AQ3.3: Activity Question 3 - Not graded

**This assignment will not be graded and is only for practice.**

**Note : This activity is for your practice purpose only. Your score in this will not count towards the Qualifier Process.**

***1 point***

Select all statements that form a part of the hypothesis that was discussed in this lecture. It is a Multiple Select Question (MSQ).

High frequency words are short

Low Frequency words are short

Low frequency words are long

High frequency words are long

***1 point***

What are the criteria used in the lecture to decide if a given word is short or long? It is a Multiple Select Question (MSQ).

A word is long if it takes a long time to type

A word is short if can be read in under half a second

A word is long if its *letter count* is six or above

A word is short if it has at most five characters

Consider the following table that has been generated from a paragraph with 100 words:  
  
 A close-up of a white background

AI-generated content may be incorrect.

***1 point***

What percentage of the high frequency words confirm the hypothesis that high frequency words are short?

75

25

60

40

***1 point***

What percentage of the low frequency words reject the hypothesis that low frequency words are long?

90

10

60

40

Consider the following pseudocode to find the average word length for the “Paragraph words” dataset:  
  
 A white screen with black text

AI-generated content may be incorrect.

***1 point***

At any intermediate stage in the execution of the above code, what does the variable **avgI** store?

It stores the average length of all words in the paragraph

It stores the average length of all words that have been moved to Pile-2

It stores the average length of all words that are still remaining in Pile-1

It does not store anything meaningful

***1 point***

After the code has been executed, which of the following expressions concerning the relationship between the variables **avgI** and **avgF** is True?

**avgI** > **avgF**

**avgI** < **avgF**

**avgI** == **avgF**

***1 point***

We plan to have a three way classification for both frequency and length. There is now a medium frequency in addition to high and low; also, there is a medium length in addition to long and short. Given this setup, how many variables would we need to maintain to keep track of the word-counts for the different combinations that would come up?

4

6

8

9