes the association between students’ average marks and stress levels.

Using stress levels as the independent/predictor (x) variable and average mark as the dependent (y) variable:

* Provide the linear equation (mark = b0 + b1 x stress) for:
  + ACU students as a whole
  + Male ACU students
  + Female ACU students
* With particular reference to the r and r2 values, what is your interpretation of the results?
* What is the predicted mark for:
  + A male student with a stress level of 75?
  + A female student with a stress level of 30?

**VARIABLES**

|  |  |
| --- | --- |
| **Name** | AGE |
| **Definition** | Participant’s age at time of enrolment at ACU |
| **Values** | NA – continuous data |
| **Notes and related variables** | Age derived from date of birth (not shown) |

|  |  |  |
| --- | --- | --- |
| **Name** | GENDER | |
| **Definition** | Student gender | |
| **Values** | 0 | Male |
| 1 | Female |
| **Notes and related variables** |  | |

|  |  |  |
| --- | --- | --- |
| **Name** | LIVING\_ARRANGE | |
| **Definition** | Self-reported initial living arrangements for students during the first year of study | |
| **Values** | 1 | At home |
| 2 | College or other student residential environment |
| 3 | independently |
| **Notes and related variables** | Self-reported data obtained from T0 survey information. Definitions:   * At home: students living in their family home. * College or other residential environment: students living in specific ACU (or other University) managed accommodation (“colleges”) or in specific residential facilities operated by private companies exclusively for students (“student accommodation”). * Independently: students living in private accommodation under standard rental arrangements or share-houses (not exclusively for students). | |

|  |  |  |
| --- | --- | --- |
| **Name** | FACULTY | |
| **Definition** | Faculty student enrolled in | |
| **Values** | 1 | Arts and Sciences |
| 2 | Education |
| 3 | Health Sciences |
| 4 | Theology and Philosophy |
| 5 | Business |
| **Notes and related variables** | Data derived from original degree course student admitted to at time of enrolment. For double degree students studying degrees from different faculties, the Faculty of the first-named degree is considered (e.g. Bachelor of Nursing/Bachelor of Business Administration = Faculty of Health Sciences).  Note 1: Faculty of Theology and Philosophy established in 2009; no data available prior to this date (courses previously available from Faculty of Arts and Sciences).  Note 2: Faculty of Business established in 2010; no data available prior to this date (courses previously available from Faculty of Arts and Sciences). | |

|  |  |  |
| --- | --- | --- |
| **Name** | METRO | |
| **Definition** | Geographic background of students prior to enrolment | |
| **Values** | 1 | Metropolitan |
| 2 | Non-metropolitan |
| **Notes and related variables** | Data available for **domestic students only**. Location (metropolitan or non-metropolitan) derived from address of students prior to enrolment using postcode assignment to either metropolitan or non-metropolitan areas. Does not reflect geographic location during study. | |

|  |  |  |
| --- | --- | --- |
| **Name** | DEGREE\_TYPE | |
| **Definition** | Type of degree undertaken by student | |
| **Values** | 0 | Single degree |
| 1 | Double degree |
| **Notes and related variables** | Data derived from original degree course student admitted to at time of enrolment. | |

**NEW VARIABLES**

|  |  |
| --- | --- |
| **Name** | AVERAGE\_MARK |
| **Definition** | Average of student marks for first year, first semester units |
| **Values** | NA – continuous data |
| **Notes and related variables** | * Data represents average of all units completed during semester 1 of first year; irrespective of full or part time study * Range: 0 to 100 |

|  |  |
| --- | --- |
| Name | STRESS |
| Definition | Stress level, as measured by totally fictional standardised survey instrument |
| Values | NA – continuous data |
| Notes and related variables | * Self-reported stress level; represents final score from survey questions * Range: 0 (complete absence of stress to 100 (completely debilitated by stress) * Population norm: 50 |

**Multiple Linear Regression**

Work through the lecture note examples on multiple linear regression. After completing the lecture examples, open the file called “strokeMSData.sav” and complete the exercises below. **Do not try to use Excel for these problems, just use SPSS.**

The data in strokeMSData.sav is a modified file from the statistical science website statsci.org.

This study compared three occupational therapy programs to help patients recover from a stroke. Eight patients were assigned to each of the three groups:

Group (E) was given an experimental program developed by the investigator from a model of intervention for stroke rehabilitation.

Group (F) was given a pre-existing program.

Group (G) was a non-treatment program.

Each program lasted for eight weeks. All subjects were evaluated at the start of the program and at weekly intervals. We are only using Weeks 1, 2, 4 and 8 evaluations.

Data Example:

|  |  |
| --- | --- |
| **Subject:** | Subject ID |
| **Group:** | E, F, G |
| **Gender:** | Male (M) or Female (F) |
| **Side:** | Side of brain affected (L, R) |
| **Age:** | Age of subject in years |
| **Lapse:** | Time lapse from stroke to start of program in weeks |

**UE1, UE2, UE4, UE8**: Upper extremities scores at weeks 1, 2, 4 and 8

**HW1, HW2, HW4, HW8**: Hand-wrist scores at weeks 1, 2, 4 and 8

**LE1, LE2, LE4, LE8**: Lower extremities scores at weeks 1, 2, 4 and 8

Stroke Data

Multiple linear regression

1. Find the line of best fit for UE8 based on the UE1, HW1, LE1 values. First do this using the Enter method.
2. Make sure you assess all assumptions of multiple linear regression and provide your justifications on whether they have been satisfied.
3. Experiment with using the Forward Selection and Backwards Elimination methods and see if your results differ from the Enter method.