COMP 3850 Final Report

Contributors

|  |  |
| --- | --- |
| Max Grant | 44910118 |
| Matthew Oxley | 44991738 |
| Alexandra Ellison | 45888949 |
| Tajkinur Rahman | 46124225 |
| Afeef Ahmed | 45833814 |

Contents

[Data Results 4](#_Toc118728298)

[NIST Results of ALL Observations 4](#_Toc118728299)

[NIST Results of Every5th and Pulsar4 Tenths 5](#_Toc118728300)

[Every5th 5](#_Toc118728301)

[Pulsar4 10ths 6](#_Toc118728302)

[6](#_Toc118728303)

[RandTest Results 7](#_Toc118728304)

[Interpretation of Data results 7](#_Toc118728305)

[RandTest Results Interpretation 7](#_Toc118728306)

[Focusing on the primary 5 NIST Results 7](#_Toc118728307)

[Summary Tables of Results 9](#_Toc118728308)

[Pulsar 1 9](#_Toc118728309)

[Pulsar 2 9](#_Toc118728310)

[Pulsar 3 9](#_Toc118728311)

[Pulsar 4 10](#_Toc118728312)

[Pulsar 5 11](#_Toc118728313)

[Pulsar 6 11](#_Toc118728314)

[Final Complete analysis 11](#_Toc118728315)

[Complete Summary Table 12](#_Toc118728316)

[Recommendations 12](#_Toc118728317)

[Further Pulsar Research 12](#_Toc118728318)

[Other Avenues of Randomness 12](#_Toc118728319)

[Appendix (Individual Pulsar Results) 12](#_Toc118728320)

[Pulsar 1 13](#_Toc118728321)

[Pulsar 2 16](#_Toc118728322)

[Pulsar 3 19](#_Toc118728323)

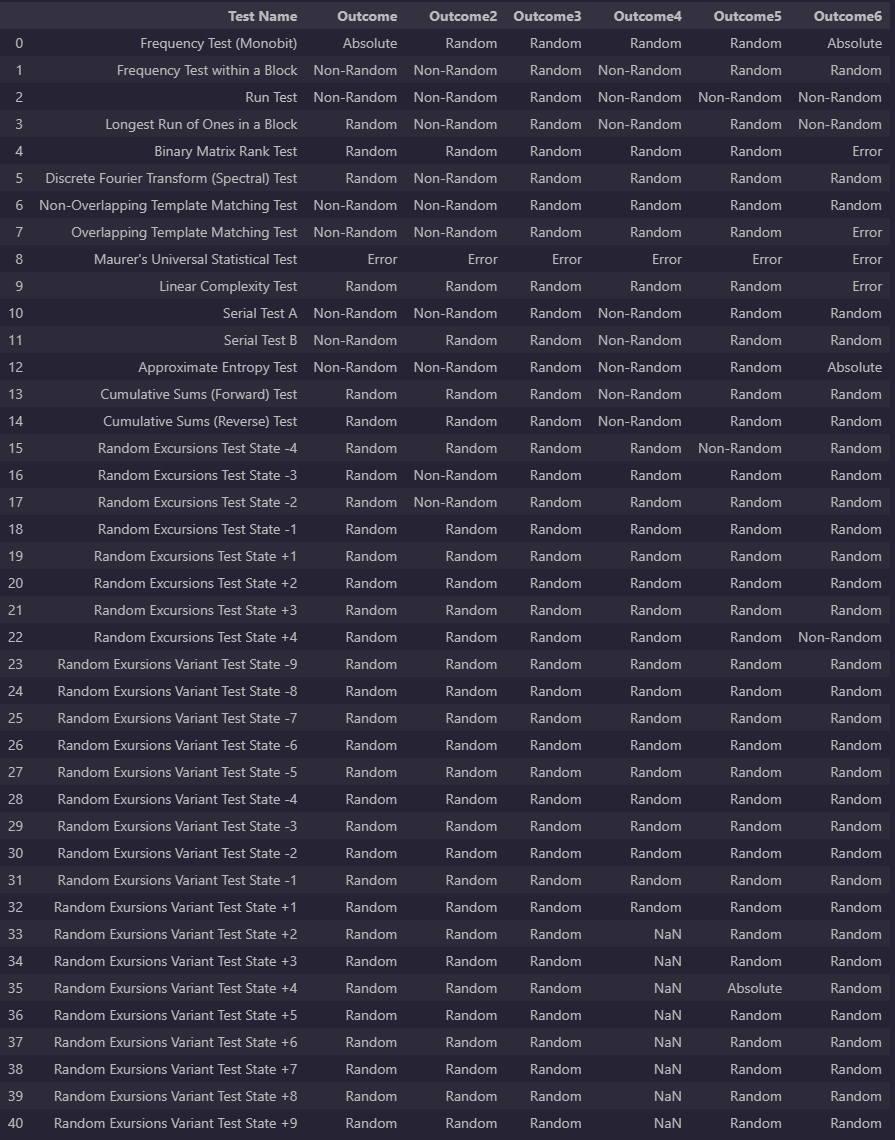
[Pulsar 4 22](#_Toc118728324)

[Pulsar 5 26](#_Toc118728325)

[Pulsar 6 29](#_Toc118728326)

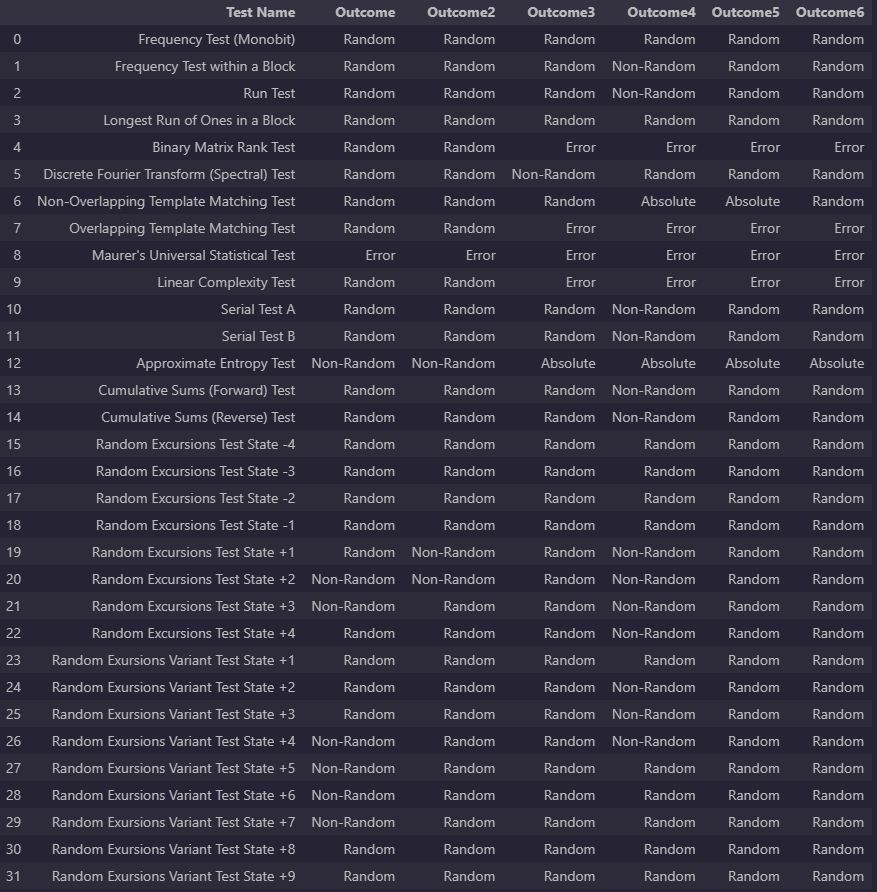
# Data Results

## NIST Results of ALL Observations



## NIST Results of Every5th and Pulsar4 Tenths

### Every5th



### Pulsar4 10ths

### Text Description automatically generated

## RandTest Results

A screenshot of a computer

Description automatically generated with medium confidence

# Interpretation of Data results

## RandTest Results Interpretation

The RandTest suite uses an algorithm called “exponentially-decaying moment prediction” to determine the net deviation between predicted and actual elements in sequences. It evaluated whether or not a sequence of numbers is random or not. It has a stated accuracy of 99.85% for nonrandom sequences and 96.82% for random sequences. This means the test will typically sustain a high degree of accuracy especially at a 0.05 cut off. But does falter at meeting the cyber security standard of 0.01. This means we will use this in conjunction with the overall statement from the NIST suite on the specific tests as extra evidence to prove or disprove randomness. In the case above we can clearly see that RandTest labels all binary sequences as random with no prediction this means we can confidently say with 96% certainty of this truth, this meets most standards of prediction. But fails to meet the industry standard of 99% accuracy with cybersecurity.

## Focusing on the primary 5 NIST Results

The first five tests are interpreted as the most important.

1. The Frequency (Monobit) Test is the first test in the NIST suite, and the remainder of the tests are dependent upon this test retaining the null hypothesis. The test determines the closeness of the fraction of 1’s to 1/2. For the test to pass, the number of 1’s and 0’s in the sequence should be similar. For the datasets containing every observation, the binary sequence has been created based on the median of the dataset, so by the nature of median (the middle value), the number of 0’s and 1’s will be equal (or different by 1 in the case of an odd number of observations), hence running the test on these sequences will return a p-value of 1 (or very close to 1 for odd number of observations), providing strong evidence to retain the null hypothesis, the expected outcome. Of more interest is the same test on the datasets of every 5th observation (pulsars 1, 2, 3, 4 ,5, 6) and the test on every 10th observation (pulsar 4), where there isn’t the same guarantee of an equal amount of 0’s and 1’s. Analyzing the results of each test, every number sequence passes. The lowest p-value observed is 0.096872 for pulsar 1 for every 5th observation, however it is still well above the 0.01 significance level. Now that each dataset has passed this test, further testing can be undertaken.

2. The next test, Frequency Test within a Block, applies a similar logic to the previous test, but instead tests the proportion of 1’s in a bit block of length M, expecting the frequency of 1’s to be close to M/2 for the test to pass. For the datasets with every observation, pulsars 1,2 and 4 fail this test (extremely small p-values). On the other hand, pulsars 3 and 6 provide extremely strong evidence (high p-value) that the fraction of 1’s with the M length bit blocks is close to 1/2. While pulsar 5 has provides moderate evidence, p-value = 0.0398, just over the alpha of 0.01, that there is an equal proportion of 1’s and 0’s. Looking to the datasets with every 5th observation, pulsars 1,2,3,5 and 6 pass with fairly strong evidence, while pulsar 4 fails. Then for pulsar 4 with every 10th observation it just passes with a p-value of 0.0216.

3. The Runs Test focuses on the number of runs in a sequence, where a run is the number of identical bits in a row. E.g., a run of length 25 is a sequence of 25 zeroes (or ones) in a row. The test determines if the number of runs is as would be expected in a random sequence. For datasets containing every observation, pulsar 1,2,4,5 and 6 fail to pass this test, while pulsar 3 passes. For the datasets on every 5th observation, there is a large shift in the number of sequences that pass, pulsar 1, 2, 3, 5 and 6 all pass, while only pulsar 4 fails. For every 10th observation on pulsar 4 the test also fails.

4. The Test for the Longest Run of Ones in a Block provides a similar approach as the previous Runs Test; however, it instead focuses on the longest run of ones within M-bit blocks and compares it with what would be expected of a random sequence. For the datasets containing all observations, pulsar 1, 3 and 5 passes pulsar 2, 4 and 6 fails. When we look at the datasets containing every 5th observation, all six pulsars pass. In addition, the test on every 10th observation of pulsar 4 also passes.

5. The fifth test is the Binary Matric Rank Test; this test breaks the binary sequence down into 32x32 matrices. Then it tests for linear dependence within the matrices. Given the nature of this test, the datasets need to be quite large for the test to be able to run, some of the datasets, especially those with every 5th or 10th observation may not have had enough data to run. If this situation occurs, the test will return a p-value of –1, which is an indicator that the test was unable to run, based on too little data. For pulsars with every observation included pulsars 1, 2, 3, 4 and 5 pass and pulsar 6 doesn’t have enough data to run (hence p-value = -1). For the datasets containing every 5th observation, 1 and 2 pass while pulsars 3, 4, 5 and 6 can’t run due to too little data

## Summary Tables of Results

### Pulsar 1

Graphical user interface, table

Description automatically generated with medium confidence

### Pulsar 2

A picture containing chart

Description automatically generated

## Pulsar 3

Graphical user interface, application, table

Description automatically generated

### Pulsar 4

Graphical user interface, table

Description automatically generated with medium confidence

Table

Description automatically generated

### Pulsar 5

Graphical user interface, application, table

Description automatically generated

### Pulsar 6

Graphical user interface, application

Description automatically generated

## Final Complete analysis

We can see here that most of the tests passed the first 2 most crucial tests except for Pulsar 4 which immediately disqualifies it from further evaluation. The same expectation can be said for the Pulsar binary datasets that do fail the Runs Test. However, Pulsar 5 is seen as a bit of an exception since it is the only pulsar that fails the Runs test but none of the other tests outside of the Random Excursions test at state -4. We pass 5 as random due to the sheer weighting of all the other evidence stacked against the runs test, this strongly indicates that the pulsar 5 should be more deeply investigate for a true determination. We can come to the summary conclusion that for all emissions, only Pulsar 3 is truly random, and Pulsar 5 is seen to be random pending further investigation. For Auto-Correlated observations at every 5th we can determine that the random pulsars are all of them except for 4 so this means Pulsars 1,2,3,5,6 at every 5th emission is random. At ACF 10th lags we find that the Pulsar 4 observations indeed become random, but due to the distance from the original emissions it is still seen as a non-random pulsar with some underlying structure that makes it uniform.

### Complete Summary Table

Table

Description automatically generated

# Recommendations

## Further Pulsar Research

There are many different approaches that would be beneficial in a secondary further study of the material covered in this project by Group 15. In list format we will explain.

* More Pulsars and more observations of pulsars to both categorize and eliminate structure of the pulsars
* Analysis with more testing methods such as TestU01 and Ent Utility.
* Application of the uncertainty variable to eliminate and false categorizations or conclusions

## Other Avenues of Randomness

There are many areas that immutable randomness for humans may exist some avenues of research could be found by examining these 3 primary locations or systems.

* Stellar surface with sunspots and/or Coronal Mass Ejections
* Black hole accretion disks with the collision of matter and their emissions
* Radioactive decay in the stochastic process on what atom decays not the rate of decay.

# Appendix (Individual Pulsar Results)

Order goes

All Observations -> Every5th -> (where applied) Every10th

## Pulsar 1

Text, calendar

Description automatically generated

Text

Description automatically generated

## Pulsar 2

Text, calendar

Description automatically generated

Graphical user interface, text

Description automatically generated

## Pulsar 3

Text, calendar

Description automatically generated

Text, calendar

Description automatically generated

## Pulsar 4

Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generated

Text

Description automatically generated

## Pulsar 5

Text, calendar

Description automatically generated

Text, calendar

Description automatically generated

## Pulsar 6

Text, calendar

Description automatically generated 