

```

Book: pride.txt
Number of lines      = 13427
Number of characters = 704192
Number of vowels     = 209731
Number of consonants = 341984
Number of letters    = 551715
% vowels              = 38.01%

```

```

Book: swann.txt
Number of lines      = 16581
Number of characters = 1012445
Number of vowels     = 363531
Number of consonants = 438475
Number of letters    = 802006
% vowels              = 45.33%

```

Actual:

Book	Consonants	Vowels
pride.txt	341984	209731
swann.txt	438475	363531

Expected:

Book	Consonants	Vowels
pride.txt	318079.53	233635.47
swann.txt	462379.47	339626.53

chi-square = 7160.61

The null hypothesis is that the text in the two books is drawn from the same population.

Chi-square is 7160.61 with $df = 1$. That is above the cutoff for $p < .001$, which is 10.828. Therefore the null hypothesis is rejected, and there is a significant difference in the percentage of vowels in the two texts.

The short way of writing this is:

Chi-square = 7160.61, $df=1$, which is significant at $p < .001$

However you write it, that means that there is only 1 chance in 1000 that this result happened by chance, i.e., by accident.

A 5% level, i.e., that there is only 1 chance in 20 that the result appears to be real but really isn't, is sufficient for publication in this type of work. That's what I suggested for the homework, and that's fine.

For publication, if you have a .01 or .001 level of significance, you would usually use that.