Problem 1

Given the subset of data from the Worcester Heart Attack Study (WHAS), steps were taken to fit a parametric model to the data.

Graphical checking for both the Weibull and lognormal survival time model (Figure 1) was performed, ignoring the effects of covariates (gender, chf, ord), using the appropriate transformation of the estimated cumulative hazard rates under each distribution.

![Hazard plots versus log time for the Weibull (solid line) and lognormal (dashed line) survival time models.](image)

**Figure 1.** Hazard plots versus log time for the Weibull (solid line) and lognormal (dashed line) survival time models.

The lognormal survival time model (dashed line) appears to be a better fit than the Weibull survival time model. The lognormal model conforms to a line with constant slope better than the Weibull model. But both models look like lack of fit. This can also been observed from the probability plots (Figure 3) in Appendix.

Based on these diagnostic plots alone, the lognormal model appears to be a better fit.
Problem 2

Assumption of lognormal survival times. Fitting the accelerated failure time model under the assumption of lognormal-distributed survival times, the model fit was obtained from LIFEREG SAS procedure (Table 1).

The fitted model (gender, chf and ord are claimed under CLASS) is

\[ \log(\hat{Y}) = 4.7774 + 0.8539(1 - \text{gender}) + 2.9658(1 - \text{chf}) + 1.0233(1 - \text{ord}) + 3.6091W \]

which simplifies to (you get this model directly if gender, chf and ord are not claimed under CLASS)

\[ \log(\hat{Y}) = 9.6204 - 0.8539(\text{gender}) - 2.9658(\text{chf}) - 1.0233(\text{ord}) + 3.6091W \]

where \( W \) follows a standard normal distribution \( N(0,1) \).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.3908</td>
<td>4.0114 5.5434</td>
<td>149.44</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>gender</td>
<td>0</td>
<td>0.3744</td>
<td>0.1200 1.5878</td>
<td>5.20</td>
<td>0.0226</td>
</tr>
<tr>
<td>gender</td>
<td>1</td>
<td>0.0000</td>
<td>. . .</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>chf</td>
<td>0</td>
<td>0.3801</td>
<td>2.2209 3.7107</td>
<td>60.90</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>chf</td>
<td>1</td>
<td>0.0000</td>
<td>. . .</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>ord</td>
<td>0</td>
<td>0.3811</td>
<td>0.2764 1.7702</td>
<td>7.21</td>
<td>0.0072</td>
</tr>
<tr>
<td>ord</td>
<td>1</td>
<td>0.0000</td>
<td>. . .</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Scale</td>
<td>1</td>
<td>0.1756</td>
<td>3.2808 3.9703</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Table 1. Partial SAS output for an AFT model fit assuming lognormal-distributed survival times.

Assumption of Weibull survival times. Fitting the accelerated failure time model under the assumption of Weibull-distributed survival times, the model fit was obtained from the LIFEREG SAS procedure (Table 2).

The fitted model is

\[ \log(\hat{Y}) = 6.1405 + 0.6856(1 - \text{gender}) + 2.6066(1 - \text{chf}) + 1.0344(1 - \text{ord}) + 2.5878W \]

which simplifies to

\[ \log(\hat{Y}) = 10.4671 - 0.6856(\text{gender}) - 2.6066(\text{chf}) - 1.0344(\text{ord}) + 2.5878W \]

where \( W \) follows an extremevalue(0,1) distribution.
### Analysis of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>6.1405</td>
<td>0.3216</td>
<td>5.5102 – 6.7709</td>
<td>364.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>gender 0 to 1</td>
<td>1</td>
<td>0.6856</td>
<td>0.3328</td>
<td>0.0333 – 1.3378</td>
<td>4.24</td>
<td>0.0394</td>
</tr>
<tr>
<td>chf 0 to 1</td>
<td>1</td>
<td>2.6066</td>
<td>0.3514</td>
<td>1.9179 – 3.2953</td>
<td>55.03</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ord 0 to 1</td>
<td>1</td>
<td>1.0344</td>
<td>0.3347</td>
<td>0.3783 – 1.6904</td>
<td>9.55</td>
<td>0.0020</td>
</tr>
<tr>
<td>Scale</td>
<td>1</td>
<td>2.5878</td>
<td>0.1434</td>
<td>2.3215 – 2.8848</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull Shape</td>
<td>1</td>
<td>0.3864</td>
<td>0.0214</td>
<td>0.3466 – 0.4308</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Partial SAS output for an AFT model fit assuming Weibull-distributed survival times.

**Assumption of generalized gamma survival times.** Fitting the accelerated failure time model under the assumption of generalized gamma-distributed survival times, the model fit was obtained from the LIFEREG SAS procedure (Table 3).

The fitted model is

\[
\log(\tilde{Y}) = 4.8139 + 0.8490(1 - \text{gender}) + 2.9640(1 - \text{chf}) + 1.0274(1 - \text{ord}) + 3.5865W
\]

which simplifies to

\[
\log(\tilde{Y}) = 9.6542 - 0.8490(\text{gender}) - 2.9640(\text{chf}) - 1.0274(\text{ord}) + 3.5865W
\]

where \(W\) follows a log-gamma distribution.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>9.6542</td>
<td>0.3424</td>
<td>8.9832 – 10.3252</td>
<td>795.17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>gender</td>
<td>1</td>
<td>-0.8491</td>
<td>0.3737</td>
<td>-1.5815 – -0.1166</td>
<td>5.16</td>
<td>0.0231</td>
</tr>
<tr>
<td>chf</td>
<td>1</td>
<td>-2.9640</td>
<td>0.3794</td>
<td>-3.7076 – -2.2204</td>
<td>61.03</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ord</td>
<td>1</td>
<td>-1.0274</td>
<td>0.3802</td>
<td>-1.7725 – -0.2823</td>
<td>7.30</td>
<td>0.0069</td>
</tr>
<tr>
<td>Scale</td>
<td>1</td>
<td>3.5866</td>
<td>0.1749</td>
<td>3.2596 – 3.9464</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>0.0279</td>
<td>0.0000</td>
<td>0.0279 – 0.0279</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Partial SAS output for an AFT model fit assuming gamma-distributed survival times.

Note that the maximum likelihood estimates using the gamma distribution may not be valid in light of the following SAS warning when PROC LIFEREG was used:

**WARNING:** The negative of the Hessian is not positive definite. The convergence is questionable.

**WARNING:** The procedure is continuing in spite of the above warning. Results shown are based on the last maximum likelihood iteration. Validity of the model fit is questionable.
Problem 3

The log likelihood values of the three preceding model fits are reported below and were obtained from SAS output (Table 4).

<table>
<thead>
<tr>
<th>Probability distribution</th>
<th>Log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>-818.6479848</td>
</tr>
<tr>
<td>Weibull</td>
<td>-826.1900361</td>
</tr>
<tr>
<td>Generalized gamma</td>
<td>-818.6415518</td>
</tr>
</tbody>
</table>

Table 4. Log likelihood values for AFT model fits under three distributions.

The likelihood ratio test is appropriate for nested model comparisons. In this case, the initial model is the generalized gamma model. Comparing the initial model to the lognormal model, the null and alternative hypotheses are

\[ H_0: \beta = \infty \text{ (lognormal model)} \]
\[ H_a: \beta \neq \infty \text{ (generalized gamma model)} \]

Test at the \( \alpha = .05 \) significance level (two-sided).

Calculating the log-likelihood ratio test statistic manually,

\[ \Lambda = 2(L_1 - L_0) = 2(-818.6415517 - (-818.6479848)) = 0.0128662 \]

This test statistic has residual degrees of freedom \( df = 3 - 2 = 1 \) with corresponding \( p \)-value .9096902. Therefore, conclude that the lognormal model, the reduced model fits the data as equally well as does the generalized gamma model. The lognormal model, which has fewer parameters than the generalized gamma model and therefore is more convenient to use, appears to be better for the given data.

Comparing the initial model to the Weibull model, the null and alternative hypotheses are

\[ H_0: \beta = 1 \text{ (Weibull model)} \]
\[ H_a: \beta \neq 1 \text{ (generalized gamma model)} \]

Test at the \( \alpha = .05 \) significance level (two-sided).

Calculating the log-likelihood ratio test statistic manually,

\[ \Lambda = 2(L_1 - L_0) = 2(-818.6415517 - (-826.1900361)) = 15.09697 \]

This test statistic also has residual degrees of freedom \( df = 3 - 2 = 1 \) with corresponding \( p \)-value .000102. Therefore, conclude that the reduced model fits the data significantly worse than the generalized gamma model. Overall, among the three models, the lognormal model appears to be the most suitable one.
**Problem 4**

Based on LR testing of nested models, the initial model chosen is the lognormal model. Each covariate under this initial model is significant at the $\alpha = .05$ level (Table 4). Therefore, using a backward model selection procedure, the full model is the final model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Limits</th>
<th>Chisquare</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.7774</td>
<td>0.3908</td>
<td>4.0114</td>
<td>5.5434</td>
<td>149.44</td>
</tr>
<tr>
<td>gender</td>
<td>0</td>
<td>0.8539</td>
<td>0.3744</td>
<td>0.1200</td>
<td>1.5878</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>chf</td>
<td>0</td>
<td>2.9658</td>
<td>0.3801</td>
<td>2.2209</td>
<td>3.7107</td>
<td>60.90</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>ord</td>
<td>0</td>
<td>1.0233</td>
<td>0.3811</td>
<td>0.2764</td>
<td>1.7702</td>
<td>7.21</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Scale</td>
<td>1</td>
<td>3.6091</td>
<td>0.1756</td>
<td>3.2808</td>
<td>3.9703</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4. Partial SAS output for an AFT model fit assuming lognormal-distributed survival times.*

A Cox-Snell residual analysis for the final model (Figure 2) indicates that the model fits the data well except the large end of the survival curve since the residuals below 1.5 more or less follow the line with the slope 1.

*Figure 2. Cumulative hazard plot versus Cox-Snell residuals of the lognormal model.*
The fitted model is
\[
\log(\hat{y}) = 4.7774 + 0.8539(1 - gender) + 2.9658(1 - chf) + 1.0233(1 - ord) + 3.6091W
\]
which simplifies to
\[
\log(\hat{y}) = 9.6204 - 0.8539(gender) - 2.9658(chf) - 1.0233(ord) + 3.6091W
\]
where \( W \) follows \( N(0,1) \).

Controlling for other covariates, the typical time to death for female admitted patients is
\[
\exp(\hat{\beta}_{age}) = \exp(-0.8539) = 0.43
\]
times that for male admitted patients.

Controlling for other covariates, the typical time to death for patients with previous heart failure complications is
\[
\exp(\hat{\beta}_{chf}) = \exp(-2.9658) = 0.05
\]
times that for patients with no previous heart failure complications.

Controlling for other covariates, the typical time to death for patients with at least one previous heart attack (recurrent) is
\[
\exp(\hat{\beta}_{ord}) = \exp(-1.0233) = 0.36
\]
times that for patients without a previous heart attack.
Appendix

Figure 3: Probability Plots for Lognormal and Weibull distributions
SAS code for problem set:

```sas
data whas;
infile 'c:\_6652\ps3.dat' firstobs=2;
input id gender chf ord time stat;
run;

proc lifetest data=whas outsurv=estimate;
time time*stat(0);
run;

data check;
set estimate;
s=survival;
logneglog=log(-log(s)); /* Weibull */
lnorm=probit(1-s); /* log-normal */
ltime=log(time);
run;

symbol1 color=black line=1 i=join;
symbol2 color=red line=2 i=join;
legend1 label=none shape=symbol(4,2) position=(top center inside) mode=share;
proc gplot data=check;
plot logneglog*ltime lnorm*ltime / overlay legend=legend1;
run;

proc lifereg data=whas;
class gender chf ord;
model time*stat(0) = gender chf ord / dist=lnormal; /* log-normal */
output out=r cdf=f;
probplot;
run;

proc lifereg data=whas;
class gender chf ord;
model time*stat(0) = gender chf ord / dist=weibull; /* Weibull */
probplot;
run;

proc lifereg data=whas;
class gender chf ord;
model time*stat(0) = gender chf ord / dist=gamma; /* generalized gamma */
run;

data resid;
set r;
e = -log(1-f);
run;

proc lifetest data=resid plots=(ls) graphics;
time e*stat(0);
symbol1 v=none;
run;
```