**Final Project**

**Harini Peddi**

**U00885938**

**1.**

|  |  |  |
| --- | --- | --- |
| **Continuous Variable** | **Mean** | **Standard Deviation** |
| Time | 153.573 | 179.7434 |
| Kps | 61.1512 | 19.5267 |
| Diagtime | 8.5169 | 8.3742 |
| Age | 57.4607 | 10.7914 |
|  |  |  |
| **Categorical Variable** | **Frequency** | **Percent** |
| Therapy |  |  |
| Standard | 78 | 43.82 |
| New Test | 100 | 56.18 |
| Cell |  |  |
| Adeno | 54 | 30.14 |
| Large | 48 | 26.97 |
| Squamous | 76 | 42.70 |
| Status |  |  |
| Dead (1) | 166 | 93.26 |
| Censored (0) | 12 | 6.74 |
| Prior |  |  |
| No | 118 | 66.29 |
| Yes | 60 | 33.71 |

**2.**

**Step1:** H0: There is no connection between the type of therapy and cell type. vs H1: There is a relationship between the type of therapy and cell type.

**Step2:** Select significance level, α=0.05.

**Step3:** SAS is used to conduct the chi-square test. The SAS output shows:

Statistic Value Prob

Chi-Square 9.7990 0.0075

A screenshot of a table of statistics

Description automatically generated

**Step4:** The SAS results reveal a Chi-square value of 9.7990 with 2 degrees of freedom, and a p-value of 0.0075.

**Step5:** With a p-value of 0.0075, lower than the significance level of 0.05, we reject the null hypothesis. Sufficient evidence supports the alternative hypothesis, indicating a significant link between the type of therapy and cell type.

**Sas Code:**

**data** valung;

infile="C:\Users\ hpeddi\New folder\SAS\valung\_formatted.csv" out=mydata dbms=csv replace;

input therapy cell time Status kps diagtime age prior;

**run**;

**proc** **freq** data=mydata;

tables therapy \* cell / chisq;

**run**;

3.

**Step1:** Null hypothesis (H0): cell\*kps=0 Vs Alternative Hypothesis (H1):

cell\*kps! =0

**Step2:** Select significance level, α=0.05.

**Step3:** It is recommended to use SAS for data analysis. The SAS output provides the following information.

A screenshot of a computer

Description automatically generated

**A screenshot of a computer

Description automatically generated**

**Step4:** The table indicates a p-value of 0.0113, with a degree of freedom of 2.

**Step5:** Since the p-value is less than 0.05, we reject the null hypothesis, indicating a significant association between cell type and Karnofsky performance status.

**Sas Code:**

**proc** **glm** data=mydata;

class cell;

model kps = cell;

lsmeans cell / cl adjust=bon;

**run**;

**4.**

**Crude:**

A screenshot of a computer screen

Description automatically generated

**Age:** Age has a negligible impact on the interaction.

A screenshot of a computer

Description automatically generated

**Check for confounding variables** - not a confounder, around 6%.

A screenshot of a computer screen

Description automatically generated

**Cell Type:**

**A screenshot of a graph

Description automatically generated**

**Final Model:**

**A screenshot of a test results

Description automatically generated**

adenocarcinoma cases: The average Karnofsky Performance Score (kps) of the standard treatment exceeds that of the experimental therapy by 7.22 points.

In cases of large cell carcinoma: The average kps of the standard treatment exceeds that of the experimental therapy by 15.44 points.

In squamous cell carcinoma cases: The average kps of the standard treatment is 3.7898 points lower than that of the experimental treatment.

**Sas Code:**

**proc** **glm** data=mydata plots=diagnostics;

class therapy;

model kps = therapy age therapy\*age/solution;

**run**;

**proc** **glm** data=mydata plots=diagnostics;

class therapy;

model kps = therapy/solution;

**run**;

**proc** **glm** data=mydata plots=diagnostics;

class therapy;

model kps = therapy age /solution;

**run**;

**proc** **glm** data=mydata plots=diagnostics;

class therapy;

model kps = therapy cell therapy\*cell/solution;

**run**;

**5**

**Step 1:** The null hypothesis (H0) states that the product of cell and kpsc is equal to zero, while the alternative hypothesis (H1) states that the product of cell and kpsc is not equal to zero.

**Step 2**: A level of significance, α=0.05, is chosen.

**Step 3**: It is advisable to utilize SAS software, which provides the following output.

A screenshot of a cell procedure

Description automatically generated

A screenshot of a table

Description automatically generated

**Step 4:** The obtained p-value from the table is 0.0217, with 2 degrees of freedom.

**Step 5**: Since the p-value is below 0.05, we can reject the null hypothesis, indicating a significant association between cell type and kpsc.

**Sas Code:**

**data** mydata;

set mydata;

if kps > **60** then kpsc = 'high';

else kpsc = 'low';

**run**;

**proc** **freq** data=mydata;

tables cell \* kpsc / chisq;

**run**;

**6.**

**Step1:** Null Hypothesis (H0): After adjusting for age, cell type, diagnosis time, and prior treatment, no impact is observed on KPSC. Vs Alternative Hypothesis (H1): After adjusting for age, cell type, diagnosis time, and prior treatment, an impact on KPSC is observed.

**Step2:** Selecting a level of significance, α=0.05.

**Step3:** It is recommended to utilize SAS software. The output generated by SAS presents the following results.

A screenshot of a screenshot of a test results

Description automatically generated

A screenshot of a table

Description automatically generated

**Step4:** The obtained p-value from the table is less than 0.0001, with 6 degrees of freedom.

**Step5:** Since the p-value is below 0.05, we can reject the null hypothesis, indicating a significant effect on KPSC after controlling for age, cell type, diagnosis time, and prior treatment.

**Sas Code:**

**SAS Code:**

**data** mydata;

set mydata;

if kps > **60** then kpsc = **1**;

else kpsc = **0**;

**run**;

**proc** **logistic** data=mydata;

class therapy (ref="test") prior (ref="yes") cell (ref="Squamous")/param=ref;

model kpsc(event='1') = age diagtime prior therapy cell;

**run**;