Feedback — Assignment 5

Help

You submitted this quiz on **Fri 22 Nov 2013 10:48 AM PST**. You got a score of **9.00** out of **9.00**.

In this assignment, you need to apply a set of metrics to two data sets, Asgn5-posthoc-data-v2-set1.csv and Asgn5-posthoc-data-v2-set2.csv

These are not real data sets, but are simulated data, generated for the purpose of this assignment.

You can use any tool you want to complete this assignment. This includes Microsoft Excel, and statistical or data mining package you choose, and other tools available on the internet.

The goal of this assignment is to conduct post-hoc methods to determine which of a large set of correlations can be trusted – and in the process to understand if these post-hoc methods are useful in educational data mining.

Question 1

Data set 1 represents a set of distinct studies conducted on small populations of students – you might see these types of results if you administered a survey to the students of just one classroom teacher. Within data set 1, how many correlations are statistically significant (according to the customary p

You entered:

10

Your Answer		Score	Explanation
10	✔	1.00	
Total		1.00 / 1.00	

Question 2

If you apply a post-hoc Bonferroni control to these results, how many correlations remain

statistically significant?

You entered:

5

Your Answer		Score	Explanation
5	✔	1.00	
Total		1.00 / 1.00	

Question 3

If you apply Benjamini & Hochberg's FDR Correction to these results, how many correlations remain statistically significant?

You entered:

9

Your Answer		Score	Explanation
9	✔	1.00	
Total		1.00 / 1.00	

Question 4

What is the correlation with the lowest p-value that comes up significant for B&H but not for Bonferroni?

Your Answer		Score	Explanation
0.288032855			
0.237749612	✔	1.00	
0.245614914			
0.231753705			

Total	1.00 / 1.00

Question 5

Data set 2 represents a larger set of correlations within data from a larger population of students – for example, the entire population of students using a medium-sized online learning environment. Within data set 2, how many of the 1,112 correlations are NOT statistically significant (according to the customary p

You entered:

19

Your Answer		Score	Explanation
19	✔	1.00	
Total		1.00 / 1.00	

Question 6

If you apply a post-hoc Bonferroni control to these results, how many correlations are now NOT statistically significant?

You entered:

28

Your Answer		Score	Explanation
28	✔	1.00	
Total		1.00 / 1.00	

Question 7

If you apply Benjamini & Hochberg's FDR Correction to these results, how many correlations are now NOT statistically significant?

You entered:

19

Your Answer		Score	Explanation
19	✔	1.00	
Total		1.00 / 1.00	

Question 8

What is the lowest correlation that is still statistically significant, according to Bonferroni's test?

Your Answer		Score	Explanation
○ 0.05965262			
○ 0.1976639			
○ 0.26227631			
◉ 0.01609085	✔	1.00	
Total		1.00 / 1.00	

Question 9

Now do you see why Professor Baker says that statistical significance doesn't matter much for really big data sets? (and this is NOT a big data set by reckoning in other fields)

Your Answer		Score	Explanation
○ No. I think the answer to Question 8 is a fine correlation, perfectly likely to represent a large effect size.			
◉ Yes – Bonferroni is ridiculously conservative with 1,112 tests, and yet correlations that are absurdly small still come up statistically significant.	✔	1.00	

No. Big data is a FAD. No one should ever expect to work with a data set over 150 data points, after the societal collapse predicted by James Howard Kunstler occurs.

Total	1.00 / 1.00