Question #1:

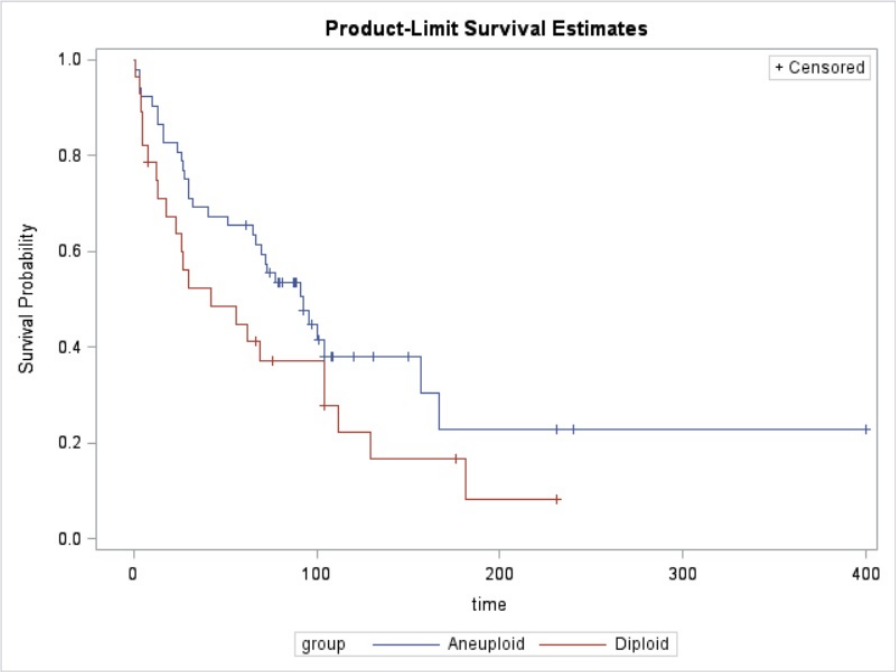
Solution 1. *Exercise* 7.4*:*

We run log-rank test for part (a) *and wilcoxon test for part* (b) *since in part* (b) *we are giving more weight to earlier part of the survival.*

We have the following SAS code:

```
dm 'log; clear; output; clear;';

proc format;
    value status 0 = "censored"
                 1 = "event";
    value group  1 = "Aneuploid"
                 2 = "Diploid";
;

data Ex7_4;
    input group time status @@;
    format group group.
           status status.;
    datalines;
    1 1 1   1 3 1   1 3 1   1 4 1   1 10 1  1 13 1  1 13 1  1 16 1  1 16 1
    1 24 1  1 26 1  1 27 1  1 28 1  1 30 1  1 30 1  1 32 1  1 41 1  1 51 1
    1 65 1  1 67 1  1 70 1  1 72 1  1 73 1  1 77 1  1 91 1  1 93 1  1 96 1
    1 100 1 1 104 1 1 157 1 1 167 1 1 61 0  1 74 0  1 79 0  1 80 0  1 81 0
    1 87 0  1 87 0  1 88 0  1 89 0  1 93 0  1 97 0  1 101 0 1 104 0 1 108 0
    1 109 0 1 120 0 1 131 0 1 150 0 1 231 0 1 240 0 1 400 0 2 1 1   2 3 1
    2 4 1   2 5 1   2 5 1   2 8 1   2 12 1  2 13 1  2 18 1  2 23 1  2 26 1
    2 27 1  2 30 1  2 42 1  2 56 1  2 62 1  2 69 1  2 104 1 2 104 1 2 112 1
    2 129 1 2 181 1 2 8 0   2 67 0  2 76 0  2 104 0 2 176 0 2 231 0
    ;
run;

proc print data=Ex7_4;
run;

/*we use log ran test for part a and wilcoxon test for part b*/
proc lifetest data= Ex7_4 plots = (s) cs = none;
    time time*status(0);
    strata group/test = (logrank wilcoxon);
run;
```

with output:



Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	2.7897	1	0.0948707
Wilcoxon	3.3055	1	0.0690486

However both the log-rank test and wilcoxon test results are not significant given significance level $\alpha = 0.05$*(p value* 0.09 *for log-rank and* 0.07 *for wilcoxon).*

So we conclude that there is not enough evidence showing that the survival rates of patients with cancer of tongue are different for patients with aneuploid and diploid tumors, using both log-rank test and wilcoxon test.

Exercise 7.12*:*

We input the data and run the Renyi statistic as the class example:

```
proc format;
    value status 0 = "censored"
                 1 = "event";
    value group  1 = "Aneuploid"
                 2 = "Diploid";
;

data Ex7_4;
    input group time status @@;
    format group group.
           status status.;
    datalines;
    1 1 1   1 3 1   1 3 1   1 4 1   1 10 1  1 13 1  1 13 1  1 16 1  1 16 1
    1 24 1  1 26 1  1 27 1  1 28 1  1 30 1  1 30 1  1 32 1  1 41 1  1 51 1
    1 65 1  1 67 1  1 70 1  1 72 1  1 73 1  1 77 1  1 91 1  1 93 1  1 96 1
    1 100 1 1 104 1 1 157 1 1 167 1 1 61 0  1 74 0  1 79 0  1 80 0  1 81 0
    1 87 0  1 87 0  1 88 0  1 89 0  1 93 0  1 97 0  1 101 0 1 104 0 1 108 0
    1 109 0 1 120 0 1 131 0 1 150 0 1 231 0 1 240 0 1 400 0 2 1 1   2 3 1
    2 4 1   2 5 1   2 5 1   2 8 1   2 12 1  2 13 1  2 18 1  2 23 1  2 26 1
    2 27 1  2 30 1  2 42 1  2 56 1  2 62 1  2 69 1  2 104 1 2 104 1 2 112 1
    2 129 1 2 181 1 2 8 0   2 67 0  2 76 0  2 104 0 2 176 0 2 231 0
    ;
run;

data sec1_6;
    infile "C:\akira\data\burn.txt";
    input obs Treatment Gender Race Percentage Head
          buttock trunk upper_leg lower_leg resp_tract type
          time_to_excision excision_ind time_to_prophy
          prophy_ind time_to_straphy straphy_ind;
    tre = treatment + 1;
run;

/*7.12a, use Renyi statistic to run log rank test*/
%_renyi(sec1_6, tre, time_to_straphy, straphy_ind, 0, LOGRANK, 0, 0);

%_renyi(Ex7_4, group, time, status, 0, Gehan, 0, 0);
```

For part (a) we have the following output:

```
************************************************************
**********RENYI STATISTICS**********************************
************************************************************

Time that maximizes Z(t): 51.0000
LOGRANK statistic (numerator): 7.5744065818
Z(t) Statistic: 7.5744065818
Z(t) Standard Deviation: 3.3947455273
Renyi Test Statistic: 2.2312148351
 Approximate P-value: 0.0513338035

************************************************************
```

we report the p value as not significant given the level $\alpha = 0.05$, but it is really close. We can conclude that we fail to reject the null hypothesis and think there is no enough evidence showing a significant difference of hazard rate, but addressing that p value is close to the significance level. and for part (b) we have:

```
****************************************************************
*********************RENYI STATISTICS*************************
****************************************************************

Time that maximizes Z(t): 69.000
GEHAN statistic (numerator):         -332
Z(t) Statistic: 366
Z(t) Standard Deviation: 180.05832859
Renyi Test Statistic: 2.0326746497
 Approximate P-value: 0.084170806

****************************************************************
```

The p value is insignificant and we fail to reject the null hypothesis. So we conclude that there is not enough evidence supporting that there is a different hazard rate between the two groups of patients.

Question #2:

Solution 2. *We use log-rank test for part (a) and Fleming test with $p < q$ for part (b). Because we are more interested in long term efficacy of the treatment and fleming test with $p < q$ put more weights on later event time. After a few trial we particularly choose $p = 0$ and $q = 0.5$. The code is:*

```
proc format;
    value status 0 = "censored"
                 1 = "event";
    value group  1 = "AZT + ddc"
                 2 = "AZT + ddc + saquinivir";
;

data Q2;
    input group time status @@;
    format group group.
           status status.;
    datalines;
    1 4 0   1 6 1   1 11 1  1 12 1  1 32 1  1 35 1
    1 38 0  1 39 1  1 45 1  1 49 1  1 75 1  1 80 1
    1 84 1  1 85 1  1 87 1  1 102 1 1 180 0 2 2 1
    2 3 1   2 4 1   2 12 1  2 22 1  2 48 1  2 51 0
    2 56 0  2 80 1  2 85 1  2 90 1  2 94 0  2 160 1
    2 171 1 2 180 1 2 180 0 2 238 1
    ;
run;

/*part (a): logrank test*/
/*part (b): wilcoxon test*/

proc lifetest data=Q2;
    time time*status(0);
    strata /group = group test = (logrank fleming(0, 0.5));
run;
```

and the output is:

Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	2.0491	1	0.1522982
Fleming(0,0.5)	3.5432	1	0.0597911

For logrank test the p value is not significant. We fail to reject the null hypothesis and conclude that there is not enough evidence showing that the AZT + ddC + saquinvavir is more effective (provides longer survival,or smaller hazard rate) than the AZT + ddc treatment.

For Flemming test with the $p = 0$ and $q = 0.5$, the p value is marginally significant ($p = 0.06$) which shows stronger evidence than the logrank test that AZT+ddc+saquinvair is more effective in the long term.

For part (c) and part (d) we fit separately an exponential AFT and weibull AFT model with a single dichotomous covariate representing treatment group. We have the following code:

```
/*part (c): exponential AFT model*/
/*part (d): weibull AFT model*/
proc lifereg data=Q2;
    class group;
    model time*status(0) = group
    /dist = exponential;
run;

proc lifereg data=Q2;
    class group;
    model time*status(0) = group
    /dist = weibull;
run;
```

For the exponential model, our output is:

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
group	1	1.6859	0.1941414

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.7321	0.2774	4.1885	5.2757	291.11	<.0000001
group	AZT + ddc	1	-0.5001	0.3852	-1.2550	0.2548	1.69	0.1941414
group	AZT + ddc + saquinivir	0	0.0000	.	.	.	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		
Weibull Shape		0	1.0000	0.0000	1.0000	1.0000		

Lagrange Multiplier Statistics		
Parameter	Chi-Square	Pr > ChiSq
Scale	0.1270	0.7215239

with AZT + ddc + saquinivir as the reference group, the regression coefficient for AZT + ddc is -0.5001 *(this makes sense since we expect the triple drug combinations to enhance survival). But the p value is* 0.19 *and not significant, which means we fail to reject the null hypothesis and conclude that there is not enough evidence showing different effectiveness between groups. Also, the p value for testing* $H_0 : \sigma = 1$ *is* 0.72 *and highly insignificant, so we fail to reject that* $H_0 : \sigma = 1$*, and this support model assumption for exponential model.*

For the Weibull model, our output is:

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
group	1	1.8894	0.1692721

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.7386	0.2631	4.2229	5.2542	324.37	<.0000001
group	AZT + ddc	1	-0.5008	0.3643	-1.2149	0.2133	1.89	0.1692721
group	AZT + ddc + saquinivir	0	0.0000	.	.	.	.	.
Scale		1	0.9459	0.1520	0.6904	1.2961		
Weibull Shape		1	1.0572	0.1699	0.7716	1.4484		

The regression coefficient is -0.5008 *with p value* 0.17*. It is not significant and we conclude that there is not enough evidence supporting the different effectiveness between treatment group and we faill to reject the null hypothesis that different treatment groups have the same hazard rate.*

The estimate for scale is 0.9459 *indicating that the hazard rate increases at a decreasing rate.However it is pretty close to* 1*, and also the* 95% *confidence interval includes* 1 *still, so we may for now keep the exponential model as a candidate.*

For part (e)*, to perform a likelihood ratio test, notice that exponential model is nested in the weibull model, and the difference of the degree of freedom is:*

$$df_w - df_e = 2 - 1 = 1$$

From the code we gave above for (c) *and* (d)*, we can separately get :*

$$LL_e = -49.213007$$
$$LL_w = -49.1545919$$

So we have:

$$-2(LL_e - LL_w) = 98.426 - 98.309 = 0.117$$

So we can compute the p value with the following code:

$$pvalue = 1 - probchi(0.117, 1) = 0.73$$

a quick comment here, we can also use the unlogged response from the model to compute the fit statistics (−2 Log Likelihood), which is in the output of SAS proc lifereg as well, and we will have the same p value.

As we can see that the p value is highly insignificant, hence we fail to reject the null hypothesis and conclude that the exponential model fit just as well as the weibull.

We may also need to check if these two models are adequate compared to a more general model, which is the generalized gamma model. In the following code, we fit the generalized gamma model and run a likelihood formal test for both exponential and weibull models compared to gamma:

```
proc lifereg data=Q2;
     class group;
     model time*status(0) = group
     /dist = gamma;
 run;

data fit3;
     pvalue1 =  1 - probchi(98.426 - 96.584, 2);
     pvalue2 = 1 - probchi(98.309 - 96.584, 1);
 run;

proc print data=fit3;
 run;
```

The p value for the χ^2 tests are:

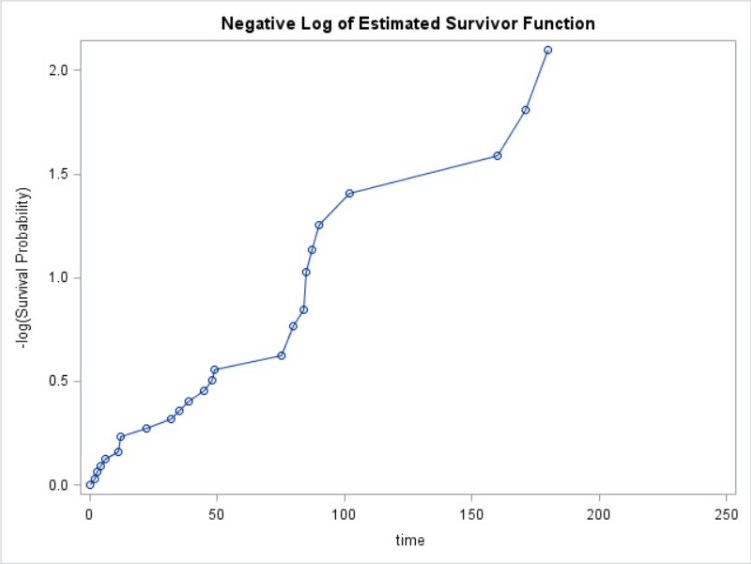
Obs	pvalue1	pvalue2
1	0.39812	0.18905

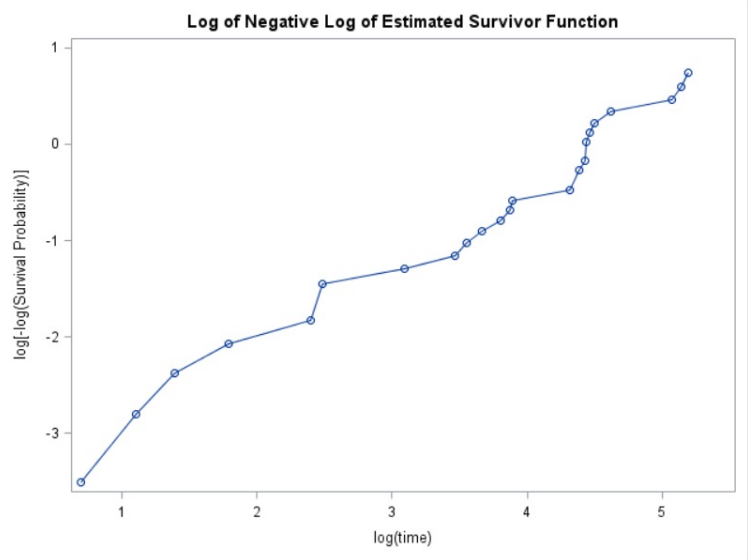
Both as insignificant, so we conclude that the weibull and exponential models are adequate compared to generalized gamma.

For part (f):

We sketch the negative log survival versus time and log-log survival versus log time with the following code and output:

```
/*(f) plot log survival and log log survival to check goodness of fit*/
proc lifetest data=Q2 plots(only) = ls;
    time time*status(0);
run;

proc lifetest data=Q2 plots(only) = lls;
    time time*status(0);
run;
```





The (LS) plot apparently does not line up as a straightline, which indicates that the exponential model is not adequate. This is not consistent with the conclusion from part (e). This might be due to the small sample size from the data.

The (LLS) plot gives a sort-of straightline, but still not very perfect. We would reserve our opinion and consider weibull distribution as candidate also for now for our final model.

For part (g):

We fit the lognormal and log-logistic model with the following SAS code:

```
/*(g) fit lognormal and log-logistic and report log-likelihood values*/
proc lifereg data=Q2;
    class group;
    model time*status(0) = group
    /dist = lnormal;
run;

proc lifereg data=Q2;
    class group;
    model time*status(0) = group
    /dist = llogistic;
run;
```

The output for log-normal model is:

Model Information	
Data Set	WORK.Q2
Dependent Variable	Log(time)
Censoring Variable	status
Censoring Value(s)	0
Number of Observations	34
Noncensored Values	27
Right Censored Values	7
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	3
Name of Distribution	Lognormal
Log Likelihood	-52.20433928

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
group	1	0.0535	0.8171233

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.0296	0.3526	3.3386	4.7206	130.63	<.0000001
group	AZT + ddc	1	-0.1151	0.4976	-1.0904	0.8602	0.05	0.8171233
group	AZT + ddc + saquinivir	0	0.0000	.	.	.	.	.
Scale		1	1.3909	0.1921	1.0610	1.8232		

The log-likelihood value is -52.20433928.

The chi-square test for the group effectiveness has a p value of 0.82 *which is highly insignificant, and we fail to reject the null and conclude that there is not enough evidence showing that there is a difference of effectiveness between groups.*

The output for log-logistic model is:

Model Information	
Data Set	WORK.Q2
Dependent Variable	Log(time)
Censoring Variable	status
Censoring Value(s)	0
Number of Observations	34
Noncensored Values	27
Right Censored Values	7
Left Censored Values	0
Interval Censored Values	0
Number of Parameters	3
Name of Distribution	LLogistic
Log Likelihood	-51.82056127

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
group	1	0.5476	0.4592864

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.2803	0.3511	3.5921	4.9684	148.63	<.0000001
group	AZT + ddc	1	-0.3478	0.4700	-1.2689	0.5733	0.55	0.4592864
group	AZT + ddc + saquinivir	0	0.0000	.	.	.	.	.
Scale		1	0.7681	0.1254	0.5578	1.0578		

The log-likelihood value is -51.82056127.

The chi-square test for the group effectiveness has a p value of 0.46 *which is insignificant, and we fail to reject the null and conclude that there is not enough evidence showing that there is a difference of effectiveness between groups.*

So far among all the models we have fit (exponential, weibull, log-normal and log-logistic), none of them is showing significance for the differenct effectivess between groups. So it seems that the effectiveness of treatment does not depend on the distributional assumptions.

For part (h)*:*

To compare between models that are not nested, we need to look at the AIC value. From the code above we have:

$$\begin{aligned} AIC_{exponential} &= 102.426 \\ AIC_{weibull} &= 104.309 \\ AIC_{lognormal} &= 110.409 \\ AIC_{loglogistic} &= 109.641 \end{aligned}$$

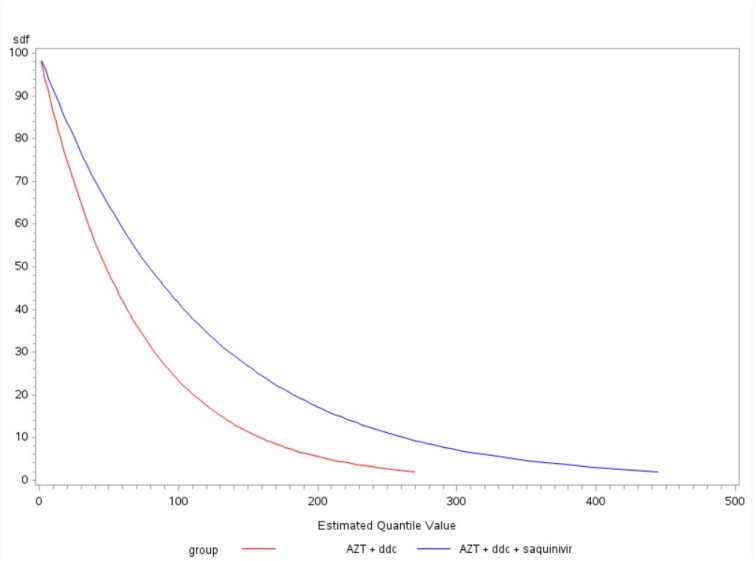
The exponential model has the smallest AIC, and previously the formal test in part (e) *also says that exponential is adequate enough compared to weibull. Although the negative log survival is against this statement, but we pointed out that this might be due to the small sample size. From the output of weibull model we see that both the scale and shape parameters have an estimate close to* 1*. Hence we consdier the exponential model as appropraite.*

Between exponential and weibull, although both can be used, we are in favor of exponential. First reason is that we have done likelihood ratio test and the formal test shows that exponential model is adequate compared to weibull. Secondly, exponential model has the smallest AIC value among all four models we have fun. Third reason is that whenever possible we prefer to choose the simpler model than the more complicated ones.

For part (i)*:*

The following code plots the survival curves for the two treatment under exponential model:

```
/*(i) sketch the survival curves*/
data one;
    C = 1; group = 1; output;
    C = 1; group = 2; output;
run;

data Q2_2;
    set Q2 one;
run;

proc lifereg data=Q2_2;
    class group;
    model time*status(0) = group
    /dist = exponential;
    output out = surv_est quantiles = 0.02 to 0.98 by 0.02
    predicted = pred control = C;
run;

data surv_est2;
    set surv_est; sdf = 100*(1 - _prob_);
run;

proc gplot data=surv_est2;
    plot sdf*pred=group;
    symbol1 I = spline color = red L= 1;
    symbol2 I = spline color = blue L = 1;
run;
```



Question #3.

Solution 3. *For part (a):*

For (i) we have:

$$F_{X^\gamma}(x) = P(X^\gamma \leq x) = P(X \leq x^{\frac{1}{\gamma}}) = F_X(x^{\frac{1}{\gamma}})$$

So

$$\begin{aligned} f_{X^\gamma}(x) &= f_X(x^{\frac{1}{\gamma}}) \cdot \frac{1}{\gamma} \cdot x^{\frac{1}{\gamma}-1} \\ &= \frac{\cancel{\gamma}}{\beta} \cdot \Big(x^{\frac{1}{\gamma}}\Big)^{\gamma-1} \cdot \exp\Big(-(x^{\frac{1}{\gamma}})^\gamma/\beta\Big) \cdot \frac{1}{\cancel{\gamma}} \cdot x^{\frac{1}{\gamma}-1} \\ &= \frac{1}{\beta} \cdot x^{1-\frac{1}{\gamma}} \cdot \exp\Big(-x/\beta\Big) \cdot x^{\frac{1}{\gamma}-1} \\ &= \frac{1}{\beta} \cdot \exp\Big(-\frac{x}{\beta}\Big) \sim \exp(\beta) \end{aligned}$$

For (ii) we have:

$$F_{cX} = P(cX \le x) = P(X \le \frac{x}{c}) = F_X(\frac{x}{c})$$

So

$$\begin{aligned} f_{cX} &= f_X(\frac{x}{c}) \cdot \frac{1}{c} \\ &= \frac{\gamma}{\beta} \cdot \Big(\frac{x}{c}\Big)^{\gamma-1} \cdot \exp\Big(-(\frac{x}{c})^\gamma/\beta\Big) \cdot \frac{1}{c} \\ &= \frac{\gamma}{\beta c^\gamma} \cdot x^{\gamma-1} \exp\Big(-x^\gamma/(\beta c^\gamma)\Big) \sim Weibull(\gamma, \beta c^\gamma) \end{aligned}$$

For (iii) we have:

$$F_{X^{\frac{1}{\gamma}}} = P(X^{\frac{1}{\gamma}} \le x) = P(X \le x^\gamma) = F_X(x^\gamma)$$

So

$$\begin{aligned} f_{X^{\frac{1}{\gamma}}} &= f_X(x^\gamma) \cdot \gamma \cdot x^{\gamma-1} \\ &= \frac{\gamma}{\beta} \cdot \Big(x^\gamma\Big)^{\gamma-1} \cdot \exp\Big(-(x^\gamma)^\gamma/\beta\Big) \cdot \gamma \cdot x\gamma - 1 \\ &= \frac{\gamma}{\beta} \cdot x^{\gamma(\gamma-1)} \cdot \exp\Big(-x^{\gamma^2}/\beta\Big) \cdot \gamma \cdot x^{\gamma-1} \\ &= \frac{\gamma^2}{\beta} \cdot x^{\gamma^2-1} \cdot \exp\Big(-x^{\gamma^2}/\beta\Big) \sim Weibull(\gamma^2, \beta) \end{aligned}$$

For (iv),

Suppose the median of X is denoted as M, then we have:

$$\begin{aligned} \frac{1}{2} = S_X(M) &= \exp\Big(-\frac{1}{\beta} \cdot M^\gamma\Big) \\ \implies -\frac{1}{\beta} \cdot M^\gamma &= \log\frac{1}{2} \\ \implies M^\gamma &= -\beta \log\frac{1}{2} = \beta \log 2 \\ \implies M &= \beta^{\frac{1}{\gamma}} \cdot (\log 2)^{\frac{1}{\gamma}} \end{aligned}$$

The mode of a distribution is the value of x at which the probability density function reaches maximum.

We have:

$$f(x|\gamma,\beta) = \frac{\gamma}{\beta}\cdot x^{\gamma-1}\cdot \exp\Big(-x^{\gamma}/\beta\Big)$$

If $\gamma > 1$, we have:

$$\begin{aligned}\frac{d}{dx}f(x|\gamma,\beta) &= \frac{\gamma}{\beta}\cdot(\gamma-1)\cdot x^{\gamma-2}\cdot\exp\Big(-x^{\gamma}/\beta\Big) + \frac{\gamma}{\beta}\cdot x^{\gamma-1}\cdot\exp\Big(-x^{\gamma}/\beta\Big)\cdot(-\frac{\gamma}{\beta})\cdot x^{\gamma-1}\\ &= \underbrace{\frac{\gamma}{\beta}\cdot x^{\gamma-2}\cdot\exp\Big(-\frac{x^{\gamma}}{\beta}\Big)}_{(1)}\cdot\underbrace{\Big[(\gamma-1)-x^{\gamma}\cdot\frac{\gamma}{\beta}\Big]}_{(2)}\end{aligned}$$

apparently $(1) > 0$ regardless the value of x. If we set equation $\frac{d}{dx}f(x|\gamma,\beta) = 0$, we then have:

$$\begin{aligned}&(\gamma-1)-x^{\gamma}\cdot\frac{\gamma}{\beta} = 0\\ \Longrightarrow\ &x^{\gamma} = \frac{(\gamma-1)\cdot\beta}{\gamma} = (1-\frac{1}{\gamma})\beta\\ \Longrightarrow\ &x = (1-\frac{1}{\gamma})^{\frac{1}{\gamma}}\cdot\beta^{\frac{1}{\gamma}}\end{aligned}$$

and it is also easy to observe that when $x < (1-\frac{1}{\gamma})^{\frac{1}{\gamma}}\cdot\beta^{\frac{1}{\gamma}}$, we have $\frac{d}{dx}f(x|\gamma,\beta) > 0$ and hence f is increasing, and when $x > (1-\frac{1}{\gamma})^{\frac{1}{\gamma}}\cdot\beta^{\frac{1}{\gamma}}$, we have $\frac{d}{dx}f(x|\gamma,\beta) < 0$ and hence f is decreasing. So f reaches maximum at

$$x = (1-\frac{1}{\gamma})^{\frac{1}{\gamma}}\cdot\beta^{\frac{1}{\gamma}}$$

or we say, the mode when $\gamma > 1$ is $x = (1-\frac{1}{\gamma})^{\frac{1}{\gamma}}\cdot\beta^{\frac{1}{\gamma}}$.

Now when $\gamma \le 1$, we have the same expression for $\frac{d}{dx}f(x|\gamma,\beta) = (1)\times(2)$ but only this time $(2) < 0$ regardless the value of x. So f is always decreasing and hence the maximum is approached when x approaches 0 from the right hand side, and hence $x = 0$ is the mode when $\lambda \le 1$.

For part (b):

Let $Y = \frac{x^{1/\beta}}{\theta}$ then we have:

$$F_Y(x) = P(\frac{x^{1/\beta}}{\theta}\le x) = P(X\le(\theta x)^{\beta}) = F_X((\theta x)^{\beta})$$

So

$$\begin{aligned} f_Y(x) &= f_X\Big((\theta x)^\beta\Big) \cdot \theta^\beta \cdot \beta \cdot x^{\beta-1} \\ &= \frac{\Big((\theta x)^\beta\Big)^{\alpha-1} \cdot \exp\Big(-\frac{(\theta x)^\beta}{\beta}\Big)}{\Gamma(\alpha)\cdot\beta^\alpha} \cdot \theta^\beta \cdot \beta \cdot x^{\beta-1} \\ &= \frac{\theta^{\beta\alpha} \cdot \beta \cdot x^{\beta\alpha-1} \cdot \exp\Big(-\frac{\theta^\beta \cdot x^\beta}{\beta}\Big)}{\Gamma(\alpha)\cdot\beta^\alpha} \\ &= \frac{\beta \cdot x^{\alpha\beta-1} \cdot \exp\Big[-\Big(\frac{x}{\beta^{1/\beta}/\theta}\Big)^\beta\Big]}{\Gamma(\alpha)\cdot\Big(\frac{\beta^{1/\beta}}{\theta}\Big)^{\beta\alpha}} \end{aligned}$$

So this is in the form of a generalized gamma distribution with shape parameter α, β and scale parameter $\frac{\beta^{1/\beta}}{\theta}$ as formatted in part 2 of lecture notes for chapter 5 on page 9.

Question #4.

Solution 4. *The following SAS code input the data and run the logrank test on the matched pairs:*

```
title "carcinogenesis study for paired rats";

proc format;
    value censor 0 = "censored"
                 1 = "event";
    value group  1 = "exposed"
                 2 = "control";
run;

data Q4;
    input pair group time censor @@;
    datalines;
        1 1 101.0 0     1 2 49.0 1      2 1 104.0 0     2 2 102.0 0     3 1 104.0 0     3 2 104.0 0
        4 1 77.0 0      4 2 97.0 0      5 1 89.0 0      5 2 104.0 0     6 1 88.0 1      6 2 96.0 1
        7 1 104.0 1     7 2 94.0 0      8 1 95.9 1      8 2 104.0 0     9 1 82.0 0      9 2 77.0 0
        10 1 70.0 1     10 2 104.0 0    11 1 88.9 1     11 2 91.0 0     12 1 91.0 0     12 2 70.0 0
        13 1 39.0 1     13 2 45.0 0     14 1 102.9 1    14 2 69.0 0     15 1 93.0 0     15 2 104.0 0
        16 1 85.0 0     16 2 72.0 0     17 1 104.0 0    17 2 63.0 0     18 1 104.0 0    18 2 104.0 0
        19 1 81.0 0     19 2 104.0 0    20 1 67.0 1     20 2 104.0 0    21 1 104.0 0    21 2 104.0 0
        22 1 104.0 0    22 2 104.0 0    23 1 104.0 0    23 2 83.0 0     24 1 87.0 0     24 2 104.0 0
        25 1 104.0 0    25 2 104.0 0    26 1 89.0 0     26 2 104.0 0    27 1 78.0 0     27 2 104.0 0
        28 1 104.0 0    28 2 81.0 1     29 1 86.0 1     29 2 55.0 1     30 1 34.0 1     30 2 104.0 0
        31 1 76.0 0     31 2 87.0 0     32 1 102.8 1    32 2 73.0 1     33 1 101.9 1    33 2 104.0 0
        34 1 79.9 1     34 2 104.0 0    35 1 45.0 1     35 2 79.0 0     36 1 94.0 1     36 2 104.0 0
        37 1 104.0 0    37 2 104.0 0    38 1 104.0 0    38 2 101.0 1    39 1 76.0 0     39 2 84.0 1
        40 1 80.0 1     40 2 80.9 1     41 1 72.0 1     41 2 95.0 0     42 1 72.9 1     42 2 104.0 0
        43 1 92.0 1     43 2 104.0 0    44 1 104.0 0    44 2 98.0 0     45 1 55.0 0     45 2 104.0 0
        46 1 49.0 0     46 2 83.0 0     47 1 89.0 1     47 2 104.0 0    48 1 88.0 0     48 2 79.0 0
        49 1 103.0 1    49 2 91.0 0     50 1 104.0 0    50 2 104.0 0
    ;
run;

proc lifetest data=Q4;
    time time*censor(0);
    strata pair/group = group test=logrank;
run;
```

The output is:

Stratified Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.7619	1	0.0163773

The p value is 0.016 *which is significant, so we reject the null and conclude that there is a significant difference at level of* $\alpha = 0.05$ *between the exposed and controlled group on the time to tumor occurrence.*

Question #5

Solution 5. *I confirm that I have read the paper and understand the derivation on page* #246.