1. The survival times of 11 patients with Acute Myelogenous Leukemia are given below:
   \[ 9, 13, 13^+, 18, 23, 28^+, 31, 34, 45^+, 48, 161^+, \]
where censored data are marked by “+” in the superscript.
   (a) Calculate the Kaplan-Meier estimate of \( S(20) \).
   (b) Obtain the standard error of \( \log \hat{S}(20) \).
   (c) Assuming an exponential distribution, estimate \( S(20) \).
   (d) Obtain the standard error of \( \log \hat{S}(20) \) for the exponential estimate.

2. [Methods qualifying exam, August 2005: use paper and pencil.] Consider the survival model characterized by the following piecewise-constant hazard function:
   \[
   \lambda(t) = \begin{cases} 
   \lambda_1, & 0 \leq t < \pi_1 \\
   \lambda_2, & \pi_1 \leq t < \pi_2 \\
   \lambda_3, & \pi_2 \leq t 
   \end{cases}
   \]
where \( \pi_1 \) and \( \pi_2 \) are known constants. Based on right-censored data \((x_i, \delta_i), i = 1, \ldots, n\) derive the MLEs of \( \lambda_1, \lambda_2, \) and \( \lambda_3 \) and their standard errors.

3. [Methods qualifying exam, January 2010: use paper and pencil.] Researcher in a cancer center exposed 8 rats to carcinogen, and recorded the following time to mortality in days:
   \[
   \begin{array}{cccccccc}
   x_i = \min(t_i, c_i) & 143 & 188 & 190 & 206 & 213 & 220 & 216 & 244 \\
   x_i^2 & 20449 & 35344 & 36100 & 42436 & 45369 & 48400 & 46656 & 59536 \\
   \delta_i = I_{[t_i \leq c_i]} & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
   \text{Sum} & 1620 & 334290 
   \end{array}
   \]
In the table, \( x_i = \min(t_i, c_i) \), with \( t_i \) and \( c_i \) being the death time and censoring time of the \( i \)th rat, respectively, and \( \delta_i \) indicates whether the observation was censored. “Sum” indicates the sum of all the values in the row.

   (a) The researcher considers modeling these data with a family of lifetime distributions, which has a hazard function \( \lambda(t) = 2t/\beta^2 \). Write down the survival function \( S(t) \) for this family of models.
   (b) Obtain the MLE \( \hat{\beta} \) of \( \beta \).
(c) Obtain the standard error of $\hat{\beta}$.

4. [Methods qualifying exam, January 2011: use paper and pencil.] A clinical trial studies the survival of lung cancer patients, where the patients were randomly assigned to an experimental treatment or to a placebo. The following table reports survival times of patients in months. ' + ' indicates a censored observation.

<table>
<thead>
<tr>
<th></th>
<th>Survival times $t_i$, months</th>
<th>$\sum_{i=1}^{16} t_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>1 1 1+ 2 2 3 3 3+</td>
<td>16</td>
</tr>
<tr>
<td>Treatment</td>
<td>1 2 2+ 2 3 3 4 5 5+</td>
<td>27</td>
</tr>
</tbody>
</table>

(a) The researchers would like to obtain the Kaplan-Meier estimate of the survival function. Compute the estimate of the survival function for the treatment group at 2 months $\hat{S}_{trt}^{KM}(2)$.

(b) The researchers now assume that the survival function for each group follows an exponential distribution.
   i. State the assumptions of the model, and interpret the parameters.
   ii. State the estimated survival function for each group, according to the Maximum Likelihood estimation procedure.

(c) To determine the effectiveness of the treatment, the researchers fit the accelerated failure time model given in the partial output below. In the output, the variable $trt=0$ for placebo, and $trt=1$ for the treatment group.

Call:
```
survreg(formula = Surv(time, obs) ~ trt, data = x, dist = "exponential")
```

```
Value Std. Error
(Intercept) 0.981 0.408
trt 0.369 0.556
```

Scale fixed at 1
Exponential distribution
Loglik(model)= -28.3 Loglik(intercept only)= -28.6
Chisq= 0.43 on 1 degrees of freedom, p= 0.51

   i. State the model, and interpret the parameters.
   ii. State the null and the alternative hypotheses regarding the in terms of the model parameters, and conclude at the confidence level of 95% whether the treatment is effective.

(d) Finally, the researchers fit the model below

Call:
```
coxph(formula = Surv(time, obs) ~ trt, data = x, method = "exact")
```

```
  coef exp(coef) se(coef)
trt -0.8198 0.4410 0.7376
```

   i. State the model and the assumptions.
   ii. Test whether the treatment is effective. State the null and the alternative hypotheses, and your conclusions at the confidence level of 95%.
   iii. Discuss how you can assess the plausibility of the assumption of “proportional hazard” for this dataset.
5. [Methods qualifying exam, August 2010: use paper and pencil.] A randomized clinical trial studies survival times of leukemia patients, where the patients are randomly assigned to an experimental treatment or to a placebo. The following table reports survival times of the patients in weeks, where ‘+’ indicates a censored observation.

<table>
<thead>
<tr>
<th></th>
<th>Survival times $t_i$, $i = 1, \ldots, 16$</th>
<th>$\sum_{i=1}^{16} t_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>10 10+ 12 12+ 13 14 17+ 20 26 34+ 35 37 38 39 40 40+</td>
<td>397</td>
</tr>
<tr>
<td>Placebo</td>
<td>7 9 9+ 10 11 11+ 13 14 14+ 15 16 17 19 21+ 22 27+</td>
<td>235</td>
</tr>
</tbody>
</table>

(a) Compute the Kaplan-Meier estimates of the survival at time 12 for both groups.
(b) Compute the confidence intervals associated with the estimates of survival above.
(c) Based on these results, does the treatment have an effect on the survival?
(d) Now assume that the survival function follows an exponential distribution. Compute model-based estimates of the survival at time 12 for both groups.
(e) Discuss the assumptions underlying these two approaches, and the reasons for the discrepancy between these two estimates of survival.

6. [Methods qualifying exam, August 2008: use paper and pencil.] To assess the gender risk of death from coronary heart disease (CHD), a study was performed to control for genetic factors with twins consisting of a male and female. The age at which a male/female twin died of CHD was recorded. The dataset `twin` contains the following samples from this study.

<table>
<thead>
<tr>
<th>gender = &quot;male&quot;</th>
<th>age</th>
<th>cens</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>49</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>61</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>74</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>gender = &quot;female&quot;</th>
<th>age</th>
<th>cens</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>52</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>73</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>74</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>0</td>
</tr>
</tbody>
</table>

(a) Calculate the Kaplan-Meier estimates of the survival functions for the male group.
(b) Although the evidence is substantial that males are at higher risk than females, the role of genetic factors versus the gender factor in CHD is still largely unknown. Using the output below, state the hypothesis that gender is a risk factor of death from CHD, and test the hypothesis while controlling for genetic factors.

```r
> survdiff(Surv(age,cens)~gender,data=twin);
```

```
          N  Observed Expected (O-E)^2/E (O-E)^2/V
gender=female 12 7   7.91    0.105  0.299
gender=male   12 8   7.09    0.117  0.299
```

(c) An accelerated failure time model is fitted to the data as shown below. Using the output below, predict the probability for a female with CHD to survive past 75.

```r
> aftmod <- survreg(Surv(age,cens)~gender,dist="exponential",data=twin)
> summary(aftmod)
```
Value Std. Error  z   p
(Intercept)  4.772   0.378 12.625 1.53e-36
gendermale -0.174   0.518 -0.337 7.36e-01

(d) A proportional hazards model is fitted to the data as shown below. For a twin consisting of a male and female, what is the effect of gender on the survival?

> coxmod <- coxph(Surv(age,cens)~gender,data=twin)
> summary(coxmod)

    coef exp(coef) se(coef)      z   p
gendermale  0.26    1.30    0.522 0.499 0.62

    exp(coef) exp(-coef) lower .95 upper .95
gendermale    1.30      0.771   0.467 3.60

7. Consider the dataset mgus in the library survival (use ?mgus for information on the dataset). Our goal is to model the survival time of these patients using Cox proportional hazard model (ignore variables pcdx and pctime, which have a large proportion of missing data).

Verify the plausibility of a proportional hazard model for this dataset. Select important predictors in the model, and their functional form. Evaluate the quality of fit. Interpret model parameters and discuss the results.