

AB Final Exercise F: Leukaemia Data from Lee (1980)

```
data one;
  input age smear infiltrt label ablasts hightemp response survtime status @@; datalines;
20 78 39 7 0.6 990 1 18 0 25 64 61 16 35.0 1030 1 31 1 26 61 55 12 7.5 982 1 31 0
26 64 64 16 21.0 1000 1 31 0 27 95 95 6 7.5 980 1 36 0 27 80 64 8 0.6 1010 0 1 0
28 88 88 20 4.8 986 1 9 0 28 70 70 14 10.0 1010 1 39 1 31 72 72 5 2.3 988 1 20 1
33 58 58 7 5.7 986 0 4 0 33 92 92 5 2.6 980 1 45 1 33 42 38 12 2.5 984 1 36 0
34 26 26 7 7.0 982 0 12 0 36 55 55 14 4.5 986 1 8 0 37 71 71 15 4.4 1020 0 1 0
40 91 91 9 35.0 986 1 15 0 40 52 49 12 2.1 988 1 24 0 43 74 63 4 0.1 986 0 2 0
45 78 47 14 4.2 980 1 33 0 45 60 36 10 0.6 992 1 29 1 45 82 32 10 28.1 1016 0 7 0
45 79 79 4 1.1 1030 0 0 0 47 56 28 2 0.9 990 0 1 0 48 60 54 10 2.2 1002 0 2 0
50 83 66 19 11.6 996 1 12 0 50 36 32 14 4.5 992 1 9 0 51 88 70 8 0.5 982 0 1 0
52 87 87 7 10.3 986 0 1 0 53 75 68 13 2.3 980 1 9 0 53 65 65 6 2.3 982 0 5 0
56 97 92 10 16.0 992 1 27 1 57 87 83 19 21.6 1020 0 1 0 59 45 45 8 1.1 999 0 13 0
59 36 34 5 0.0 1038 0 1 0 60 39 33 7 0.9 988 0 5 0 60 76 53 12 0.4 982 0 1 0
61 46 37 4 1.4 1006 0 3 0 61 39 8 8 0.3 990 0 4 0 61 90 90 1 9.9 990 0 1 0
62 84 84 19 115.0 1020 1 18 0 63 42 27 5 0.3 1014 0 1 0 65 75 75 10 20.0 1004 0 2 0
71 44 22 6 0.3 990 0 1 0 71 63 63 11 10.0 986 1 8 0 73 33 33 4 0.5 1010 0 3 0
73 93 84 6 38.0 1020 0 4 0 74 58 58 10 2.4 1002 1 14 0 74 32 30 16 6.7 988 0 3 0
75 60 60 17 8.2 990 1 13 0 77 69 69 9 1.5 986 1 13 0 80 73 73 7 1.5 986 0 1 0
;
*Cox PH Model;
proc phreg;
  model survtime*status(1)= age smear infiltrt label ablasts hightemp /
    selection=s slentry=0.10 slstay=0.10;
run;
```

The PHREG Procedure Model Information

Data Set	WORK.TWO
Dependent Variable	survtime
Censoring Variable	status
Censoring Value(s)	1
Ties Handling	BRESLOW

Number of Observations Read	51
Number of Observations Used	51

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
51	45	6	11.76

Step 1. Variable age is entered. The model contains the following explanatory variables: age

Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	291.106	280.289
AIC	291.106	282.289
SBC	291.106	284.096

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	10.8162	1	0.0010
Score	11.1522	1	0.0008
Wald	10.5684	1	0.0012

NOTE: No (additional) variables met the 0.1 level for entry into the model.

Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
age	1	0.03106	0.00955	10.5684	0.0012	1.032

Summary of Stepwise Selection						
Step	Entered	Variable Removed	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	age		1	11.1522	.	0.0008

```

data three; set one; agegrp=(age lt 45);
proc lifetest plot=(s);
  time survtime*status(1);
  strata agegrp;
run;

```

The LIFETEST Procedure						
Stratum 1: agegrp = 0						
Product-Limit Survival Estimates						
survtime	Survival	Failure	Survival Standard Error	Number Failed	Number Left	
0.0000	1.0000	0	0	0	33	
0.0000	0.9697	0.0303	0.0298	1	32	
1.0000	.	.	.	2	31	
1.0000	.	.	.	3	30	
1.0000	.	.	.	4	29	
1.0000	.	.	.	5	28	
1.0000	.	.	.	6	27	
1.0000	.	.	.	7	26	
1.0000	.	.	.	8	25	
1.0000	.	.	.	9	24	
1.0000	.	.	.	10	23	
1.0000	0.6667	0.3333	0.0821	11	22	
2.0000	.	.	.	12	21	
2.0000	0.6061	0.3939	0.0851	13	20	
3.0000	.	.	.	14	19	
3.0000	.	.	.	15	18	
3.0000	0.5152	0.4848	0.0870	16	17	
4.0000	.	.	.	17	16	
4.0000	0.4545	0.5455	0.0867	18	15	
5.0000	.	.	.	19	14	
5.0000	0.3939	0.6061	0.0851	20	13	
7.0000	0.3636	0.6364	0.0837	21	12	
8.0000	0.3333	0.6667	0.0821	22	11	
9.0000	.	.	.	23	10	
9.0000	0.2727	0.7273	0.0775	24	9	
12.0000	0.2424	0.7576	0.0746	25	8	
13.0000	.	.	.	26	7	
13.0000	.	.	.	27	6	
13.0000	0.1515	0.8485	0.0624	28	5	
14.0000	0.1212	0.8788	0.0568	29	4	
18.0000	0.0909	0.9091	0.0500	30	3	
27.0000*	.	.	.	30	2	
29.0000*	.	.	.	30	1	
33.0000	0	1.0000	0	31	0	

NOTE: The marked survival times are censored observations.

Summary Statistics for Time Variable survtime			
Quartile Estimates			
	Point	95% Confidence Interval	
Percent	Estimate	[Lower	Upper)
75	12.0000	5.0000	14.0000
50	4.0000	2.0000	8.0000
25	1.0000	1.0000	3.0000
	Mean	Standard Error	
	7.7576	1.6457	

Stratum 2: agegrp = 1
Product-Limit Survival Estimates

Survival					
survtime	Survival	Failure	Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	18
1.0000	.	.	.	1	17
1.0000	0.8889	0.1111	0.0741	2	16
2.0000	0.8333	0.1667	0.0878	3	15
4.0000	0.7778	0.2222	0.0980	4	14
8.0000	0.7222	0.2778	0.1056	5	13
9.0000	0.6667	0.3333	0.1111	6	12
12.0000	0.6111	0.3889	0.1149	7	11
15.0000	0.5556	0.4444	0.1171	8	10
18.0000	0.5000	0.5000	0.1179	9	9
20.0000*	.	.	.	9	8
24.0000	0.4375	0.5625	0.1185	10	7
31.0000	.	.	.	11	6
31.0000	0.3125	0.6875	0.1129	12	5
31.0000*	.	.	.	12	4
36.0000	.	.	.	13	3
36.0000	0.1562	0.8438	0.0964	14	2
39.0000*	.	.	.	14	1
45.0000*	.	.	.	14	0

NOTE: The marked survival times are censored observations.

Summary Statistics for Time Variable survtime			
Quartile Estimates			
	Point	95% Confidence Interval	
Percent	Estimate	[Lower	Upper)
75	36.0000	24.0000	.
50	21.0000	9.0000	36.0000
25	8.0000	1.0000	18.0000
	Mean	Standard Error	
	20.5139	3.3638	

NOTE: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.

Summary of the Number of Censored and Uncensored Values

Stratum	agegrp	Total	Failed	Censored	Percent Censored
1	0	33	31	2	6.06
2	1	18	14	4	22.22

Total		51	45	6	11.76

Testing Homogeneity of Survival Curves for survtime over Strata

Rank Statistics			
agegrp	Log-Rank	Wilcoxon	
0	9.8204	309.00	
1	-9.8204	-309.00	
Covariance Matrix for the Log-Rank Statistics			
agegrp	0	1	
0	8.79140	-8.79140	
1	-8.79140	8.79140	
Covariance Matrix for the Wilcoxon Statistics			
agegrp	0	1	
0	10469.9	-10469.9	
1	-10469.9	10469.9	
Test of Equality over Strata			
			Pr >
Test	Chi-Square	DF	Chi-Square
Log-Rank	10.9698	1	0.0009
Wilcoxon	9.1196	1	0.0025
-2Log(LR)	14.8910	1	0.0001

```
proc logistic descending data=one;
  model response=age smear infiltrt label ablasts hightemp;
run;
```

The LOGISTIC Procedure		
Model Information		
Data Set	WORK.ONE	
Response Variable	response	
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	
Number of Observations Read	51	
Number of Observations Used	51	
Response Profile		
Ordered Value	response	Total Frequency
1	1	24
2	0	27
Probability modeled is response=1.		
Model Convergence Status		
Convergence criterion (GCONV=1E-8) satisfied.		
Model Fit Statistics		
	Intercept Only	Intercept and Covariates
Criterion		
AIC	72.524	53.275
SC	74.456	66.798
-2 Log L	70.524	39.275

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	31.2493	6	<.0001
Score	22.3265	6	0.0011
Wald	11.7946	6	0.0667

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	108.3	41.8431	6.7025	0.0096
age	1	-0.0623	0.0275	5.1485	0.0233
smear	1	-0.00469	0.0400	0.0137	0.9068
infiltrtrt	1	0.0310	0.0379	0.6712	0.4127
label	1	0.3728	0.1325	7.9203	0.0049
ablasts	1	0.0327	0.0461	0.5034	0.4780
hightemp	1	-0.1116	0.0426	6.8557	0.0088

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits
age	0.940	0.890 0.992
smear	0.995	0.920 1.077
infiltrtrt	1.032	0.958 1.111
label	1.452	1.120 1.882
ablasts	1.033	0.944 1.131
hightemp	0.894	0.823 0.972

Association of Predicted Probabilities and Observed Responses

Percent Concordant	90.1	Somers' D	0.806
Percent Discordant	9.6	Gamma	0.808
Percent Tied	0.3	Tau-a	0.409
Pairs	648	c	0.903