

2012

Period Life Tables: A Resource for Quantitative Literacy

Thomas J. Pfaff

Ithaca College, tpfaff@ithaca.edu

Stanley Seltzer

Ithaca College, seltzer@ithaca.edu

Follow this and additional works at: <https://scholarcommons.usf.edu/numeracy>



Part of the [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

Pfaff, Thomas J., and Stanley Seltzer. "Period Life Tables: A Resource for Quantitative Literacy." *Numeracy* 5, Iss. 1 (2012): Article 5. DOI: <http://dx.doi.org/10.5038/1936-4660.5.1.5>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

Period Life Tables: A Resource for Quantitative Literacy

Abstract

A period life table provides an estimate of the probability that a person will die at a particular age. Using data available online, we examine tables of expected years to live for males and females against age for three populations: the United States in 2007, the U.S. at the turn of the twentieth century, and the Roman Empire. Scatter plots of males and females for each population show how life expectancy increases with age (e.g., U.S. 2007: 50 year-old female > 40 year-old female > 45 year-old male). The three data sets allow historical comparisons (e.g., of gender disparity, larger now; of infant mortality, smaller now). Regression lines for the linear portion of the plots (ages 5 to 70) show the annual increase in the years to live (e.g., U.S. 2007: 0.11 years for men, 0.07 years for women). Residual plots show that, even though the coefficients of determination of the line exceed 0.99, a concave-up, decreasing function would be a better model. The residual plots also reveal a curious inflection for the males that is not evident for the females. Such examples from period life tables might be presented in a discussion of life expectancy; alternatively, one or more could add to an introduction to regression, particularly illustrating the value of residual plots in understanding a data set.

Keywords

Regression Analysis, Residual Plots, Life Expectancy

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Cover Page Footnote

Tom Pfaff is an associate professor of mathematics at Ithaca College. His main teaching interest is incorporating sustainability and quantitative literacy issues into mathematics courses. Pfaff is the lead-PI of the Multidisciplinary Sustainability Education Project, and he has given several presentations on sustainability in the curriculum. He also maintains a page of sustainability curriculum materials for Mathematics at <http://www.ithaca.edu/tpfaff/sustainability.htm>.

Stan Seltzer is an associate professor and chair of the mathematics department at Ithaca College. He is a co-author of *Calculus: An Active Approach with Projects* (an e-book published by the Mathematical Association of America; available at <http://www.maa.org/ebooks/crm/ICP.html>). His current interests include voting and apportionment.

Introduction

The ability to use quantitative tools and understand quantitative information can greatly deepen one's understanding of a topic as well as generate interesting questions to research. Here we will use basic regression analysis to examine life expectancy of three populations: the United States in 2007, the United States at the turn of the twentieth century, and the Roman Empire. Our goal here is to engage in a quantitative discussion of life expectancy based on period life tables while also touching on a range of topics from historical comparisons to health policies and leaving the reader with interesting questions gleaned from the analysis. Hopefully the examples that we will present will be used in context, such as a history course or a health course, where the quantitative information is matched with the perspective of an expert to bring together richer and deeper understandings of life expectancy.

Period Life Tables

A *period life table* provides an estimate of the probability that a person will die at a particular age. From that information we calculate the life expectancy for each age. For example, for the United States in 2007, on average a 5 year old is expected to live another 71.03 years or until age 76.03.

Now, let's begin with a question. Who is expected to die at the oldest age, a 40 year-old female, a 45 year-old male, or a 50 year-old female? Before reading further, stop and see if you can estimate each person's expected age at death.

This question, of course, is not clearly defined. You likely assumed that it refers to three people living now, presumably in the United States. We will up the stakes and ask the same question for people living in the U.S. in 1900 or in the Roman Empire rather than in the U.S. today. These questions can be answered by looking at period life tables. The ones we used for the U.S. were obtained from the Social Security Administration and were derived by statistical sampling (Social Security Online 2011; Bell and Miller 2005, Table 6); for people living in Roman Empire we have a data set that was constructed based on historical information (Wikipedia 2011).¹ These period life tables are quantitatively rich data sets which can be useful in many courses. Each data set itself is interesting, but as one goes further by looking at scatter plots, then regression lines, and finally residual plots (with a twist) there are ample opportunities to make connections to the real world and history; and what could be more interesting than examining how long we are expected to live?

¹ All of the data are in the appendix: A, Roman Empire; B, U.S., 1900; C. U.S., 2007..

Answering the First Question

To answer our first question, Figure 1 shows scatter plots of life expectancy (i.e., expected number of years until death) versus age, where points corresponding to females are circles and those corresponding to men are squares. The scatter plots show data from Appendix C, and they tell us that, for example, our 40 year-old female, 45-year old male, and 50-year old female are expected to live 41.91, 33.33, and 32.69 more years or until 81.91, 78.33, and 82.69 years of age, respectively. (Clearly we cannot read the scatter plots this accurately; these values come from Appendix C.) At this point we see that the 50 year-old female is expected to live the longest and the 45 year-old male the least. Note that each of these people are expected to live longer than what is often referred to as “life expectancy,” roughly 75 for males and 80 for females in the U.S, but is better referred to as life expectancy at birth. Lesson 1 (perhaps obvious): as one lives longer one's expected age at death increases; this change isn't the same for males and females as we see in our example above, and it has changed over time. One number, such as “life expectancy” is meant as a simple measurement of how long people live, but it really doesn't give a complete picture (as one number usually doesn't).

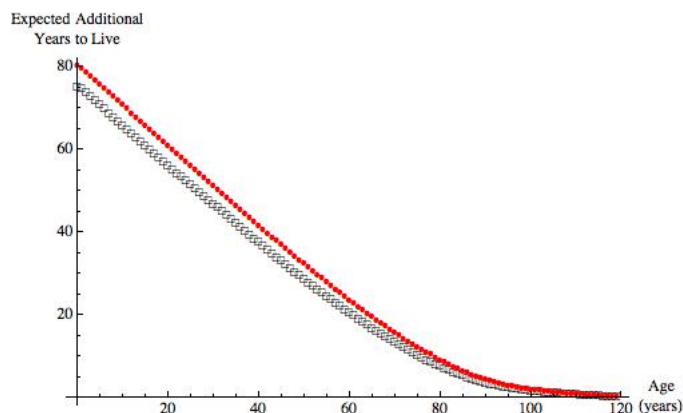


Figure 1. Scatter plot for United States females and males: expected years to live at a given age in 2007; females are circles, men are squares. Data from Social Security Online (2011).

Digging Deeper

Now take a look at Figures 2 and 3 and answer our original question before reading further. From these models we see that a 40 year-old female, a 45 year-old male, and a 50 year-old female could have expected to live to (again all values from Appendix C) 68.69, 68.99, and 71.52 years, respectively, in the United States in 1900 and to 61.1, 61.4, and 65.6 years, respectively, during the Roman Empire.

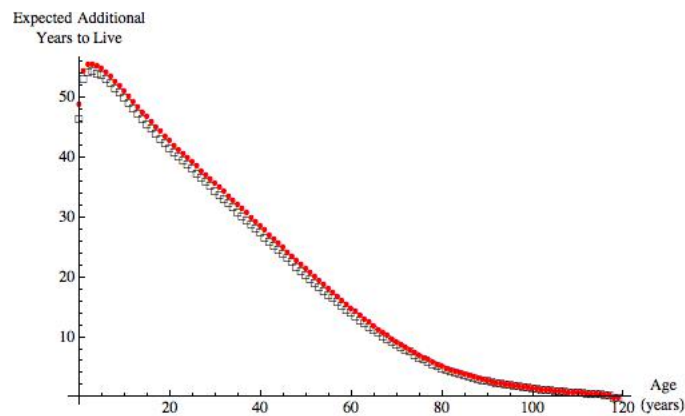


Figure 2. Scatter plot for United States females and males: expected years to live at a given age in 1900; females are circles, men are squares. Data from Bell and Miller (2005).

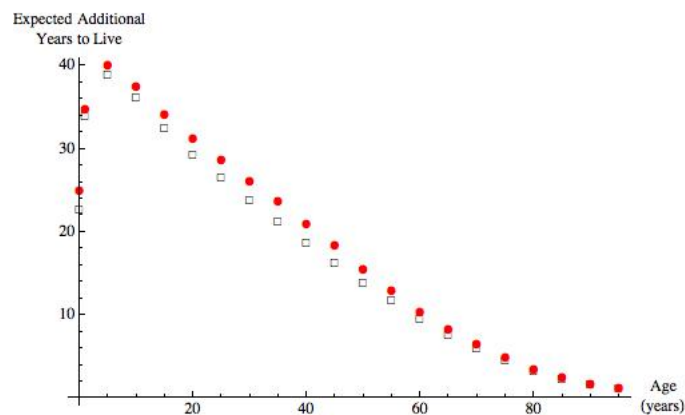


Figure 3. Scatter plot for females and males: expected years to live at a given age for those living in the Roman Empire; females are circles, men are squares. Data from Wikipedia (2011).

Interestingly, note the gap between men and women in Figure 1 as compared to Figures 2 and 3; males appear to have fared better relative to females in 1900 in the U.S. and in the Roman Empire, compared with today. But, the most striking feature of these scatter plots is the first two data points. During the Roman Empire, males at birth are expected to live to only 22.8 years, but if they survive the initial year, their life expectancy jumps to 35.1 years, and if they live to 5, they are likely to reach 44. Similarly, females start out with a life expectancy of 25 years and this jumps to 35.9 and 45.1 at ages 1 and 5, respectively. In the U.S. in 1900 men and women gained 12.37 and 10.93 years by making it to age 5. Today, surviving the first year of life gets you a bit of a jump, as seen in Figure 1, but it is nowhere near as drastic.

The data are showing us that one of the biggest increases in life expectancy today comes from reduced infant mortality. In other words, keeping a person from

dying in the first few years of life does more for the average life expectancy at birth than keeping an older person alive for a few extra months. There are various health policy questions one can ask based on this observation alone such as how we spend our medical research dollars. For instance, if one only wants to increase a society's life expectancy at birth then they should focus on reducing infant mortality. However, using the single measurement of life expectancy at birth ignores, for example, quality-of-life issues.

We should pause for a moment to talk about the reliability of our results. For the United States in 2007 we have excellent computing power and ability to obtain data through the United States Census Bureau. We should expect that the life expectancies presented are fairly accurate. Similarly, the U.S. 1900 data should also be good given our record keeping. On the other hand, the data for the Roman Empire are for a longer period of time and collected through historical techniques. We should expect that these data are significantly less accurate than the U.S. data and only provide an approximate view of life expectancy during that time. Still, even an approximation can be interesting and provide some insight.

Regression Lines

Back to the data. At this point it seems reasonable to model each set of data with a regression line since all six sets appear linear until somewhere between age 60 and 70. Based on the scatter plots we expect good fits. We will use data between ages 5 through 70, removing the early outliers. The regression lines and r^2 values are given in Table 1.

Table 1

Slopes, intercepts, and r^2 values of regression lines for expected years to live from ages 5 to 70 for men and women.

	Men			Women		
	Slope	Intercept	r^2	Slope	Intercept	r^2
Roman	-0.51	40.06	0.992	-0.52	42.12	0.998
U.S. 1900	-0.70	56.08	0.997	-0.71	57.42	0.998
U.S. 2007	-0.89	74.28	0.998	-0.93	79.82	0.999

The intercepts of these lines are higher than the observed data points, consistent with our observation about infant mortality. The more interesting information that the regression line provides is its slope, which tells us how many fewer years an individual lives (on average) by virtue of having survived one additional year. Since expected life span is current age plus expected number of years to live, the increased expected life span from one year to the next is 1 (the year survived) plus the slope (the negative amount by which remaining years changes). For example, each year that a United States male lives between ages 5 and

70 adds (approximately) 0.11 years (about 40 days) to his expected life span; for a woman, this figure is 0.07 years. The corresponding results for men and women in the United States in 1900 are 0.30 and 0.29, respectively and in the Roman Empire they are 0.49 years and 0.48 years, respectively.

The differences are quite striking. In the U.S. today, women gain about 25 days for each year they live, but men gain half again that much for each year they live. Why? On the other hand, both men and women living in the Roman Empire gain almost six months for each year they live, between the ages of 5 and 70. What this is reflecting is the likelihood of death. Living in the Roman Empire was much more dangerous than today and getting through a year is meaningful and more or less the same for men as for women. Again, one should ask why the historical slopes are about the same for men and women, but today men gain more by each year they live than women. A partial answer is that females live longer, so one additional year is a greater percentage of expected remaining life span for a male than for a female.

Residual Plots

Now, any good statistician will check residual plots even though the fits are so good.² We begin with the residuals from the Roman Empire in Figure 4. Here we see a concave-up shape for both men and women, which we expect since there must be some

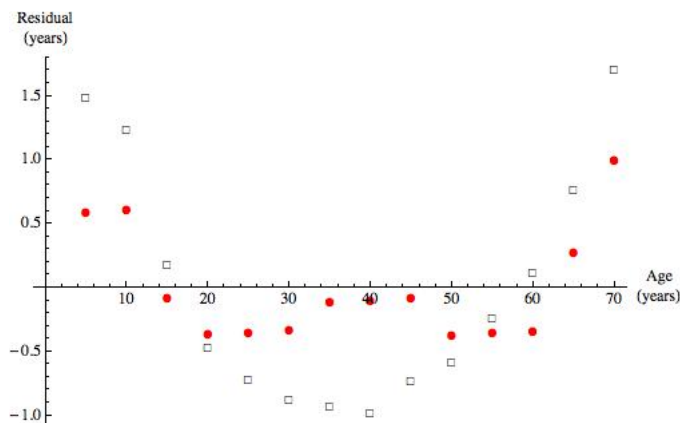


Figure 4: Residuals vs. age for Roman Empire females (circles) and Roman Empire males (squares).

concavity in our original data set. Between ages 20 and 50 there is a different pattern which is interesting, and we have no explanations but it seems something is different for men and women in terms of life expectancy for this era.

The residual plots for U.S. men and women are given in Figure 5. Here we encourage the reader to pause for a moment and try to interpret these graphs.

² Recall that the *residual* of a data point (x, y) is $y - \hat{y}$, the vertical (directed) distance from the point on the regression line (x, \hat{y}) to the actual observed point (x, y) . The residual plot shows these residuals plotted against the x -values.

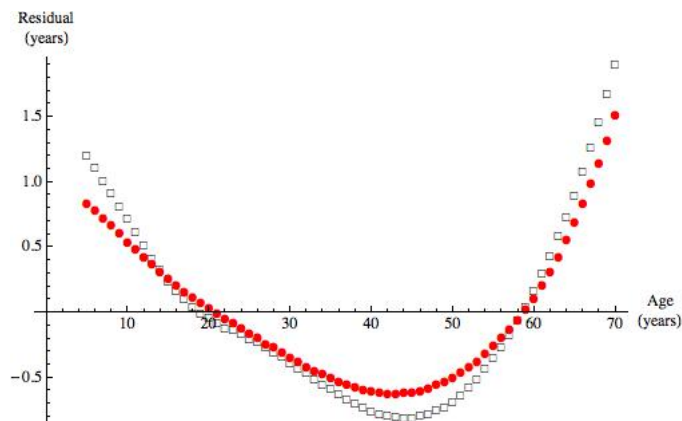


Figure 5. Residuals vs. age, for the regression line from ages 5 to 70 for U.S. females (circles) and U.S. males (squares) in 2007.

Not only do we have a pattern in the residual plots for both, but the patterns aren't quite the same. First note that a positive residual means that the data value is higher than the regression line; i.e., the probability of death is smaller (relative to the regression line). Conversely, a negative residual means that probability of death is greater. Also one should note that, for example, a residual of -0.1 means that the life expectancy is about one month less than the line predicts.

More importantly though, why is there an inflection point (a change in concavity – from bending up to bending down) in the male residual plot which is preceded by steeper declines that aren't in the female residual plot? We suspect the steeper decline in the residual plot is due to generally riskier behavior by males around age 20, which is supported by data on mortality (Miniño et al.

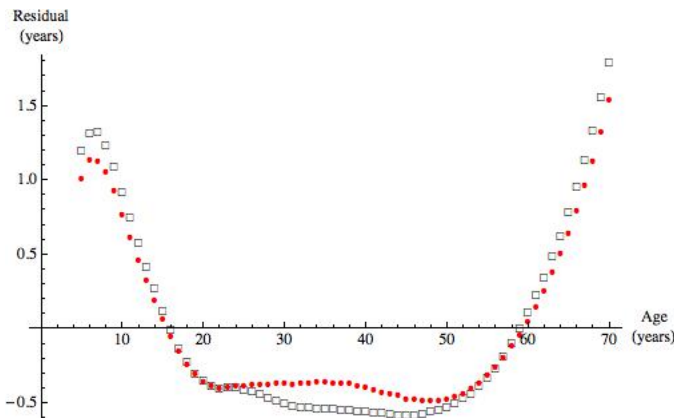


Figure 6. Residuals vs. Age, for the regression line from ages 5 to 70 for U.S. females (circles) and U.S. males (squares) in 1900.

2010, Table 1, Page 9). Still this doesn't explain the inflection point and leveling out of the residual plot, which also exists but to a lesser extent in the 1900 data seen in Figure 6. Possibly this is due to causes of death for men becoming more similar to that of women. Again the quantitative analysis brings forth some interesting questions.

At this point we realize that our choice of cutoff at age 70 is somewhat arbitrary, so we will run the same analysis with a cutoff at age 40 and at age 80. We give only the residual plots in Figure 7 for these two scenarios. Our conclusion is that the age cutoff does not change the analysis.

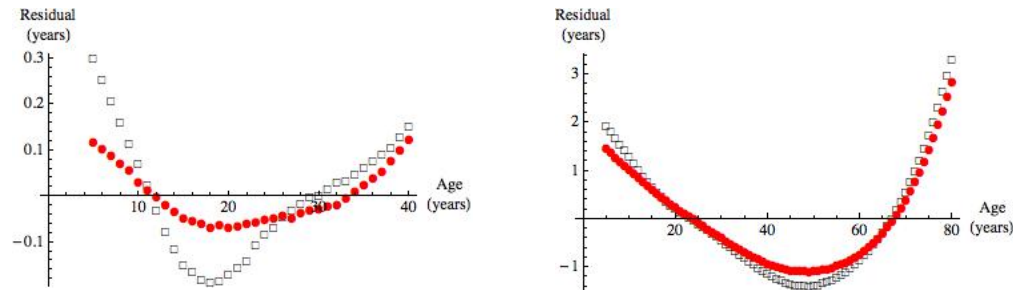


Figure 7. Residuals vs. Age for females (circles) and males (squares) using a cutoff of 40 years (left panel) and 80 years (right panel).

The residual plots with age cutoffs at 40 years and 80 years share the same features as the first residual plots we considered, but note the difference of the two vertical scales. The residuals using a cutoff of 80 years are notably larger since – as noted above – the graphs in Figure 1 start to bend up somewhere around 65. (We note that we get r^2 values well above 0.99 for both females and males, even in the through-age-80 case.)

In the end, the residual plot tells us that a line is not appropriate, which really does make sense. The data must have a horizontal asymptote at $y = 0$ and hence the data must “bend.” The pattern in the residual plot confirms the general shape of the data suggesting, despite r^2 values close to 1, that we should be using a fit with a function that is concave up and decreasing. We cannot explain the bumps, but the benefit of the residual plot is that it provided us with information we would have missed by considering only the scatter plot.

Concluding Remarks

These period life tables reaffirmed information we knew: women live longer today than men, and we live longer than Romans; but they also provided us with quantitative information that help us understand differences in life expectancy: how it changes over the course of one's life, how it differs from men and women, and how it has changed over time. In review, the changes in infant mortality are clear in the scatter plots. We know that today men increase their age at death by half again as much as women for each year they live between the ages of 5 and 70. This increase in life expectancy is much less than in our two historical examples, and in

the historical examples these were the same for men and women. Again, we can only conjecture as to why, but clearly our quantitative analysis has been fruitful.

As we noted in the introduction, this collection of examples might be presented in a discussion of life expectancy. Alternatively, one or more of these examples could add to an introduction to regression, particularly illustrating the value of residual plots in understanding a data set.

References

- Bell F. C., and M. L. Miller. 2005. *Life Tables for the United States Social Security Area 1900-2100*. Actuarial Study No. 120. Social Security Administration, Office of the Chief Actuary. SSA Pub. No. 11-11536.
http://www.ssa.gov/OACT/NOTES/pdf_studies/study120.pdf. Accessed 6 July 2011.
- Miniño A. M., J. Xu, and K. D. Kochanek. 2010. Deaths: Preliminary data for 2008. *National Vital Statistics Reports*, v. 59, no. 2, Dec. 9, 2010.
http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_02.pdf Accessed 9 December 2011.
- Social Security Online. 2011. Actuarial Publications, Period Life Table.
<http://www.ssa.gov/OACT/STATS/table4c6.html>. Accessed 6 July 2011.
- Wikipedia. 2001. "Demography of the Roman Empire." Wikimedia Foundations, Inc.
http://en.wikipedia.org/wiki/User:G.W./Demography_of_the_Roman_Empire
. Accessed 6 July 2011.

Appendix: Period Life Tables

A. Roman Empire

Age	Expected Years to Live	
	Male	Female
0	22.8	25.0
1	34.1	34.9
5	39.0	40.1
10	36.2	37.5
15	32.6	34.2
20	29.4	31.3
25	26.6	28.7
30	23.9	26.1
35	21.3	23.7
40	18.7	21.1
45	16.4	18.5
50	14.0	15.6
55	11.8	13.0
60	9.6	10.4
65	7.7	8.4
70	6.1	6.5
75	4.6	4.9
80	3.4	3.6
85	2.4	2.5
90	1.7	1.8
95	1.2	1.2

Source: Wikipedia (2011)

B. United States (1900)

Age	Expected Years to Live		Age	Expected Years to Live		Age	Expected Years to Live		Age	Expected Years to Live	
	Male	Female		Male	Female		Male	Female		Male	Female
0	46.41	48.96	30	34.57	35.80	60	14.18	14.96	90	2.80	2.87
1	53.28	54.57	31	33.86	35.09	61	13.59	14.35	91	2.65	2.69
2	54.38	55.60	32	33.15	34.39	62	13.01	13.75	92	2.50	2.52
3	54.45	55.60	33	32.45	33.68	63	12.45	13.16	93	2.37	2.37
4	54.19	55.32	34	31.74	32.98	64	11.89	12.58	94	2.25	2.25
5	3.78	54.89	35	31.04	32.27	65	11.35	12.01	95	2.13	2.13
6	53.20	54.31	36	30.34	31.55	66	10.82	11.45	96	2.02	2.02
7	52.51	53.60	37	29.63	30.84	67	10.30	10.91	97	1.91	1.91
8	51.72	52.82	38	28.93	30.13	68	9.80	10.37	98	1.81	1.81
9	50.87	51.98	39	28.22	29.41	69	9.32	9.86	99	1.71	1.71
10	50.00	51.11	40	27.52	28.69	70	8.85	9.36	100	1.61	1.61
11	49.13	50.25	41	26.81	27.96	71	8.40	8.89	101	1.52	1.52
12	48.26	49.39	42	26.11	27.24	72	7.97	8.44	102	1.43	1.43
13	47.40	48.54	43	25.40	26.52	73	7.55	8.01	103	1.35	1.35
14	46.55	47.70	44	24.69	25.80	74	7.15	7.59	104	1.26	1.26
15	45.70	46.87	45	23.99	25.07	75	6.75	7.18	105	1.19	1.19

16	44.87	46.04	46	23.29	24.36	76	6.37	6.77	106	1.11	1.11
17	4.05	45.23	47	22.60	23.64	77	5.99	6.37	107	1.04	1.04
18	43.26	44.43	48	21.91	22.93	78	5.63	5.98	108	0.97	0.97
19	42.48	43.66	49	21.22	22.22	79	5.29	5.62	109	0.91	0.91
20	41.73	42.90	50	20.54	21.52	80	4.98	5.29	110	0.84	0.84
21	40.99	42.16	51	9.87	20.83	81	4.69	4.99	111	0.78	0.78
22	40.28	41.44	52	19.20	20.14	82	4.42	4.72	112	0.72	0.72
23	39.58	40.74	53	8.53	19.47	83	4.18	4.47	113	0.67	0.67
24	38.88	40.04	54	17.88	18.80	84	3.95	4.24	114	0.62	0.62
25	8.17	39.33	55	17.24	18.14	85	3.73	4.00	115	0.57	0.57
26	37.46	38.63	56	6.60	17.49	86	3.52	3.76	116	0.52	0.52
27	36.74	37.92	57	15.98	16.84	87	3.32	3.52	117	0.50	0.50
28	36.01	37.21	58	15.37	16.21	88	3.14	3.29	118	0.00	0.00
29	35.29	36.51	59	14.77	15.58	89	2.96	3.08	119	0.00	0.00

Source: Bell and Miller (2005, Table 6)

C. United States (2007)

Age	Expected Years to Live		Age	Expected Years to Live		Age	Expected Years to Live		Age	Expected Years to Live	
	Male	Female		Male	Female		Male	Female		Male	Female
0	75.38	80.43	30	47.13	51.50	60	20.92	23.97	90	3.92	4.69
1	74.94	79.92	31	46.20	50.53	61	20.16	23.14	91	3.64	4.36
2	73.98	78.95	32	45.27	49.56	62	19.40	22.31	92	3.38	4.04
3	73.00	77.97	33	44.33	48.60	63	18.66	21.49	93	3.15	3.76
4	72.02	76.99	34	43.40	47.64	64	17.92	20.69	94	2.93	3.50
5	71.03	76.00	35	42.47	46.68	65	17.19	19.89	95	2.75	3.26
6	70.04	75.01	36	41.54	45.72	66	16.48	19.10	96	2.58	3.05
7	69.05	74.02	37	40.61	44.76	67	15.77	18.32	97	2.44	2.87
8	68.06	73.03	38	39.68	43.81	68	15.08	17.55	98	2.30	2.70
9	67.07	72.04	39	38.76	42.86	69	14.40	16.79	99	2.19	2.54
10	66.08	71.04	40	37.84	41.91	70	13.73	16.05	100	2.07	2.39
11	65.09	70.05	41	36.93	40.97	71	13.08	15.32	101	1.96	2.25
12	64.09	69.06	42	36.02	40.03	72	12.44	14.61	102	1.85	2.11
13	63.10	68.07	43	35.12	39.10	73	11.82	13.91	103	1.75	1.98
14	62.12	67.08	44	34.22	38.17	74	11.21	13.22	104	1.66	1.86
15	61.14	66.09	45	33.33	37.24	75	10.62	12.55	105	1.56	1.74
16	60.18	65.11	46	32.45	36.32	76	10.04	11.90	106	1.47	1.62
17	59.22	64.13	47	31.57	35.41	77	9.48	11.26	107	1.39	1.52
18	58.27	63.15	48	30.71	34.50	78	8.94	10.63	108	1.30	1.41
19	57.33	62.18	49	29.84	33.59	79	8.41	10.03	109	1.22	1.31
20	56.40	61.20	50	28.99	32.69	80	7.90	9.43	110	1.15	1.22
21	55.47	60.23	51	28.15	31.80	81	7.41	8.86	111	1.07	1.13
22	54.54	59.26	52	27.32	30.91	82	6.94	8.31	112	1.00	1.05
23	53.63	58.29	53	26.49	30.02	83	6.49	7.77	113	0.94	0.97
24	52.71	57.32	54	25.68	29.14	84	6.06	7.26	114	0.87	0.89
25	51.78	56.35	55	24.87	28.27	85	5.65	6.77	115	0.81	0.82
26	50.86	55.38	56	24.06	27.40	86	5.26	6.31	116	0.75	0.75
27	49.93	54.40	57	23.26	26.53	87	4.89	5.87	117	0.70	0.70
28	49.00	53.44	58	22.48	25.67	88	4.55	5.45	118	0.64	0.64
29	48.07	52.47	59	21.69	24.82	89	4.22	5.06	119	0.59	0.59

Source: Social Security Online, 2011.