

Stat 574 Biostatistical Methods Homework 6

Erik Erhardt

May 1, 2005

In this problem, we will use the data on cancer of the tongue (see Section 1.11). Recall that the data set includes survival times of patients who have either an aneuploid or a diploid DNA profile. Fit the exponential, Weibull, log logistic, log normal, and generalized gamma models to this data, and use these models to answer the following questions:

Handout 6.1 *Which model fits the data best?*

Using the AIC criterion, $AIC = -2\ln L + 2(p + k)$, $p = \#covar$, $k = \#params$, in the table below, we choose the Log Logistic model having the smallest AIC value of 253.2076. However, this is only marginally better than the Log Normal and the Weibull model, which are essentially equivalent in terms of AIC. Additionally, the Gamma fits best having the largest Log Likelihood, it is because it is a three parameter model that it come in forth using the AIC.

Dist	Log Likelihood	AIC
LLOGISTIC	-123.6038176	253.2076
LNORMAL	-123.6057011	253.2114
WEIBULL	-123.8016154	253.6032
GAMMA	-123.1824481	254.3649
EXPONENT	-125.7248870	257.4498

Handout 6.2 *For the Weibull model:*

1. Find the MLEs of lambda and alpha and their standard errors.

From the fit of the model below, we observe (intercept) $\hat{\mu} = 5.64052$ with $\text{Var}(\hat{\mu}) = 0.070286$ and (scale) $\hat{\sigma} = 1.24085$ with $\text{Var}(\hat{\sigma}) = 0.020600$ and $\text{Cov}(\hat{\mu}, \hat{\sigma}) = -0.002485$. Using these values and the delta method described on p. 376 (1st edition) we can calculate $\hat{\lambda} = 0.01061284$ with $\text{Var}(\hat{\lambda}) = 0.00003917062$ or $\text{SE}(\hat{\lambda}) = 0.006258643$ and $\hat{\alpha} = 0.8058992$ with $\text{Var}(\hat{\alpha}) = 0.008689406$ or $\text{SE}(\hat{\alpha}) = 0.09321698$.

Analysis of Parameter Estimates						
Variable	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Label
Intercept	1	5.64052	0.52804	114.1064	<.0001	Intercept
group	1	-0.66896	0.35092	3.6341	0.0566	
Scale	1	1.24085	0.14353			Extreme value scale

Estimated Covariance Matrix				
	Intercept	group1	Scale	
Intercept	0.070286	-0.071009	-0.002485	
group1	-0.071009	0.123143	0.008478	
Scale	-0.002485	0.008478	0.020600	

2. Find a point estimate and 95% CI for the relative risk of death for an individual with an aneuploid vs. diploid tumor.

From the PH model, $h(x|z) = h_0(x)e^{\beta z}$, the 95% CI for the RR of Aneuploid vs. Diploid Tumor is $e^{\hat{\beta} \pm 1.96\text{SE}(\hat{\beta})} = e^{(\hat{\gamma} \pm 1.96\text{SE}(\hat{\gamma}))/\sigma}$ where $\hat{\beta} = -\hat{\gamma}/\sigma = \text{group}/\text{scale} = 0.66896/1.24085 = 0.5391143$ giving an estimate of $e^{0.5391143} = 1.714488$ with CI (0.984929, 3.129792). (Note, the group1 variable is coded 1=Aneuploid Tumor, 2=Diploid Tumor, so the sign is reversed from the estimate in the table above.)

3. Find a point estimate and 95% CI for the acceleration factor and interpret.

The median lifetime for a Aneuploid vs. Diploid Tumor patient is $e^{-\hat{\gamma}} = e^{-0.66896} = 0.512241$ with a 95% CI $e^{-\hat{\gamma} \pm 1.96\text{SE}(\hat{\gamma})}$ giving (0.2574929, 1.0190219).

Handout 6.3 *For the log logistic model:*

1. Find a point estimate and 95% CI for the relative odds and interpret.

From the PH model, $h(x|z) = h_0(x)e^{\beta z}$, the 95% CI for the RR of Aneuploid vs. Diploid Tumor is $e^{\hat{\beta} \pm 1.96SE(\hat{\beta})} = e^{(\hat{\gamma} \pm 1.96SE(\hat{\gamma}))/\sigma}$ where $\hat{\beta} = -\hat{\gamma}/\sigma = \text{group}/\text{scale} = 0.79059/0.95954 = 0.823926$ giving an estimate of $e^{0.823926} = 2.279431$ with CI (0.9982698, 5.0740681). (Note, the group1 variable is coded 1=Aneuploid Tumor, 2=Diploid Tumor, so the sign is reversed from the estimate in the table above.)

Analysis of Parameter Estimates						
Variable	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Label
Intercept	1	5.26012	0.58232	81.5948	<.0001	Intercept
group	1	-0.79059	0.40421	3.8255	0.0505	
Scale	1	0.95954	0.11140			Logistic scale

Estimated Covariance Matrix			
	Intercept	group	Scale
Intercept	0.339101	-0.221861	0.007188
group	-0.221861	0.163388	-0.003143
Scale	0.007188	-0.003143	0.012410

2. Find a point estimate and 95% CI for the acceleration factor and interpret.

The median lifetime for a Aneuploid vs. Diploid Tumor patient is $e^{-\hat{\gamma}} = e^{-0.79059} = 0.4535771$ with a 95% CI $e^{-\hat{\gamma} \pm 1.96SE(\hat{\gamma})}$ giving (0.2053906, 1.0016630).

3. Compare your results to the results in part (2).

Both the Weibull and Log Logistic models have RR estimates around 2 and have 95% CI including 1, though barely. Both have AFT estimates around 1/2 and the 95% CI including 1, though barely.

Handout 6.4 *Check the fit of the first four models graphically. Interpret all four plots.*

The plots in Figure 1 on page 4 give the Nelson-Aalen estimator of the cumulative hazard function to compare what's expected under the parametric model, under a variety of models.

Each of the four plots shows roughly parallel lines for each group, which is an indication the model fits similarly for both groups. For the four plots, the slope and intercept is what is given from the model fits. For example, for the exponential model, the intercept is at 0.

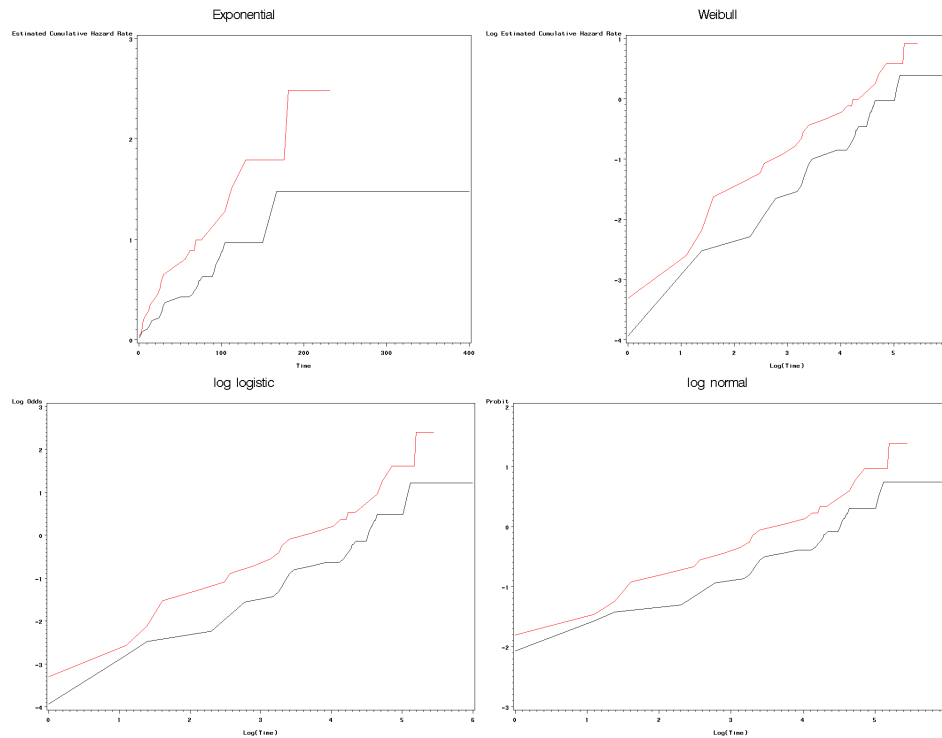


Figure 1: 6.4: Graphical model fits under four models.

Appendix

SAS code used for the above analysis

```

options nocenter linesize=256 pagesize=70 nodate nonumber ;
/*****
/*
Tumor DNA profile (1=Aneuploid Tumor, 2=Diploid Tumor)
Time to death or on-study time, weeks
Death indicator (0=alive, 1=dead)
*/
data tongue;
  infile 'F:\USERS\Erik\UNM\biostat_stat574\tongue.txt';
  input group time death;
  group1=(group=1);
  group2=(group=2);
run;
proc print;run;
/* Can only use one model statement at a time */
/* The others are listed below */
/* 2 and 3 */
proc lifereg data=tongue;
  *model time*death(0) = group1 / dist=weibull covb;
  *model time*death(0) = group / dist=exponential covb;
  model time*death(0) = group / dist=llogistic covb;
  *model time*death(0) = group / dist=lnormal covb;
  *model time*death(0) = group / dist=gamma covb;
run;
/*
# R code to compute estimates and standard errors of lambda and alpha
#####
# problem 2
m = 5.64052
s = 1.24085
m.v = 0.070286
s.v = 0.020600
ms.cv = -0.002485
l = exp(-m/s)
a = 1/s
l.v = exp(-2*m/s)*(m.v/s+m^2*s.v/s^4-2*m*ms.cv/s^3)
a.v = s.v/s^4
l.se = sqrt(l.v)
a.se = sqrt(a.v)
g = 0.66896
g.se = 0.35092
b.h = exp(g/s)
b.ci = c(exp((g-1.96*g.se)/s),exp((g+1.96*g.se)/s))
# aft
aft = exp(-g)
aft.ci= c(exp(-g-1.96*g.se),exp(-g+1.96*g.se))
#####
# problem 3
m = 5.26012
s = 0.95954
m.v = 0.339101
s.v = 0.007188
ms.cv = -0.002485
l = exp(-m/s)
a = 1/s
l.v = exp(-2*m/s)*(m.v/s+m^2*s.v/s^4-2*m*ms.cv/s^3)
a.v = s.v/s^4
l.se = sqrt(l.v)
a.se = sqrt(a.v)
# prop odds
g = 0.79059
g.se = 0.40421
b.h = exp(g/s)
b.ci = c(exp((g-1.96*g.se)/s),exp((g+1.96*g.se)/s))
# aft
aft = exp(-g)
aft.ci= c(exp(-g-1.96*g.se),exp(-g+1.96*g.se))
*/
/* 4 */
/* plots */
data tongue;

```

```

    set tongue;
    cons=1;
run;
proc phreg data=tongue;
    model time*death(0) = cons;
    where group=1;
    output out=out_tongue1 survival=survival1 logsurv=logsurv1 loglogs=loglogs1;
run;
proc phreg data=tongue;
    model time*death(0) = cons;
    where group=2;
    output out=out_tongue2 survival=survival2 logsurv=logsurv2 loglogs=loglogs2;
run;
proc print data=out_tongue1;run;
proc print data=out_tongue2;run;
proc sort data=out_tongue1;by time;run;
proc sort data=out_tongue2;by time;run;
data out_tongue_all;
    merge out_tongue1 out_tongue2;
    haz1=-logsurv1;                * Exponential;
    haz2=-logsurv2;
    loghaz1=log(haz1);            * Weibull;
    loghaz2=log(haz2);
    logtime=log(time);
    lohaz1=log(exp(haz1)-1);      * loglogistic;
    lohaz2=log(exp(haz2)-1);
    ln haz1=probit(1-exp(-haz1)); * lognormal;
    ln haz2=probit(1-exp(-haz2));
    by time;
run;
proc print data=out_tongue_all;run;
title "Exponential";
proc gplot data = out_tongue_all;
    symbol i=j;
    plot haz1*time = 1 haz2*time = 2 / overlay;
    label haz1 = "Estimated Cumulative Hazard Rate";
    label time = "Time";
run;
title "Weibull";
proc gplot data = out_tongue_all;
    symbol i=j;
    plot loghaz1*logtime = 1 loghaz2*logtime = 2 / overlay;
    label loghaz1 = "Log Estimated Cumulative Hazard Rate";
    label logtime = "Log(Time)";
run;
title "log logistic";
proc gplot data = out_tongue_all;
    symbol i=j;
    plot lohaz1*logtime = 1 lohaz2*logtime = 2 / overlay;
    label lohaz1 = "Log Odds";
    label logtime = "Log(Time)";
run;
title "log normal";
proc gplot data = out_tongue_all;
    symbol i=j;
    plot ln haz1*logtime = 1 ln haz2*logtime = 2 / overlay;
    label ln haz1 = "Probit";
    label logtime = "Log(Time)";
run;

```