

# 2008 Excellence in Mathematics Contest Team Project A



CHANDLER-GILBERT COMMUNITY COLLEGE



School Name:

---

Group Members:

---

---

---

---

---

---

---

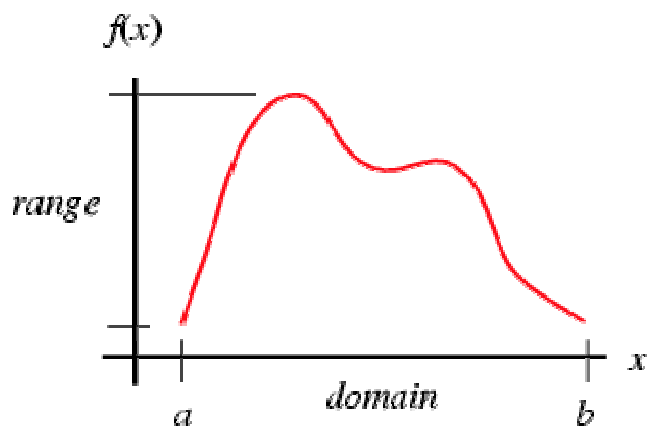
# Reference Sheet

**Relative Frequency** is the ratio of the absolute frequency to the total number of data points in a frequency distribution.

For example, if the Arizona Diamondbacks win 96 out of 162 games, we say that they won  $\frac{96}{162}$  or, equivalently,  $\frac{16}{27}$  of their games. We can also say that the relative frequency is  $\frac{16}{27} \approx 0.59$  or that the Diamondbacks won 59% of their games.

**Domain** is a term is used to describe the set of values for which a function is defined.

**Range** is a term used to describe the set of values that a particular function can take on over a given domain.



**Coefficient of Determination**,  $r^2$  is a value that describes the strength of fit of a linear regression model to a set of data. The closer the value of  $r^2$  is to 1, the stronger the fit.

In **statistics**, the **residual** is the observed value minus the predicted value. The bigger the residual, the poorer the model used to make the prediction.



## Scrabble and Mathematical Modeling

*How many of each letter should there be and how much should each letter be worth?*

*Adapted from Richardson & Gabrosek (2004 NCTM "Mathematics Teacher" Journal)*

Eleven students randomly chose a starting point in a newspaper article, internet resource, or book and counted out the next 300 letters. They tallied the number of each of the letters found in their particular selection. The 11 students then pooled the results together. The results are shown in the table and will be used in Activity I.

Table 1

Letter	Total
A	434
B	133
C	175
D	169
E	563
F	147
G	192
H	292
I	307
J	7
K	63
L	186
M	146
N	300
O	326
P	118
Q	35
R	266
S	307
T	417
U	158
V	81
W	102
X	20
Y	103
Z	26
<b>Total</b>	<b>5073</b>

## Background

During the Great Depression, an out-of-work architect named Alfred Mosher Butts decided to invent a board game. He did some market research and concluded that games fall into three categories: number games, such as dice and bingo; move games, such as chess and checkers; and word games, such as anagrams.

Butts wanted to create a game that combined the vocabulary skills of crossword puzzles and anagrams, with the additional element of chance. The game was originally named Lexico, but Butts eventually decided to call the game "Criss-Cross Words."

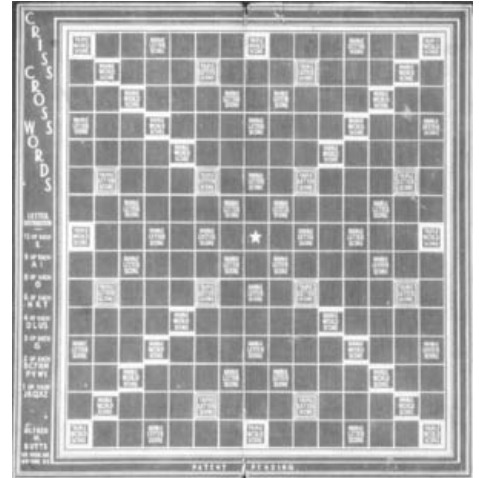


### How did he do it?

Butts studied the front page of The New York Times to calculate how often each of the 26 letters of the English language was used. He discovered that vowels appear far more often than consonants, with *E* being the most frequently used vowel. After figuring out frequency of use, Butts assigned different point values to each letter and decided how many of each letter would be included in the game. The letter *S* posed a problem. While it's frequently used, Butts decided to include only four *S*'s in the game, hoping to limit the use of plurals. After all, he didn't want the game to be too easy!

Butts got it just right. His basic cryptographic analysis of our language and his original tile distribution have remained valid for almost three generations and for billions of games played.

The boards for the first Criss-Cross Words game were hand drawn with his architectural drafting equipment, reproduced by blueprinting and pasted on folding checkerboards. The tiles were similarly hand-lettered, then glued to quarter-inch balsa and cut to match the squares on the board.



### Purpose of Activity 1

The purpose of this activity is to examine the relationship between a letter's relative frequency in English and the percent of Scrabble tiles for the letter.

Use the following table to record the frequency for the letters of the alphabet in English text using the data from Table 1 on Page 3. The actual percentage of Scrabble tiles containing the letters are recorded except for the letter *L* and *W*. We wish to use this information to examine the relationship between a letter's relative frequency and its percent of Scrabble tiles.

# *Percent of Scrabble Tiles*

Table 2

Letter	Relative Frequency from Table 1	Percent of Scrabble Tiles	Letter	Relative Frequency from Table 1	Percent of Scrabble Tiles	Letter	Relative Frequency from Table 1	Percent of Scrabble Tiles
A		9.18	J		1.02	S		4.08
B		2.04	K		1.02	T		6.12
C		2.04	L		?	U		4.08
D		4.08	M		2.04	V		2.04
E		12.24	N		6.12	W		?
F		2.04	O		8.16	X		1.02
G		3.06	P		2.04	Y		2.04
H		2.04	Q		1.02	Z		1.02
I		9.18	R		6.12			

1. Is the association between the percent of scrabble tiles and the relative frequency of a letter positive or negative? Why?

2. Using your graphing calculator or Excel, make a scatterplot with the percent of Scrabble tiles for each letter on the vertical axis and the relative frequency of the letter on the horizontal axis. **Provide a printout of the graph when you turn in this project.**



7. Plot the regression line on your scatterplot.

8. Interpret the slope of the regression line in the context of this problem.

9. Use your regression line to predict the percent of Scrabble tiles for the letters *L* and *W*. Complete the following table.

Table 3

Letter	Relative Frequency in English Text	Actual Percent of Scrabble Tiles	Predicted Percent of Scrabble Tiles	Residual
<i>L</i>		4.08		
<i>W</i>		2.01		

- 10.** In Scrabble, there are some blank tiles used as “wild cards”. That is, they can represent any letter that the player chooses. According to your regression line model, what percent of the Scrabble tiles should be blank? Explain.
- 11.** Does question 10 provide any information that could be used in describing the domain and range of your regression line model? Explain.



## Purpose of Activity 2

The purpose of this activity is to examine the relationship between a letter's relative frequency in English text and the Scrabble-tile point value of the letter.

In the following table, record the class frequency for the letters of the alphabet in English text. The corresponding Scrabble-tile point values are recorded, except for the letters *C* and *N*.

### *Scrabble-Tile Point Values*

Table 4

Letter	Relative Frequency from Table 1	Scrabble Tile Points	Letter	Relative Frequency from Table 1	Scrabble Tile Points	Letter	Relative Frequency from Table 1	Scrabble Tile Points
A		1	J		8	S		1
B		3	K		5	T		1
C		?	L		1	U		1
D		2	M		3	V		4
E		1	N		?	W		4
F		4	O		1	X		8
G		2	P		3	Y		4
H		4	Q		10	Z		10
I		1	R		1			

1. Do you think that the association between the Scrabble-tile point value and the letter's relative frequency will be positive or negative? Why?
  
2. Using your graphing calculator or Excel, make a scatterplot with the Scrabble-tile point value for each letter on the vertical axis and the relative frequency of the letter on the horizontal axis. ***Provide a printout of the graph when you turn in this project.***



8. Use your calculator to fit the exponential model  $y = ab^x$  to the data. Plot your fitted-model equation on the scatterplot in question 2.

9. Choose the best model (linear, quadratic, cubic, or exponential) and explain why you think it is the best. Using this model, predict the Scrabble-tile point value for the letters C and N. Complete the following table.

Table 5

Letter	Relative Frequency in English Text	Actual Scrabble-Tile Points	Predicted Scrabble-Tiles Points	Residual
<i>C</i>		3		
<i>N</i>		1		

**10.** Use the model you chose in question 9 to determine how many points the blank, “wild card” tiles should be worth. How does this compare to the actual game where the blank tiles are worth 0 points? Explain.

**11.** Discuss the domain and range for your chosen model.

**Scrabble - TEAM PROJECT A**  
**2008 Excellence in Mathematics Contest**

---

*The Team Project is a group activity in which the students are presented a series of mathematical problems relating to a specific theme. The team members are to solve the problems and write a narrative about the theme which answers all the mathematical questions posed. Teams are graded on accuracy of mathematical content, clarity of explanations, and creativity in their narrative.*

---

## Scoring Sheet

The Team Project is a group activity in which the students are presented a series of mathematical problems relating to a specific theme. The team members are to solve the problems and write a narrative about the theme which answers all the mathematical questions posed. Teams are graded on accuracy of mathematical content, clarity of explanations, and creativity in their narrative. A holistic scoring approach should be used to judge the team project. For each project, assign a score to each of the major areas:

School Name: \_\_\_\_\_ -

Accuracy of mathematical content:      0            1            2            3            4  
Comments:

Clarity of Explanations:                      0            1            2            3            4  
Comments:

Creativity in Narrative:                      0            1            2            3            4  
Comments:

Overall Presentation:                      0            1            2            3            4  
Comments:

