

## John Graunt's Life Table

In 1661 A.D., some 3,064 years after the first census of Moses, an obscure haberdasher, late a captain in the loyalist army of Charles II, published an analysis on data originally collected by Thomas Cromwell, 127 years earlier, dealing with age at the time of death in London. The data had been collected at the request of the merchants of London who were carrying out the most basic kind of marketing research: i.e., whether potential customers (i.e., live people) were on the increase or decrease. Interestingly, enough, the question originally arose because of the fact that the "Plague" (perhaps the Bubonic Plague, perhaps a collection of various contagious diseases) had been endemic in England for many years. At times, there would be an increase of the incidence of the disease, at other times a decrease. It was a matter of sufficient importance to be attended to by Chancellor Thomas Cromwell (also "Master of the Rolls"). Without any central data bank, a merchant might put a shop in an area where the decline in population had eliminated any potential opportunity, due to market saturation.

Cromwell's data base consisted in records of births and deaths from the Church of England to be carried out and centrally located by the clergy. Before Graunt, all analyses of the data had suffered the usual "can't see the forest for the trees" difficulty. The records were not kept in EXCEL. Each parish priest (and there were hundreds) had his own way of recording births, deaths, marriages, etc. The data base was included in ledgers without any good sense of a common taxonomy.

Graunt solved this problem, and started modern statistics by creating the following table:

Age Interval	Prop. Deaths in Interval	Prop. Surviving til start of Interval
0-6	0.36	1.00
7-16	0.24	0.64
17-26	0.15	0.40
27-36	0.09	0.25
37-46	0.06	0.16
47-56	0.04	0.10
57-66	0.03	0.06
67-76	0.02	0.03
77-86	0.01	0.01

Graunt did a little too much smoothing, for we only know the death proportion in each of the intervals. But Graunt could still make some rough sense of what was going on. The people who were in the interval from 0 to 6 years could be taken to have a roughly equal death rate for each of the years. So, we could say that about  $.36/6 = .06$  of the people who had been born died at age less than one year,  $.06$  from one year to two years, . . . ,  $.06$  from five years to six years. That is to say,

$$P(\text{die in } [0,1) \text{ given born}) = \frac{.36}{6 \times 1.00} = 0.0600 \quad (1)$$

Then, of the people who survived their sixth birthday, there was a proportion of  $1.00 - 0.36 = 0.64$ . So, then for the next ten years, we could say that of the people who had lived past their sixth birthday, there was a proportion of  $0.64$ . Graunt then divided the proportion of deaths in the interval by  $10$  to get the per year death rate. But he also divided by  $.64$ , since only  $64\%$  of the population had survived until their sixth birthday. So, he came up with the proportion of people who died in the interval from  $6$  to  $7$ , given that they had survived at least until age six to be

$$P(\text{die in } [6,7) \text{ given live until } 6) = \frac{.24}{10 \times .64} = 0.0375 \quad (2)$$

And, of course, he approximated the probability of death in each of the one year intervals from seven through  $16$  also to be equal to  $0.0375$ .

Then, for the next ten one year intervals, he would have

$$P(\text{die in } [16,17] \text{ given live until } 16) = \frac{.15}{10 \times 0.40} = 0.0375 \quad (3)$$

Next, for the next ten one year intervals, we see that

$$P(\text{die in } [26,27] \text{ given live until } 26) = \frac{.09}{10 \times 0.25} = 0.0360 \quad (4)$$

For the next ten one year intervals, we see that

$$P(\text{die in } [36,37] \text{ given live until } 36) = \frac{.06}{10 \times 0.16} = 0.0375 \quad (5)$$

Carrying on in this fashion

$$P(\text{die in } [46,47] \text{ given live until } 46) = \frac{.04}{10 \times 0.10} = 0.0400 \quad (6)$$

$$P(\text{die in } [56,57] \text{ given live until } 56) = \frac{.03}{10 \times 0.06} = 0.0500 \quad (7)$$

$$P(\text{die in } [66,67] \text{ given live until } 66) = \frac{.02}{10 \times 0.03} = 0.0667 \quad (8)$$

$$P(\text{die in } [76,77] \text{ given live until } 66) = \frac{.01}{10 \times 0.01} = 0.1000 \quad (9)$$

Graunt did not know much about graphical presentation. But he made a great discovery concerning the prevalence of the plague. A 50 year old person had about as much chance of dying in the next year as did a 20 year old person. This indicates that people were dying largely due to causes which are not age related. These were very possibly plague (and here we could mean a variety of diseases) related. So Graunt could go to the Royal Society of London and give a rather convincing argument that the answer to Thomas Cromwell's question posed over a century earlier as to whether plague continued to be a problem in London. The answer was "Yes." King Charles was so impressed that he ordered the newly created Royal Society of London to take Graunt in as a member. There was much grumbling about this, for Graunt was not from the squirearchy, and, like King Charles himself, he had Catholic leanings. He was taken in nonetheless. So a statistician was one of the first members of

the Royal Society of London. Graunt is spoken of several times in Pepy's *Diary*. When the King became ill and was not looking to business very closely, Graunt was dropped from the membership of the RSL. Though he had performed well for the Royal Cause in the War with the Roundheads, he made his living as a haberdasher and died broke.<sup>1</sup>

Graunt was not mathematically sophisticated. Probably anyone reading this chapter knew more math than Graunt did. He was simply a practical intuitivist, and intuition is worth tons of sophistication. Let us observe that he had practically discovered one of the most basic concepts in statistics: the *cumulative distribution function*. We define the cdf  $F$  as

$$F(x) = \text{Probability a variable } X \text{ is less than or equal to } x. \quad (10)$$

Of course, he actually discovered  $1-F(x)$ , for he did not tabulate the probability of death by  $x$  but the probability one lived longer than  $x$ . From his table, Graunt computed the life expectancy of a Londoner via

$$\begin{aligned} E(X) &= 3 \times .36 + 10.5 \times .24 + 20.5 \times .15 + 30.5 \times .09 + 40.5 \times .06 \\ &\quad + 50.5 \times .04 + 60.5 \times .03 + 70.5 \times .02 + 80.5 \times .01 = 18. \end{aligned}$$

Now, although Graunt did not do graphs, there is nothing to stop us from graphing his data as we do in Figure 1.

---

<sup>1</sup>Following Graunt's death, there was a serious attempt made to credit his *Essay on the Bills of Mortality* to the well born and politically correct William Petty. Petty had become wealthy auctioning off the property of the Irish who had opposed the rule of Oliver Cromwell. Since Charles II had the policy of not retaliating against the Cromwellians, Petty was able to continue in London following the Restoration, where he lived in luxury and was one of the founding members of the Royal Society of London. It just was too much against the grain of the RSL to leave the discovery of life tables as within the capabilities of a humble tradesman (and a Catholic to boot). So, since Petty had had lots of experience in the tabulation of properties seized by the state in Ireland, it became current, following Graunt's death, to claim that the work of Graunt really should be ascribed to Petty. Graunt's authorship is no longer disputed today, but one can find Petty given credit in books written even during the latter part of the Twentieth Century.

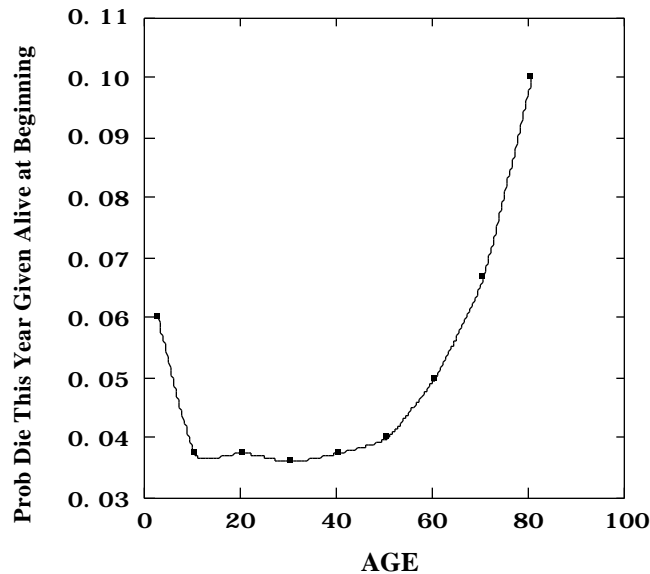


Figure 1. Graunt's Life Table Graphed.

What sort of progress had Graunt made since the censuses of Moses and David? A very great deal indeed. Note that Graunt ordered his data. A ten year old was more likely to be similar to an eleven year old than to a 50 year old. This is to us intuitively obvious, but it took over 3,000 years for somebody to make this step in the analysis of data. Had Graunt taken the step of simply graphing  $F(x)$ , he might well have looked at tangents to the curve and discovered derivatives. That, unfortunately, he had to leave to Newton and Leibnitz. But we should not forget that it was Graunt who made sense of temporally indexed quantitative and probabilistic data for the first time (so far as we know) in human history.

Anybody who feels “mathematically challenged” might do well to picture the sight of this humble tradesman, pouring over hundreds of ledgers in his “copious leisure time.” There are undoubtedly many such obscure persons in the history of science whose contributions were not recorded. Happily, thanks to Pepys and others, John Graunt has not escaped our attention.