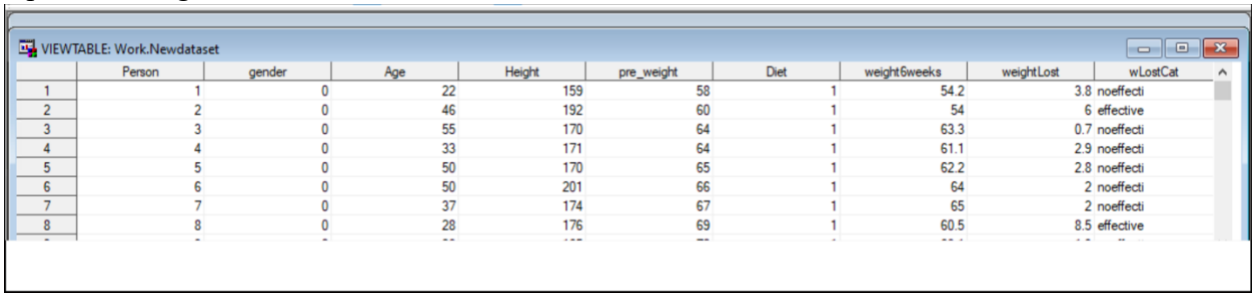
Biostatistical Methods
Project

This diet data set contains information about a study that investigated 78 people using one of three diets. There are seven variables in this dataset. The name and a brief variable description are listed below.

Variable name	Variable
Person	Participant number
gender	Gender, 1 = male, 0 = female
Age	Age (years)
Height	Height (cm)
pre.weight	Weight before the diet (kg)
Diet	Diet, 1 =control diet, 2 =test diet BeautyA 3= test diet BeautyB
weight6weeks	Weight after 6 weeks (kg)

1) The aim of this study was to investigate the weight loss after three different diets. Please create a new variable called **weightlost**, which represents the difference between weight before the diet treatment and the weight after 6 weeks of diet treatment. Furthermore, the researcher thinks that a weight loss of more than 5 kg is considered an effective response to the diet treatment. Please create another variable called **wlostcat**, which has two levels: *effective* if the weight loss is larger than 5kg; *noeffective* if the weight loss is smaller or equal than 5kg.


VIEWTABLE: Work.Newdataset

	Person	gender	Age	Height	pre_weight	Diet	weight6weeks	weightLost	wLostCat
1	1	0	22	159	58	1	54.2	3.8	noeffecti
2	2	0	46	192	60	1	54	6	effective
3	3	0	55	170	64	1	63.3	0.7	noeffecti
4	4	0	33	171	64	1	61.1	2.9	noeffecti
5	5	0	50	170	65	1	62.2	2.8	noeffecti
6	6	0	50	201	66	1	64	2	noeffecti
7	7	0	37	174	67	1	65	2	noeffecti
8	8	0	28	176	69	1	60.5	8.5	effective

```
data newdataset;
        set midterm;    weightLost =
pre_weight-weight6weeks;
```

```
        if weightLost > 5 then wLostCat = 'effective';
        else if weightLost <= 5 then wLostCat =
'noeffective'; run;
```

2) Please provide two appropriate descriptive statistics for *all* variables (The six original variables and the two new created variables, excluding Person ID). Please presenting your results in a single summary table. You will need to manually create this table and not just attached SAS output.

The SAS System

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
gender	78	0.4487179	0.5005824	0	1.0000000
Age	78	39.1538462	9.8152769	16.0000000	60.0000000
Height	78	170.8205128	11.2766206	141.0000000	201.0000000
pre_weight	78	72.5256410	8.7233443	58.0000000	103.0000000
Diet	78	2.0384615	0.8129201	1.0000000	3.0000000
weight6weeks	78	68.6807692	8.9245038	53.0000000	103.0000000
weightLost	78	3.8448718	2.5514777	-2.1000000	9.2000000

```
data newdatset1;
        set newdataset;
        drop Person;
run;

proc means data=newdatset1;
        var gender Age Height pre_weight Diet weight6weeks weightLost; run;
```

The SAS System

The FREQ Procedure

wLostCat	Frequency	Percent	Cumulative Frequency	Cumulative Percent
effective	25	32.05	25	32.05
noeffecti	53	67.95	78	100.00

```
proc freq data=newdatset1;
        table wLostCat;
run;
```

3) Is wlostcat associated with diet treatment? Conduct a hypothesis test to answer this question.

Ans.

Step1: Hypothesis

- H0: The wLostCat and Diet are not Associated vs
- H1: The wLostCat and Diet are Associated

Step2:

- Significant level alpha = 0.05 Step3:

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of wLostCat by Diet			
	Diet			
wLostCat	1	2	3	Total
effective	4 5.13 16.00 16.67	6 7.69 24.00 22.22	15 19.23 60.00 55.56	25 32.05
noeffecti	20 25.64 37.74 83.33	21 26.92 39.62 77.78	12 15.38 22.64 44.44	53 67.95
Total	24 30.77	27 34.62	27 34.62	78 100.00

Statistics for Table of wLostCat by Diet

Statistic	DF	Value	Prob
Chi-Square	2	10.6551	0.0049
Likelihood Ratio Chi-Square	2	10.5248	0.0052
Mantel-Haenszel Chi-Square	1	8.9767	0.0027
Phi Coefficient		0.3696	
Contingency Coefficient		0.3467	
Cramer's V		0.3696	

```
proc freq data=newdatset1;
table wLostCat*Diet/chisq; run;
```

- We are using the Chi-square for the Hypothesis test.

Step4:

- From the results above, the value of Chi-square is 10.6551 with degrees of freedom 2 and the p-value is 0.0049.

Step5:

- The p-value is 0.0049, which is less than the alpha value which is 0.005. So, we reject the null hypothesis. There is sufficient evidence to support the alternative hypotheses that wLostCat and Diet are associated.

4) Is wlostcat associated with gender? Conduct a hypothesis test to answer this question.

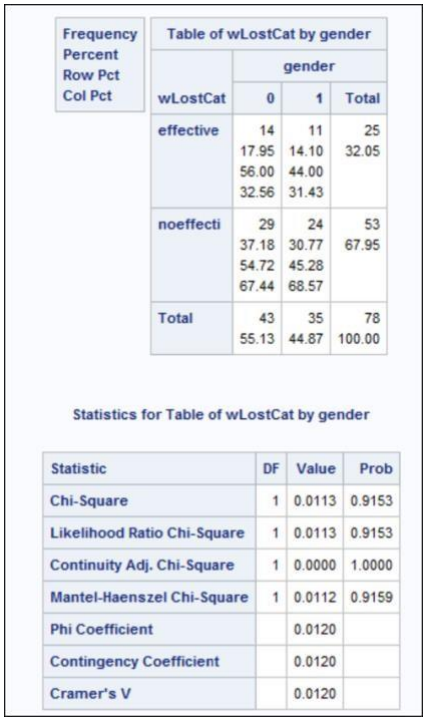
Ans.

Step1: Hypothesis

- H0: The wLostCat and gender are not Associated vs
- H1: The wLostCat and gender are Associated

Step2:

- Significant level alpha = 0.05 Step3:



Frequency
Percent
Row Pct
Col Pct

Table of wLostCat by gender

wLostCat	gender 0	gender 1	Total
effective	14 17.95 56.00 32.56	11 14.10 44.00 31.43	25 32.05
noeffecti	29 37.18 54.72 67.44	24 30.77 45.28 68.57	53 67.95
Total	43 55.13	35 44.87	78 100.00

Statistics for Table of wLostCat by gender

Statistic	DF	Value	Prob
Chi-Square	1	0.0113	0.9153
Likelihood Ratio Chi-Square	1	0.0113	0.9153
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0112	0.9159
Phi Coefficient		0.0120	
Contingency Coefficient		0.0120	
Cramer's V		0.0120	

```
proc freq data=newdatset1;
table wLostCat*gender/chisq;
run;
```

- We are using the Chi-square for the Hypothesis test.

Step4:

- From the results above, the value of Chi-square is 0.0113 with degrees of freedom 1 and the p-value is 0.9153.

Step5:

- The p-value is 0.9153, which is greater than the alpha value which is 0.005. So, we fail to reject the null hypothesis. There is no sufficient evidence to support the alternative hypotheses that wLostCat and gender are associated.

5) For subjects on BeautyA diet, was there a significant decrease in weight from baseline (pre.weight) to after diet treatment (weight6weeks)? Conduct a hypothesis test to answer this question.

Ans.

Step1: Hypothesis

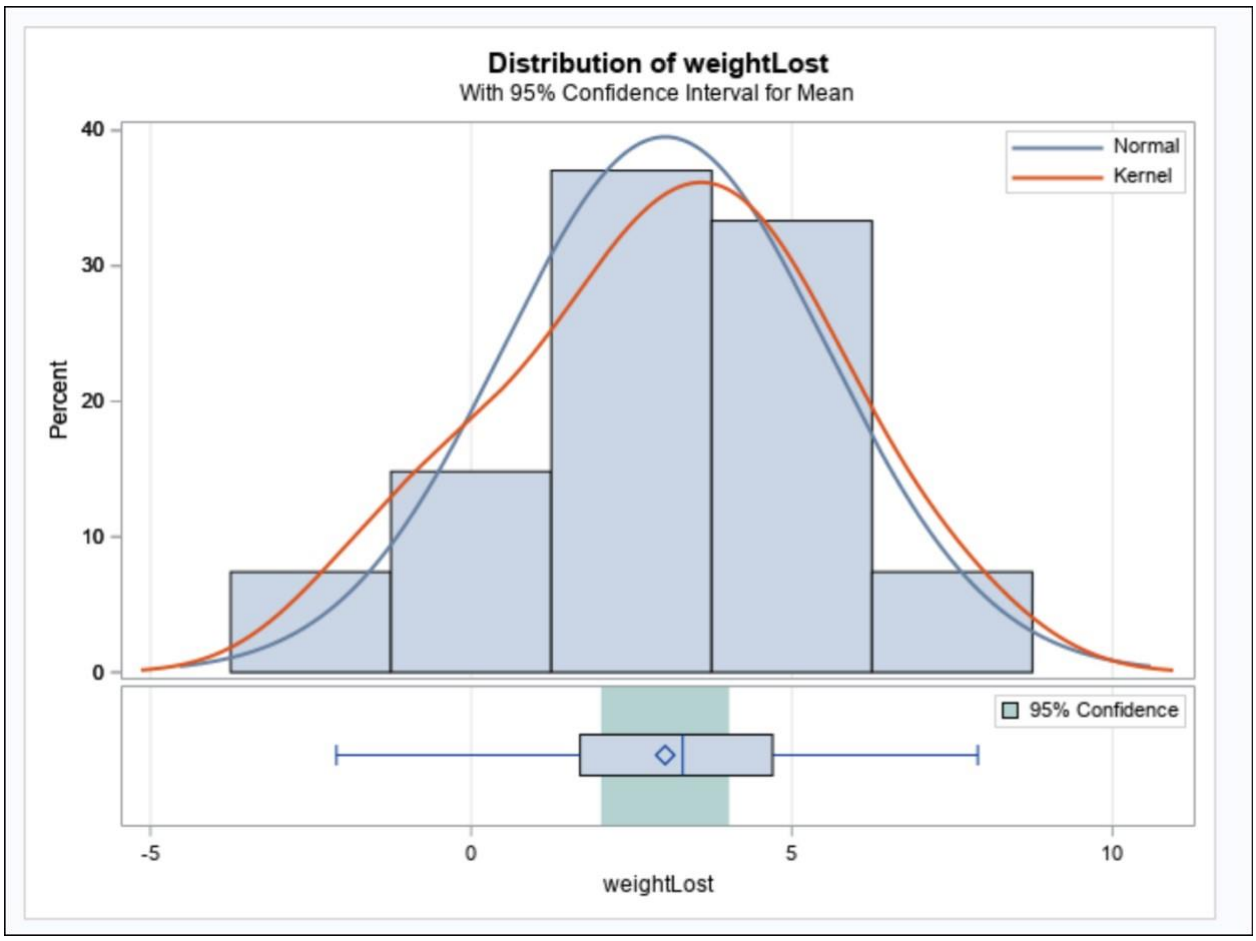
- H0: uafter - ubefore = 0

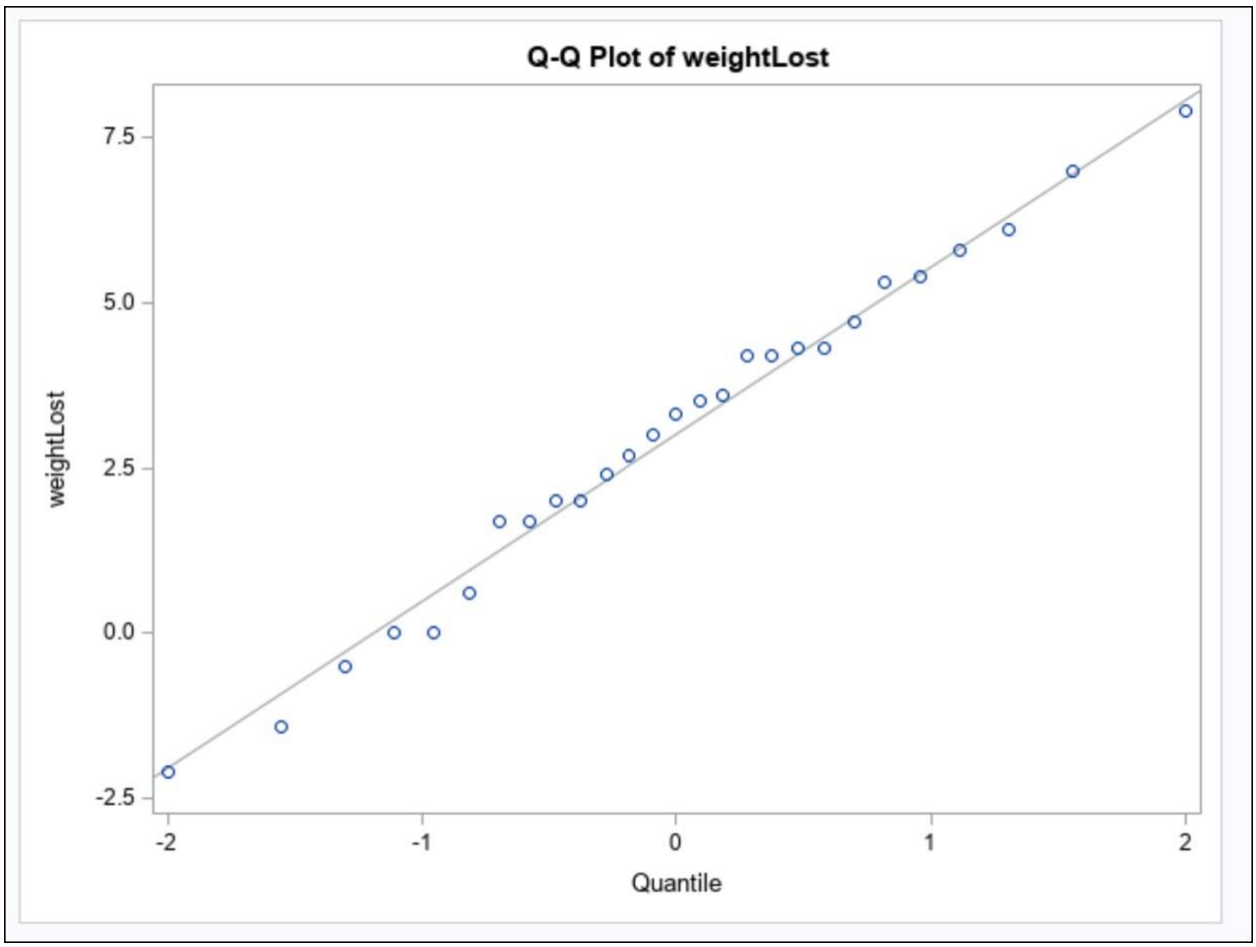
vs

- H1: uafter - ubefore not equals to 0

Step2:

- Significant level alpha = 0.05 Step3:


Distribution of weightLost
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
40
30
20
10
0
95% Confidence
-5
0
5
10
weightLost



- To define which test we need to use, we take a look at the above graphs. Just want to confirm that the values are distributed normally. From the above QQ plot graph, we see that dots are aligned in a straight line.We can conclude that it is a unimodal and subjects are distributed normally. So we can use t-test. Step4:

The SAS System

The TTEST Procedure

Variable: weightLost

N	Mean	Std Dev	Std Err	Minimum	Maximum
27	3.0259	2.5234	0.4856	-2.1000	7.9000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
3.0259	2.0277	4.0241	2.5234	1.9872	3.4581

DF	t Value	Pr > |t|
26	6.23	<.0001

```
proc ttest data=newdatset1 h0=0 alpha=0.05;
 var weightLost;
where Diet = 2; run;
```

- From the results above, when n=78, with 26 degrees of freedom, the p-value is < 0.0001 and the t value is 6.23.

Step5:

- The p-value is < 0.0001, which is less than the alpha value which is 0.05. So, we reject the null hypothesis. There is sufficient evidence to support the alternative hypothesis that there is the decrease in weight when compared with before and after the diet for BeautyA. 6) Was there a difference in weight loss (weightlost) between the two diet groups (Beauty A and Beauty B)? Conduct a hypothesis test to answer this question.

Ans.

Step1: Hypothesis

- H0: uBeautyA = uBeautyB

vs

- H1: uBeautyA not equals uBeautyB

Step2:

- Significant level alpha = 0.05 Step3:

The TTEST Procedure

Variable: weightLost

Diet	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2		27	3.0259	2.5234	0.4856	-2.1000	7.9000
3		27	5.1481	2.3956	0.4610	0.5000	9.2000
Diff (1-2)	Pooled		-2.1222	2.4603	0.6696		
Diff (1-2)	Satterthwaite		-2.1222		0.6696		

Diet	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2		3.0259	2.0277	4.0241	2.5234	1.9872	3.4581
3		5.1481	4.2005	6.0958	2.3956	1.8865	3.2830
Diff (1-2)	Pooled	-2.1222	-3.4659	-0.7786	2.4603	2.0651	3.0441
Diff (1-2)	Satterthwaite	-2.1222	-3.4660	-0.7785			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	52	-3.17	0.0026
Satterthwaite	Unequal	51.86	-3.17	0.0026

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	26	26	1.11	0.7930

```
proc ttest data=newdatset1 h0=0 sides=2 alpha=0.005;
      class Diet;       var weightLost;          where Diet
in (2,3); run;
```

- As the dataset size is more than 30 observations, we are using ttest for the hypothesis.

Step4:

- From the results above, if we check the equality of variances, the p-value is 0.7930 which is greater than the alpha value 0.005.

Step5:

- So, we conclude that even though both the groups have equal variances, we consider the Satterthwaite method, where the t value is -3.17 in the confidence interval of (-4.0856, 0.1591) and the p-value is 0.0026 which is less than 0.05.
- So, we reject the null Hypothesis, and the results show that there is sufficient evidence that the mean score of BeautyB weight loss is different frm that of Beauty A, which in turn is greater than that of BeautyA.
- If we check the mean of the two groups BeautyA (3.0259) is less than BeautyB (5.1481) giving supportive evidence for the statement above.

7) Similarly, is there difference in weight loss between the three types of diet (control group, Beauty A and Beauty B)? Conduct a hypothesis test to answer this question.

Ans.

Step1: Hypothesis

- H0: uControl = uBeautyA = uBeautyB

  vs

- H1: Atleast one group has difference in weight loss from given diet groups.

Step2:

- Significant level alpha = 0.05 Step3:

The SAS System

The GLM Procedure

Dependent Variable: weightLost

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	71.0936895	35.5468447	6.20	0.0032
Error	75	430.1792593	5.7357235		
Corrected Total	77	501.2729487			

ANOVA table proc glm data=newdatset1; class Diet; model weightLost=Diet; run; - I'm going to use ANOVA procedure for this hypothesis. So, we build an ANOVA table.

Step4:

- From the results above, we see that the F-statistic is 6.20 and p-value is 0.0032 Step5:
- Since the p-value is less than the alpha value 0.05 (0.0032 < 0.05), we reject our null hypothesis and conclude that one of the group results is different from other diet groups.
  - To identify which pairs are different we perform pairwise comparision using Tukey's method when alpha = 0.05.

The SAS System

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

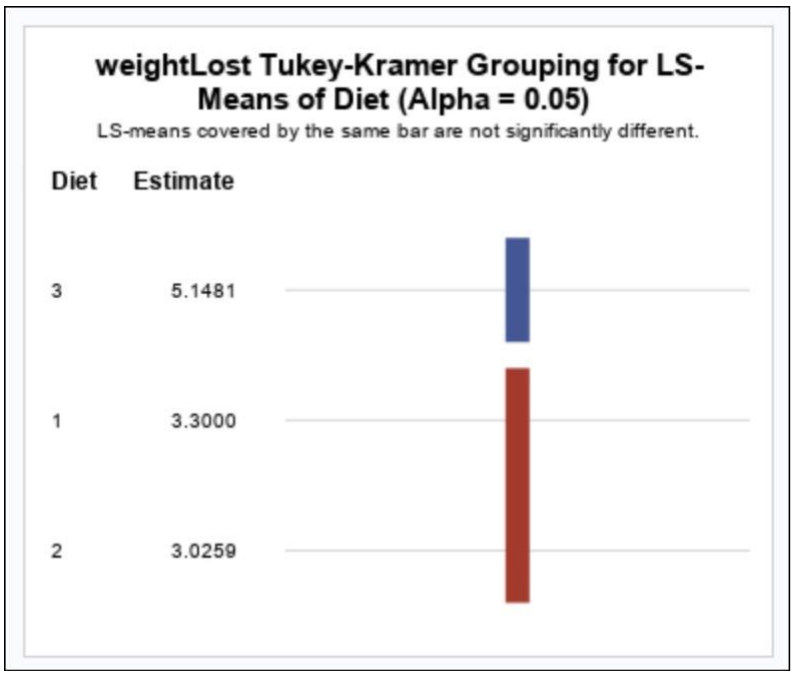
Diet	weightLost LSMEAN	LSMEAN Number
1	3.30000000	1
2	3.02592593	2
3	5.14814815	3

Least Squares Means for effect Diet
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: weightLost

i/j	1	2	3
1		0.9125	0.0201
2	0.9125		0.0048
3	0.0201	0.0048	

- Comparision between pairs:

1. From the above screenshot, when we compare the p-values between the control group (1)and BeautyA (2), the p-value is 0.9125 which is greater than the alpha value 0.05. From which we can conclude that there is no difference of weight loss between Control group and BeautyA.
2. When we compare the p-values between groups BeautyA (2) and BeautyB (3), the pvalue is 0.0048 which is less than the alpha value 0.05. From which we can conclude that there is a significant difference in weight loss between BeautyA and BeautyB.
3. When we compare the p-values between groups Control group (1) and BeautyB (3), the p-value os 0.0201 which is less than the alpha value 0.05. From which we can conclude that there is significant difference of weight loss between BeautyB and Control group.




```
proc glm data=newdatset1;
class Diet; model
weightLost=Diet;
    lsmeans Diet/ adjust=tukey;
run;
```

4. The above graph also gives sufficient evidence for the above-mentioned conclusions about the different groups.

8) To further assess the amount of weight lost, we would also like to examine the differences across the diets after adjusting for gender. Is there a difference between the diets? Conduct a hypothesis test to answer this question.

Ans.

As we would like to find if there is a difference between the diet groups after adjudting the gender, we want to check if there is an interaction that exists or not.

The GLM Procedure

Dependent Variable: weightLost

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	96.3699194	19.2739839	3.43	0.0078
Error	72	404.9030293	5.6236532		
Corrected Total	77	501.2729487			

R-Square	Coeff Var	Root MSE	weightLost Mean
0.192250	61.67759	2.371424	3.844872

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Diet	2	71.09368946	35.54684473	6.32	0.0030
gender	1	0.13552633	0.13552633	0.02	0.8771
Diet*gender	2	25.14070363	12.57035182	2.24	0.1143

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Diet	2	62.24180512	31.12090256	5.53	0.0058
gender	1	0.06660309	0.06660309	0.01	0.9136
Diet*gender	2	25.14070363	12.57035182	2.24	0.1143

proc glm data=newdatset1; class Diet gender;
model weightLost=Diet gender Diet*gender; run;

From the above results, we see that p-value is 0.1143 which is greater than alpha value 0.05, which states that interaction is insignificant.

- We use Two-way ANOVA without interaction

Step1: Hypothises

- H0: uControl = uBeautyA = uBeautyB when gender is adjusted vs
- H1: Atleast one group has difference in weight loss from given diet groups and when gender is adjusted.

Step2:

- Significant level alpha = 0.05 Step3:
- As we have confirmed that the interaction is insignificant, we use two-way ANOVA without interaction. Step4:

The SAS System

The GLM Procedure

Dependent Variable: weightLost

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	71.2292158	23.7430719	4.09	0.0097
Error	74	430.0437329	5.8114018		
Corrected Total	77	501.2729487			

R-Square	Coeff Var	Root MSE	weightLost Mean
0.142097	62.69871	2.410685	3.844872

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Diet	2	71.09368946	35.54684473	6.12	0.0035
gender	1	0.13552633	0.13552633	0.02	0.8790

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Diet	2	71.00703119	35.50351559	6.11	0.0035
gender	1	0.13552633	0.13552633	0.02	0.8790

```
proc glm data=newdatset1; class Diet gender; model weightLost=Diet gender;
    lsmeans Diet gender/adjust=tukey; run;
```

- The F-statistic is 6.12 and p-value is 0.0035 which is less than the alpha value 0.05.

Step5:

- Since the p-value is less than the alpha value, we reject our null hypothesis and conclude that atleast one group has difference in the weight loss when compared with othe diet groups. - To identify the pairs which are different, we perform pairwise comparision and adjusted using Tukey's adjustment at alpha = 0.05.

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

Diet	weightLost LSMEAN	LSMEAN Number
1	3.29300610	1
2	3.02437172	2
3	5.14348555	3

Least Squares Means for effect Diet
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: weightLost

i/j	1	2	3
1		0.9170	0.0210
2	0.9170		0.0052
3	0.0210	0.0052	

- Comparision between pairs:

1. From the above screenshot, when we compare the p-values between the control group (1) and BeautyA (2), the p-value is 0.9170 which is greater than the alpha value 0.05. From which we can conclude that there is no difference in weight loss between Control group and BeautyA when the gender is adjusted.
2. When we compare the p-values between groups BeautyA (2) and BeautyB (3), the p-value is 0.0052 which is less than the alpha value 0.05. From this, we can conclude that there is a significant difference in weight loss between BeautyA and BeautyB when the gender is adjusted.
3. When we compare the p-values between groups Control group (1) and BeautyB (3), the pvalue os 0.0210 which is less than the alpha value 0.05. From this, we can conclude that there is significant difference of weight loss between BeautyB and Control group when the gender is adjusted.