MATH 105 Extra Credit Project- Decoding Using Letter Frequencies

DUE: December 10, 2007 by 12:00 pm

Part I-Decoding English text.

On the next page you are given a coded English text. The text was coded by mixing up the letters in the alphabet and creating a one-to-one correspondence between the usually alphabet and the new "mixed-up" version.

For example, a possible correspondence could be the following:

Old	Α	В	С	D	Е	F	G	н	I	J	Κ	L	М	Ν	0	Ρ	Q	R	S	Т	U	V	W	X	Y	Ζ
New	Н	S	R	V	0	Ζ	Т	U	P	W	Ι	Q	Y	L	А	F	Ν	В	G	Х	С	D	K	М	Ε	J

According to the table above the phrase "he was on top of the hill" would be coded as "uo khg al xaf az xuo upqq."

Your task is to use the letter frequencies of the English language to help you create the table that will decode the coded text in the following page.

The project will be graded according to the following criteria:

- 1) Correct alphabet table.
- 2) Correctly decoded text.
- 3) Clear and organized presentation.

To gain full points for the presentation, you must include frequency histograms, and explanations of what you are doing. Full credit will not be awarded if such explanations are not included or if the presentation is not clean. If you have any questions about this, please ask your instructor.

For your convenience, you are given a sample of the English language consisting of two news articles taken from the New York Times on November 11, 2007. These articles, and the coded text can be found in electronic form at:

www.math.cornell.edu/~web105

Part II-Identifying the language of a coded text.

In this part of the project you are given a coded text, but you do not know what language the original text was in. You only know that it can be one of the following three choices: *Italian, Hungarian, or German.*

By using frequency histograms you are expect to determine what language the original message was in. NOTE: you are not expected to come up with an alphabet table as in part 1.

The project will be graded according to the following criteria:

- 1) Correctly identifying the coded language.
- 2) Clear and organized presentation.

To gain full points for the presentation, you must include frequency histograms, and explanations of what you are doing. Full credit will not be awarded if such explanations are not included or if the presentation is not clean. If you have any questions about this, please ask your instructor.

For your convenience, you are given a sample of the Italian, Hungarian and German languages consisting of news articles taken from various sources on November 11, 2007. These articles, and the coded text can be found in electronic form at:

www.math.cornell.edu/~web105

Hints:

In order to count the letter frequencies one can use Microsoft Word. For example, to count the number of appearances of the letter "a" go to the Edit menu and select Replace. A window will pop up. In the section where says "Find what:" put "a". In the section where it says "Replace with:" you may put anything (although it might be easier to insert an uncommon symbol such as # .) When you hit the button "Replace all" a window will pop up letting you know how many replacements were made, this number will be the number of the letter "a" in the text.

Once you have collected all the data on the frequency of the letters it is highly recommended that you use Microsoft Excel (or any similar program) to help you make the frequency distribution charts. Although this is not necessary, it will make your analysis much simpler and cleaner.

If you have any questions about how to use any of these programs please contact your instructor.

Grading:

If you chose to do this project, it will be graded as another exam (100 points), and your total points for the course will be out of 600 points and not 500. This extra credit project is only meant to help your grade in the class, so if you do not get enough points on the extra credit to improve your average the project will be dropped and your grade will be calculated out of 500 points.

You are expected to work on your own; projects which are noticeably the same will not be awarded full points. The amount of points taken off will be at the grader's discretion.

Good Luck and have fun!