Bio Statistical Methods
Project

The following data are collected from a Veterans Affairs (VA) hospital on lung cancer patients undergoing a placebo controlled randomized trial. The goal of the treatment is to lower the Karnofsky performance status score such that the patients are eligible for surgery and will have a better outcome post-surgery. The treatment given to each patient was double-blinded (neither the doctors nor patients knew what drug the patient was receiving) at the first scheduled visit for that patient following the decision of the monitoring committee.

Additional data for each case was provided such as age, cell histology, diagnosis waiting time (not the same as follow up time) and prior treatment status. After surgery, some patients passed away. Thus, the vital status and time to death after the surgery (in days) was also recorded.

LIST OF VARIABLES:

Variable	**Description**
Therapy	Type of treatment therapy: standard, new test
Cell	Cell histology: Adeno, Large, Squamous
Time	Follow up time (in days) from surgery
Status	Vital status, 1: Dead; 0: Censored
kps	Karnofsky performance status
diagtime	The waiting time for patients diagnosed to start therapy
Age	Age in years
Prior	Prior treatment status: no, yes

1) Please provide **two** appropriate summary statistics for all variables listed in the above table by creating a ***single*** summary table. Please manually create the summary table yourself and do not simply use the SAS output as the table (to keep it more organized since too many SAS outputs stack together in Midterm project). You should only use SAS to help you calculate the statistics.

   Ans:

Variables	Mean	Standard Deviation
Continuous Variables		
Time	153.5730337	179.743407
Kps	61.1516854	19.5267316
Diagtime	8.5168539	8.3741766
Age	57.4606742	10.7913506
Categorical Variables	**Frequency**	**Percent (%)**
Therapy		
Standard	78	43.82
Test	100	56.18
Cell		
Adeno	54	30.34
Large	48	26.97
Squamous	76	42.7
Status		
Dead (0)	12	6.74
Vital Status (1)	166	93.26
Prior		
No	118	66.29
Yes	60	33.71

Above is the summarized data of all the variables.

```
proc means data=finalproject mean
std; var time kps diagtime age; run;

proc freq data=finalproject;
table therapy; table cell;
table status; table prior;
run;
```

2) Is there a significant association between type of therapy and cell type? Use a hypothesis test to answer this question (5-step procedure is needed).

Ans:

Here we are going to test if there is an association between type of therapy and cell type. We are going to use the logistic regression.

Step1: Hypothesis

- H0: There is no association between type of therapy and cell type vs
- H1: There is an association between type of therapy and cell type Step2:
- Significant level alpha = 0.05 Step3:

-

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	246.034	240.218
SC	249.216	249.764
-2 Log L	244.034	234.218

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.8161	2	0.0074
Score	9.7990	2	0.0075
Wald	9.4398	2	0.0089

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
cell	2	9.4398	0.0089

```
proc logistic data=finalproject;
        class cell(ref="Large")/param=ref;
        model therapy(event="test")=cell;
run;
```

- From the result above, the p-value is 0.0089 - We have taken cell type Large as the reference.

Step4:

- From the results above, the Wald chi-sq value is 9.4398 with degrees of freedom 2 and the associated p-value is 0.0089 which is less than the alpha value 0.05.

Step5:

- As the p-value is 0.0089 is less than the alpha value of 0.05, we reject the null hypothesis. We can conclude that there is a significant association between the type of therapy and cell type.

Now we are going to test the specific association of Cell (AdenoVsLarge) and Therapy.

Step1: Hypothesis

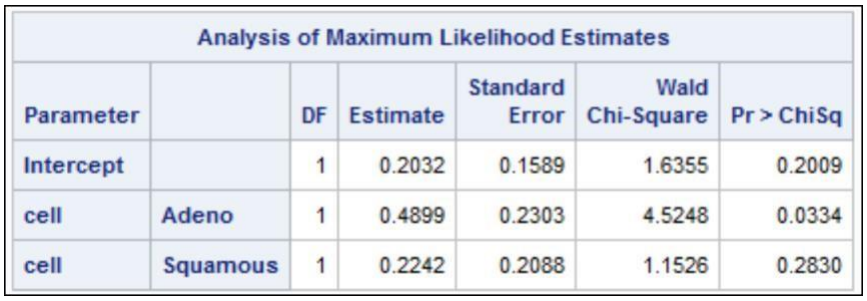
- H0: BcellAvsL = 0

vs

- H1: BcellAvsL not equals to zero

Step2:

- Significant level alpha = 0.05 Step3:

- 

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.2032	0.1589	1.6355	0.2009
cell	Adeno	1	0.4899	0.2303	4.5248	0.0334
cell	Squamous	1	0.2242	0.2088	1.1526	0.2830

proc logistic data=finalproject;

class cell(ref="Large")/param=ref;

model therapy(event="test")=cell;

run;

- From the result above, the p-value is 0.0334 Step4:
- From the results above, the Wald chi-sq value is 4.5248 with degrees of freedom 1 and the associated p-value is 0.0334 which is less than the alpha value 0.05.

Step5:

- As the p-value is 0.0334 is less than the alpha value of 0.05, we reject the null hypothesis. We can conclude that the log odds of the association of therapy and cell type Adeno is different from cell type Large.

Now we are going to test the specific association of Cell (SquamousVsLarge) and Therapy.

Step1: Hypothesis

- H0: BcellSvsL = 0

vs

- H1: BcellSvsL not equal to zero Step2:
- Significant level alpha = 0.05 Step3:

-

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.2032	0.1589	1.6355	0.2009
cell	Adeno	1	0.4899	0.2303	4.5248	0.0334
cell	Squamous	1	0.2242	0.2088	1.1526	0.2830

proc logistic data=finalproject;

class cell(ref="Large")/param=ref;

model therapy(event="test")=cell;
run;

- From the result above, the p-value is 0.2830 Step4:
- From the results above, the Wald chi-sq value is 1.1526 with degrees of freedom 1 and the associated p-value is 0.2830 which is greater than the alpha value 0.05.

Step5:

- As the p-value is 0.2830 is greater than the alpha value of 0.05, we fail to reject the null hypothesis. We can conclude that the log odds of the association of therapy and cell type Squamous are not so different from cell type Large.

From the above conclusions, we can say that (Interpretation of regression coefficients),

The log odds of therapy is 0.4899 more in patients with Cell Type Adeno compared to Cell Type Large.
The log odds of therapy is 0.2242 more in patients with Cell Type Squamous compared to Cell Type Large.

3) Does cell type have an effect on the Karnofsky performance status? Use a hypothesis test to answer this question. If there is a significant effect, please indicate which cell pairs are different. (5-step procedure is needed).

Ans:

Here we are going to test if there is a significant effect of Cell type on KPS and we are going to use logistic regression.

Step1: Hypothesis

- H0: There is no significant effect of Cell type on KPS. vs
- H1: There is a significant effect of Cell type on KPS.

Step2:

- Significant level alpha = 0.05

Step3:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	898.257	893.196
SC	945.984	947.287
-2 Log L	868.257	859.196

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.0606	2	0.0108
Score	8.8693	2	0.0119
Wald	8.9222	2	0.0115

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
cell	2	8.9222	0.0115

-

```
proc logistic data=finalproject;
        class cell(ref="Large")/param=ref;
        model kps=cell;
run;
```

- From the result above, the p-value is 0.0115 - We have taken cell type Large as the reference.

Step4:

- From the results above, the Wald chi-sq value is 8.9222 with degrees of freedom 2 and the associated p-value is 0.0115 which is less than the alpha value 0.05.

Step5:

- As the p-value is 0.0115 is less than the alpha value of 0.05, we reject the null hypothesis. We can conclude that there is a significant effect of Cell type on KPS.

Now we are going to test the specific effect of Cell (AdenoVsLarge) and KPS.

Step1: Hypothesis

- H0: BcellAvsL = 0

vs

- H1: BcellAvsL not equals to zero Step2:
- Significant level alpha = 0.05

Step3:

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	10	1	-4.5785	0.7117	41.3830	<.0001
Intercept	20	1	-3.1534	0.3647	74.7780	<.0001
Intercept	25	1	-2.7237	0.3028	80.8901	<.0001
Intercept	30	1	-2.1023	0.2372	78.5815	<.0001
Intercept	35	1	-1.8985	0.2207	74.0041	<.0001
Intercept	40	1	-1.4864	0.1938	58.7995	<.0001
Intercept	45	1	-1.3800	0.1882	53.7752	<.0001
Intercept	50	1	-0.8270	0.1663	24.7453	<.0001
Intercept	55	1	-0.6396	0.1615	15.6924	<.0001
Intercept	60	1	-0.0109	0.1544	0.0050	0.9438
Intercept	65	1	0.0834	0.1546	0.2913	0.5894
Intercept	70	1	0.8849	0.1686	27.5376	<.0001
Intercept	75	1	1.1770	0.1799	42.8178	<.0001
Intercept	80	1	2.2940	0.2593	78.2693	<.0001
Intercept	85	1	2.3613	0.2663	78.6340	<.0001
cell	Adeno	1	0.4651	0.1937	5.7661	0.0163
cell	Squamous	1	0.1118	0.1767	0.4005	0.5268

- proc logistic data=finalproject;

class cell(ref="Large")/param=ref;

model kps=cell;

run;

- From the result above, the p-value is 0.0163 Step4:
- From the results above, the Wald chi-sq value is 5.7661 with degrees of freedom 1 and the associated p-value is 0.0163 which is less than the alpha value 0.05.

Step5:

- As the p-value is 0.0163 is less than the alpha value of 0.05, we reject the null hypothesis. We can conclude that the log odds of association of KPS for patients with Cell Type Adeno is significantly different from Cell Type Large.

Now we are going to test the specific association of Cell (SquamousVsLarge) and KPS.

Step1: Hypothesis

- H0: BcellSvsL = 0

vs

- H1: BcellSvsL not equal to zero Step2:
- Significant level alpha = 0.05 Step3:

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	10	1	-4.5785	0.7117	41.3830	<.0001
Intercept	20	1	-3.1534	0.3647	74.7780	<.0001
Intercept	25	1	-2.7237	0.3028	80.8901	<.0001
Intercept	30	1	-2.1023	0.2372	78.5815	<.0001
Intercept	35	1	-1.8985	0.2207	74.0041	<.0001
Intercept	40	1	-1.4864	0.1938	58.7995	<.0001
Intercept	45	1	-1.3800	0.1882	53.7752	<.0001
Intercept	50	1	-0.8270	0.1663	24.7453	<.0001
Intercept	55	1	-0.6396	0.1615	15.6924	<.0001
Intercept	60	1	-0.0109	0.1544	0.0050	0.9438
Intercept	65	1	0.0834	0.1546	0.2913	0.5894
Intercept	70	1	0.8849	0.1686	27.5376	<.0001
Intercept	75	1	1.1770	0.1799	42.8178	<.0001
Intercept	80	1	2.2940	0.2593	78.2693	<.0001
Intercept	85	1	2.3613	0.2663	78.6340	<.0001
cell	Adeno	1	0.4651	0.1937	5.7661	0.0163
cell	Squamous	1	0.1118	0.1767	0.4005	0.5268

- proc logistic data=finalproject;
  class cell(ref="Large")/param=ref;
  model kps=cell;

run;

- From the result above, the p-value is 0.5268 Step4:
- From the results above, the Wald chi-sq value is 0.4005 with degrees of freedom 1 and the associated p-value is 0.5268 which is greater than the alpha value 0.05.

Step5:

- As the p-value is 0.5268 is greater than the alpha value of 0.05, we fail to reject the null hypothesis. We can conclude that the log odds of association of KPS for patients with Cell Type Squamous is not significantly different from Cell Type Large.

From the above conclusions, we can say that (Interpretation of regression coefficients),

The log odds of KPS is 0.4651 more in patients with Cell Type Adeno compared to Cell Type Large.
The log odds of KPS is 0.1118 more in patients with Cell Type Squamous compared to Cell Type Large.

4) Does the type of therapy have an effect on the Karnofsky performance status after including confounders and effect modifiers? Answer this by first checking to see if age and cell type are effect modifiers or confounders (following the workflow of Page 7 Week 10). Retain only significant effect modifiers (as interaction) and confounders (as main effect) in the model. Provide the prediction equation for the final optimal model and interpret the effect of therapy on Karnofsky performance status. (note, if there are interaction terms in your final model, e.g., cell*therapy, in the interpretation, you should interpret the coefficients between therapy and Karnofsky performance when cell = Adeno, Large, and Squamous, separately).

Ans:

First, we are going to model the crude relationship between the Type of therapy and KPS. I have taken the Standard therapy type as a reference.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	64.74358974	B	2.18761649	29.60	<.0001
therapy test	-6.39358974	B	2.91864444	-2.19	0.0298
therapy standard	0.00000000	B	.	.	.

```
proc glm data=finalproject;
class therapy(ref="standard");
model kps=therapy /solution;
run;
```

From the above result, as the p-value is less than 0.05, we can conclude that there is a relation between the Type of therapy and KPS.

Next, I will check for the interaction for each potential variable.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	73.50000000	B	3.42544790	21.46	<.0001
therapy test	-15.44444444	B	5.59373300	-2.76	0.0064
therapy standard	0.00000000	B	.	.	.
cell Adeno	-12.38888889	B	5.59373300	-2.21	0.0281
cell Squamous	-15.33333333	B	4.84431488	-3.17	0.0018
cell Large	0.00000000	B	.	.	.
therapy*cell test Adeno	8.22222222	B	7.78614681	1.06	0.2924
therapy*cell test Squamous	19.23429952	B	7.11870429	2.70	0.0076
therapy*cell test Large	0.00000000	B	.	.	.
therapy*cell standard Adeno	0.00000000	B	.	.	.
therapy*cell standard Squamous	0.00000000	B	.	.	.
therapy*cell standard Large	0.00000000	B	.	.	.

```
proc glm data=finalproject;
class therapy(ref="standard") cell(ref="Large");
model kps=therapy cell therapy*cell /solution;
run;
```

As there is one significant interaction term which is BTherapy*Cell(SquamousVsLarge), so the overall interaction is significant.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	78.61381805	B	10.97458241	7.16	<.0001
therapy test	-6.36934420	B	15.69963060	-0.41	0.6855
therapy standard	0.00000000	B	.	.	.
age	-0.24521256	B	0.19015410	-1.29	0.1989
age*therapy test	0.00631170	B	0.26892992	0.02	0.9813
age*therapy standard	0.00000000	B	.	.	.

```
proc glm data=finalproject;
class therapy(ref="standard");
model kps=therapy age therapy*age /solution;
run;
```

As the p-value (0.9813) of the interaction term is greater than 0.05, the interaction is insignificant.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
therapy	1	921.391521	921.391521	2.92	0.0891
time	1	6832.447624	6832.447624	21.68	<.0001
time*therapy	1	14.970752	14.970752	0.05	0.8277

```
proc glm data=finalproject;
class therapy(ref="standard");
model kps=therapy time therapy*time /solution;
run;
```

As the p-value (0.8277) of the interaction term is greater than 0.05, the interaction is insignificant.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	79.16666667	B	7.81180009	10.13	<.0001
therapy test	-10.83333333	B	11.04755363	-0.98	0.3281
therapy standard	0.00000000	B	.	.	.
Status 1	-15.62500000	B	8.13077932	-1.92	0.0563
Status 0	0.00000000	B	.	.	.
therapy*Status test 1	5.00443262	B	11.44678802	0.44	0.6625
therapy*Status test 0	0.00000000	B	.	.	.
therapy*Status standard 1	0.00000000	B	.	.	.
therapy*Status standard 0	0.00000000	B	.	.	.

```
proc glm data=finalproject; class
therapy(ref="standard") status(ref="0"); model
kps=therapy status therapy*status /solution; run;
```

As the p-value (0.6625) of the interaction term is greater than 0.05, the interaction is insignificant.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	68.71230024	B	3.06343934	22.43	<.0001
therapy test	-7.46814867	B	4.16319311	-1.79	0.0746
therapy standard	0.00000000	B	.	.	.
diagtime	-0.41274589	B	0.22516613	-1.83	0.0685
diagtime*therapy test	0.03491932	B	0.35171942	0.10	0.9210
diagtime*therapy standard	0.00000000	B	.	.	.

proc glm data=finalproject;
class therapy(ref="standard");
model kps=therapy diagtime therapy*diagtime /solution;
run;

As the p-value (0.9210) of the interaction term is greater than 0.05, the interaction is insignificant.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	67.29166667	B	2.78608884	24.15	<.0001
therapy test	-9.43452381	B	3.61732013	-2.61	0.0099
therapy standard	0.00000000	B	.	.	.
prior yes	-6.62500000	B	4.49243327	-1.47	0.1421
prior no	0.00000000	B	.	.	.
therapy*prior test yes	8.26785714	B	6.15827380	1.34	0.1812
therapy*prior test no	0.00000000	B	.	.	.
therapy*prior standard yes	0.00000000	B	.	.	.
therapy*prior standard no	0.00000000	B	.	.	.

proc glm data=finalproject; class
therapy(ref="standard") prior(ref="no"); model
kps=therapy prior therapy*prior /solution; run;

As the p-value (0.1812) of the interaction term is greater than 0.05, the interaction is insignificant.

Now we check for cofounding

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	64.74358974	B	2.18761649	29.60	<.0001
therapy test	-6.39358974	B	2.91864444	-2.19	0.0298
therapy standard	0.00000000	B	.	.	.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	78.43532533	B	7.88949862	9.94	<.0001
therapy test	-6.00729164	B	2.90797421	-2.07	0.0403
therapy standard	0.00000000	B	.	.	.
age	-0.24205698		0.13408044	-1.81	0.0727

proc glm data=finalproject; class
therapy(ref="standard"); model
kps=therapy age /solution; run;

%change = [(-6.00729164--6.39358974)/-6.00729164)]*100 = 6.4258% < 10%
So, age is not considered as the cofounder.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	58.53859116	B	2.26535940	25.84	<.0001
therapy test	-7.27007235	B	2.67871659	-2.71	0.0073
therapy standard	0.00000000	B	.	.	.
time	0.04361055		0.00741523	5.88	<.0001

proc glm data=finalproject; class
therapy(ref="standard"); model
kps=therapy time /solution; run;

%change = [(-7.27007235--6.39358974)/-7.27007235)]*100 = 12.05% > 10%
So, time is considered as the cofounder.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	76.83594825	B	5.69672416	13.49	<.0001
therapy test	-6.17189650	B	2.88553997	-2.14	0.0338
therapy standard	0.00000000	B	.	.	.
Status 1	-13.10005505	B	5.70991613	-2.29	0.0230
Status 0	0.00000000	B	.	.	.

proc glm data=finalproject;
class therapy(ref="standard")
status(ref="0"); model kps=therapy status
/solution; run;

%change = [(-6.17189650--6.39358974)/-6.17189650)]*100 = -3.5921% < 10%
So, status is not considered as the cofounder.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	68.57469156	B	2.72420983	25.17	<.0001
therapy test	-7.17268261	B	2.90300386	-2.47	0.0144
therapy standard	0.00000000	B	.	.	.
diagtime	-0.39843459		0.17248692	-2.31	0.0221

proc glm data=finalproject; class
therapy(ref="standard"); model
kps=therapy diagtime /solution; run;

%change = [(-7.17268261--6.39358974)/-7.17268261)]*100 = 10.86% > 10%
So, diagtime is considered as the cofounder.

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	65.59941520	B	2.49034675	26.34	<.0001
therapy test	-6.58187135	B	2.93421114	-2.24	0.0261
therapy standard	0.00000000	B	.	.	.
prior yes	-2.22514620	B	3.07979660	-0.72	0.4710
prior no	0.00000000	B	.	.	.

proc glm data=finalproject;
class therapy(ref="standard")
prior(ref="no"); model kps=therapy prior
/solution; run;

%change = [(-6.58187135--6.39358974)/-6.58187135)]*100 = 2.8609% < 10%
So, prior is not considered as the cofounder.

So, the final model is

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	68.64751014	B	3.90759299	17.57	<.0001
therapy test	-14.59682900	B	5.20150035	-2.81	0.0056
therapy standard	0.00000000	B	.	.	.
cell Adeno	-8.48030518	B	5.29696827	-1.60	0.1112
cell Squamous	-11.29408749	B	4.56700058	-2.47	0.0144
cell Large	0.00000000	B	.	.	.
therapy*cell test Adeno	7.49274274	B	7.22398452	1.04	0.3011
therapy*cell test Squamous	13.21987123	B	6.70192983	1.97	0.0502
therapy*cell test Large	0.00000000	B	.	.	.
therapy*cell standard Adeno	0.00000000	B	.	.	.
therapy*cell standard Squamous	0.00000000	B	.	.	.
therapy*cell standard Large	0.00000000	B	.	.	.
time	0.03792332		0.00794645	4.77	<.0001
diagtime	-0.31188086		0.16545826	-1.88	0.0611

proc glm data=finalproject;
class therapy(ref="standard") cell(ref="Large"); model
kps=therapy cell therapy*cell time diagtime/solution; run;

The prediction equation is:
kps = 68.64-14.59therapy(TvsSt)-8.48cell(AvsL)-
11.29(SvsL)+7.49(therapy(TvsSt)*cell(AvsL))+13.21(therapy(TvsSt)*cell(SvsL))+0.03time-0.31diagtime

Interpretation of regression coefficients:
Interpreting kps for therapy type of test (non-reference level so we take it as 1):
kps = 68.64-14.59(1)-8.48cell(AvsL)-
11.29(SvsL)+7.49(1*cell(AvsL))+13.21(1*cell(SvsL))+0.03time-0.31diagtime
= 68.64-14.59+7.49+13.21-8.48cell(AvsL)-
11.29(SvsL)+7.49cell(AvsL)+13.21cell(SvsL)+0.03time-0.31diagtime
= 74.75-0.99cell(AvsL)+1.92cell(SvsL)+0.03time-0.31diagtime

For the therapy type Test, the kps for cell type Adeno is 0.99 times less than cell type Large with an increase in 0.03 days and a decrease in 0.31 waiting days.

For the therapy type Test, the kps for cell type Squamous is 1.92 times higher than cell type Large with an increase in 0.03 days and a decrease in 0.31 waiting days.

Interpreting kps for therapy type of Standard (non-reference level so we take it as 0):
kps = 68.64-14.59(0)-8.48cell(AvsL)-
11.29(SvsL)+7.49(0*cell(AvsL))+13.21(0*cell(SvsL))+0.03time-0.31diagtime
= 68.64-8.48cell(AvsL)-11.29(SvsL)+0.03time-0.31diagtime

For the therapy type Standard, the kps for cell type Adeno is 8.48 times less than cell type Large with an increase in 0.03 days and a decrease in 0.31 waiting days.

For the therapy type Standard, the kps for cell type Squamous is 11.29 times less than cell type Large with an increase in 0.03 days and a decrease in 0.31 waiting days.

5) Please create a new variable, kpsc, that dichotomizes the Karnofsky performance status such that scores over 60 are at the ‘high’ level and scores at or below 60 are at the ‘low’

level. Again, does cell type have an effect on Kpsc? Use a hypothesis test to answer this question (5-step procedure is needed).

Ans:

Creating a new field:

VIEWTABLE: Work.Finalprojectnew

	therapy	cell	time	Status	kps	diagtime	age	prior	kpsc
1	standard	Squamous	72	1	70	7	69	no	high
2	standard	Squamous	411	1	60	5	64	yes	low
3	standard	Squamous	228	1	65	3	38	no	high
4	standard	Squamous	126	1	75	9	63	yes	high
5	standard	Squamous	118	1	50	11	65	yes	low
6	standard	Squamous	10	1	20	5	49	no	low
7	standard	Squamous	82	1	55	10	69	yes	low
8	standard	Squamous	110	1	75	29	68	no	high
9	standard	Squamous	314	1	50	18	43	no	low
10	standard	Squamous	100	0	65	6	70	no	high
11	standard	Squamous	42	1	55	4	81	no	low

```
data finalprojectnew; set
finalproject; if kps > 60 then
kpsc="high"; if kps <= 60
then kpsc="low"; run;
```

Here we are going to test if there is a significant effect of Cell type on KPSC. We are going to use Logistic Regression.

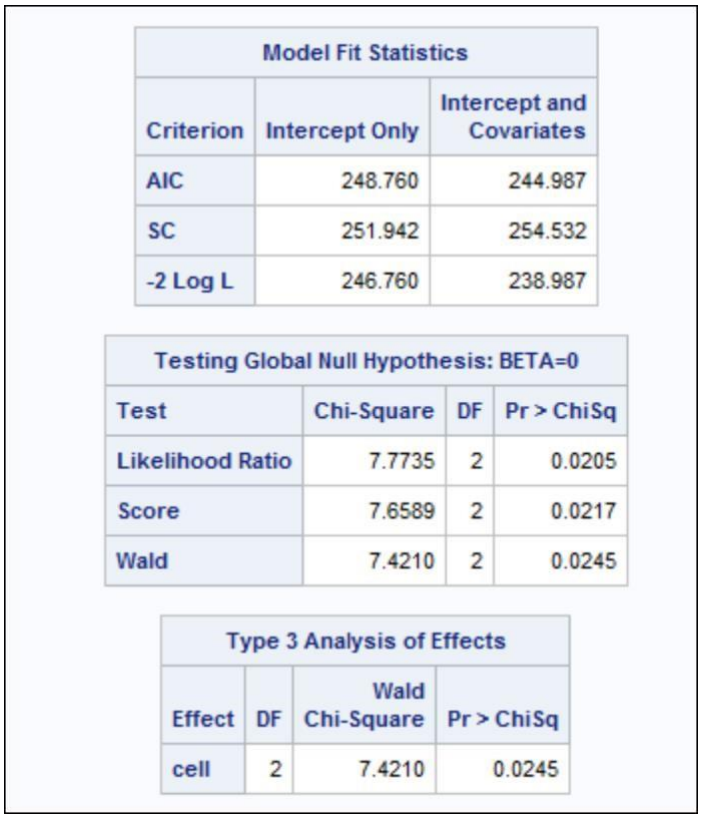
Step1: Hypothesis

- H0: There is no significant effect of Cell type on KPSC. vs
- H1: There is a significant effect of Cell type on KPSC.

Step2:

- Significant level alpha = 0.05

Step3:



Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	248.760	244.987
SC	251.942	254.532
-2 Log L	246.760	238.987

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7.7735	2	0.0205
Score	7.6589	2	0.0217
Wald	7.4210	2	0.0245

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
cell	2	7.4210	0.0245

-

```
proc logistic data=finalprojectnew;
        class cell(ref="Large")/param=ref;
        model kpsc(event="low")=cell;
run;
```

- From the result above, the p-value is 0.0245 - We have taken cell type Large as the reference.

Step4:

- From the results above, the Wald chi-sq value is 7.4210 with degrees of freedom 2 and the associated p-value is 0.0245 which is less than the alpha value 0.05.

Step5:

- As the p-value is 0.0245 is less than the alpha value of 0.05, we reject the null hypothesis. We can conclude that there is an overall significant effect of Cell type on KPSC.

Now we are going to test the specific effect of Cell (AdenoVsLarge) and KPS.

Step1: Hypothesis

- H0: BcellAvsL = 0

vs

- H1: BcellAvsL not equals to zero Step2:
- Significant level alpha = 0.05 Step3:

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.0534	0.1576	0.1148	0.7348
cell	Adeno	1	0.4281	0.2245	3.6363	0.0565
cell	Squamous	1	0.2116	0.2061	1.0540	0.3046

-

```
proc logistic data=finalprojectnew;
        class cell(ref="Large")/param=ref;
        model kpsc(event="low")=cell;
run;
```

- From the result above, the p-value is 0.0565 Step4:
- From the results above, the Wald chi-sq value is 3.6363 with degrees of freedom 1 and the associated p-value is 0.0565 which is greater than the alpha value 0.05.

Step5:

- As the p-value is 0.0565 is greater than the alpha value of 0.05, we fail to reject the null hypothesis. We can conclude that the log odds of association of KPSC for patients with Cell Type Adeno is not significantly different from Cell Type Large.

Now we are going to test the specific association of Cell (SquamousVsLarge) and KPSC.

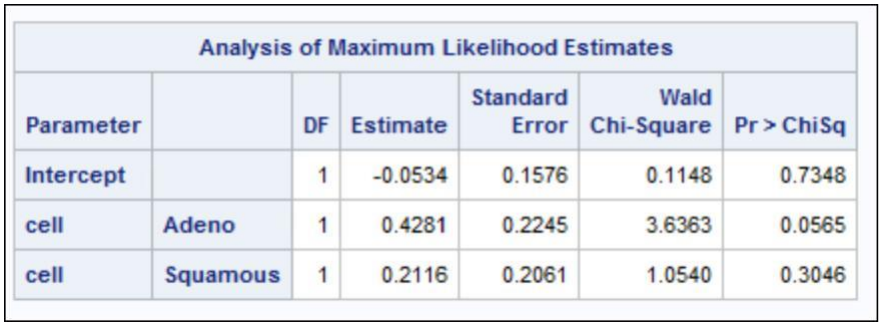
Step1: Hypothesis

- H0: BcellAvsL = 0

vs

- H1: BcellAvsL not equals to zero Step2:
- Significant level alpha = 0.05

Step3:

- 

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.0534	0.1576	0.1148	0.7348
cell	Adeno	1	0.4281	0.2245	3.6363	0.0565
cell	Squamous	1	0.2116	0.2061	1.0540	0.3046

proc logistic data=finalprojectnew;

class cell(ref="Large")/param=ref;

model kpsc(event="low")=cell;

run;

- From the result above, the p-value is 0.3046 Step4:
- From the results above, the Wald chi-sq value is 1.0540 with degrees of freedom 1 and the associated p-value is 0.3046 which is greater than the alpha value 0.05.

Step5:

- As the p-value is 0.3046 is greater than the alpha value 0.05, we fail to reject the null hypothesis. We can conclude that the log odds of association of KPSC for paitents with Cell Type Squamous is not significantly different from Cell type Large.

6) Does the type of treatment have an effect on the categorized Karnofsky performance level (Kpsc) after adjusting for age, cell type, diagnosis time (diagtime) and prior treatment? Provide appropriate hypothesis, test statistics, and p-values used to draw this conclusion. (list the model you are using, and then use the 5-step hypothesis test to draw the conclusion.)

Ans:

Here we are going to conduct a hypothesis test to check whether treatment affects KPSC using Logistic regression.

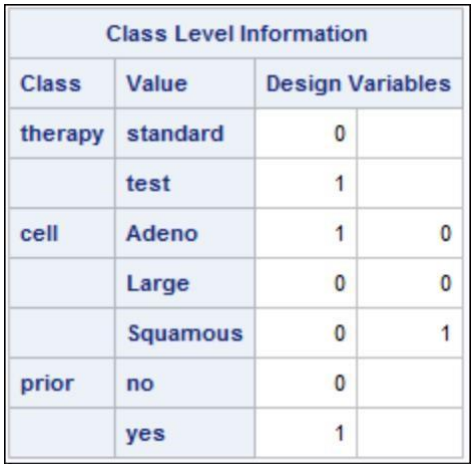
Step1: Hypothesis

- H0: There is no significant effect of treatment on KPSC. vs
- H1: There is a significant effect of treatment on KPSC.

Step2:

- Significant level alpha = 0.05 Step3: proc logistic data=finalprojectnew;

class therapy(ref="standard") cell(ref="Large") prior(ref="no")/param=ref;
model kpsc(event="low")=therapy cell age diagtime prior;
run;

- 

Class Level Information			
Class	Value	Design Variables	
therapy	standard	0	
	test	1	
cell	Adeno	1	0
	Large	0	0
	Squamous	0	1
prior	no	0	
	yes	1	

- From the result above, check the design variables
- We have already told SAS to choose standard for therapy, Large for cell, no for prior as reference groups resptectively. We can confirm this from the above screenshot.

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
therapy	1	2.7801	0.0954
cell	2	5.6806	0.0584
age	1	1.1925	0.2748
diagtime	1	2.5334	0.1115
prior	1	0.0040	0.9496

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.1413	0.9186	5.4338	0.0198
therapy	test	1	0.5421	0.3251	2.7801	0.0954
cell	Adeno	1	1.0265	0.4337	5.6031	0.0179
cell	Squamous	1	0.6361	0.3981	2.5529	0.1101
age		1	0.0160	0.0147	1.1925	0.2748
diagtime		1	0.0379	0.0238	2.5334	0.1115
prior	yes	1	0.0231	0.3658	0.0040	0.9496

-

- From the above screenshot, after adjusting the variables age, cell type, diagnosis time, prior, the p-value of therapy is 0.0954.

Step4:

- From the results above, the Wald chi-sq value is 2.7801 with degrees of freedom 1 and the associated p-value is 0.0954 which is greater than the alpha value 0.05. This is same for Btherapy(TestvsStandard).

Step5:

- As the p-value is 0.0954 is greater than the alpha value 0.05, we fail to reject the null hypothesis. We can conclude that there is no significant effect of Treatment on KPSC.