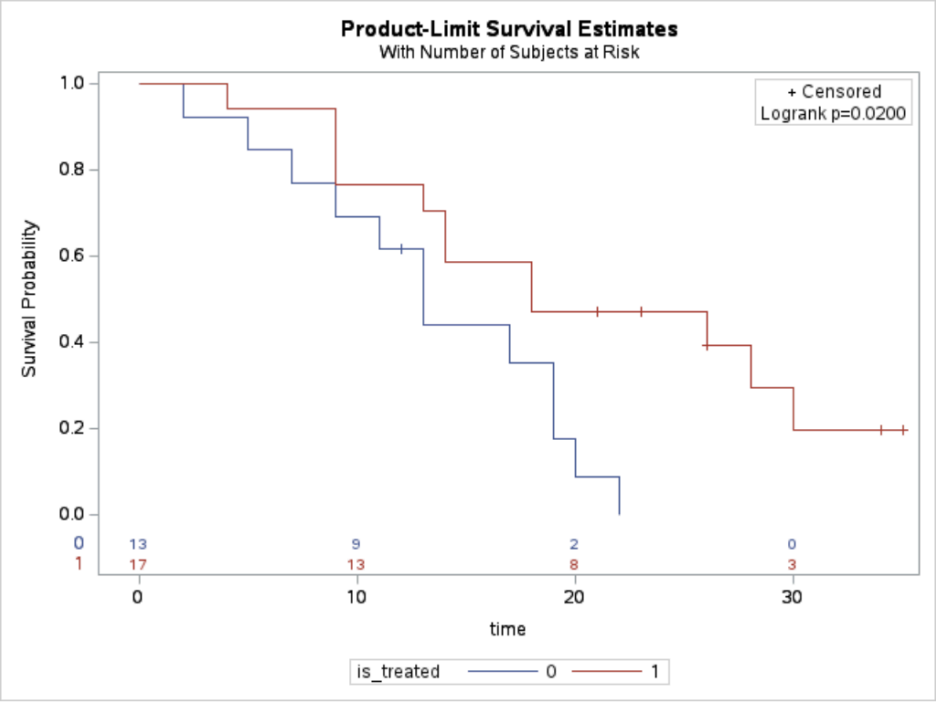
1) The log-rank p-value is 0.0200, which means that there is a statistically significant difference in remission times between the groups that did and didn't receive the treatment. From the Kaplan-Meier plot, it can be seen that the treatment group has higher survival probabilities at each timepoint than the non-treated group. Additionally, quartile estimates for those who are treated are consistently higher than for those who aren't treated. This, paired with the log-rank test p-value, demonstrates how the treatment group is better than the non-treated group in terms of extending remission time.



2)

a) Using PROC PHREG, we can use a proportional hazards model to compare remission time of Leukemia patients by whether or not they were treated by placebo or 6-MP. With this model, we get a resulting p-value of 0.0002, indicating that there is a highly significant difference in remission time between the treatment groups. Additionally, the hazard ratio is 0.221 for those who were treated with 6-MP, telling that receiving 6-MP instead of the placebo reduces risk by 78%.

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
treatment	6-MP	1	-1.50919	0.40956	13.5783	0.0002	0.221	treatment 6-MP

b) Using the Weibull proportional hazard model to compare remission times between the placebo and 6-MP treatments, we were able to find medians for each treatment and the hazard ratio between treatments. The median time for the placebo group was 7.3 months and the median for 6-MP was 25.8 months. The hazard ratio between the treatments was 0.177, with a 95% confidence interval of (-1.594, 1.948). With the point estimate, we can see how effective the 6-MP treatment is.

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	2.2484	0.1660	1.9231	2.5737	183.51	<.0001
treatment	6-MP	1	1.2673	0.3106	0.6585	1.8762	16.64	<.0001
treatment	Placebo	0	0.0000	.	.	.	.	.
Scale		1	0.7322	0.1078	0.5486	0.9772		
Weibull Shape		1	1.3658	0.2012	1.0233	1.8228		

Estimated Covariance Matrix

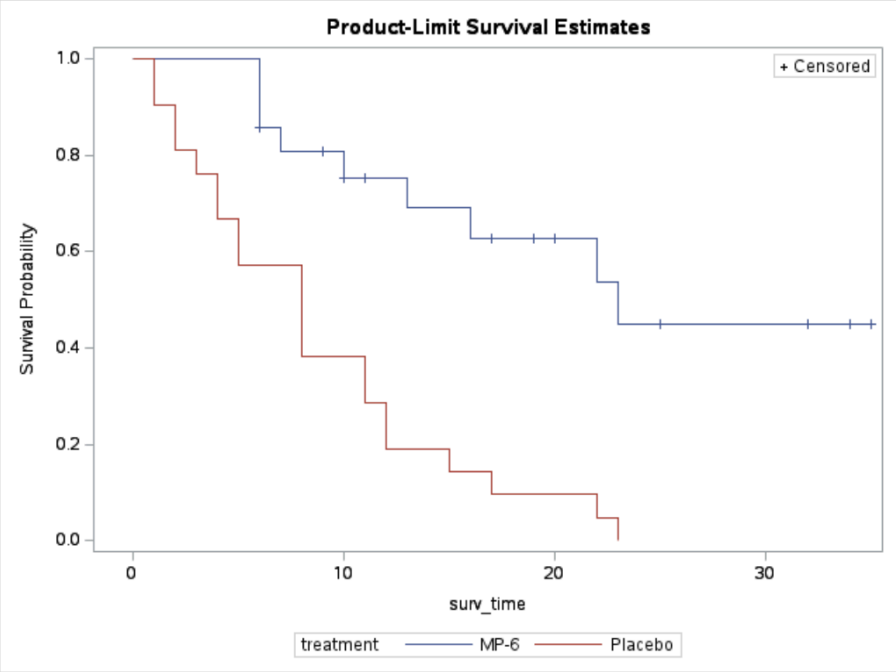
	Intercept	treatment6-MP	Scale
Intercept	0.027547	-0.030325	-0.004844
treatment6-MP	-0.030325	0.096497	0.011515
Scale	-0.004844	0.011515	0.011631

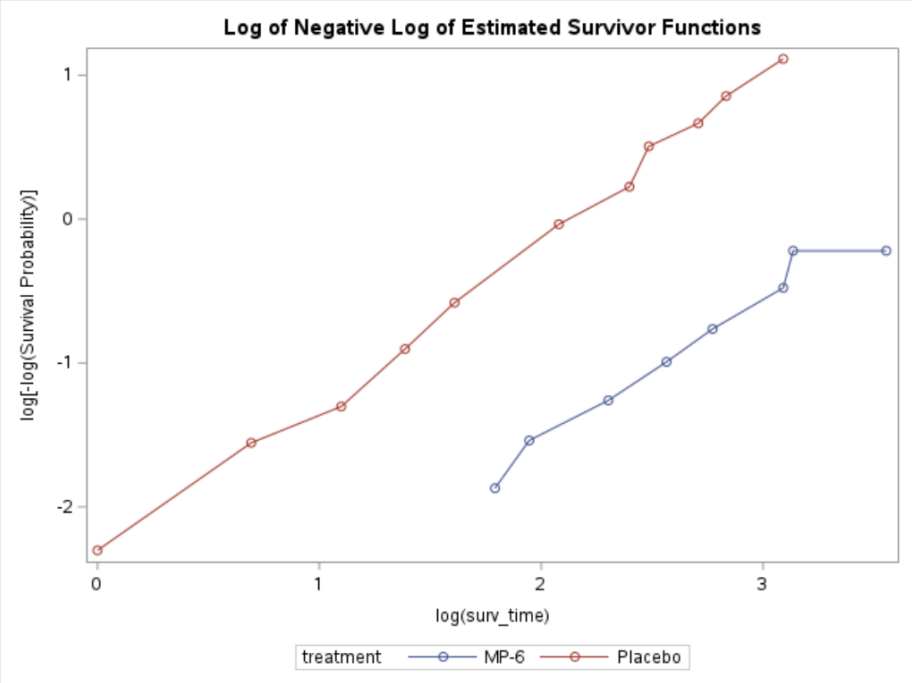
Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	2.2494	0.1678	1.9205	2.5783	179.67	<.0001
treat_num	1	0	0.0000	.	.	.	.	.
Scale		1	0.7297	0.1266	0.5192	1.0253		
Weibull Shape		1	1.3705	0.2379	0.9753	1.9259		

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	3.5194	0.2734	2.9836	4.0552	165.75	<.0001
treat_num	2	0	0.0000	.	.	.	.	.
Scale		1	0.7387	0.2057	0.4280	1.2748		
Weibull Shape		1	1.3537	0.3769	0.7844	2.3362		





c) Using a log-logistic AFT model, we can calculate the acceleration factor of the treatments. The calculated relative acceleration factor is 0.282, which means that the 6-MP treatment slows down Leukemia progression by a factor 3.5. This furthers the case that the 6-MP treatment is effective in helping patients as

compared to the placebo.

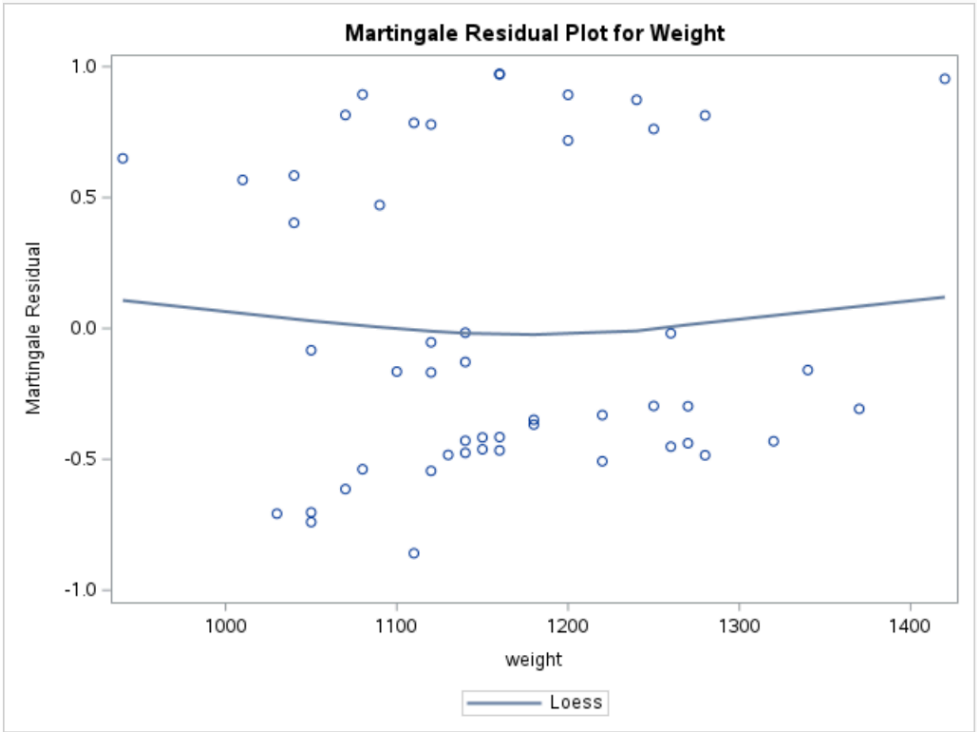
Analysis of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	1.8927	0.2076	1.4858	2.2996	83.10	<.0001
treat_num	2	1	1.2655	0.3257	0.6272	1.9037	15.10	0.0001
treat_num	1	0	0.0000	.	.	.	.	.
Scale		1	0.5466	0.0820	0.4072	0.7335		

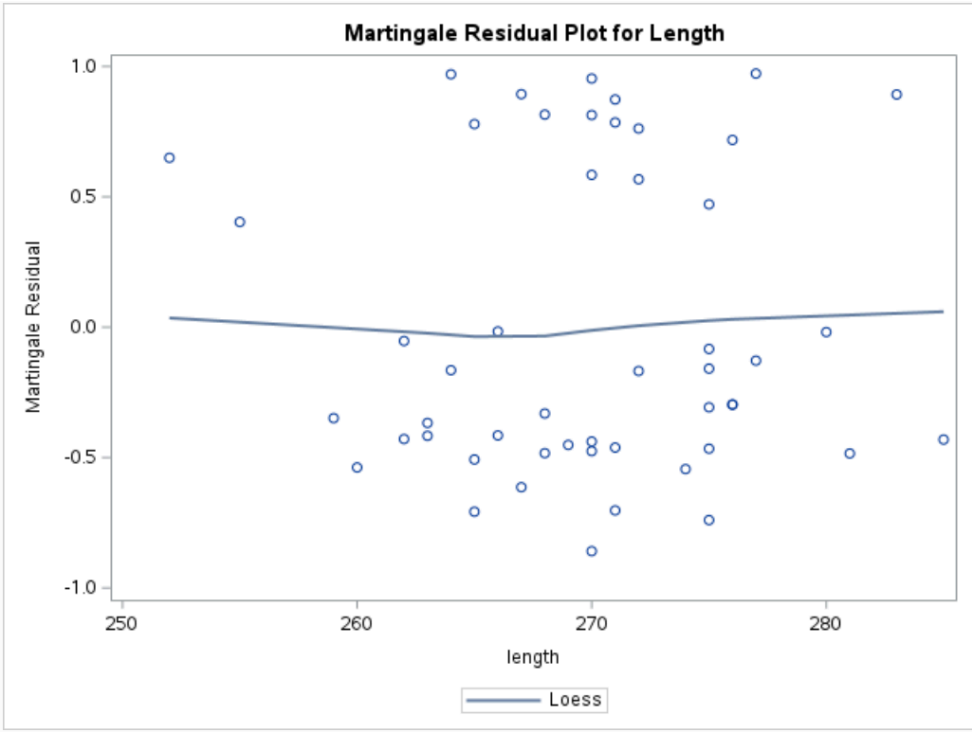
3)

a) The Cox proportional hazards model for the data returned results that indicate that none of the predictor variables are statistically significant as all their p-values are greater than 0.05. Because of this, there is no evidence to suggest that keeping/dropping specific covariates will help improve the model, so we will keep all of them.

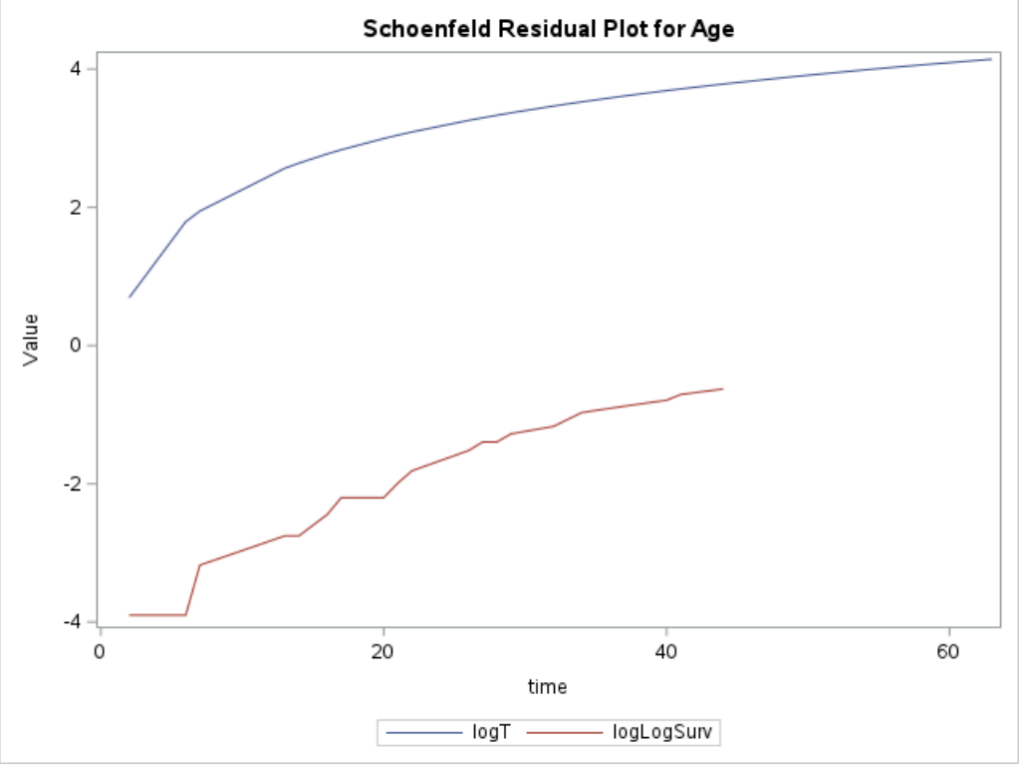
Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
age	0	1	-0.46568	0.57757	0.6501	0.4201	0.628	age 0
weight		1	-0.00420	0.00289	2.1165	0.1457	0.996	
length		1	0.01284	0.04081	0.0989	0.7531	1.013	

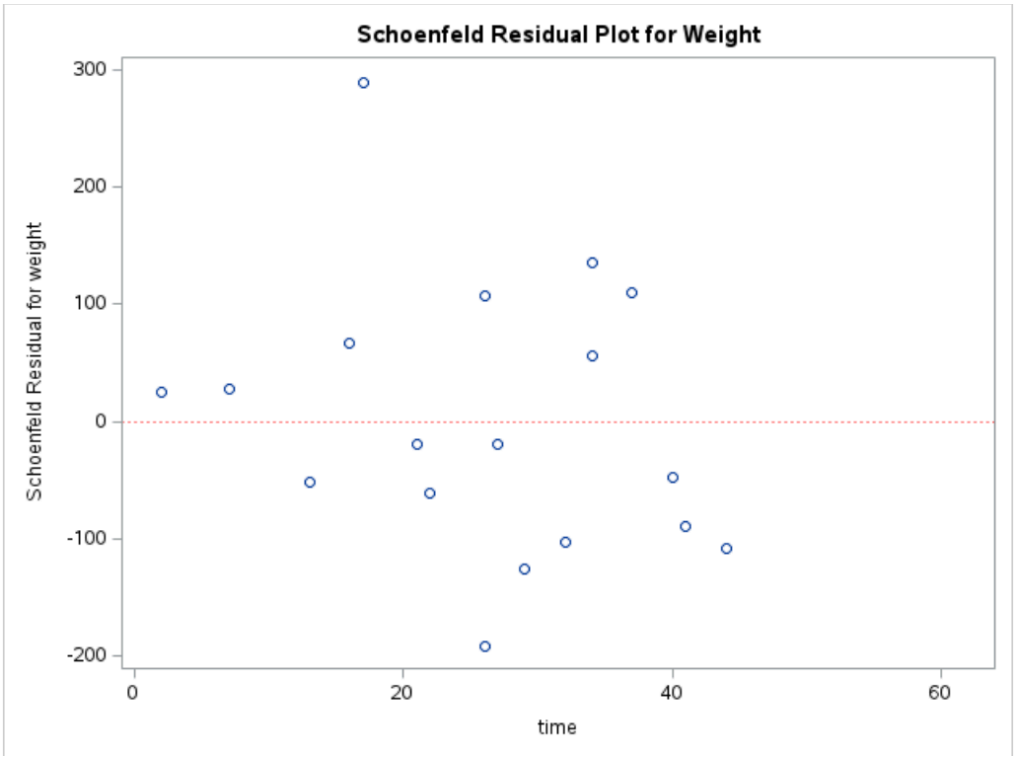
b) Checking martingale residuals allow us to check the functional forms of the continuous covariates, which are weight and length. For both of these variables, the residual plots return pretty linear loess smooth curves, indicating that there isn't any need to discretize either of them.

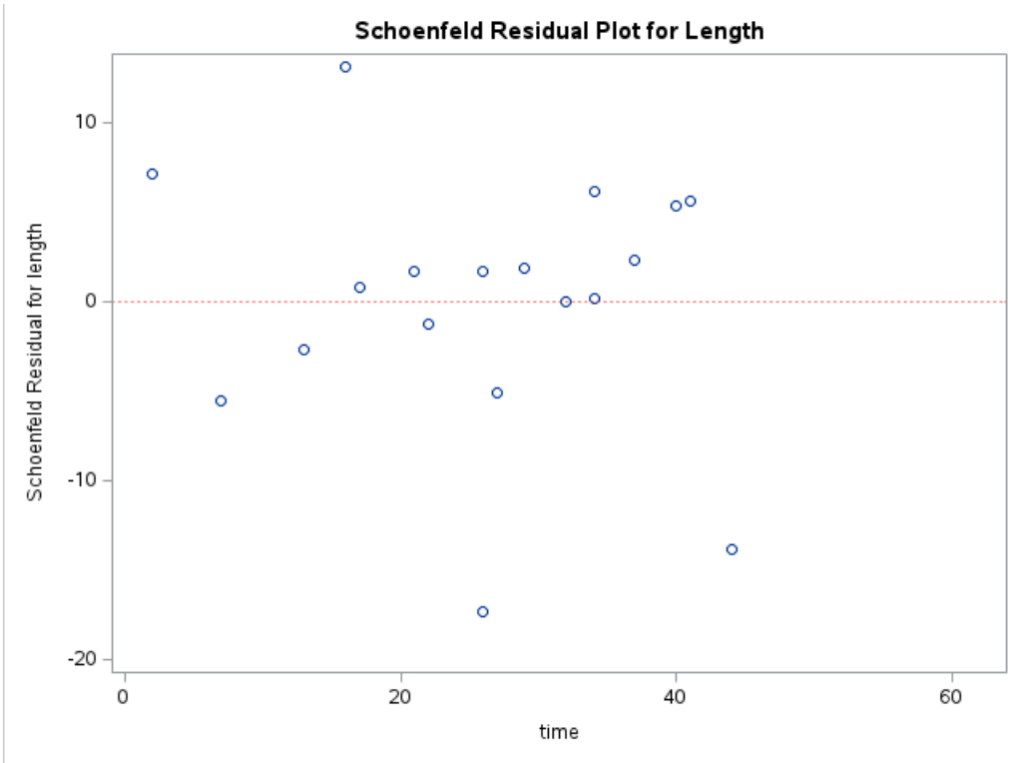

Martingale Residual Plot for Weight
Martingale Residual
1.0
0.5
0.0
-0.5
-1.0
1000
1100
1200
1300
1400
weight
Loess


Martingale Residual Plot for Length
Martingale Residual
1.0
0.5
0.0
-0.5
-1.0
250
260
270
280
length
Loess

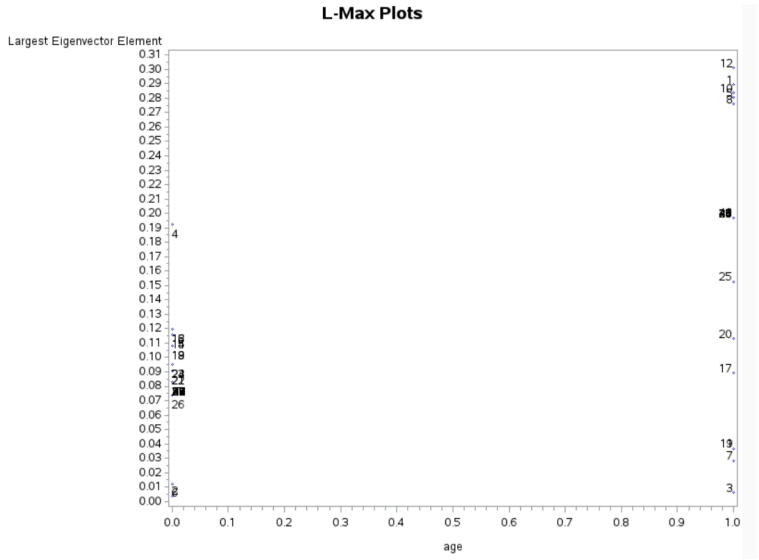
c) To check the proportional hazards assumption for each covariate, we looked at plots of the schoenfeld residuals. From them, we can see that for age, the assumption appears to hold since the lines are parallel. For the continuous variables though, it looks like the assumption may not hold for either. For weight, it looks like the residuals are diverging and getting farther from 0 as time increases. For length, while the residuals are still around zero overall, it looks like there is a general pattern of the residual values and is not random. As such, I would only keep age in the model.

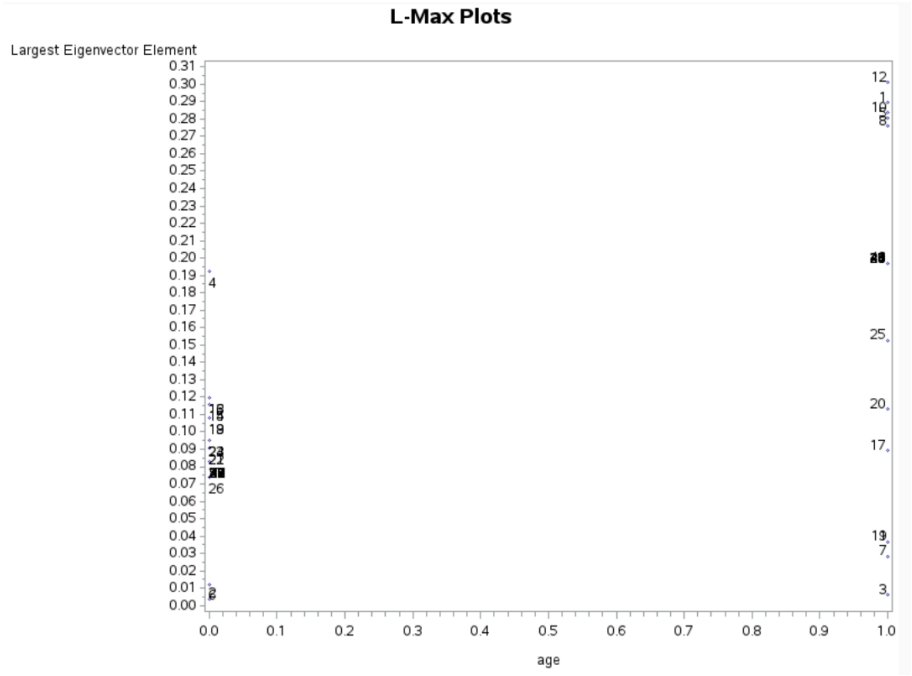


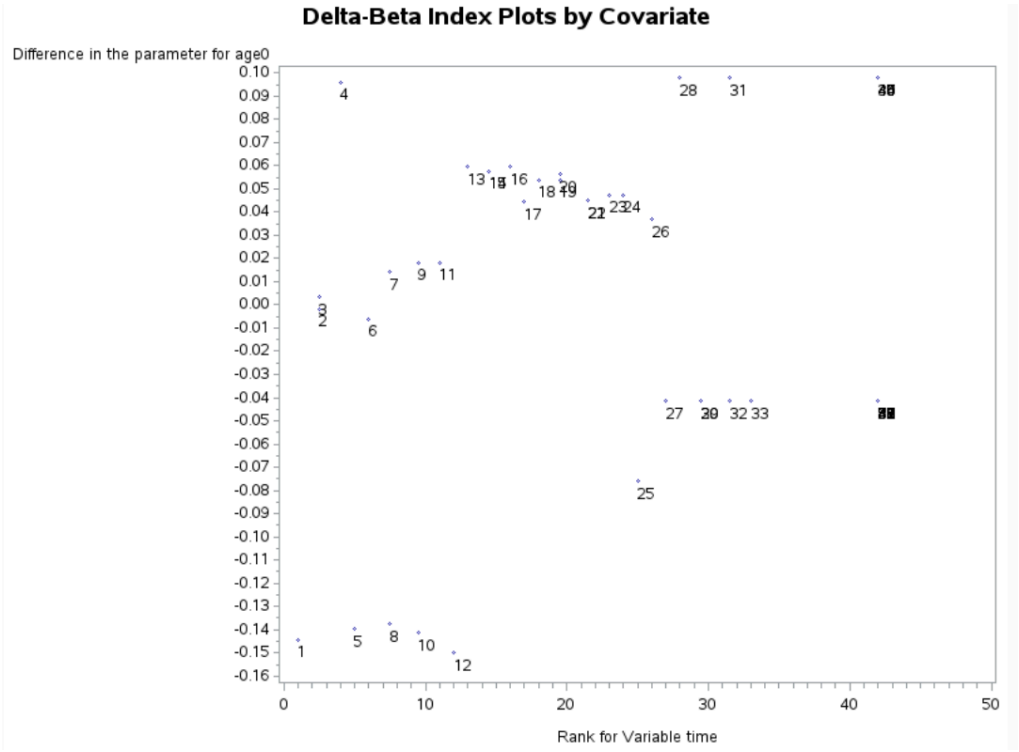

Schoenfeld Residual Plot for Weight
Schoenfeld Residual for weight
300
200
100
0
-100
-200
0
20
40
60
time


Schoenfeld Residual Plot for Length
Schoenfeld Residual for length
10
0
-10
-20
0
20
40
60
time

d) Based on the final model only containing age as a predictor variable, it looks like many observations have a decent amount of influence, which makes sense considering that there's only one predictor and it's categorical. However, there are a couple that stand out, which are ducks/observations 1, 8, 10 and 12.






Delta-Beta Index Plots by Covariate
Difference in the parameter for age0
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
-0.07
-0.08
-0.09
-0.10
-0.11
-0.12
-0.13
-0.14
-0.15
-0.16
0
10
20
30
40
50
Rank for Variable time