LETTERKENNY INSTITUTE OF TECHNOLOGY

ASSIGNMENT COVER SHEET

Lecturer’s Name: James Connolly

Assessment Title: Hypothesis Testing

Work to be submitted to: James Connolly

Date for submission of work: 29-08-2018

Place and time for submitting work: 3:00 pm

To be completed by the Student

Student’s Name: PRATEEK PARASHER

Class: Msc in Big Data

Subject/Module: Data Science

Word Count (where applicable):

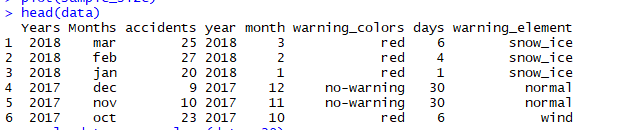
I confirm that the work submitted has been produced solely through my own efforts.

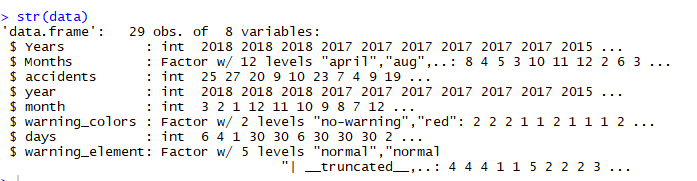
Student’s signature: PRATEEK PARASHER Date: 29/8/2018

|  |
| --- |
| **Notes**  **Penalties:** The total marks available for an assessment is reduced by 15% for work submitted up to one week late. The total marks available are reduced by 30% for work up to two weeks late. Assessment work received more than two weeks late will receive a mark of zero. [Incidents of alleged plagiarism and cheating are dealt with in accordance with the Institute’s Assessment Regulations.]  **Plagiarism:** Presenting the ideas etc. of someone else without proper acknowledgement (see section L1 paragraph 8).  **Cheating:** The use of unauthorised material in a test, exam etc., unauthorised access to test matter, unauthorised collusion, dishonest behaviour in respect of assessments, and deliberate plagiarism (see section L1 paragraph 8).  **Continuous Assessment:** For students repeating an examination, marks awarded for continuous assessment, shall normally be carried forward from the original examination to the repeat examination. |

**ABSTRACT: -** Using statistical analysis I am interested to examine my dataset, but I am not thinking there must be some correlation between weather red warnings & road accidents. But I can’t make that decision on my hypothesis or my assumption I need to perform statistical hypothesis analysis testing. I applied power test, t-test , correlation test. Correlation test is used to find the sample size to perform the power analysis with effect size of 0.5 with 80% certainty and no more than a 5% chance of inaccuracy. Then the sample of 29 records is used to do the power analysis. And the output of p-value is significantly high from 0.05. Therefore, this proves that the null hypothesis is true

**DATA DESCRIPTION**





Dataset containing year 2012-2018 data of road accident and red weather warning data with different -2 warning elements.

Type of data -> continuous

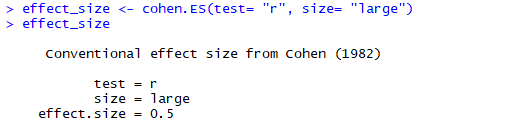
No. of sample -> two – sample

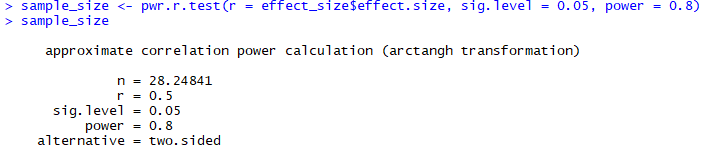
Hypothesis testing -> correlation

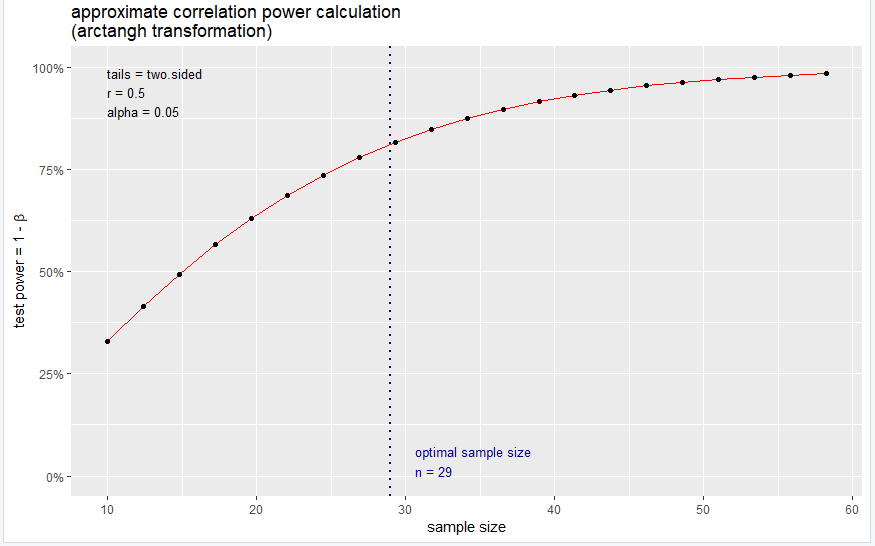
**HYPOTHESIS TESTING**

H0 = The road accident is not related to the red weather warnings

H1 = The road accident is related to the red weather warnings

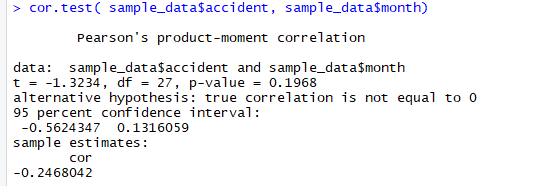






*The results suggest that we need 29 records to detect an effect size of 0.5 with 80% certainty and no more than a 5% chance of inaccuracy.*

**RESULTS**



* P value is 0.1968 in that case we have less strict cut-offs, such as 0.10, requiring less evidence

**The results are considered non-significant - fail to reject Ho.**

result it is seen that p-value is significantly high from 0.05. Therefore, this proves that the null hypothesis is true which means that the alternate hypothesis is false. That is, there is a relation between road accidents and red weather warnings.

**Conclusion :-** From this hypothesis test, I am concluding that the red weather warnings is having relation with road accident. By this result my initial null hypothesis true and I will continue my prediction and further findings with alternative hypothesis.

GitHub Link :- <https://github.com/prateekparasher/web_scrap>