

BIOSTATISTICS_FINAL EXAM

***** QUESTION 1 *****

```
proc import
    datafile = "/home/u63635623/Final Exam/BoneMarrow.xlsx"
    dbms = xlsx
    out = bone;
run;
```

Imports the dataset to SAS.

a. Descriptive stat

```
proc means
    data = bone;
    class group;
    var t;
run;
```

The MEANS Procedure

Analysis Variable : T T						
Group	N Obs	N	Mean	Std Dev	Minimum	Maximum
ALL	9	9	162.0000000	147.4364270	1.0000000	466.0000000
AML-High Risk	5	5	2319.00	215.9212820	2133.00	2640.00
AML-Low Risk	9	9	240.6666667	251.5173950	10.0000000	704.0000000

INTERPRETATION:

ALL: There are 9 observations, with a mean disease-free survival time of 162. The values range from 1 to 466, with a standard deviation of approximately 147.44.

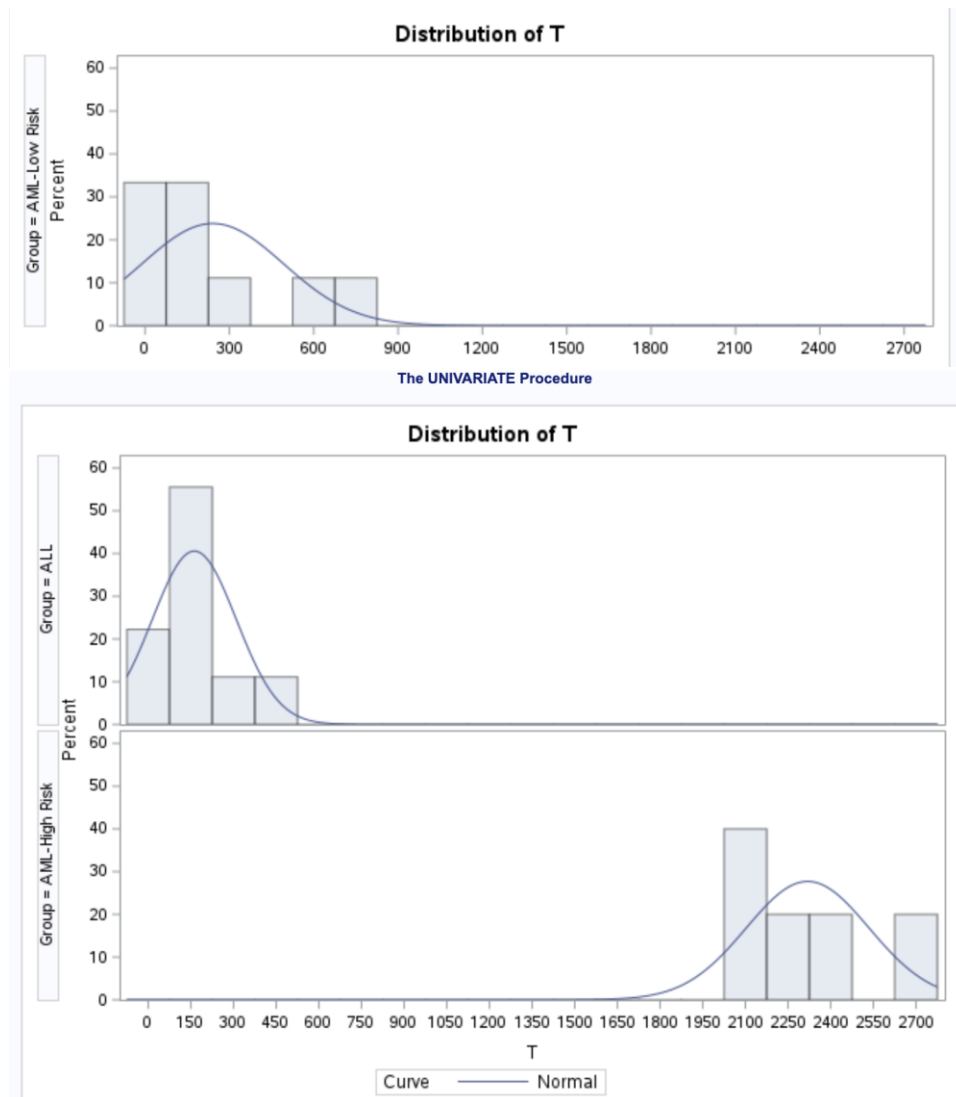
AML-High Risk: In this group, there are 5 observations. The mean disease-free survival time' value is significantly higher (2319), with values ranging from 2133 to 2640. The standard deviation is about 215.92.

AML-Low Risk: This group has 9 observations. The mean disease-free survival time' value is 240.67, with a wider range of values (from 10 to 704) compared to the other groups. The standard deviation is approximately 251.52.

(b) Check the normality assumption.

```
proc univariate data= bone normal;
class group;
```

```
var T;
histogram T/normal;
run;
```



INTERPRETATION:

Based on the graphs, they do not follow a normal distribution.

(c) Is disease-free survival time different among the three groups? State the null and alternative hypotheses. Which test do you use? Choose the most appropriate test and justify it. Conduct the test at the 0.05 level of significance. If needed, conduct post-hoc tests. What do you conclude?

Null Hypothesis: There is no difference in the disease-free survival time between the 3 groups: ALL (acute lymphoblastic leukemia), AML-Low Risk (acute myelocytic leukemia, low risk), and AML-High Risk.

Alternative hypothesis : There is difference in disease-free survival time between at least one of the 3 groups: ALL (acute lymphoblastic leukemia), AML-Low Risk (acute myelocytic leukemia, low risk), and AML-High Risk.

We use anova test in this case. ANOVA is used when comparing the means of three or more groups for a single variable. It helps determine if there's a statistically significant difference in disease-free survival times among the groups (ALL, AML-Low Risk, AML-High Risk). ANOVA allows for this simultaneous comparison while considering the assumption of equal variances among the groups.

```
proc anova data=bone;
class group;
model T =group;
means group/hovtest;
run;
```

The ANOVA Procedure

Dependent Variable: T T

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	17575938.87	8787969.43	202.84	<.0001
Error	20	866476.00	43323.80		
Corrected Total	22	18442414.87			

R-Square	Coeff Var	Root MSE	T Mean
0.953017	31.45611	208.1437	661.6957

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Group	2	17575938.87	8787969.43	202.84	<.0001

/* With a p value of <.0001 and an alpha of 0.05, We reject the null hypothesis. This suggests that there's evidence that at least one of the groups (ALL, AML-Low Risk, AML-High Risk) differs in disease-free survival time. */

/* posthoc */

The results of the ANOVA test indicated that at least one of the groups (ALL, AML-Low Risk, and AML-High Risk) had a different disease-free survival time. Similar to Tukey's HSD, the post hoc test is performed to identify which particular groups show these significant differences.

This test helps to clarify which pairs of groups differ in disease-free survival time, enhancing our understanding beyond the general difference highlighted by ANOVA.

```
PROC GLM DATA=bone;
CLASS Group;
MODEL T = Group;
LSMEANS Group / DIFF ADJUST=TUKEY;
RUN;
```

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Group	T LSMEAN	LSMEAN Number
ALL	162.00000	1
AML-High Risk	2319.00000	2
AML-Low Risk	240.66667	3

Least Squares Means for effect Group
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: T

i/j	1	2	3
1		<.0001	0.7063
2	<.0001		<.0001
3	0.7063	<.0001	

Interpretation:

Comparing 'ALL' with 'AML-High Risk' and 'AML-Low Risk' showed significant differences in disease-free survival time ($p < 0.0001$).

There was also a significant difference between 'AML-High Risk' and 'AML-Low Risk' ($p < 0.0001$).

However, there doesn't appear to be a significant difference between 'ALL' and 'AML-Low Risk' ($p = 0.7063$), based on these pairwise comparisons.

/* *****QUESTION 2***** */

```
data clinic;
  input Treatment $ Response $ Count;
  datalines;
Placebo Satisfactory 2
Placebo Unsatisfactory 3
```

```

Test Satisfactory 4
Test Unsatisfactory 1
;
proc print data= clinic;
run;

```

- Create a 2x2 table and print it using SAS.
- Is there any association between treatment and response? State the null and alternative hypotheses. Which test do you use? Choose the most appropriate test and justify it. Conduct the test at the 0.05 level of significance. What do you conclude?

Null hypothesis: There is no association between treatment and response.
Alternative hypothesis: Not Null hypothesis

```

proc freq data= clinic;
table Treatment*response/chisq;
Weight count;
run;

```

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of Treatment by Response			
	Treatment	Response		Total
		Satisfac	Unsatisf	
	Placebo	2 20.00 40.00 33.33	3 30.00 60.00 75.00	5 50.00
	Test	4 40.00 80.00 66.67	1 10.00 20.00 25.00	5 50.00
	Total	6 60.00	4 40.00	10 100.00

- This is 2x2 table.
-

Statistics for Table of Treatment by Response

Statistic	DF	Value	Prob
Chi-Square	1	1.6667	0.1967
Likelihood Ratio Chi-Square	1	1.7261	0.1889
Continuity Adj. Chi-Square	1	0.4167	0.5186
Mantel-Haenszel Chi-Square	1	1.5000	0.2207
Phi Coefficient		-0.4082	
Contingency Coefficient		0.3780	
Cramer's V		-0.4082	
WARNING: 100% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	2
Left-sided Pr <= F	0.2619
Right-sided Pr >= F	0.9762
Table Probability (P)	0.2381
Two-sided Pr <= P	0.5238

Sample Size = 10

INTERPRETATION:

The Chi-square test is used to check the relationship between treatment and response. It is used to assess the association between treatment and response because both variables are categorical. This test determines if there's a significant relationship between categorical variables.

Since more than 20% of the cells had expected counts below 5, indicating low counts, the Chi-square test might be unreliable. Therefore, Fisher's exact test is preferred. With a p-value of 0.5238 at a significance level of 0.05, we fail to reject the null hypothesis. Therefore, based on this analysis, there isn't enough evidence to claim an association between the treatment administered and the observed response.

```
/* ***** QUESTION 3 ***** */
proc import
    datafile = "/home/u63635623/Final Exam/hypoxia.xlsx"
```

```
dbms = xlsx
out = hypo;
run;
```

- a. Compute descriptive statistics for TWA_HR (time-weighted average of heart rate during surgery), age, gender (female), sleeptime (time spent at Sao2<90%), AHI (Apnea/Hypnea index).

```
proc means
data = hypo maxdec=2;
var TWA_HR age female sleeptime AHI;
run;
```

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
TWA_HR	TWA_HR	203	76.10	10.16	54.22	109.09
Age	Age	281	47.29	11.12	16.40	73.80
Female	Female	281	0.72	0.45	0.00	1.00
Sleeptime	Sleeptime	281	18.71	23.73	0.00	99.60
AHI	AHI	278	2.87	1.08	1.00	4.00

Interpretation:

TWA_HR: The average time-weighted heart rate during surgery was approximately 76 beats per minute, ranging from 54 to 109 beats per minute.

Age: The average age in the dataset was around 47 years, with ages ranging from 16 to 73 years.

Gender (Female): The mean value of 0.72 suggests that, on average, about 72% of the dataset consists of females.

Sleeptime: On average, patients spent approximately 18.71% of their time at a Sao2 level below 90%, with a wide variation from 0% to almost 100%.

AHI: The average Apnea/Hypopnea Index was 2.87, ranging from 0 to 16.40, indicating the frequency of apnea and hypopnea events during sleep.

- b. Is TWA_HR different among AHI? State null and alternative hypotheses. Which test do you use? Conduct the test at the 0.05 level of significance. Interpret the results.

Null Hypothesis: TTWA_HR is different among AHI

Alternative Hypothesis: Not null

Anova test is used here.

```
proc anova data=hypo;
class AHI;
model TWA_HR = AHI;
means AHI/hovtest;
run;
```

The ANOVA Procedure					
Dependent Variable: TWA_HR TWA_HR					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	774.77128	258.25709	2.53	0.0586
Error	196	20027.85868	102.18295		
Corrected Total	199	20802.62996			

R-Square	Coeff Var	Root MSE	TWA_HR Mean
0.037244	13.28056	10.10856	76.11545

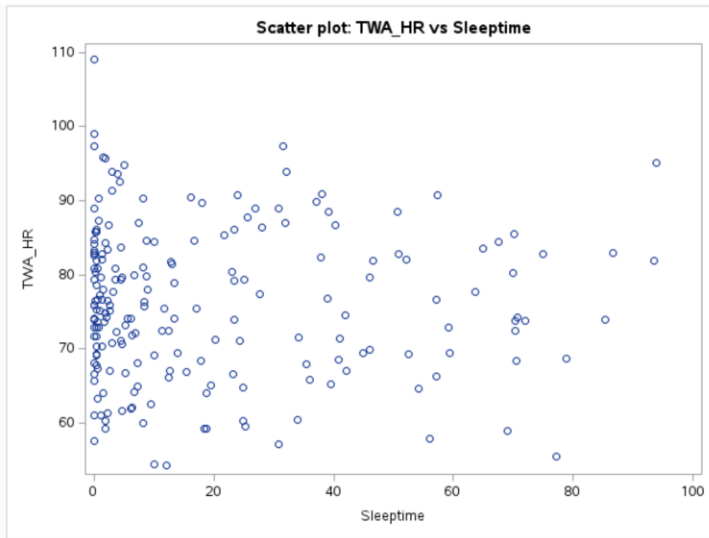
Source	DF	Anova SS	Mean Square	F Value	Pr > F
AHI	3	774.7712785	258.2570928	2.53	0.0586

With a p value of 0.0586 and an alpha of 0.05, We do not reject the null hypothesis. There isn't enough evidence to conclude that TTWA_HR significantly differs among different AHI levels.

c. Produce a scatter plot of TWA_HR versus Sleep time. Produce another scatter plot of TWA versus minimum nocturnal Sao2. Conduct a correlation analysis for these three variables and interpret the results.

Scatter plot of TWA_HR versus Sleep time:

```
title "Scatter plot: TWA_HR vs Sleep time";
proc sgplot data= hypo;
scatter y=TWA_HR x=Sleep time;
run;
title "";
```

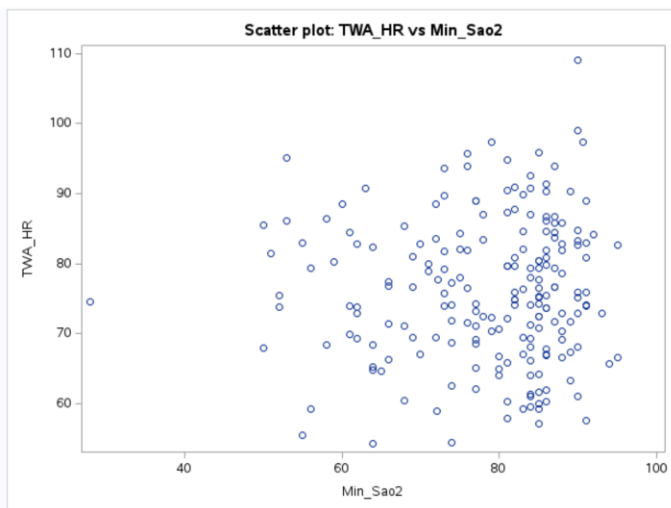



Scatter plot of TWA versus minimum nocturnal Sao2:

```

title "Scatter plot: TWA_HR vs Min_Sao2";
proc sgplot data= hypo;
scatter y=TWA_HR x=Min_Sao2 ;
run;
title "";

```



Correlation analysis for these three variables and interpret the results.

Null Hypothesis: The population correlation coefficient (ρ) is 0 and there is no correlation between TWA_HR and AHI.

Alternative Hypothesis: Not null hypothesis.

Null Hypothesis: The population correlation coefficient (ρ) is 0 and there is no correlation between TWA_HR and Min_Sao2.

Alternative Hypothesis: Not null hypothesis.

Null Hypothesis: The population correlation coefficient (ρ) is 0 and there is no correlation between AHI and Min_Sao2.

Alternative Hypothesis: Not null hypothesis.

```
proc corr data=hypo plots=matrix(histogram);  
var TWA_HR AHI Min_Sao2;  
run;
```

Pearson Correlation Coefficients Prob > r under H0: Rho=0 Number of Observations			
	TWA_HR	AHI	Min_Sao2
TWA_HR TWA_HR	1.00000 203	-0.07266 0.3066 200	0.03874 0.5832 203
AHI AHI	-0.07266 0.3066 200	1.00000 278	-0.50245 <.0001 278
Min_Sao2 Min_Sao2	0.03874 0.5832 203	-0.50245 <.0001 278	1.00000 281

With a p value of 0.3066 and alpha of 0.05, we fail to reject the null hypothesis. So, there's insufficient evidence to claim a correlation between TWA_HR and AHI.

With a p value of <0.0001 and alpha of 0.05, we reject the null hypothesis. So, there is no correlation AHI and Min_Sao2.

With a p value of 0.5832 and alpha of 0.05, we fail to reject the null hypothesis. So, there's insufficient evidence to claim a correlation between TWA_HR and Min_Sao2.