1. (Life table analysis)

(a) Three approaches

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>d</th>
<th>w</th>
<th>q</th>
<th>p</th>
<th>( \hat{S} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>146</td>
<td>27</td>
<td>3</td>
<td>0.185</td>
<td>0.815</td>
<td>0.815</td>
</tr>
<tr>
<td>1-2</td>
<td>116</td>
<td>18</td>
<td>10</td>
<td>0.155</td>
<td>0.845</td>
<td>0.689</td>
</tr>
<tr>
<td>2-3</td>
<td>88</td>
<td>21</td>
<td>10</td>
<td>0.239</td>
<td>0.761</td>
<td>0.524</td>
</tr>
<tr>
<td>3-4</td>
<td>57</td>
<td>9</td>
<td>3</td>
<td>0.158</td>
<td>0.842</td>
<td>0.441</td>
</tr>
<tr>
<td>4-5</td>
<td>45</td>
<td>1</td>
<td>3</td>
<td>0.022</td>
<td>0.972</td>
<td>0.432</td>
</tr>
</tbody>
</table>

(b) Use Greenwoods formula

\[
se \left\{ \hat{S}(t) \right\} = \hat{S}(t) \left\{ \sum_{j=1}^{t} \frac{d_j}{(n_j - w_j/2)(n_j - d_j - w_j/2)} \right\}^{1/2}
\]

and 95% confidence intervals for \( S(t) \) is \( \hat{S}(t) \pm 1.96 \times se[\hat{S}(t)] \).
2. Exponential distributions.

(a) Straightforward

(b) Likelihood function

\[ L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)^{\Delta_i} S(X_i; \theta)^{1-\Delta_i} \]

and the MLE can be solved by

\[ \hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta) = \arg \max_{\theta \in \Theta} [\log L(\theta)] = \frac{\sum_i \Delta_i}{\sum_i X_i}. \]

It is the person-year analysis estimate. A variance estimate is

\[ \text{vár}(\hat{\theta}) = \left[ -\frac{\partial^2 \log L(\hat{\theta})}{\partial \theta^2} \right]^{-1} = \frac{\hat{\theta}^2}{\sum_i \Delta_i} \]

Key assumptions are needed: (1) subjects are independent, (2) non-informative censoring.

3. Use computer to do this exercise
Supplement for Solution of Problem Set 1

Question 1-(b)

95% CI calculation for Approach 3:

<table>
<thead>
<tr>
<th>Years</th>
<th>( S(t) )</th>
<th>( V_j )</th>
<th>( \sum V_j )</th>
<th>( \left( \sum V_j \right)^{1/2} )</th>
<th>( S(t) \left( \sum V_j \right)^{1/2} )</th>
<th>95% CI LB</th>
<th>95% CI UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>0.813</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.040</td>
<td>0.032</td>
<td>0.750</td>
<td>0.877</td>
</tr>
<tr>
<td>1-2</td>
<td>0.681</td>
<td>0.0017</td>
<td>0.0033</td>
<td>0.058</td>
<td>0.039</td>
<td>0.604</td>
<td>0.758</td>
</tr>
<tr>
<td>2-3</td>
<td>0.509</td>
<td>0.0041</td>
<td>0.0074</td>
<td>0.086</td>
<td>0.044</td>
<td>0.423</td>
<td>0.595</td>
</tr>
<tr>
<td>3-4</td>
<td>0.426</td>
<td>0.0035</td>
<td>0.0109</td>
<td>0.104</td>
<td>0.045</td>
<td>0.339</td>
<td>0.514</td>
</tr>
<tr>
<td>4-5</td>
<td>0.417</td>
<td>0.0005</td>
<td>0.0114</td>
<td>0.107</td>
<td>0.045</td>
<td>0.329</td>
<td>0.504</td>
</tr>
<tr>
<td>5-6</td>
<td>0.393</td>
<td>0.0017</td>
<td>0.0131</td>
<td>0.115</td>
<td>0.045</td>
<td>0.305</td>
<td>0.481</td>
</tr>
<tr>
<td>6-7</td>
<td>0.347</td>
<td>0.0052</td>
<td>0.0184</td>
<td>0.135</td>
<td>0.047</td>
<td>0.255</td>
<td>0.439</td>
</tr>
<tr>
<td>7-8</td>
<td>0.325</td>
<td>0.0042</td>
<td>0.0225</td>
<td>0.150</td>
<td>0.049</td>
<td>0.230</td>
<td>0.421</td>
</tr>
<tr>
<td>8-9</td>
<td>0.263</td>
<td>0.0224</td>
<td>0.0449</td>
<td>0.212</td>
<td>0.056</td>
<td>0.154</td>
<td>0.373</td>
</tr>
<tr>
<td>9-10</td>
<td>0.158</td>
<td>0.1333</td>
<td>0.1783</td>
<td>0.422</td>
<td>0.067</td>
<td>0.027</td>
<td>0.289</td>
</tr>
</tbody>
</table>

\[ V_j = \frac{d_j}{\left( n_j - \frac{w_j}{2} \right) \left( n_j - d_j - \frac{w_j}{2} \right)} \]

Plot of three survival functions

Three survival estimates are quite similar, however, the first approach produces higher estimates, while the second approach leads to lower estimates which is more conservative. The 95% CIs of the third one cover other two estimates, this approach is widely used in epidemiological studies with right censored data (e.g., in a longitudinal study, when failure event occurred in the interval between two follow-up visits).
Question 2-(a)

\[ T \sim \exp(\theta) \]
\[ \therefore \text{pdf: } f(t; \theta) = \theta e^{-\theta t} \quad (t > 0) \quad \text{(see lecture -1, last two slides)} \]
\[ \therefore \lambda(t) = \theta \]
\[ \text{and } \lambda(t) = \frac{f(t)}{S(t)} \]
\[ \therefore \text{survival function: } S(t) = \frac{f(t)}{\lambda(t)} = \frac{\theta e^{-\theta t}}{\theta} = e^{-\theta t} \]
\[ \text{median of } T: S(t) = e^{-\theta t} = 0.5 \quad \Rightarrow \quad t = \frac{\log 0.5}{-\theta} = \frac{\log 2}{\theta} \]

Question 2-(b)

(1) the likelihood function:

\[ L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)^{\delta_i} S(X_i; \theta)^{1-\delta_i} \]
\[ = \prod_{i=1}^{n} \frac{f(X_i; \theta)^{\delta_i}}{S(X_i; \theta)^{\delta_i}} S(X_i; \theta) \]
\[ = \prod_{i=1}^{n} \lambda(X_i; \theta)^{\delta_i} S(X_i; \theta) \]
\[ = \prod_{i=1}^{n} \theta^{\delta_i} e^{-\theta X_i} \]

(2) the log-likelihood:

\[ l(\theta) = \log \left[ \prod_{i=1}^{n} \theta^{\delta_i} e^{-\theta X_i} \right] \]
\[ = \sum_{i=1}^{n} \left[ \log \theta^{\delta_i} + \log e^{-\theta X_i} \right] \]
\[ = \sum_{i=1}^{n} \left[ \delta_i \log \theta - \theta X_i \right] \]
(3) take first derivative:
\[
\frac{\partial}{\partial \theta} l(\theta) = \sum_{i=1}^{n} \left[ (\delta_i \log \theta)' - (\theta X_i)' \right] - \sum_{i=1}^{n} \frac{\delta_i}{\theta} - X_i \]
\[
= \sum_{i=1}^{n} \frac{\delta_i}{\theta} - \sum_{i=1}^{n} X_i \]

(4) set \( \frac{\partial}{\partial \theta} l(\theta) = 0 \), solve for MLE of \( \theta \):
\[
\sum_{i=1}^{n} \frac{\delta_i}{\theta} - \sum_{i=1}^{n} X_i = 0 \quad \Rightarrow \quad \frac{\sum_{i=1}^{n} \delta_i}{\theta} = \sum_{i=1}^{n} X_i \quad \Rightarrow \quad \hat{\theta}_{MLE} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} X_i} \]

(5) Fisher information (see lecture 1, the last two slides):
\[
I(\theta) = E \left[ -\frac{\partial^2}{\partial \theta^2} \log L(\theta) \right] = - \frac{\partial}{\partial \theta} \left( \frac{\sum_{i=1}^{n} \delta_i}{\theta} - \sum_{i=1}^{n} X_i \right) \]
\[
= E \left[ \left( \sum_{i=1}^{n} \frac{\delta_i}{\theta} - \sum_{i=1}^{n} X_i \right)' \right] = E \left[ - \left( \frac{\sum_{i=1}^{n} \delta_i}{\theta^2} \right) \right] = E \left[ \frac{\sum_{i=1}^{n} \delta_i}{\theta^2} \right] = \frac{\sum_{i=1}^{n} \delta_i}{\hat{\theta}_{MLE}^2} \]
\[
\text{(6) Var}(\hat{\theta}) = \frac{1}{I(\theta)} = \frac{\hat{\theta}_{MLE}^2}{\sum_{i=1}^{n} \delta_i} \]

Key assumption:
(1) subjects are independent;
(2) non-informative censoring (ie, failure time is independent of censoring time).
Question 2- (c).

```
. gen z01=z1
. replace z01=0 if z1==2
. streg z01, d(exponential) nohr
```

```
|      _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval] |
|----------+------------------------------------------|
|   z01 |   .0450511   .1790287     0.25   0.801    -.3058388    .3959409 |
|   _cons |  -8.541941   .1290994   -66.17   0.000    -8.794971   -8.288911 |
```

```
. * hazard rate for two treatment groups
. predictnl haz = predict(hazard), ci(haz_lb haz_ub)
```

```
|      _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval] |
|----------+------------------------------------------|
|  _t       |      haz     haz_lb     haz_ub   z1 |
|  ------------------------------- |
| 1. |  400 .0002041 .0001545 .0002537 1 |
| 2. | 4500 .0002041 .0001545 .0002537 1 |
| 3. | 1012 .0002041 .0001545 .0002537 1 |
| 4. | 1925 .0002041 .0001545 .0002537 1 |
| 5. | 1504 .0001951 .0001457 .0002445 2 |
| 6. | 2503 .0001951 .0001457 .0002445 2 |
| 7. | 1832 .0001951 .0001457 .0002445 2 |
| 8. | 2466 .0001951 .0001457 .0002445 2 |
| 9. | 2400 .0002041 .0001545 .0002537 1 |
| 10. |   51 .0001951 .0001457 .0002445 2 |
```

```
. * plot survival functions
. stcurve, survival at1(z01=1) at2(z01=0)
```

![Exponential regression graph](attachment:image.png)
Plot of exponential survival curve versus Kaplan Meier survival curve

- * generate Kaplan Meier survival estimate using
  - `sts gen st= s`
- * plot exponential survival curve and Kaplan Meier survival curve in one graph
  - `stcurve, survival addplot(line st _t, sort)`
Question 3-b

```
. stsg graph, by(z1) yline(0.5, lpattern(dash))
```

![Kaplan-Meier survival estimates graph]

```
. stci, by(z1)
```

<table>
<thead>
<tr>
<th>z1</th>
<th>no. of subjects</th>
<th>50%</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>158</td>
<td>3282</td>
<td>272.7255</td>
<td>2540     4191</td>
</tr>
<tr>
<td>2</td>
<td>154</td>
<td>3428</td>
<td>174.1291</td>
<td>3090     3853</td>
</tr>
</tbody>
</table>

```
| total | 312 | 3395 | 151.7444 | 3086 | 3839 |
```

```
. * estimate median time based on exponential regression model
. predict mt, median time
. list mt z1 in 1/10
```

```
+-----------------+
<table>
<thead>
<tr>
<th>mt   z1</th>
</tr>
</thead>
</table>
1. 3396.08   1  |
2. 3396.08   1  |
3. 3396.08   1  |
4. 3396.08   1  |
5. 3552.576  2  |
6. 3552.576  2  |
7. 3552.576  2  |
8. 3552.576  2  |
9. 3396.08   1  |
10. 3552.576  2  |
+-----------------+