One_sample_t-test

```
* one sample t-test ;
proc ttest data=heart sides=2 alpha=0.05 h0=64.8 ;
var height;
run;
```

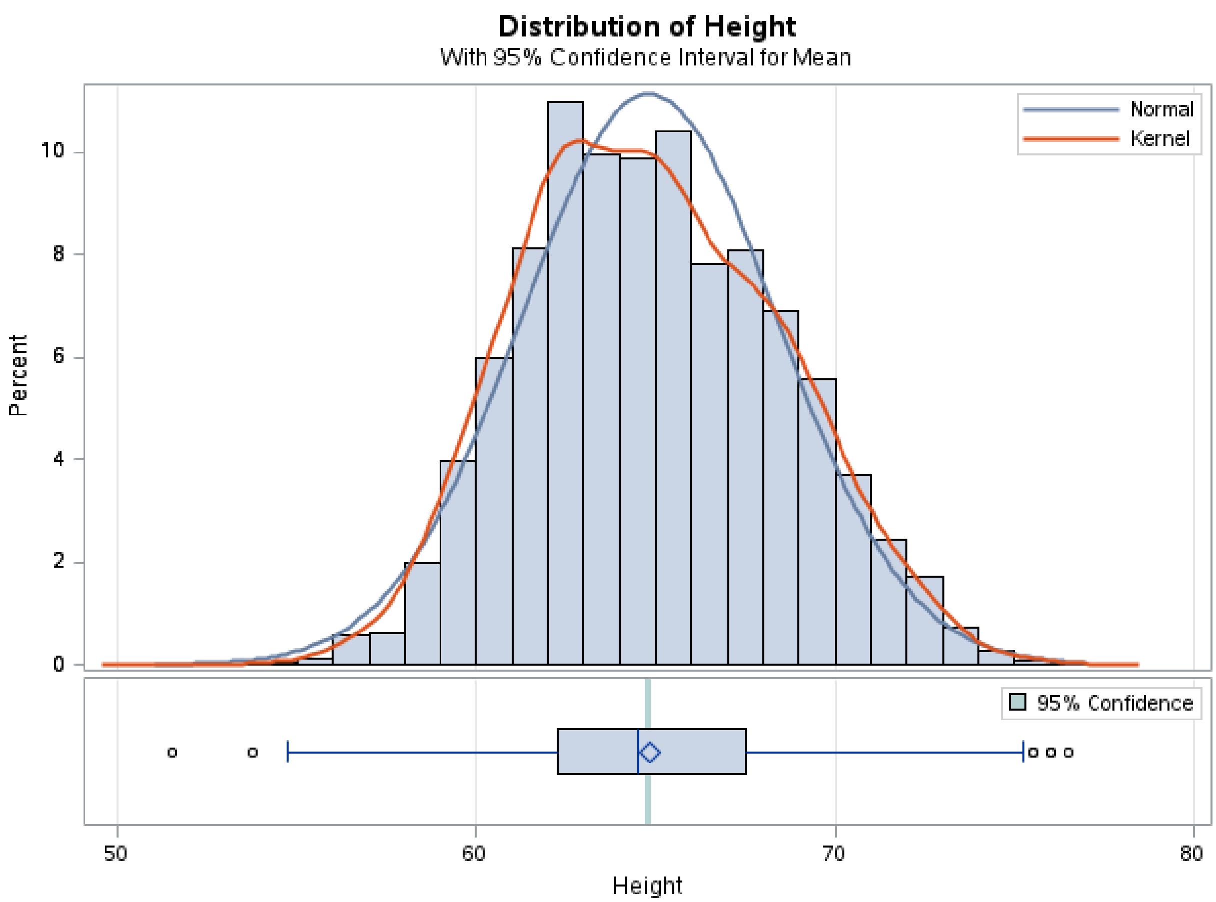
Variable: Height

N	Mean	Std Dev	Std Err	Minimum	Maximum
5203	64.8132	3.5827	0.0497	51.5000	76.5000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
64.8132	64.7158	64.9106	3.5827	3.5152	3.6529

DF	t Value	Pr > |t|
5202	0.27	0.7907

Here, h0 is used to define the population mean which is 64.8 for this data
In second table as we see Confidence Interval for mean which is(64.7,64.91)
P_value 0.7907 > 0.05 we can say that
Mean weight of height from the data and the given population height is equal.



Two-sample t-test

```
data heart ;
set sashelp.heart;
run;
proc print data=heart ;
run;
*  compare weight of the patient according to there Sex ;
proc ttest data=heArt SIDES=2 alpha=0.5 plots=summary ;
class Sex ;
var weight;
run;
/* sides is used for defining the tail
alpha is significance level,
class is categorical variable that divide weight into two independent group;
var is dependent continuouis variable */;
```

Output:

Hypothesis: -

Var equal variance: H_01 : there is no significant difference among the variation of weight according to sex

The TTEST Procedure

Variable: Weight

Sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		2869	141.4	26.2880	0.4908	67.0000	300.0
Male		2334	167.5	25.2907	0.5235	99.0000	276.0
Diff (1-2)	Pooled		-26.0775	25.8454	0.7204		
Diff (1-2)	Satterthwaite		-26.0775		0.7176		

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5201	-36.20	<.0001
Satterthwaite	Unequal	5057.9	-36.34	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2868	2333	1.08	0.0503

For equal variance we check the p_value for last one i.e. Equality of variances;

0.0503 > 0.05 ; we accept the H_01 for this

Since, variance for the Sex is equal then we Pooled the variance for the both groups

For this check the p_value at the front of Pooled

Which is 0.0001 < 0.05 ;

We reject the H_02 ;

And conclude that there is significant difference in both the groups.

ONE_WAY_ANOVA

```
* One_Way_Anova;
proc glm data=heart ;
class Weight_Status      ;
model Cholesterol = Weight_Status    ;
means Weight_Status / tukey ;
run;
```

Cholesterol is dependent continuous variable
Weight_Status is an independent categorical grouping variable containing more than two groups.
Tukey is used to perform paired-wise comparison.

The GLM Procedure

Dependent Variable: Cholesterol

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	259397.60	129698.80	65.92	<.0001
Error	5048	9932551.25	1967.62		
Corrected Total	5050	10191948.85			

R-Square	Coeff Var	Root MSE	Cholesterol Mean
0.025451	19.50614	44.35788	227.4047

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Weight_Status	2	259397.6031	129698.8016	65.92	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Weight_Status	2	259397.6031	129698.8016	65.92	<.0001

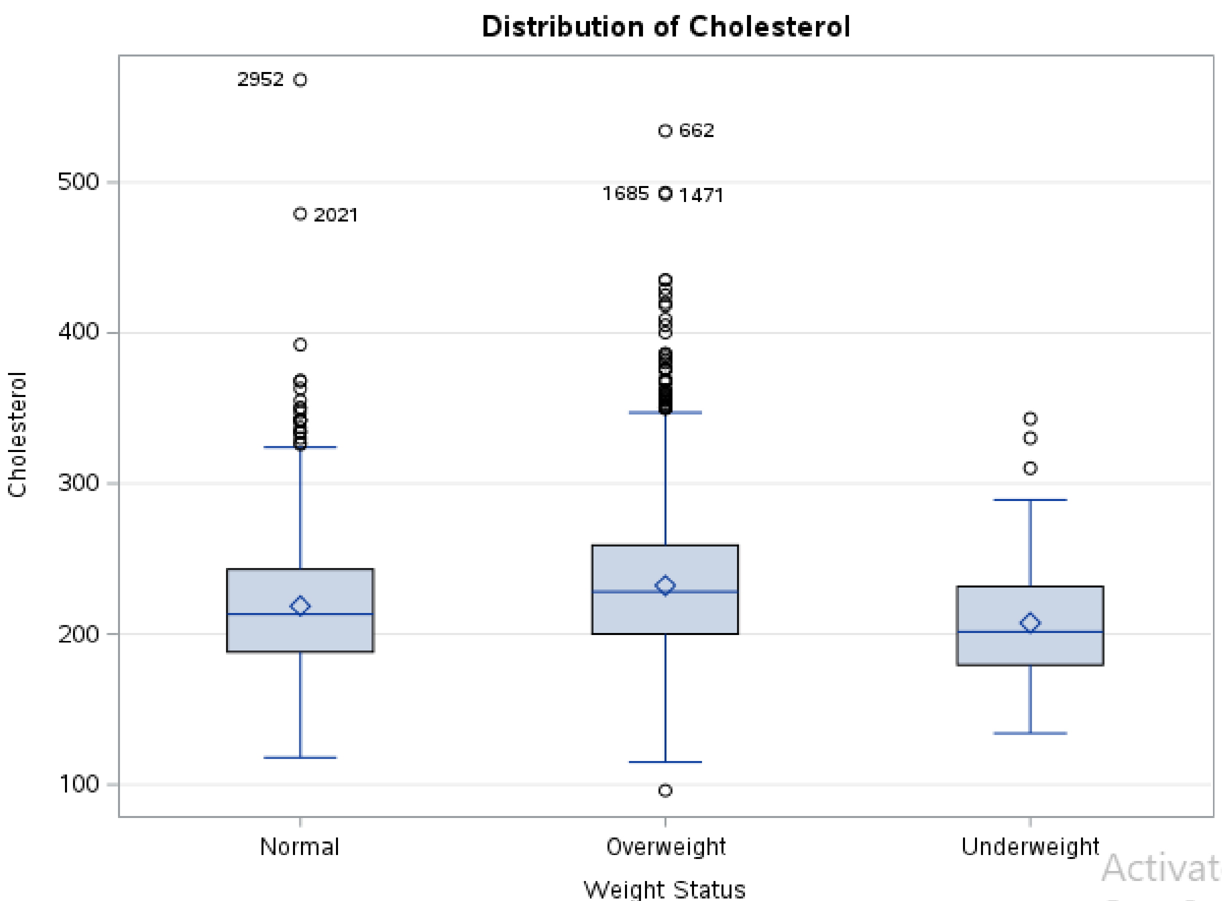
Here R^2 is 0.025 which variation explained by the overall model is 2.5% which is closer to zero,
And error explained by the model is 43.35 which is overall 50%
So we can say that the model is Bad fit.

From the below box plot we can see Normal and Overweight Having more Outliers

Pairwise_comparison: -

Comparisons significant at the 0.05 level are indicated by ***.				
Weight_Status Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Overweight - Normal	13.548	10.277	16.820	***
Overweight - Underweight	24.833	16.796	32.869	***
Normal - Overweight	-13.548	-16.820	-10.277	***
Normal - Underweight	11.285	2.978	19.592	***
Underweight - Overweight	-24.833	-32.869	-16.796	***
Underweight - Normal	-11.285	-19.592	-2.978	***

(***) tells us that the respective group means are different



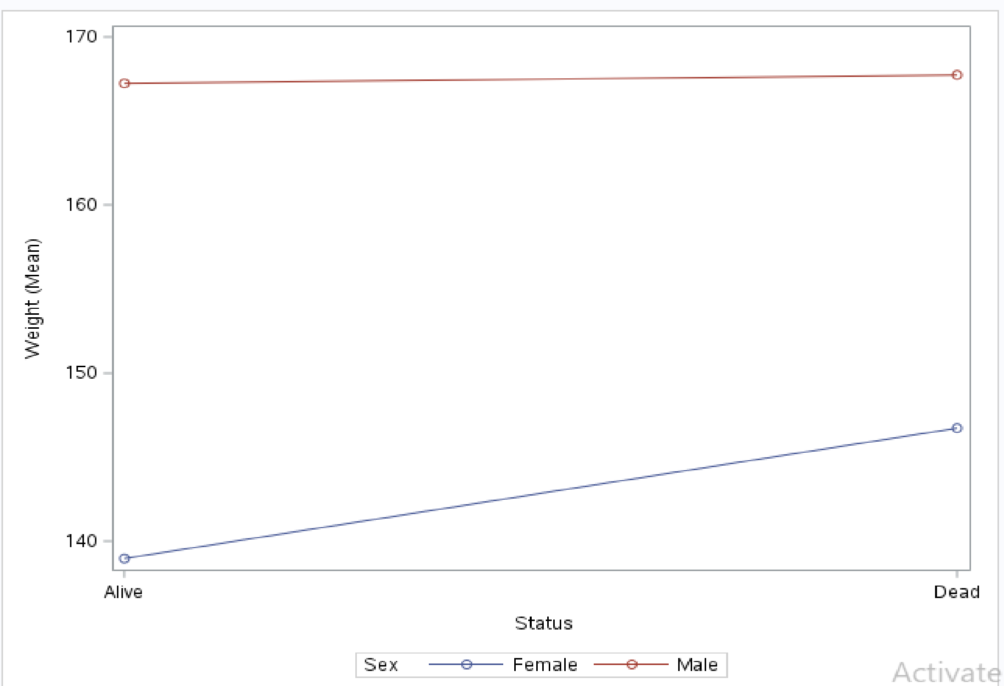
preliminary view of data:

```
* preliminary view of data;
proc means data=heart mean var std nway;
class Status Sex;
var weight;
run;
```

The MEANS Procedure

Analysis Variable : Weight					
Status	Sex	N Obs	Mean	Variance	Std Dev
Alive	Female	1977	138.9701114	593.2109663	24.3559226
	Male	1241	167.2312651	593.5166346	24.3621968
Dead	Female	896	146.7229050	866.3817455	29.4343633
	Male	1095	167.7328454	692.4230685	26.3139330

```
* plot ;
proc sgplot data=heart;
vline Status/ group=Sex stat=mean response=Weight markers;
run;
```



Two_Way_Anova:

```
proc glm data=Heart;
class Status Sex;
model Weight= Status Sex Status*Sex;
LSMEANS Status*Sex/diff;
run;
```

LSMEANS (least square means) * for interaction

The GLM Procedure

Dependent Variable: Weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	912364.772	304121.591	460.03	<.0001
Error	5199	3437037.135	661.096		
Corrected Total	5202	4349401.907			

R-Square	Coeff Var	Root MSE	Weight Mean
0.209768	16.79557	25.71178	153.0867

Here, R^2 = 0.209 i.e. total variation explained by the model is 20.9%
P_vlaue for the overall model is 0.0001 < 0.005 ,
We can say that there is a significant difference between the Wight in Different types of Sex and Status.
Since, we are working on Two-Anova so check for interaction for that we check the table of type3 ss

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Status	1	20371.4353	20371.4353	30.81	<.0001
Sex	1	725834.7735	725834.7735	1097.93	<.0001
Status*Sex	1	15720.8029	15720.8029	23.78	<.0001

For this all the p_value is less than 0.05 so say that
No interaction is present in the data
Status having mean weight is also significantly different
Also Sex having mean weight is also significantly different.

The GLM Procedure
Least Squares Means

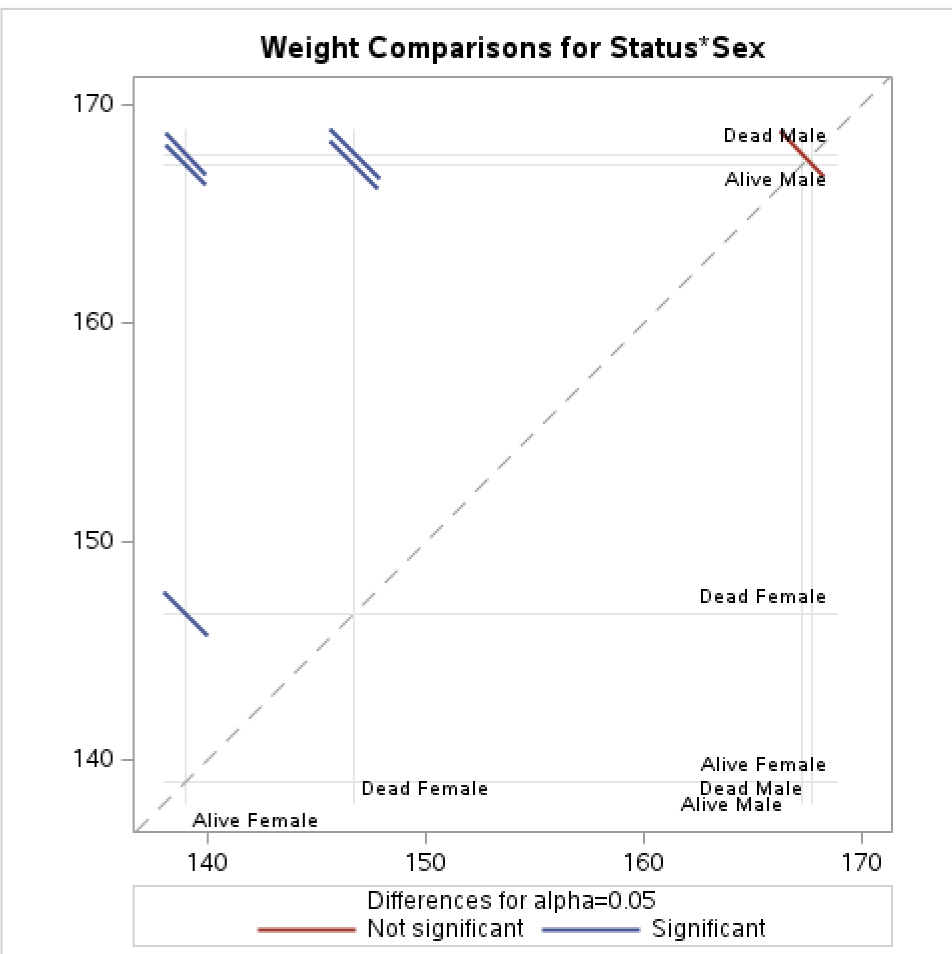
Status	Sex	Weight LSMEAN	LSMEAN Number
Alive	Female	138.970111	1
Alive	Male	167.231265	2
Dead	Female	146.722905	3
Dead	Male	167.732845	4

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

Dependent Variable: Weight

i/j	1	2	3	4
1		<.0001	<.0001	<.0001
2	<.0001		<.0001	0.6382
3	<.0001	<.0001		<.0001
4	<.0001	0.6382	<.0001	

Only the pair (4,2) are significant 0.6382 > 0.05



For equal variance:

```
* test for equal variance Homogenity;
proc glm data=heart plots=diffplot;
    class Weight_Status;
    model Cholesterol=Weight_Status;
    means Weight_Status / tukey hovtest=levene;
    run;
```

The GLM Procedure

Levene's Test for Homogeneity of Cholesterol Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Weight_Status	2	1.0832E8	54157538	3.31	0.0367
Error	5048	8.268E10	16379706		

Here the 0.0367 < 0.05 homogeneity is no present in

Normality:

```
* normality ;
proc UNIVARIATE data=heart normal ;
class Status Sex;
var Weight;
run;
```

Sex>Female>Status>Alive

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.942948	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.080058	Pr > D	<0.0100
Cramer-von Mises	W-Sq	3.284532	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	19.90022	Pr > A-Sq	<0.0050

Sex>Male>Status>Alive

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.986975	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.040345	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.464748	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3.274516	Pr > A-Sq	<0.0050

Sex>Female>Status>Dead

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.962263	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.067472	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.070914	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	6.442368	Pr > A-Sq	<0.0050

Sex>Male>Status>Dead

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.991009	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.034347	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.16309	Pr > W-Sq	0.0174
Anderson-Darling	A-Sq	1.260595	Pr > A-Sq	<0.0050