**INTEREST IN THE SUMMER SCHOOL**

**=========================================================================**

**Note: The following subsections are designed to aid the directors in determining the appropriateness of your application to the school. Please take great care in answering these questions.**

**Part 1 :- About you**

Describe briefly your area of research (Maximum 1000 characters)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Describe briefly your experience with research data, including usage, scientific software development and/or experience in open science (Maximum 1500 characters)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Describe briefly how you anticipate the research Data Science skills you will learn at this workshop will improve your ability to carry out your research project. (Maximum 1500 characters)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What specific skills do you wish to acquire or improve during this school? (Maximum 1000 characters)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What broader impact to your work environment do you anticipate from participating in the workshop?

(Maximum 1000 characters)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Part 2 - Statistics**

Compute the sample mean and sample variance (to 2 decimal places) of the following data

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.30 | 1.31 | 1.11 | 1.08 | 1.09 | 1.13 | 1.65 | 1.26 | 1.26 | 1.18 |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Given the following paired data set

|  |  |
| --- | --- |
| x | y |
| 7.5 | 28.66 |
| 4.48 | 20.37 |
| 8.60 | 22.33 |
| 7.73 | 26.35 |
| 5.28 | 22.29 |
| 4.25 | 21.74 |
| 6.99 | 23.11 |
| 6.31 | 23.13 |
| 9.15 | 24.68 |
| 5.06 | 21.89 |

Insert a scatter plot of x and y here.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Compute (to 2 decimal places) the sample linear Pearson correlation coefficient and equivalent covariance between x and y.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Why is a correlation coefficient often more useful than a covariance?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Two different Universities record data on students who are unable to attend classes due to illness. University 1 recorded absences over ten consecutive days. This data is recorded as N1 below. University 2 recorded absences over six consecutive days. This data is recorded as N2 in below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **N1** | **9** | **9** | **5** | **5** | **5** | **6** | **16** | **8** | **8** | **7** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **N2** | **13** | **11** | **14** | **13** | **12** | **11** |

Using a two sampled t-test (assuming equal variances), determine whether there is a difference between means. Clearly state your hypothesis.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Comment on the appropriateness of the test.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In a related study, you observe that when students are ill, they take on average five days off. How would this affect the appropriateness of using a t-test described above.

**Part 3 – Technical skills (note – these are non-essential requirements but will help us determine the level of ability of the applicants)**

Knowledge of Programming/Script Languages (mark all/any that apply):

excellent good fair poor

R/Rstudio ☐ ☐ ☐ ☐

Mysql or other SQL ☐ ☐ ☐ ☐

C ☐ ☐ ☐ ☐

C++ ☐ ☐ ☐ ☐

Python ☐ ☐ ☐ ☐

Shell Script (Bash) ☐ ☐ ☐ ☐

Other: \_\_\_\_\_\_\_\_ ☐ ☐ ☐ ☐ (optional)

Knowledge of Linux/Unix (mark all/any that apply):

excellent good fair poor

Command line interface ☐ ☐ ☐ ☐

Bash shell scripting ☐ ☐ ☐ ☐

Batch system configuration ☐ ☐ ☐ ☐

Compilation of scientific ☐ ☐ ☐ ☐

software

Knowledge of Revision Control Software (mark all/any that apply):

excellent good fair poor

git ☐ ☐ ☐ ☐

SVN ☐ ☐ ☐ ☐ Other:\_\_\_\_\_\_\_\_\_ ☐ ☐ ☐ ☐

User experience of cloud computing (mark all/any that apply):

extensive some never used it

Amazon EC2 ☐ ☐ ☐

Microsoft Azure ☐ ☐ ☐ 

OpenStack ☐ ☐ ☐ 

Other:\_\_\_\_\_\_\_\_\_ ☐ ☐ ☐  

Knowledge of Data Deposition tools (mark all/any that apply):

excellent good fair poor

figshare☐ ☐ ☐ ☐

zenodo☐ ☐ ☐ ☐

Other:\_\_\_\_\_\_\_\_\_ ☐ ☐ ☐ ☐

Editor/Tools For Programming (mark all/any that apply):

frequently sometimes never

Eclipse ☐ ☐ ☐

Emacs/XEmacs ☐ ☐ ☐

Gedit ☐ ☐ ☐

Kdevelop ☐ ☐ ☐

Vi/Vim ☐ ☐ ☐

Visual Studio ☐ ☐ ☐  

Eclipse ☐ ☐ ☐

Other: \_\_\_\_\_\_\_\_ ☐ ☐ ☐ (optional)