



## Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship

Claude S. Brinegar

*Journal of the American Statistical Association*, Vol. 58, No. 301. (Mar., 1963), pp. 85-96.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28196303%2958%3A301%3C85%3AMTATQC%3E2.0.CO%3B2-V>

*Journal of the American Statistical Association* is currently published by American Statistical Association.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# MARK TWAIN AND THE QUINTUS CURTIUS SNODGRASS LETTERS: A STATISTICAL TEST OF AUTHORSHIP

CLAUDE S. BRINEGAR  
Pasadena, California

Mark Twain is widely credited with the authorship of 10 letters published in 1861 in the *New Orleans Daily Crescent*. The adventures described in these letters, which were signed "Quintus Curtius Snodgrass," provide the historical basis of a main part of Twain's presumed role in the Civil War. This study applies an old, though little used statistical test of authorship—a word-length frequency test—to show that Twain almost certainly did not write these 10 letters. The statistical analysis includes a visual comparison of several word-length frequency distributions and applications of the  $\chi^2$  and two-sample *t* tests.

## 1. INTRODUCTION AND SUMMARY

MARK TWAIN'S role in the Civil War has long been a subject of dispute—a dispute that has at times flared even beyond the realm of scholarly debate.<sup>1</sup> In 1940, for example, Twain was accused on the floor of the House of Representatives of being a Confederate deserter, and was defended, in turn, in the editorial pages of the *New York Times*, as a young man of mixed emotions and loyalties.<sup>2</sup>

Twain's own public comments on the matter were confined to a few humorous remarks 20 or so years after the fact and to the partly fictional, partly defensive "The Private History of a Campaign that Failed."<sup>3</sup> A. B. Paine likewise failed to contribute much in the way of solid fact in *Mark Twain, A Biography*,<sup>4</sup> for it is evident that Paine was strongly influenced by "The Private History" and by Twain's personal reminiscences from 1906 until his death in 1910. Although these reminiscences are enormously valuable as literary source material, they should not be mistaken for history. Twain was not unaware of this, for as he noted at one point: "When I was younger I could remember anything, whether it happened or not; but now I am getting old and soon I shall remember only the latter."<sup>5</sup>

Essentially all that we know for sure of Twain's Civil War experiences was recently stated in four sentences by Edward Wagenknecht:

Mark Twain's connection with the Civil War was both brief and loose. He may have had some military connection in New Orleans; if so, we know none of the details. Nor do we know a great deal more about his adventures with an informally organized Confederate group in Missouri, for "The Private History of a Campaign that Failed" mingles fact with fiction. Nevertheless the fact remains that Lieutenant Clemens shortly mustered himself out of Confederate service and went to Nevada with his

<sup>1</sup> Good summaries can be found in Fred W. Loreh, "Mark Twain and the Campaign that Failed," *American Literature*, XII (January, 1941), 254-270, and John Gerber, "Mark Twain's 'Private Campaign,'" *Civil War History*, I (March, 1955), 37-60.

<sup>2</sup> *Congressional Record*, Vol. 86, Part 1 (January 25, 1940), p. 698; *New York Times* (February 7, 1940), p. 20.

<sup>3</sup> First published in *Century Magazine*, December, 1885; also included in *Merry Tales*, New York: Harper & Brothers, 1892, pp. 9-50.

<sup>4</sup> A. B. Paine, *Mark Twain, A Biography*, New York: Harper & Brothers, 1912, 2 Volumes. See especially Volume I, Chapter XXX.

<sup>5</sup> Mark Twain, *Autobiography*, New York: Harper & Brothers, 1924, p. xii of Paine's introduction.

brother Orion, a staunch Union man, whom President Lincoln had appointed Secretary of the Territory.<sup>6</sup>

This study is concerned only with the evidence regarding Twain's possible military connection in New Orleans. The evidence is drawn entirely from the content of 10 letters published in the *New Orleans Daily Crescent* in early 1861. In these letters, which have been widely credited to Twain, the writer described his military adventures while serving as a "High Old Private of Louisiana Guard." The letters were signed "Quintus Curtius Snodgrass." Based on the results of applying a statistical test of authorship both to the 10 Quintus Curtius Snodgrass letters and to known contemporary Twain writings, we conclude that Twain was *not* the author of the disputed letters. It follows, consequently, that the evidence supporting the claim that Twain served with a New Orleans Confederate garrison must likewise be discounted.

## 2. QUINTUS CURTIUS SNODGRASS LETTERS

The Quintus Curtius Snodgrass (QCS) letters were first brought to light in 1934 by Miss Minnie M. Brashear, in *Mark Twain, Son of Missouri*.<sup>7</sup> She reprinted one letter and described three others, and then commented:

None of the *Crescent* letters suggests the later Mark Twain as satisfactorily as one wishes they did. The peculiar type of exaggeration, anti-climax, and irony which gave flavor to his sentences is discoverable in the earlier Snodgrass letters,<sup>8</sup> but not here—though the satire at the expense of the military profession was often a theme during the earlier period of American humor. And yet, if it can be proved that they are authentic, they are immensely significant as a link in Mark Twain's development as a humorist.<sup>9</sup>

Miss Brashear then argues that they should be recognized as Twain's, and that the difference in style reflects the fledgling writer's efforts to achieve a more consciously literary manner.

The remaining six QCS letters were uncovered by Ernest E. Leisy and published, together with the first four, in 1946.<sup>10</sup> Leisy's careful study of parallels provides the strongest case for admitting the QCS letters to the Mark Twain canon. Leisy also emphasized the literary significance of the discovery:

I believe that scholars and collectors will rejoice to have these letters reprinted, because if they are by Mark Twain they throw new light on the young soldier and developing writer. . . .<sup>11</sup>

Biographers and literary researchers have reacted in various ways toward the QCS letters.

Lorch, in his 1941 study of "The Private History," relegated the possibility

<sup>6</sup> Edward Wagenknecht, *Mark Twain: The Man and His Work*. Norman, Oklahoma: University of Oklahoma Press, New and Revised Edition, 1961, p. 11.

<sup>7</sup> Minnie M. Brashear, *Mark Twain, Son of Missouri*. Chapel Hill, North Carolina: University of North Carolina Press, 1934.

<sup>8</sup> This reference is to three letters Twain wrote in 1857 for the Keokuk (Iowa) *Saturday Post*, which he signed "Thomas Jefferson Snodgrass." This earlier use of "Snodgrass" is one of the main clues that led to attributing the Quintus Curtius Snodgrass letters to Twain. See Mark Twain, *The Adventures of Thomas Jefferson Snodgrass*, Chicago: Pascal Covici, Publisher, Inc., 1928, edited by Charles Honce.

<sup>9</sup> Brashear, *op. cit.*, pp. 191-2.

<sup>10</sup> Mark Twain, *The Letters of Quintus Curtius Snodgrass*. Dallas, Texas: Southern Methodist University Press, 1946, edited by Ernest E. Leisy.

<sup>11</sup> *Ibid.*, p. v.

that Twain was the author to a questioning footnote.<sup>12</sup> De Lancey Ferguson, on the other hand, writing in 1943 concluded: "There is little reason to doubt that they [the QCS letters] were the work of Sam Clemens."<sup>13</sup> This belief was echoed by Gerber in 1955.<sup>14</sup> However, in 1961 Wagenknecht noted that Twain's proposed authorship was still only a "conjecture."<sup>15</sup>

It is clear that considerable doubt remains about Mark Twain's hypothesized authorship of the QCS letters.

### 3. STATISTICAL TEST OF AUTHORSHIP

A related problem, though one on a far grander scale, is the long time controversy about the authorship of Shakespeare's works.<sup>16</sup> In 1955 Calvin Hoffman, in *The Murder of the Man Who Was "Shakespeare,"*<sup>17</sup> called attention to two statistical studies of authorship by T. C. Mendenhall, published in 1887 and 1901,<sup>18</sup> that had apparently gone unnoticed by modern statisticians. Subsequent to Hoffman's discussion, Mendenhall's work was reviewed and several of his charts redrawn by C. B. Williams in *Biometrika*.<sup>19</sup> Mendenhall's method forms the basis of the statistical test we here apply to the QCS letters.

The method, although laborious, is simple in concept and application: By forming a frequency distribution of a great many words classified according to their length, it is possible to obtain a graphic representation (called a "characteristic curve of composition" by Mendenhall) that would "become a means of identification, at least by exclusion."<sup>20</sup> The use of this method assumes that every author unconsciously uses words that, at least in the long run, could be considered as random drawings from a fixed frequency distribution of word lengths.<sup>21</sup> (This assumption stands up surprisingly well, as subsequently shown.) Although it is conceivable that every author's word frequency distribution is as unique as his fingerprints, it is (1) often difficult to obtain a large enough sample (Mendenhall suggested that at least 100,000 words would be needed—although this is probably far too large) and (2) even with such a large sample the statistical tests that can be used to distinguish between frequency distributions that closely resemble each other are far less sensitive than fingerprint tests. Probably a better analogy is to consider the test as similar to the blood test in a paternity case—a test that by itself can exonerate but cannot convict. If the large-sample word frequency distributions for two sets of writings differ greatly, this should be considered as strong evidence that the two

<sup>12</sup> Lorch, *op. cit.*, p. 462n.

<sup>13</sup> De Lancey Ferguson, *Mark Twain: Man and Legend*. New York: The Bobbs-Merrill Company, 1943, p. 59.

<sup>14</sup> Gerber, *op. cit.*, p. 37n.

<sup>15</sup> Wagenknecht, *op. cit.*, p. 231.

<sup>16</sup> In his late years Twain himself developed considerable interest in this controversy. See Mark Twain, *Is Shakespeare Dead?* New York: Harper & Brothers, 1909.

<sup>17</sup> Calvin Hoffman, *The Murder of the Man Who Was "Shakespeare,"* New York: Julian Messner, Inc., 1955. Reprinted in 1960 in paperback by Grosset and Dunlap. Hoffman's nominee for authorship is Christopher Marlowe.

<sup>18</sup> T. C. Mendenhall, "The Characteristic Curve of Composition," *Science*, IX (March 11, 1887), 237-49; "A Mechanical Solution of a Literary Problem," *The Popular Science Monthly*, LX (December, 1901), 97-105.

<sup>19</sup> C. B. Williams, "Studies in the History of Probability and Statistics," *Biometrika*, 43 (1956), 248-56. Figures 4 and 5, showing the "characteristic curve" for Shakespeare, contain separate plotting errors.

<sup>20</sup> Mendenhall, *Popular Science Monthly*, p. 97.

<sup>21</sup> This should be true at least for writings of a related type over a reasonable span of years. Consciously different styles (for example, fiction heavily laden with dialect conversation as contrasted with the description of a trip to Europe) might be expected to show statistically significant variation.

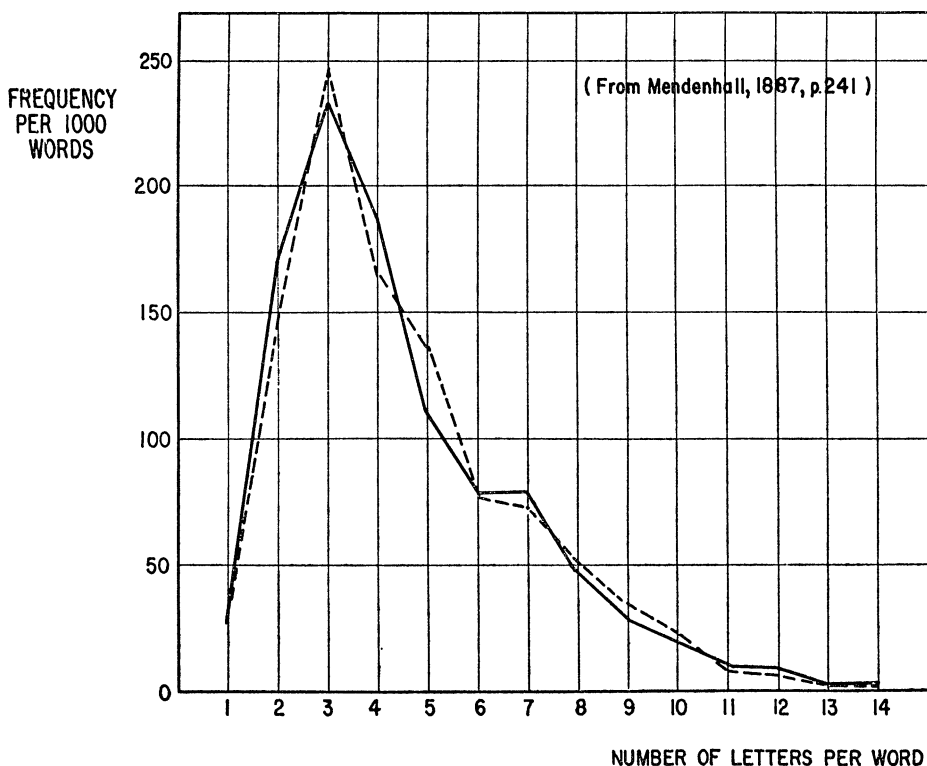


CHART 1. Word frequencies for two samples of 1000 words from "Vanity Fair."

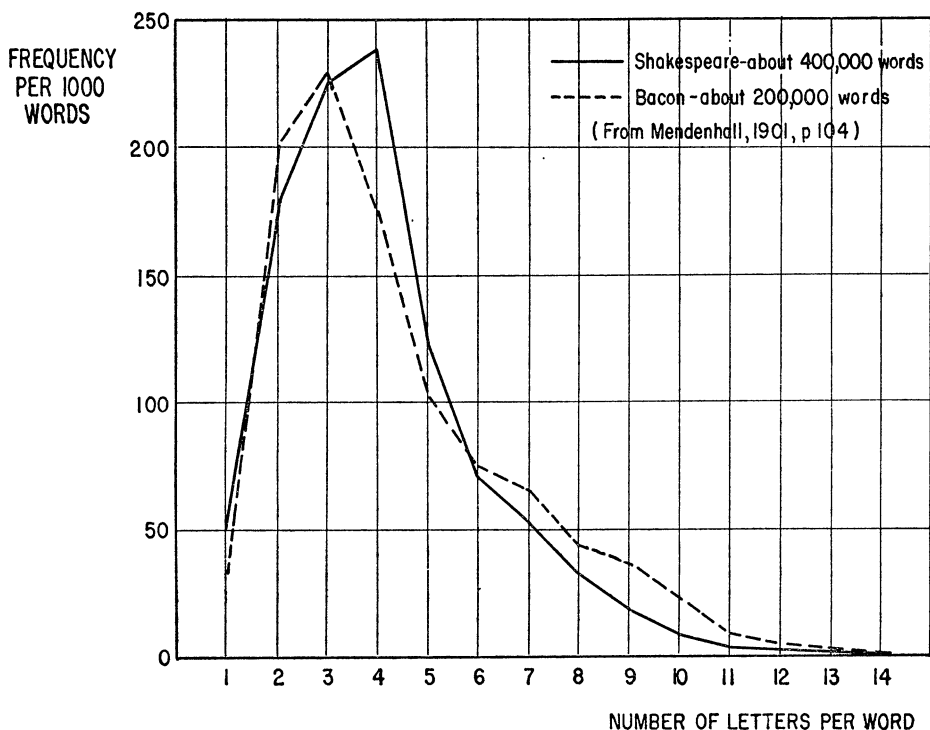


CHART 2. Estimated word frequencies for large samples from works of Shakespeare and Bacon.

sets were not written by the same person; if the distributions are quite similar, this supports but does not prove the hypothesis that they could have been the work of the same person.

Chart 1, redrawn from Mendenhall (1887), illustrates the method by plotting the distributions for two small samples from Thackeray's *Vanity Fair*. Mendenhall's study contains numerous such examples. Near the end of the 1887 study he commented:

Many interesting applications of the process will suggest themselves to every reader; the most notable, of course, being the attempt to solve questions of disputed authorship, such as exist in reference to the letters of Junius, the plays of Shakespeare, and other less widely known examples.<sup>22</sup>

Mendenhall's 1901 study was largely devoted to a report on the work in the intervening years, during which, thanks "to the liberality of Mr. Augustus Heminway, of Boston," a staff—two ladies from Worcester, Massachusetts—had been paid to "count and classify nearly two million words." Mr. Heminway was evidently seeking to resolve the Bacon-Shakespeare controversy once and for all. The results of counting 400,000 words from Shakespeare and 200,000 words from Bacon are shown in Chart 2. These two curves were not encouraging to the Baconians.

#### 4. APPLICATION OF TEST TO THE QCS LETTERS

Applying Mendenhall's method to the QCS letters required that we initially determine Mark Twain's frequency distribution before and after 1861, and also if it was consistent within itself. For this purpose the following word counts were made from writings that are indisputably his:

1. A letter, known as the "Sergeant Fathom" letter, published in the New Orleans *True Delta* in 1859, in which Twain burlesqued an old-time river pilot (Captain Isaiah Sellers); a letter to his brother Orion in early 1861 describing his visit to a New Orleans fortune-teller, known as Madame Caprell. These two letters total about 1,900 words.<sup>23</sup>
2. The first four letters in which he is known to have used the name "Mark Twain," published in 1863 in the *Territorial Enterprise*.<sup>24</sup> These letters total about 6,100 words.
3. The first letter to the *Alta California*, published in 1867, describing his adventures with the "Innocents Abroad."<sup>25</sup> This letter, covering the visit to the Azores, contains about 3,000 words.

These three items—coming before, during, and after the time of the QCS letters—form the control group for the test. Although 11,000 words falls far short of Mendenhall's standard, the consistency of the three suggests that a sample of this size can be satisfactory.

The counting was done by the tally method, with the count limited to text

<sup>22</sup> Mendenhall, *Science*, p. 245.

<sup>23</sup> Both are included in Paine, *Mark Twain, A Biography*.

<sup>24</sup> Collected by Henry Nash Smith in *Mark Twain of the Enterprise*. Berkeley, California: University of California Press, 1957.

<sup>25</sup> Collected by Daniel Morley McKeithan in *Traveling with the Innocents Abroad*. Norman, Oklahoma: University of Oklahoma Press, 1958. These letters were the raw material for Twain's first big success, *Innocents Abroad*.

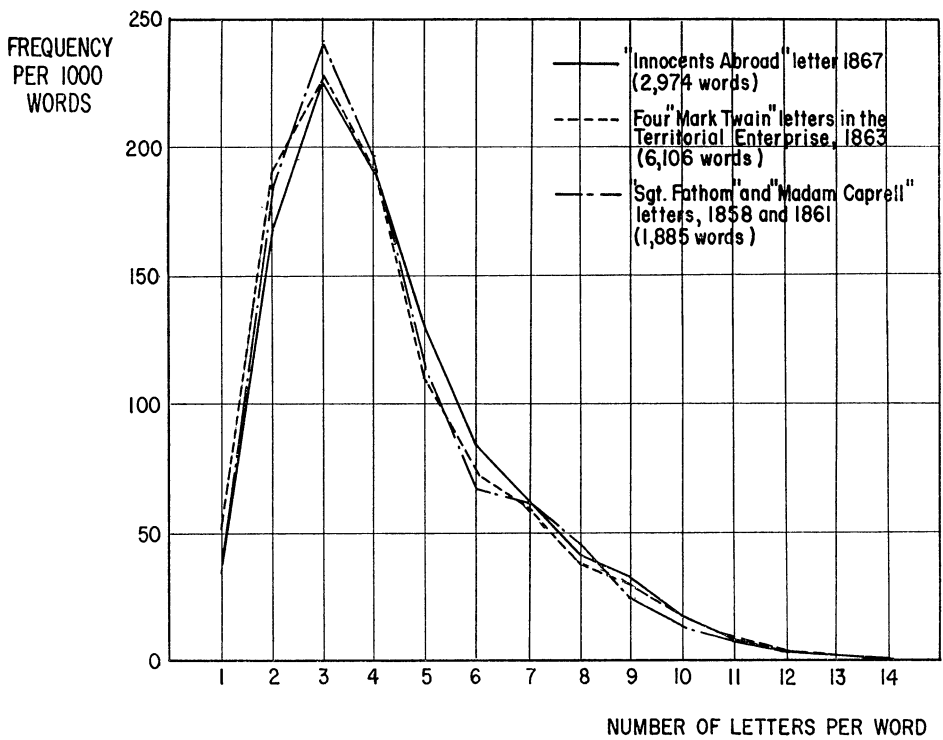


CHART 3. Word frequencies for known Mark Twain Writings, 1858-1867.

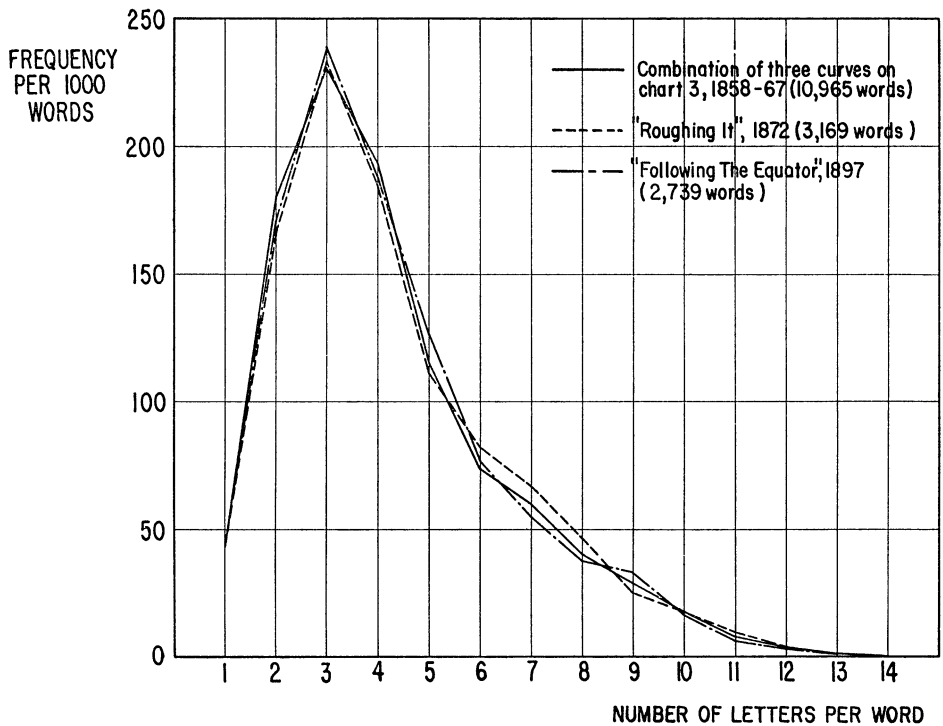


CHART 4. Word frequencies for known Mark Twain Writings 1858-1867, 1872, 1897.

words. Arbitrarily, it was decided to omit headings, proper names, direct quotes, foreign words, abbreviations, and involved dialect spellings.<sup>26</sup> Hyphenated words were counted as separate words.

The three frequency distributions for the control group—expressed as frequency per 1,000 words—are shown in Chart 3. Considering that the samples are individually fairly small, the consistency of the pattern within the samples is encouraging.

As a check on the consistency over a long span of time, two additional samples of Twain's work were counted:

1. Four passages, selected randomly, from *Roughing It* (1872), totaling about 3,200 words.<sup>27</sup>
2. Three passages, selected randomly, from *Following the Equator* (1897), totaling about 2,700 words.<sup>28</sup>

These two samples, together with the 11,000 word composite of the first three samples, are shown in Chart 4. Based on the visual evidence of Charts 3 and 4, we conclude that the word frequency distribution for Twain's "travel letters" maintained a high degree of consistency over a span of 40 years.

The 10 QCS letters were counted in three groups:

1. Letters I, II, and III. About 5,000 words.
2. Letters IV, V, and VI. About 4,000 words.
3. Letters VII, VIII, IX, and X. About 4,000 words.

Chart 5 shows the frequency distribution for each of the three QCS segments. The two principal conclusions are (1) the samples are consistent within themselves and (2) the curves of the QCS letters are quite unlike those of Charts 3 and 4. These differences are shown more clearly in Chart 6, where the 11,000 words of the Twain control group are compared to the 13,000 words of the QCS letters. Particularly noticeable are the differences in frequencies in words of 1, 2, 3, and 4 letters, and in the tendency for Twain to use relatively fewer words of 7 or more letters.

From the standpoint of visual inspection of the results of this literary blood test, we conclude that Mark Twain could *not* have fathered the QCS letters.<sup>29</sup>

##### 5. $\chi^2$ GOODNESS OF FIT TEST

It is also possible to test the agreement of the various samples to the control group by using the  $\chi^2$  "goodness of fit" test.<sup>30</sup> Although we would not expect a given author's output of words to always conform strictly to the hypothesis of

<sup>26</sup> This was done in an effort to eliminate any unusual characteristics of the subject matter. For example, after counting Dickens' *Christmas Carol* Mendenhall observed that there was an excess of seven-letter words, resulting from the repeated appearance of "Scrooge." See Mendenhall, *Science*, p. 244.

<sup>27</sup> Mark Twain, *Roughing It*. Hartford, Connecticut: American Publishing Company, 1872.

<sup>28</sup> Mark Twain, *Following the Equator*. Hartford, Connecticut: American Publishing Company, 1897.

<sup>29</sup> This does not mean, of course, that he could not have somehow offered suggestions to whoever did write them. Some of the unusual parallels cited by Leisy could have occurred in this way. For example, the name "Snodgrass" and the "Brown" to whom they were addressed, both of which are known to have been literary favorites of the young Twain, might have been suggested by him.

<sup>30</sup> See, for example, Wilfrid J. Dixon and Frank J. Massey, Jr., *Introduction to Statistical Analysis*. New York: McGraw-Hill Book Company, Inc., 1951, pp. 184 ff.

random drawings from a fixed frequency distribution,<sup>31</sup> the  $\chi^2$  test nevertheless should be useful as a ranking procedure (i.e., the results should provide evidence such as "probably the same," "possibly the same," "virtually impossible to be the same," etc.). In any case, until more detailed studies of this use of the test are made, the following results should be considered only in the terms of a tentative exploration.

The word counts for the three segments that make up the control group and for the two samples taken from Twain's later writings are summarized in Table 1. Table 2 summarizes the word counts for the 10 QCS letters. Since tallying is an obviously monotonous and error-prone job, it is unlikely that the counts are as precise as the numbers suggest. Random double checking, however, suggests that counting errors are small and unimportant.

TABLE 1. WORD COUNTS FOR KNOWN MARK TWAIN WRITINGS

| Word Length | Sergeant Fathom and Madame Caprell Letters (1858; 1861) | Four "Mark Twain" letters from <i>Territorial Enterprise</i> (1863) | First <i>Innocents Abroad</i> Letter (1867) | Sample from <i>Roughing It</i> (1872) | Sample from <i>Following the Equator</i> (1897) |
|-------------|---------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------|---------------------------------------|-------------------------------------------------|
| 1           | 74                                                      | 312                                                                 | 116                                         | 138                                   | 122                                             |
| 2           | 349                                                     | 1,146                                                               | 496                                         | 532                                   | 466                                             |
| 3           | 456                                                     | 1,394                                                               | 673                                         | 741                                   | 653                                             |
| 4           | 374                                                     | 1,177                                                               | 565                                         | 591                                   | 517                                             |
| 5           | 212                                                     | 661                                                                 | 381                                         | 357                                   | 343                                             |
| 6           | 127                                                     | 442                                                                 | 249                                         | 258                                   | 207                                             |
| 7           | 107                                                     | 367                                                                 | 185                                         | 215                                   | 152                                             |
| 8           | 84                                                      | 231                                                                 | 125                                         | 150                                   | 103                                             |
| 9           | 45                                                      | 181                                                                 | 94                                          | 83                                    | 92                                              |
| 10          | 27                                                      | 109                                                                 | 51                                          | 55                                    | 45                                              |
| 11          | 13                                                      | 50                                                                  | 23                                          | 30                                    | 18                                              |
| 12          | 8                                                       | 24                                                                  | 8                                           | 10                                    | 12                                              |
| 13 and over | 9                                                       | 12                                                                  | 8                                           | 9                                     | 9                                               |
| Total       | 1,885                                                   | 6,106                                                               | 2,974                                       | 3,169                                 | 2,739                                           |

The  $\chi^2$  test was carried out by comparing the frequency distribution of the Mark Twain control group to each of the two later Mark Twain samples and to the QCS letters. The results are summarized below:<sup>32</sup>

| Control group compared to:               | Value of $\chi^2$ | $P(\chi^2 > \text{Value})$ —<br>12 d.f. |
|------------------------------------------|-------------------|-----------------------------------------|
| Sample from <i>Roughing It</i>           | 12.4              | .41                                     |
| Sample from <i>Following the Equator</i> | 8.5               | .73                                     |
| QCS letters (in total)                   | 127.2             | Nil                                     |

<sup>31</sup> An analogy might be to a statistical quality control chart for variables that would show a few out-of-control points because of an occasional deliberate change in the production process. Similarly, an author could consciously change his normal style enough to produce segments that would result in large values of  $\chi^2$ .

<sup>32</sup> The computation procedure is outlined in Dixon and Massey, *op. cit.*, p. 190.

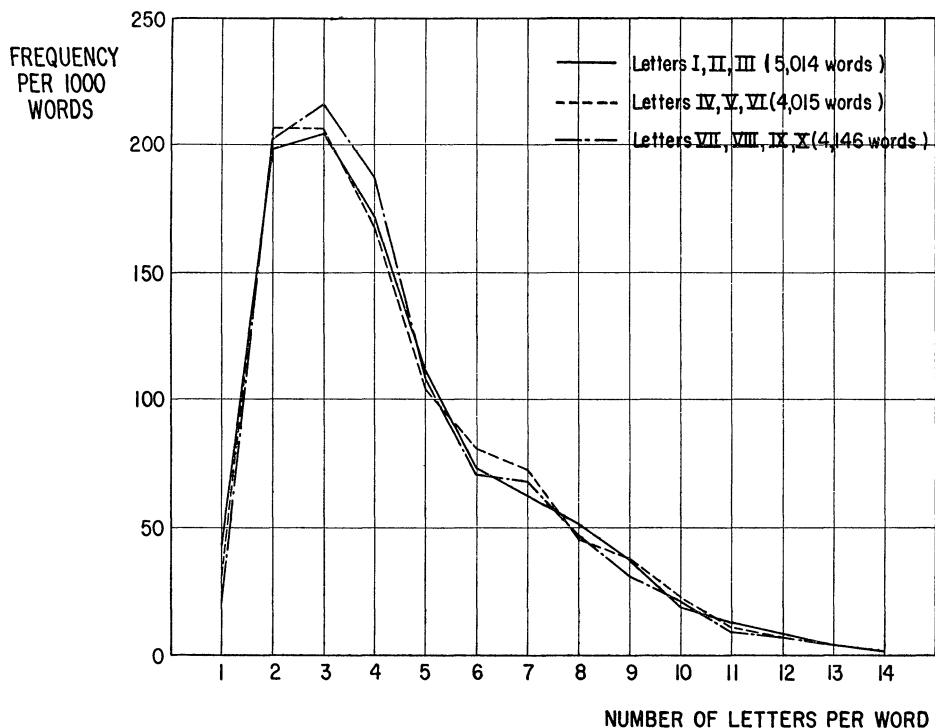


CHART 5. Word frequencies for Quintus Curtius Snodgrass Letters, 1861.

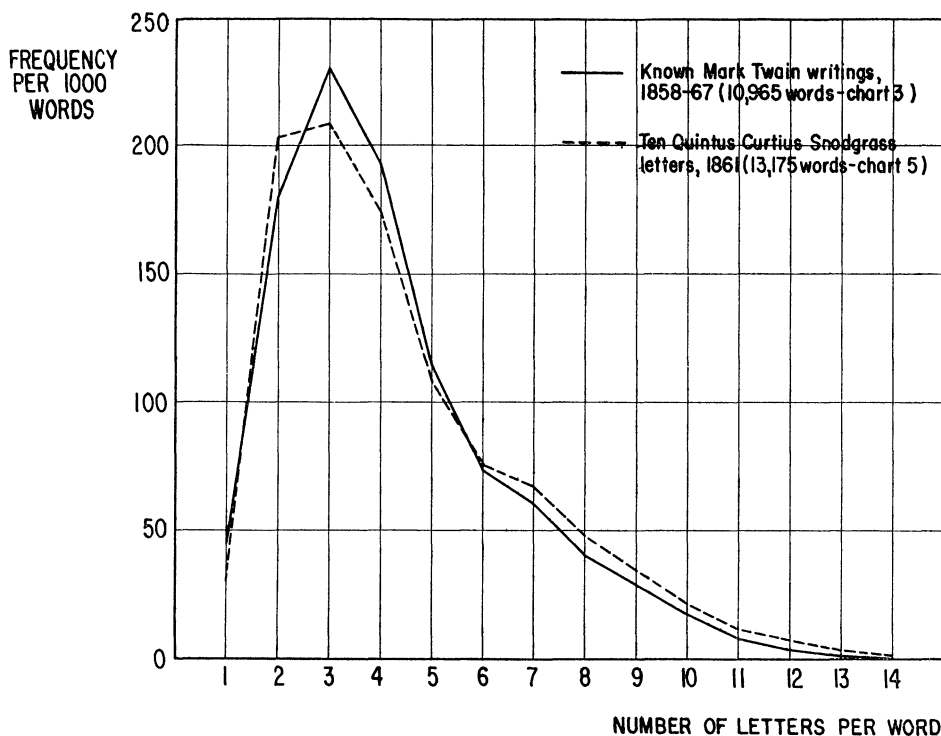


CHART 6. Word frequencies for known Mark Twain Writings and Quintus Curtius Snodgrass Letters.

TABLE 2. WORD COUNTS FOR QUINTUS CURTIUS SNODGRASS LETTERS

| Word Length | Letter Number |       |       |       |       |       |     |       |       |       | Total  |
|-------------|---------------|-------|-------|-------|-------|-------|-----|-------|-------|-------|--------|
|             | I             | II    | III   | IV    | V     | VI    | VII | VIII  | IX    | X     |        |
| 1           | 100           | 71    | 45    | 27    | 33    | 57    | 25  | 20    | 21    | 25    | 424    |
| 2           | 492           | 208   | 297   | 266   | 230   | 335   | 168 | 239   | 242   | 208   | 2,685  |
| 3           | 504           | 247   | 275   | 266   | 254   | 308   | 196 | 250   | 235   | 217   | 2,752  |
| 4           | 423           | 205   | 228   | 201   | 212   | 256   | 158 | 188   | 193   | 238   | 2,302  |
| 5           | 266           | 139   | 160   | 125   | 150   | 145   | 84  | 134   | 113   | 115   | 1,431  |
| 6           | 162           | 96    | 108   | 113   | 112   | 101   | 60  | 86    | 75    | 79    | 992    |
| 7           | 152           | 81    | 85    | 91    | 107   | 95    | 63  | 77    | 81    | 64    | 896    |
| 8           | 114           | 60    | 84    | 62    | 64    | 57    | 41  | 59    | 48    | 49    | 638    |
| 9           | 93            | 43    | 50    | 57    | 44    | 49    | 27  | 27    | 29    | 46    | 465    |
| 10          | 43            | 25    | 28    | 22    | 26    | 46    | 29  | 15    | 16    | 26    | 276    |
| 11          | 32            | 16    | 15    | 16    | 11    | 22    | 11  | 14    | 10    | 5     | 152    |
| 12          | 17            | 10    | 15    | 11    | 11    | 8     | 6   | 12    | 4     | 7     | 101    |
| 13 and over | 10            | 5     | 10    | 8     | 6     | 11    | 6   | 2     | 1     | 2     | 61     |
| Total       | 2,408         | 1,206 | 1,400 | 1,265 | 1,260 | 1,490 | 874 | 1,123 | 1,068 | 1,081 | 13,175 |

The small  $\chi^2$  values for the samples from *Roughing It* and *Following the Equator* confirm our conclusion about the consistency of Mark Twain's writings reached by visual inspection of Charts 3 and 4. The very large  $\chi^2$  for the QCS comparison confirms the conclusion drawn from Chart 6 that the QCS letters were not the work of Mark Twain.

6. TWO-SAMPLE *t* TEST

The two-sample *t* test for testing the difference between sample means can also be used as a simple statistical technique for comparing the Mark Twain writings and the QCS letters. Whereas the  $\chi^2$  test provided a measure of agreement of an entire large-sample frequency distribution to a control distribution, this test considers independently the frequency of occurrence of individual words of specified lengths. That is, sample data on two-letter words make up one test; sample data on three letter words another test; and so on. The sensitivity of this test is increased by dividing the data into a larger number of segments than used in the  $\chi^2$  test.

The data used and the results of the calculations are summarized in Table 3. As shown there, the control group (making up sample 1) was divided into eight subsamples of roughly equal size and the 10 QCS letters (making up sample 2) were treated as 10 separate subsamples (giving a total of 16 degrees of freedom). The tests were limited to 2, 3, or 4 letter words (about 60 per cent of the total).

Based on the large *t* values (shown at the bottom of Table 3) we conclude that it is highly unlikely that the 2, 3, and 4 letter words in the QCS letters came from the same distribution<sup>33</sup> that produced the words of these lengths in the Mark Twain control group.<sup>34</sup> This conclusion agrees with the one reached

<sup>33</sup> The null hypothesis has the implicit assumption of normality. When the data of Table 3 are plotted on normal probability paper, each set of subsample points falls reasonably well on a straight line.

<sup>34</sup> I am indebted to the referee for suggesting this additional test procedure. The referee's further suggestion that the values be weighted was considered to be too great a refinement for these data, particularly since most of the sample sizes are of the same order of magnitude. (Of the 18 samples in Table 3, all but three contain between one and two thousand words.) The calculation procedure is outlined in Dixon and Massey, *op. cit.*, Second Edition, 1957, p. 121.

TABLE 3. TWO SAMPLE *t* TESTS FOR FREQUENCIES OF 2, 3,  
AND 4 LETTER WORDS

| Item                                                   | Proportion of Each Work |                   |                   |
|--------------------------------------------------------|-------------------------|-------------------|-------------------|
|                                                        | 2-Letter<br>Words       | 3-Letter<br>Words | 4-Letter<br>Words |
| Sample 1: Mark Twain Group                             |                         |                   |                   |
| Sergeant Fathom Letter                                 | .214                    | .225              | .187              |
| Madame Caprell Letter                                  | .152                    | .262              | .212              |
| Mark Twain Letters in<br><i>Territorial Enterprise</i> |                         |                   |                   |
| First Letter                                           | .176                    | .217              | .196              |
| Second Letter                                          | .196                    | .240              | .179              |
| Third Letter                                           | .190                    | .230              | .189              |
| Fourth Letter                                          | .190                    | .229              | .208              |
| First <i>Innocents Abroad</i> Letter                   |                         |                   |                   |
| First Half                                             | .172                    | .235              | .195              |
| Second Half                                            | .161                    | .217              | .185              |
| Average—Sample 1                                       | .181                    | .232              | .194              |
| Sample 2: QCS Letters                                  |                         |                   |                   |
| I                                                      | .204                    | .209              | .176              |
| II                                                     | .172                    | .205              | .170              |
| III                                                    | .212                    | .196              | .163              |
| IV                                                     | .210                    | .210              | .159              |
| V                                                      | .183                    | .202              | .168              |
| VI                                                     | .225                    | .207              | .172              |
| VII                                                    | .192                    | .224              | .181              |
| VIII                                                   | .213                    | .223              | .167              |
| IX                                                     | .227                    | .220              | .181              |
| X                                                      | .192                    | .201              | .220              |
| Average—Sample 2                                       | .203                    | .210              | .176              |
| Value of <i>t</i> (16 d. f.)                           | −2.43                   | +3.88             | +2.57             |
| Approx. <i>P</i> ( $ t  > \text{value}$ )              | .028                    | .002              | .022              |

above based upon use of the  $\chi^2$  test and upon visual inspection of the frequency distribution.

7. POSTSCRIPT

At the conclusion of the above research, the author wrote to Henry Nash Smith, literary editor of the Mark Twain Papers, asking his opinion about the QCS letters. No reference was made in the author's letter to any of the above work or conclusions. The following reply was received from Frederick Anderson, assistant to Professor Smith:

Professor Henry Nash Smith has asked me to write to tell you that we are quite sure that Mark Twain did not write the Quintus Curtius Snodgrass letters. In addition to the fact that their prose style and attempts at humor are considerably more primitive than other writings by Clemens during this period, there is external evidence that he could not have been at some of the places mentioned in the letters at the time the writer of the letters was there. A graduate student [Allan Bates] at the University of Chicago is writing a dissertation on Mark Twain as a pilot and has established a chronological account of Clemens' movements during these years which supports this idea.<sup>35</sup>

Thus, there is independent corroborative evidence to support the conclusion reached in this study through Mendenhall's word frequency distribution test.

---

<sup>35</sup> Personal letter, September 6, 1961.