Chapter 14: Survival Analysis

Section 14-1: Life Tables

1. A cohort life table is a record of the actual mortality experience of a particular group, such as all people born in 1970. This would include data of the mortality of this group at each year. A period life table, on the other hand, the mortality conditions and results for different ages are based on data from one particular year.

2. Period life tables are more practical because the data comes from one year, as opposed to being recorded over the course of decades.

3. The probability of dying during the first interval would remain the same, as would the expected remaining lifetime value. However, the number surviving to the beginning of the interval, number of deaths during the interval, person-years lived, and total person years lived would all be half of the stated values.

4. The period life table assumes that condition affecting mortality remain fixed throughout the lives of everyone in the hypothetical cohort. Over the course of the years since 2000, those conditions may have changed, and so mortality rates may have changed.

5. There were 99,529 white females alive on their first birthday, and there were 38 deaths during the interval. The probability of dying in that interval was $38/99,529 = 0.000382$.

6. There were 99,491 white females alive at their second birthday, and the probability of dying between their second and third birthdays was 0.000204. The number of deaths during the interval was $99,491(0.000204) = 20$.

7. There were 100,000 white females at birth, and there were 99,491 that survived until their second birthday. The probability of surviving until the second birthday is $99,491/100,000 = 0.99491$.

8. The values would be:

0–2	$\frac{471+38}{100,000} = 0.00509$	100,000	509	$99,588+99,510 = 199,098$	8,128,871	81.3

9. From column 2, the probability of dying between the 20th birthday and the 21st birthday is 0.000744, so the probability of surviving that same period is $1-0.000744 = 0.999256$. If 5000 people live to their 20th birthday, we expect that $5000(0.999256) = 4996.28$ will survive to their 21st birthday. Using the values from columns 3 and 4,

$$1-\frac{74}{98910} = 0.999252 \text{ and } 5000(0.999252) = 4996.26.$$

10. a. The expected remaining lifetime for someone who has just reached their 60th birthday is 23.1 years. For someone reaching their 61st birthday, the expected remaining lifetime is 22.3 years.

    b. The expected age of death for someone who has just reached their 60th birthday is 83.1 years. For someone reaching their 61st birthday, the expected age of death is 82.3 years.

    c. The expected age at death is not exactly one year different because the values for total person years lived does not drop uniformly from one year to the next.

11. There were 99,121 people alive on their 16th birthday, and 83,264 people alive on their 66th birthday. The probability of surviving is $83,264/99,121 = 0.840024$.

12. There were 99,836 people alive on their 21st birthday and 57,188 alive on their 80th birthday. The probability of surviving is $57,188/99,836 = 0.572819$.

13. a. $\frac{612+43}{100,000} = 0.006550$

    b. $\frac{27+21}{99,345} = 0.000483$

    The results are so different because of the exceptionally high mortality rate at or very near birth.

14. The values would be:

0–2	$\frac{612+43}{100,000} = 0.006550$	100,000	655	$99,465+99,366 = 198,831$	7,866,027	78.7

15. The values would be:

2–4	$\frac{27+21}{99,345} = 0.000483$	99,345	48	$99,331+99,307 = 198,638$	7,667,195	77.2

16. The values would be:

0–10	$\frac{776}{100,000} = 0.007760$	100,000	776	893,779	7,866,027	78.7

17. H_0: $p = 0.000744$; H_1: $p > 0.000744$; The test statistic is $z = 1.87$, the P-value is 0.0308 (Table: 0.0307), and the critical value is $z = 1.645$. Reject H_0 and conclude that there is sufficient evidence to support the claim that the number of deaths is significantly high. [If the probability of dying is calculated from the third column of Table 14-1, use H_1: $p > 0.00074815$ to get a test statistic of $z = 1.85$, a P-value of 0.0323 (Table: 0.0322), and the same critical value and conclusions.]

$$z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{15}{12,500} - 0.000744}{\sqrt{\frac{(0.000744)(1-0.000744)}{12,500}}} = 1.87$$

18. H_0: $p = 0.004177$; H_1: $p > 0.004177$; The test statistic is $z = 1.45$, the P-value is 0.0731 (Table: 0.0735), and the critical value is $z = 1.645$. Fail to reject H_0 and conclude that there is not sufficient evidence to support the claim that the number of deaths is significantly high.

$$z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{36}{6772} - 0.004177}{\sqrt{\frac{(0.004177)(1-0.004177)}{6772}}} = 1.45$$

19. H_0: $p = \frac{99,121-98,011}{99,121}$=0.011198; H_1: $p > 0.011198$; The test statistic is $z = 4.95$, the P-value is 0.0000 (Table: 0.0001), and the critical value is $z = -1.645$. Reject H_0 and conclude that there is sufficient evidence to support the claim that the number of deaths is significantly high.

$$z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{147}{8774} - 0.011198}{\sqrt{\frac{(0.011198)(1-0.011198)}{8774}}} = 4.95$$

20. H_0: $p = \frac{88,770-78,096}{88,770} = 0.120243$; H_1: $p < 0.120243$; The test statistic is $z = -1.45$, the P-value is 0.0711 (Table: 0.0708), and the critical value is $z = 1.645$. Fail to reject H_0 and conclude that there is not sufficient evidence to support the claim that the number of deaths is significantly low.

$$z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{484}{4285} - 0.120243}{\sqrt{\frac{(0.120243)(1-0.120243)}{4285}}} = -1.47$$

Section 14-2: Kaplan-Meier Survival Analysis

1. A survivor is a subject that successfully lasted throughout a particular time period without reaching some terminating event. A survivor could be a person or an object or some other entity such as a marriage.

2. Censored data are subjects that are removed from the study because they survived past the end of the study or were no longer available to be included in the study.

3. A life table is based on fixed intervals of time, but a table of survival data and Kaplan-Meier calculations is based on times that vary according to the terminating event.

4. It is a graphical representation of the information generated using Kaplan-Meier calculations.

5. The fast walkers are more likely to be healthier with greater longevity, so they correspond to the top (green) curve. The moderate group corresponds to the middle (black) curve. The slow group of walkers is more likely to have health issues with lower longevity, so they correspond to the bottom (red) curve.

6. The event is surviving for the given time period. The probability of survival decreases with the passage of time.

7. a. 4 years b. 1 year

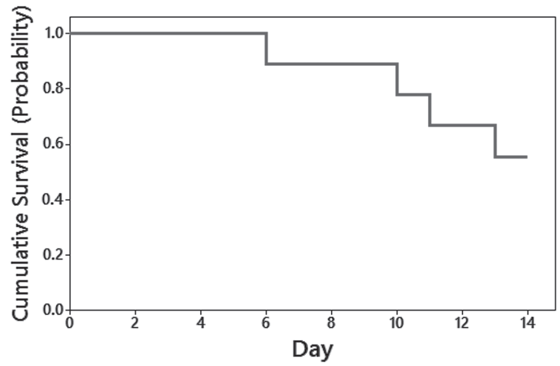
8. The red graph reaches zero, so none of the slow walkers survived past the study.

9. The five-year survival rates are 0.2 for the group of slow walkers, 0.5 for the group with moderate walking speeds, and 0.9 for the group of fast walkers. The differences are substantial, and they suggest that after five years, those with faster walking speeds have much greater survival rates, and those with slow walking speeds have much lower survival rates. The data do not necessarily suggest that we can get older people to live longer by somehow getting them to walk faster. It's very possible that walking speed is one manifestation of overall health status, so longevity and walking speed are likely both affected by one or more other extraneous variables.

10.

Day	Status 0 = Censored 1 = Failed	Number of Patients	Patients Not Requiring Retreatment	Proportion of Patients Not Requiring Retreatment	Cumulative Proportion of Patients Not Requiring Retreatment
2	0				
6	1	9	8	$8/9 = 0.889$	0.89
10	1	8	7	$7/8 = 0.875$	0.78 (0.889×0.875)
11	1	7	6	$6/7 = 0.857$	0.67 $(0.889 \times 0.875 \times 0.857)$
13	1	6	5	$5/6 = 0.833$	0.56 $(0.889 \times 0.875 \times 0.857 \times 0.833)$
14	0				
14	0				
14	0				
14	0				
14	0				

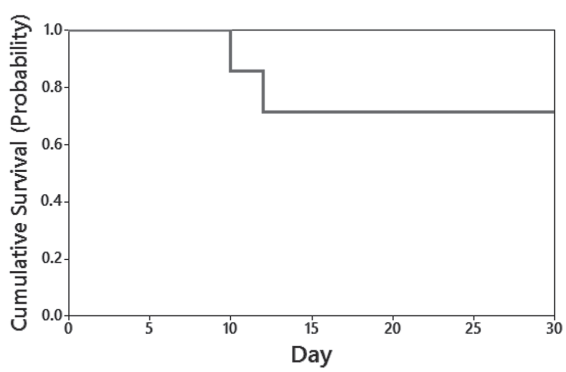
11.



12.

Day	Status 0 = Censored 1 = Failed	Number of Patients	Patients Not Requiring Retreatment	Proportion of Patients Not Requiring Retreatment	Cumulative Proportion of Patients Not Requiring Retreatment
2	0				
10	1	7	6	$6/7 = 0.857$	0.86
12	1	6	5	$5/6 = 0.833$	0.71 (0.857×0.833)
0	0				
0	0				
0	0				
0	0				
0	0				

13. The graph does not show any information about the six censored survival times. The graph shows information about the two survival times that were not censored.



14. Estimates will vary. Placebo: 0.50; Treatment: 0.80; It appears that the treatment is effective.

Chapter Quick Quiz

1. life table

2. survival table with Kaplan-Meier calculations

3. A period life table describes mortality and longevity data for a hypothetical group that would have lived with the same mortality conditions throughout their lives.

4. A cohort life table is a record of the actual observed mortality experience for a particular group.

5. false

6. true

7. false
8. true
9. The entries are 1, 0, 0, and 0.

10. $1-\dfrac{35}{98,975}=0.999345$

Review Exercises

1. $\dfrac{99,527-99,492}{99,527}=0.000352$

2. $99,527-99,492=35$

3. $\dfrac{99,492}{100,000}=0.99492$

4. The values would be:

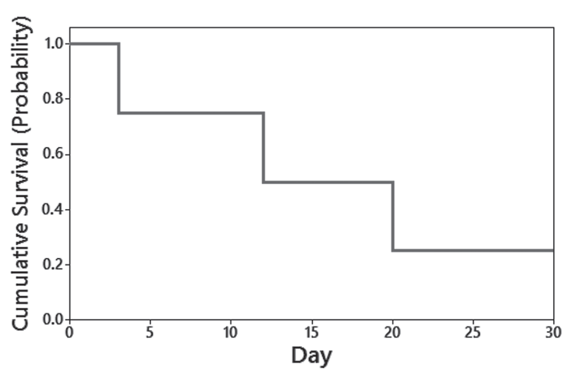
0–2	$\dfrac{473+35}{100,000}$ $=0.00508$	100,000	508	$99,588+99,510$ $=199,098$	8,382,303	83.8

5. 83.8 years; 83.2 years; the second value is less than the first value. As we age, our expected remaining lifetime steadily decreases.

6.

Day	Status 0 = Censored 1 = Failed	Number of Patients	Patients Not Requiring Retreatment	Proportion of Patients Not Requiring Retreatment	Cumulative Proportion of Patients Not Requiring Retreatment
3	1	4	3	$3/4=0.75$	0.75
12	1	3	2	$2/3=0.667$	0.50 (0.75×0.667)
20	1	2	1	$1/2=0.5$	0.25 $(0.75\times 0.667\times 0.5)$
30	0				

7.



8. The treatment group and the placebo group appear to have approximately the same behavior. The treatment does not appear to be effective.

Cumulative Review Exercises

1. The table shows that among 100,000 births, 99,492 survived to the second birthday, so the probability of dying is $1-0.99492=0.00508$. Using $H_1: p < 0.00508$, the test statistic is $z=-0.67$, the P-value is 0.2501 (Table: 0.2514), and the critical value is $z=-1.645$ (assuming a 0.05 significance level), so fail to reject the null hypothesis of $p=0.00508$ and conclude that there is not sufficient evidence to support the claim that the proportion of deaths is less than 0.00508. The program does not appear to be effective in reducing the mortality rate.

$$z=\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}=\frac{\frac{5}{1328}-0.00508}{\sqrt{\frac{(0.00508)(1-0.00508)}{1328}}}=-0.67$$

2. 95% CI: $0.99294 < p < 0.99953$; The confidence interval limits contain 0.99492, so it appears that the mortality rate has not been lowered by a significant amount.

$$z=\hat{p}\pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}=\tfrac{1323}{1328}\pm 1.96\sqrt{\frac{\left(\frac{1323}{1328}\right)\left(1-\frac{1323}{1328}\right)}{1328}}$$

3. $(1-0.004729)^3=0.986$

4. $1-(1-0.004729)^4=0.0188$; No, because being in the same family causes the events to be dependent, instead of being independent, as required by the multiplication rule. It is reasonable to expect that four Hispanic females in the same family are more likely to experience similar environmental and hereditary characteristics.

5. The graph is misleading. The vertical scale begins with a frequency of 800 instead of 0, so the difference between the "yes" and "no" responses is greatly exaggerated.

6. a. $z_{x=60}=\dfrac{60.0-68.6}{2.8}=-3.07$ and $z_{x=80}=\dfrac{80.0-68.6}{2.8}=4.07$, which have a probability of $0.9999-0.0011$ $=0.9988$, or 99.88% (Tech: 99.89%) between them.

   b. $z_{x=70}=\dfrac{70.0-68.6}{2.8/\sqrt{4}}=1.00$, which has a probability of $1-0.8413=0.1587$ to the right.