Assignment Content:

1. Open the BMT file available in the SASHELP directory. The file contains data from 137 bone marrow transplant patients. The Group variable contains data on patient classification into one of three risk categories: ALL (acute lymphoblastic leukemia), AML-Low Risk (acute myelocytic leukemia, low risk), and AML-High Risk. The dependent variable of interest is the disease-free survival time in days, T. Conduct the test to infer if the mean days of survival is different for the three disease groups and state your results.
   1. As shown below, the p-value of 0.0012 is less than the alpha value of 0.05 and therefore we reject the null and conclude that the mean days of survival is different for the three disease groups. Also, when comparing the three groups to each other, we discover that groups one and two (ALL and AML-High Risk) have a comparison p-value of 0.9890 which is greater than alpha of 0.05 so we conclude that groups 1 and 2 are not significantly different from each other. The comparison between groups 1 and 3 (ALL and AML-Low Risk), and groups 2 and 3 (AML-High Risk and AML-Low Risk) have p-values of 0.0081 and 0.0031 respectively. Therefore, we conclude groups 1 and 3, and groups 2 and 3 are significantly different from each other.

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **Group** | 3 | ALL AML-High Risk AML-Low Risk |

|  |  |
| --- | --- |
| **Number of Observations Read** | 137 |
| **Number of Observations Used** | 137 |

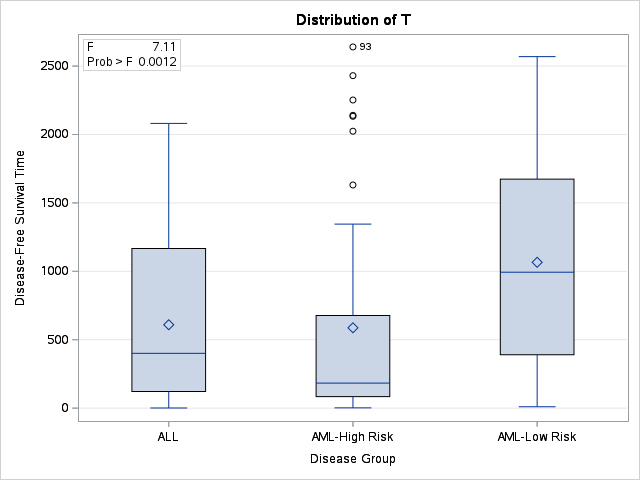
**Dependent Variable: T Disease-Free Survival Time**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 2 | 7186442.04 | 3593221.02 | 7.11 | 0.0012 |
| **Error** | 134 | 67675769.84 | 505043.06 |  |  |
| **Corrected Total** | 136 | 74862211.88 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **T Mean** |
| --- | --- | --- | --- |
| 0.095996 | 90.87433 | 710.6638 | 782.0292 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Group** | 2 | 7186442.042 | 3593221.021 | 7.11 | 0.0012 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Group** | 2 | 7186442.042 | 3593221.021 | 7.11 | 0.0012 |

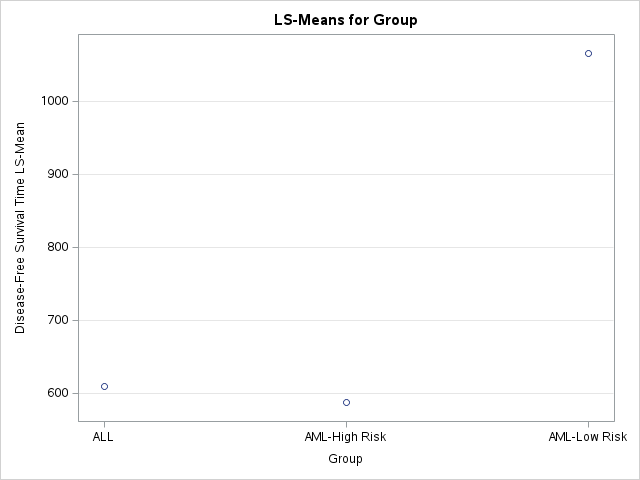


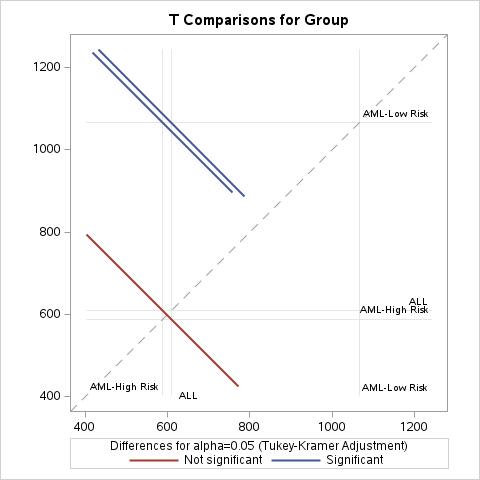
**Least Squares Means**

**Adjustment for Multiple Comparisons: Tukey-Kramer**

| **Group** | **T LSMEAN** | **LSMEAN Number** |
| --- | --- | --- |
| **ALL** | 609.42105 | 1 |
| **AML-High Risk** | 587.28889 | 2 |
| **AML-Low Risk** | 1065.77778 | 3 |

| **Least Squares Means for effect Group Pr > |t| for H0: LSMean(i)=LSMean(j)  Dependent Variable: T** | | | |
| --- | --- | --- | --- |
| **i/j** | **1** | **2** | **3** |
| **1** |  | 0.9890 | 0.0081 |
| **2** | 0.9890 |  | 0.0031 |
| **3** | 0.0081 | 0.0031 |  |





1. The heart dataset in the SASHELP directory contains data from the Framingham heart study. We would like to explore the relationship between Smoking status (5 categories) and the Age at Death variables. (Note that although the study contains data from 5209 individuals, only a subset of them are dead). Conduct the appropriate analysis and state your conclusions.
   1. As shown below, the p-value of the 1,971 observations used (those that are dead) is <0.0001 which is less than the alpha value of 0.05. Therefore, we reject the null and conclude that the relationship between Smoking status (5 categories) and the Age at Death variables is different. Furthermore, when we compare the 5 groups to each other we find that groups 1 and 3 have a p-value of 0.9616, and groups 2 and 3 have a p-value of 0.3078. These p-values are greater than the alpha value of 0.05 and therefore are not significantly different from each other. All other groups compared to each other have p-values less than the alpha value of 0.05 and therefore are significantly different from one another.

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **Smoking\_Status** | 5 | Heavy (16-25) Light (1-5) Moderate (6-15) Non-smoker Very Heavy (> 25) |

|  |  |
| --- | --- |
| **Number of Observations Read** | 5209 |
| **Number of Observations Used** | 1971 |

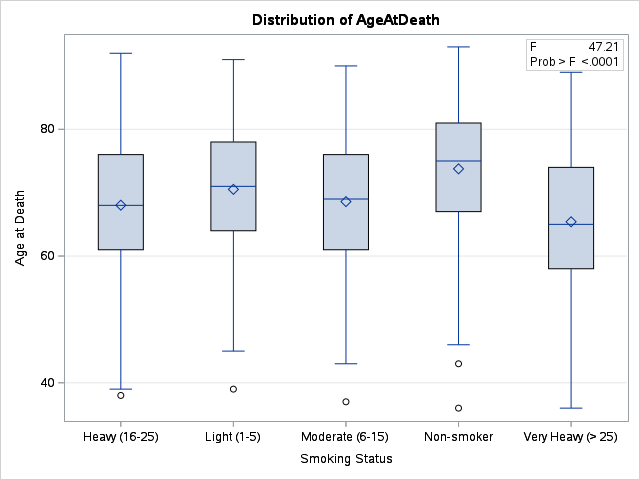
**Dependent Variable: AgeAtDeath Age at Death**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 4 | 19086.1404 | 4771.5351 | 47.21 | <.0001 |
| **Error** | 1966 | 198696.2137 | 101.0662 |  |  |
| **Corrected Total** | 1970 | 217782.3541 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **AgeAtDeath Mean** |
| --- | --- | --- | --- |
| 0.087639 | 14.23927 | 10.05317 | 70.60173 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Smoking\_Status** | 4 | 19086.14039 | 4771.53510 | 47.21 | <.0001 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Smoking\_Status** | 4 | 19086.14039 | 4771.53510 | 47.21 | <.0001 |

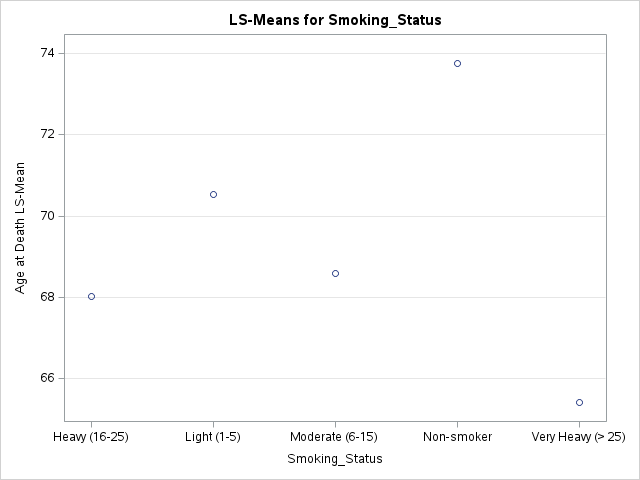


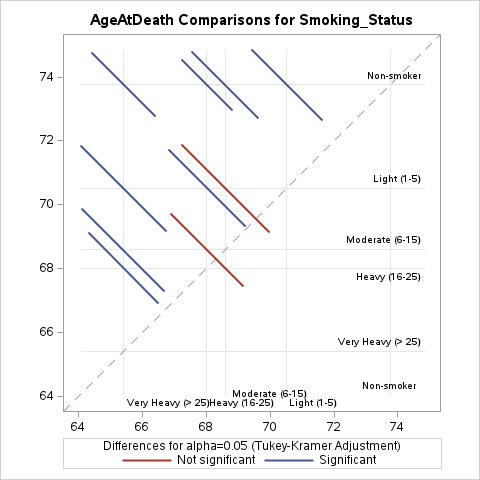
**Least Squares Means**

**Adjustment for Multiple Comparisons: Tukey-Kramer**

| **Smoking\_Status** | **AgeAtDeath LSMEAN** | **LSMEAN Number** |
| --- | --- | --- |
| **Heavy (16-25)** | 68.0248307 | 1 |
| **Light (1-5)** | 70.5240642 | 2 |
| **Moderate (6-15)** | 68.5915493 | 3 |
| **Non-smoker** | 73.7609428 | 4 |
| **Very Heavy (> 25)** | 65.4092827 | 5 |

| **Least Squares Means for effect Smoking\_Status Pr > |t| for H0: LSMean(i)=LSMean(j)  Dependent Variable: AgeAtDeath** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **i/j** | **1** | **2** | **3** | **4** | **5** |
| **1** |  | 0.0357 | 0.9616 | <.0001 | 0.0109 |
| **2** | 0.0357 |  | 0.3078 | 0.0006 | <.0001 |
| **3** | 0.9616 | 0.3078 |  | <.0001 | 0.0073 |
| **4** | <.0001 | 0.0006 | <.0001 |  | <.0001 |
| **5** | 0.0109 | <.0001 | 0.0073 | <.0001 |  |





1. Use the HBAT\_200 dataset for this question. We would like to examine the relationship between (X1) Customer Type (1= customer for less than a year, 2 = between 1 to 5 years, 3= customer for more than 5 years) and whether they would be likely to recommend (X20). Conduct the analysis and state your conclusions.
   1. As shown below, the p-value of <.0001 is less than the alpha value of 0.05 therefore we reject the null and conclude that the relationship between (X1) Customer Type and likely to recommend (X20) are different. Furthermore, the p-value between groups 2 and 3 is 0.1204 so we conclude this relationship is not significantly different. All other group comparisons have p-values <.0001 so we conclude these comparisons to be significantly different from one another.

| 1. **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **X1\_\_\_Customer\_Type** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 200 |
| **Number of Observations Used** | 200 |

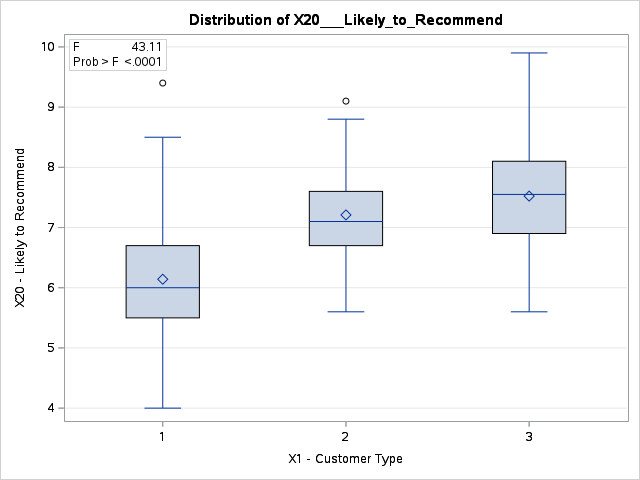
**Dependent Variable: X20\_\_\_Likely\_to\_Recommend X20 - Likely to Recommend**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 2 | 71.0427574 | 35.5213787 | 43.11 | <.0001 |
| **Error** | 197 | 162.3159926 | 0.8239390 |  |  |
| **Corrected Total** | 199 | 233.3587500 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **X20\_\_\_Likely\_to\_Recommend Mean** |
| --- | --- | --- | --- |
| 0.304436 | 13.05589 | 0.907711 | 6.952500 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **X1\_\_\_Customer\_Type** | 2 | 71.04275735 | 35.52137868 | 43.11 | <.0001 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **X1\_\_\_Customer\_Type** | 2 | 71.04275735 | 35.52137868 | 43.11 | <.0001 |



**Least Squares Means**

**Adjustment for Multiple Comparisons: Tukey-Kramer**

| **X1\_\_\_Customer\_Type** | **X20\_\_\_Likely\_to\_Recommend LSMEAN** | **LSMEAN Number** |
| --- | --- | --- |
| **1** | 6.14117647 | 1 |
| **2** | 7.20937500 | 2 |
| **3** | 7.52205882 | 3 |

| **Least Squares Means for effect X1\_\_\_Customer\_Type Pr > |t| for H0: LSMean(i)=LSMean(j)  Dependent Variable: X20\_\_\_Likely\_to\_Recommend** | | | |
| --- | --- | --- | --- |
| **i/j** | **1** | **2** | **3** |
| **1** |  | <.0001 | <.0001 |
| **2** | <.0001 |  | 0.1204 |
| **3** | <.0001 | 0.1204 |  |

