# The "Unusual Episode" Data Revisited

Robert J. MacG. Dawson
Saint Mary's University

**Key Words**: Elementary statistics teaching; Categorical datasets; Survival data; Classroom exercises.

## Abstract

A certain dataset, giving population at risk and fatalities for "an unusual episode," has been used for some time in classrooms as an elementary exercise in statistical thinking, the challenge being to deduce the context of the data. Unfortunately, the "solution" has frequently been circulated orally, with few details. Moreover, discrepancies have been found between the dataset and the "solution," which would render the exercise somewhat artificial. This paper investigates the discrepancies and includes a fully-explained version of the dataset for classroom use.

# 1. The "Unusual Episode" Exercise and Its History

1 In June, 1994, I attended a STATS ("Statistical Thinking And Teaching Statistics") workshop at Reed College, put on by the Mathematical Association of America. This workshop was intended for mathematics instructors whose duties included teaching elementary statistics, and the goal was to expose us to modern, data-oriented trends in statistical teaching and practice. It was, I should add, an extremely busy, instructive, and pleasant week!

2 Donald Bentley began his series of talks with a class exercise. The dataset shown in Table 1 was given out (Bentley 1995) and participants were invited to determine the nature of the "unusual episode." We were permitted to ask "yes/no" questions suggested by the data. Members of the group noticed such anomalies as the comparatively small number of children in the higher economic groups, their total absence from the "unknown" group, and the complicated ways in which the mortality rate was affected by all three factors.

```
                         By Economic Status and Sex
           -----------------------------------------------------------------
                Population Exposed      Number of          Deaths per 100
                     to Risk              Deaths           Exposed to Risk
  Economic      -----------------------------------------------------------------
  Status        Male  Female  Both    Male  Female  Both    Male  Female  Both
           -----------------------------------------------------------------
  I(high)       172    132    304     111     6     117      65     5     39
  II            172    103    275     150    13     163      87    13     59
  III           504    208    712     419   107     526      83    22     41
  Unknown         9     23     32       8     5      13      89    22     41
           -----------------------------------------------------------------
  Total         857    466   1323     688   131     819      80    28     62
```

```
                         By Economic Status and Age
           -----------------------------------------------------------------
                Population Exposed      Number of          Deaths per 100
                     to Risk              Deaths           Exposed to Risk
  Economic      -----------------------------------------------------------------
  Status        Adult  Child  Both   Adult  Child  Both    Adult  Child  Both
           -----------------------------------------------------------------
  I and II      560     19    579     280     0     280      50     0     48
  III           645     67    712     477    49     526      74    73     74
  Unknown        32      0     32      13     0      13      41     -     41
           -----------------------------------------------------------------
  Total        1237     86   1323     770    49     819      62    57     62
```

**Table 1**: Population at Risk, Deaths, and Death Rates for an Unusual Episode

3 After several minutes of testing theories, the intended answer was reached: the episode referred to was the sinking of the ocean liner *Titanic* after colliding with an iceberg on April 15th, 1912. "Economic status," we were told, had been determined for the dataset based on the class (first cabin, second cabin, or steerage) in which the passengers travelled. The high mortality rates among males in all classes were the result of the rule of "women and children first"; the higher mortality rate among those in Class III could be explained by the more vulnerable position of the steerage cabins, low in the hull of the ship.

4 We were told that this exercise had been used for some time as an informal exercise in statistical thinking. I used it as a warm-up activity that September for an introductory statistics course. The course was being reintroduced after many years and had been little advertised; as a result, it had only a few students, and there were not as many heads working on the problem as there had been at Reed. However, an answer was at last reached, and it did stimulate considerable conjecture and discussion. The dataset went into my "keeper" file, to be used again in the future.

5 I was more than a little surprised, a few weeks later, when one student came into class with a book, Logan Marshall's *Sinking of the Titanic and Great Sea Disasters* (Marshall 1912) written shortly after the accident, in which it was stated that not 819, but 1635, of those on board had been lost; and that the

original number had been not 1323, but 2340! Was this dataset in fact the record of another shipwreck, or was there some other reason for the discrepancy? Unfortunately, by the time I obtained my copy of the dataset, no "official solution" or author's name was attached to it, so it was not possible to ask what had originally been intended.

6 Research in the local libraries and (especially) the Provincial Archives of Nova Scotia provided a moderate amount of information on the *Titanic* sinking and other marine disasters. Interestingly enough, there have been very few marine disasters with a death toll in the neighborhood of 800 (*Encyclopaedia Americana* 1994); only 3 disasters with between 700 and 900 casualties have occurred (omitting naval vessels sunk by enemy fire). Two of these involved naval ships, and the third (the excursion steamer *Eastland* which sank in the Chicago River in 1915) seems obscure enough that it may probably be eliminated as the source of the dataset.

7 More detailed research into the *Titanic* disaster revealed some differences of opinion on the number lost. For instance, the *Encyclopaedia Americana* (1994) gives the death toll as "variously estimated as 1,490, 1,502, and 1,517." A book edited in 1912 under the pseudonym Marshall Everett gives the figure variously as 1635 and 1595 (Everett 1912); the first of these figures agrees with that found in Logan Marshall's book (Marshall 1912). However, the British Board of Trade Inquiry Report (1990), written originally in 1912, claims a death toll of 1490. Modern sources seem to agree that the true numbers are in the neighborhood of 1,500, but the exact numbers may never be known.

8 Interestingly, the Board of Trade Inquiry Report (1990, p. 42) contains a table, similar to Table 1, in which the ship's population and the survivors (not the deceased) are classified by age, sex, and economic status. Four economic groups are listed: first-class passengers, second-class passengers, third-class passengers, and crew. Table 2 is a modified version of Table 1, based on these data. (The original report also breaks the crew down into Deck Department, Engine Room Department, and Victualling Department; I have not reproduced these figures.)

### By Economic Status and Sex

| Economic Status | Population Exposed to Risk | | | Number of Deaths | | | Deaths per 100 Exposed to Risk | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | Both | Male | Female | Both | Male | Female | Both |
| I(high) | 180 | 145 | 325 | 118 | 4 | 122 | 65 | 3 | 37 |
| II | 179 | 106 | 285 | 154 | 13 | 167 | 87 | 12 | 59 |
| III | 510 | 196 | 706 | 422 | 106 | 528 | 83 | 54 | 73 |
| Other | 862 | 23 | 885 | 670 | 3 | 673 | 78 | 13 | 76 |
| Total | 1731 | 470 | 2201 | 1364 | 126 | 1490 | 80 | 27 | 67 |

### By Economic Status and Age

| Economic Status | Population Exposed to Risk | | | Number of Deaths | | | Deaths per 100 Exposed to Risk | | |
|---|---|---|---|---|---|---|---|---|---|
| | Adult | Child | Both | Adult | Child | Both | Adult | Child | Both |
| I(high) | 319 | 6 | 325 | 122 | 0 | 122 | 38 | 0 | 37 |
| II | 261 | 24 | 285 | 167 | 0 | 167 | 64 | 0 | 59 |
| III | 627 | 79 | 706 | 476 | 52 | 528 | 76 | 66 | 73 |
| Other | 885 | 0 | 885 | 673 | 0 | 673 | 76 | - | 76 |
| Total | 2092 | 109 | 2201 | 1438 | 52 | 1490 | 69 | 48 | 67 |

**Table 2**: Population at Risk, Deaths, and Death Rates for the Sinking of the *Titanic*

9 Comparing the tables, it appears that the "Unusual Incident" dataset does describe the *Titanic* disaster, but that the numbers are for passengers only. The numbers agree approximately in all cases, although there are many small discrepancies: for instance, the Board of Trade figures show 175 adult males and 5 boys in first class, whereas the "Unusual Episode" figures show 172 males including boys. Witnesses tended, when giving evidence at the inquiry, to underestimate the number of crew in the boats and to overstate the numbers of women and children saved (Board of Trade Inquiry Report 1990, p. 39). These inaccuracies may have been perpetuated in some subsequent references.

10 While there are differences in format, the similarities between the two tables are strong enough to suggest that the "Unusual Episode" sheet is in some way descended from the Board of Trade Report. However, the discrepancies are numerous, and the "Unusual Episode" table is self-consistent (with the exception of the female and combined death rates in economic group III in the upper subtable, which do not agree with the numbers to their left), so that it cannot be supposed that all the discrepancies are copying errors. It would appear that some person created a table, similar to that in the Report but based upon other figures. Was this done by coincidence, without knowledge of the table in the Report -- or was the creator of the original "Unusual Episode" table deliberately rewriting the Report's table, considering another source's figures to be more reliable?

11 The White Star Line's final (May 9, 1912) list of passengers lost and saved (most easily found in Lord's *A Night To Remember* (1955) yields numbers that are tantalizingly close to those of Table 1. The total numbers of first, second, and third-class passengers are 323, 272, and 712 respectively; the second number is close, and the third equal, to that of Table 1. The first is too high; however, in the White Star list, servants travelling in first class with their employers are identified as such, rather than by name. These comprise 8 male servants (of whom one was saved), and 22 female servants (of whom 15 were saved); these are almost certainly the "unknown" class of Table 1. After removing them from the first-class total, the remainder (293) is a little smaller than the first-class total of the table. (The 8 bandsmen, who were travelling in second-class cabins provided by the White Star line, are not separately identified either in the table or the list. The Board of Trade inquiry report, by contrast, identifies them -- possibly due to their contractual status vis-a-vis the shipping line -- and not the servants.)

12 However, the White Star list has its own quirks. In first and second class, boys are identified as "Master," but there is no sure way to distinguish girls from women. "Miss Smith," listed after "Mr. and Mrs. Smith" in second class, might be their child, their grown daughter, Mr. Smith's unmarried sister, or completely unrelated. Children in second class are sometimes listed by name, but in some cases the list merely adds "and child" or "and infant" after the

parent's name. Twelve boys and six children of unspecified sex in first and second class are so identified; however, it seems most likely that several of the 26 passengers identified as "Miss" and travelling with a married couple of the same name were also children. Thus, while we can see how the figure of 19 children in first and second class, given in Table 1, could have been derived from the White Star list, it is almost certainly too low. In third class, all passengers are listed by name; some names are followed by "(child)" or "(infant)." Sixty-seven passengers are so identified, agreeing with the count in Table 1.

13 The class system, as practiced by the White Star Line, causes another difficulty in tabulating the third-class passengers. The use of such gender-identifying honorifics as "Mr." or "Mrs." by the line appears to have been reserved for holders of first or second-class tickets, even after death. Classifying third-class passengers with certain foreign first names, then, would require considerable linguistic detective-work for most researchers, whereas the sex of those identified only by initials could only be determined from other sources.

14 Overall, there is considerable evidence that the numbers in Table 1 have been derived (directly or indirectly) from a source similar to but different from the White Star Line's final May 9 passenger list -- perhaps some earlier version of the same list. Particular clues are the identification of servants (and not bandsmen), and the low number of children in the higher classes. Indeed, the very fact that the sex and age breakdowns are done separately may reflect an awareness on the part of the original compiler that the data, as derived from such a source, did not permit an accurate determination of the number of girls in the upper classes.

# 2. Classroom Use

15 An instructor could still use this exercise with the original dataset, noting afterwards that the "group at risk" represents the passengers only. However, the omission of the crew seems a little artificial; it is hard to argue that they were not part of the group at risk. As an alternative, the revised figures in Table 2 may be used (changing, of course, the title of the table!) for the same purpose. The information in this note will, it is hoped, enable the instructor to answer any questions that may arise from marine historians in the class.

16 With a very small class, as observed above, there may be the problem of too few heads working on the exercise. I found that asking leading questions about classes of the population that appeared to be at exceptionally high or low risk helped to some extent, but the exercise was still not completed as fast as I would have liked. This will probably not be relevant most of the time, as introductory statistics courses with only a handful of students are rare.

17 This year, due to various changes, the course is much larger: there are two sections, each of over 100 students! Moreover, the students (mainly first- and second-year psychology majors) are considerably less mathematically inclined than last year's class were. At first, the students seemed a little unsure what was expected; one student asked whether they were really expected to figure out what had happened, based only on those numbers. However, once they got the idea, the students seemed to enjoy the exercise, and took part enthusiastically. Because of the large number of students, the questions came thick and fast; for the first day of class, this was a heartening reaction.

18 This time, the exercise was completed very quickly: it was only about five minutes until a woman in the front row asked "Was it a shipwreck?" and then "Was it the *Titanic*?" I had anticipated a fast solution and used another five or ten minutes asking students to come up with as many clues from the table as they could that they could recognize as typical of the *Titanic* sinking. They again participated actively, picking out most of the distinctive features of the table.

19 The following imaginary dialogue is put together from questions asked on different occasions, and others that might be asked.

"What do you notice about those numbers?"

"The death rate for females was much less than for males."

"Yes, but much more so among the rich."

"It was like that with the children, too -- none of the kids in groups I or II died, but lots of the ones in group III did."

"Was it something that could be cured but the cure cost a lot?"

"But then why didn't the wealthy men get cured?"

"Is there any disease that women and children can be cured from and men can't?"

"Maybe it was something else -- a massacre or something?"

"But who'd massacre poor women and children and all men but not rich women?"

"Hey, look at the numbers _exposed_. Look how many more men were exposed, and how few children."

"Maybe they just looked at more men?"

"No, it says `Population Exposed'."

"Those numbers are pretty exact. They must have known exactly who was exposed and who wasn't. How could they tell that so accurately with an epidemic?"

"Yes, and who were those `other' people? Notice they were all adults, and almost all guys?"

"They had the same death rate as the poor people. Were they poor?"

"Maybe they were doctors or rescue workers?"

And so on...

20 It must be said that, when I used this exercise for the first time with a really large class, the quality of the questions was not as high as I had hoped. There were many rather superficial questions such as "Was it a virus?" or "Was it smoking?" and comparatively few attempts to progressively analyze the data. This may have been due simply to the number of students in the class: a fairly small proportion of students with wild guesses, in a class of over 100, can occupy a fair amount of time and cover many categories of incident! To counteract this, I began to ask questioners why they thought that a certain question was suggested by the data. This slowed down the flow of questions a little, and seemed to encourage more thought about what to ask. Next time I use this exercise, I will do something like this from the beginning in the hopes of eliciting some really thoughtful questions based on the data.

# 3. Getting the Data

21 The file titanic.dat.txt contains the raw data. The file titanic.txt is a documentation file containing a brief description of the dataset.

## Acknowledgments

I would like to thank the Mathematical Association of America and Reed College for hospitality at the STATS workshop where this all began; Saint Mary's University for sending me there; Donald Bentley for introducing us to the original "Unusual Episode" dataset; Peter Ewert for noticing that something peculiar was going on; NSERC and the Canadian taxpayers for funding; and the JSE editors for help and encouragement.

---

### Addendum (added April 3, 1998)

The years 1997-1998 saw a sudden rise in public interest in the *Titanic*, due in large part to the film of the same name. Very detailed data about the passengers is now available on the Internet, at sites such as *Encyclopedia Titanica* [http://www.rmplc.co.uk/eduweb/sites/phind]. By means of such sources, it is possible to show that even the Board of Trade Inquiry figures are not entirely correct. (For instance, there was one juvenile fatality in first class.) Therefore, educators using Table 2 above may want to describe it as "the official report," or otherwise indicate that the figures may not be the best known.

---

# Appendix - Key to Variables in titanic.dat.txt

```
Column
   1      Class (0=crew, 1=first, 2 = second, 3 = third)
  10      Age   (1 = adult, 0 = child)
  19      Sex   (1 = male, 0 = female)
  28      Survived (1 = yes, 0 = no)
```

Values are aligned and delimited by blanks. There are no missing values.

---

# References

Bentley, D. L. (1995), "My First Days' Lectures: Past and Present," in *Education in a Research University*, eds. K. J. Arrow, B. C. Eaves, and I. Olkin, Stanford, CA: Stanford University Press, pp. 215-228.

"Marine Disasters" (1994), *Encyclopaedia Americana* (Vol. 9), Danbury, CT: Grolier, pp. 164-165.

"Report on the Loss of the `Titanic' (S.S.)" (1990), *British Board of Trade Inquiry Report* (reprint), Gloucester, UK: Allan Sutton Publishing.

Everett, M. [Henry Neil] (ed.) (1912), *The Story of the Wreck of the Titanic* (memorial edition), Chicago: Homewood Press.

Lord, W. (1955), *A Night to Remember*, New York: Holt.

-------- (1988), *The Night Lives On*, New York: Morrow.

Marshall, L. (ed.) (1912), *Sinking of the Titanic and Great Sea Disasters*, Philadelphia: The John C. Winston Co.

---

Robert J. MacG. Dawson
Department of Mathematics and Computing Science
Saint Mary's University
Halifax, Nova Scotia B3H 3C3
CANADA

*rdawson@husky1.stmarys.ca*

---