STATISTICAL PROGRAMMING FOR BUSINESS ANALYTICS

ASSIGNMENT NO.6

TANAY BHALERAO

U47707491

MARCH 27, 2015

UNIVERSITY OF SOUTH FLORIDA
MANAGEMENT INFORMATION SYSTEMS

<u>**Homework for Chapter 7 and 8**</u>

1. Refer to USEDCARS data. Suppose that you want to see if there are significant differences among the prices charged by the dealers. However, you feel that it is necessary to compensate for the annual depreciation of the cars. Some dealers may primarily stock newer cars, and, naturally, their prices will be higher. Perform an analysis of covariance by considering the price of the car to be the response, predicted by both the categorical variable for the used car dealer and a continuous variable representing the year in which the car was manufactured. Write down another variable, which may not necessarily appear in the data, that you would expect to be another significant predictor of the cost of a used car.

```
libname tan "\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\";
DATA tan.USEDCARS;
INFILE
'\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\usedcars.txt'
firstobs=2 obs=51;
INPUT YEAR 1-2 MANUFACTURER $ 9-23 MODEL $ 24-37 MILES $ 38-48 PRICE $ 49-60
DEALER $ 61-86;
PRICE_D = INPUT(PRICE,comma9.);


PROC GLM DATA=tan.USEDCARS;
CLASS DEALER;
MODEL PRICE_D = DEALER YEAR DEALER*YEAR;
TITLE "Analysis of Co-Variance";
      RUN;
```

Analysis of Co-Variance

The GLM Procedure

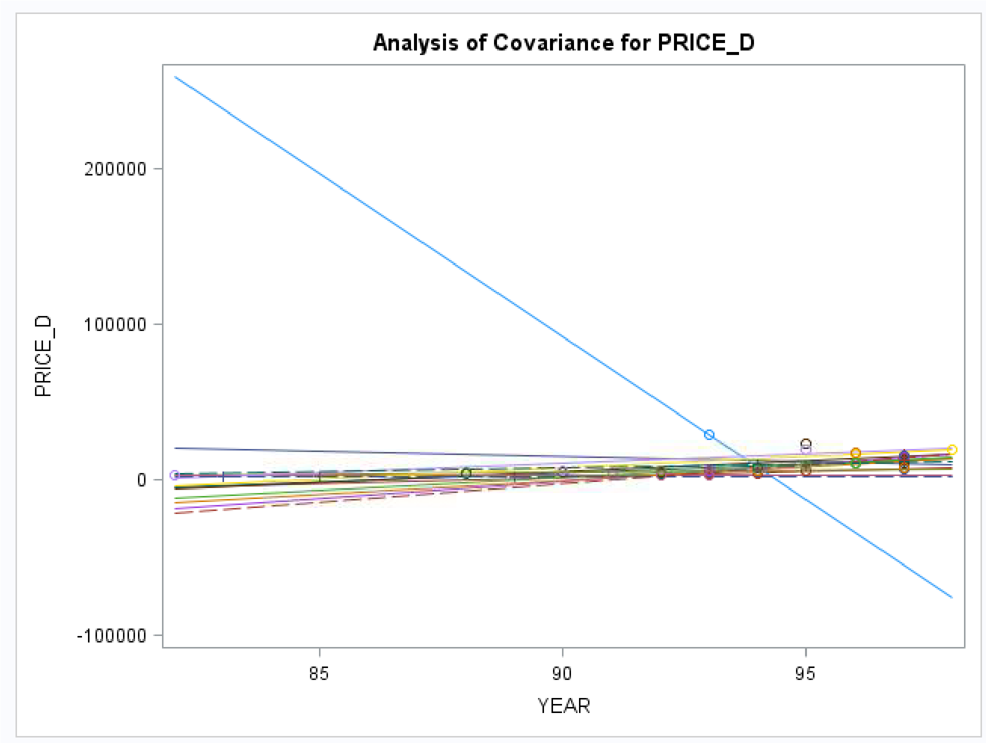
Dependent Variable: PRICE_D

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	1398960296	48240010	2.87	0.0086
Error	20	336615042	16830752		
Corrected Total	49	1735575338			

R-Square	Coeff Var	Root MSE	PRICE_D Mean
0.806050	44.67751	4102.530	9182.540

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DEALER	14	774446352.0	55317596.6	3.29	0.0077
YEAR	1	298230159.9	298230159.9	17.72	0.0004
YEAR*DEALER	14	326283784.2	23305984.6	1.38	0.2467

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DEALER	14	327337051.8	23381218.0	1.39	0.2447
YEAR	1	7789883.6	7789883.6	0.46	0.5041
YEAR*DEALER	14	326283784.2	23305984.6	1.38	0.2467



Analysis of Co-Variance(ANCOVA) is used when we want to compare two or more regression lines to each other. The regression lines can be different from each other in either slope or intercept. The class statement has the nominal variable: **Dealer** and the Model statement has the Y variable: **PRICE_D.** Now if we see the TYPE III sum of squares- the p values of the slope are insignificant(>0.05) for dealer, price_d as well as the interaction term. Hence there is significant difference in the prices charged by the dealer. Another factor which can be added as a significant predictor of the cost of the used car can be the alignment of the chasis with the engine can change significantly over a period of time which can be an added attribute to calculate the predictor of the cost. The value of the alignment can be measured from the center of gravity of the vehicle.

2. Refer to the SOCCER data. Create a new dataset which defines each player's primary position as the one listed in the data if only one is given or as the one listed first if two positions are given. (For example Kerry Doran's position is listed as MF/D, so her primary position would be MF.) Perform a one-way analysis of variance to see if the heights of the players (in inches) can be predicted by the categorical primary position variable.

   Strictly speaking, the soccer players are not a random sample from some well-defined population, so the usual assumptions about the F test are not valid. However, we can use the statistics in the ANOVA table as descriptive measures of variation between and within primary positions.

```
DATA tan.soccer;
INFILE "\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\soccer.txt";
```

```
INPUT VALUE $ /
FNAME $ 1-11 LNAME $ 12-22 /
POSN $ /
HEIGHT $ 1-22 /
CLASS $ /
EXP $
;
INPUT VALUE $ /
FNAME $ 1-11 LNAME $ 12-22 /
POSN $ /
HT_FEET 1 HT_INCH 3-4 /
CLASS $ /
EXP $
;
FINAL_POS = SCAN(POSN, 1, '/');
TOTAL_HT= 12*HT_FEET+HT_INCH;
RUN;
PROC PRINT DATA=tan.soccer;
RUN;
PROC ANOVA DATA = tan.soccer;
TITLE "Analysis of Variation";
CLASS FINAL_POS;
MODEL TOTAL_HT = FINAL_POS;
RUN;
```

Analysis of Variation

Obs	VALUE	FNAME	LNAME	POSN	HEIGHT	CLASS	EXP	HT_FEET	HT_INCH	FINAL_POS	TOTAL_HT
1	23	Erin	Baxter	MF	5-8	SR	3L	5	8	MF	68
2	8	Christie	Brady	MF	5-9	SO	1L	5	4	MF	64
3	19	Lia	Cummins	MF	5-3	FR	HS	5	6	MF	66
4	7	Kerri	Doran	MF/D	5-9	SR	2L	5	5	MF	65
5	6	Danielle	Fotopoulos	F	5-6	SR	1L	5	11	F	71
6	3	Karyn	Hall	MF	5-6	FR	HS	5	6	MF	66
7	24	Jordan	Kellgren	GK	5-8	FR	HS	5	6	GK	66
8	25	Alexis	MacKenzie	MF/F	5-6	FR	HS	5	4	MF	64
9	16	Heather	Mitts	MF/D	5-9	JR	2L	5	5	MF	65
10	2	Lisa	Olinyk	MF	5-3	SR	2L	5	7	MF	67
11	00	Lynn	Pattishall	GK	5-6	SR	3L	5	8	GK	68
12	9	Renee	Reynolds	D	5-6	SO	1L	5	5	D	65
13	18	Whitney	Singer	MF	5-7	FR	HS	6	1	MF	73
14	30	Jamie	Theil	F/D	5-6	FR	HS	5	8	F	68
15	28	Abby	Wambach	F	5-4	FR	HS	5	10	F	70
16	20	Sarah	Yohe	F	5-4	JR	2L	5	6	F	66

Analysis of Variation

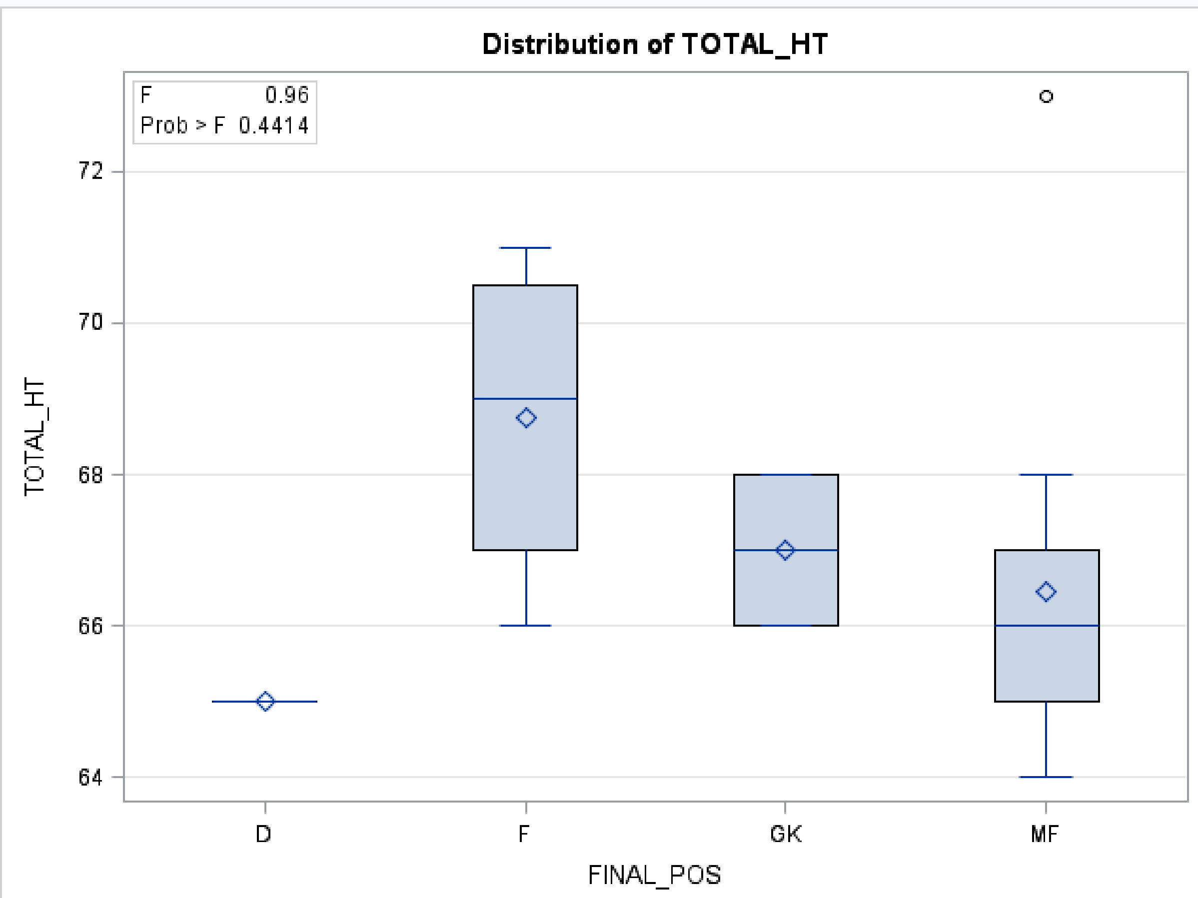
The ANOVA Procedure

Dependent Variable: TOTAL_HT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	19.02777778	6.34259259	0.96	0.4414
Error	12	78.97222222	6.58101852		
Corrected Total	15	98.00000000			

R-Square	Coeff Var	Root MSE	TOTAL_HT Mean
0.194161	3.828880	2.565350	67.00000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
FINAL_POS	3	19.02777778	6.34259259	0.96	0.4414



```
Here Prob>F 0.4414. Hence the relation between final position and height in
inches is insignificant
```

3. Refer to the LIMES data. Suppose that a grower wants to see if the juice volumes of the limes depend on the times that they are harvested. Of course, juice volumes could also vary according to size. Create a new categorical variable to denote time as “early”, “middle”, or “late” season, depending on whether the limes were picked in February, March, or April, respectively. Fit an *analysis of covariance* model to predict juice volumes from the categorical time variable and fruit diameter, regarded as a continuous variable. (Hint: If X is a SAS date, the SAS function MONTH(X) returns the month of the year, expressed as an integer from 1 to 12, for that date.)

JUICE VOL. CO-VARIANCE

The GLM Procedure

Class Level Information		
Class	Levels	Values
TIME	3	early late middle

Number of Observations Read	100
Number of Observations Used	96

The GLM Procedure

Dependent Variable: juice_vol

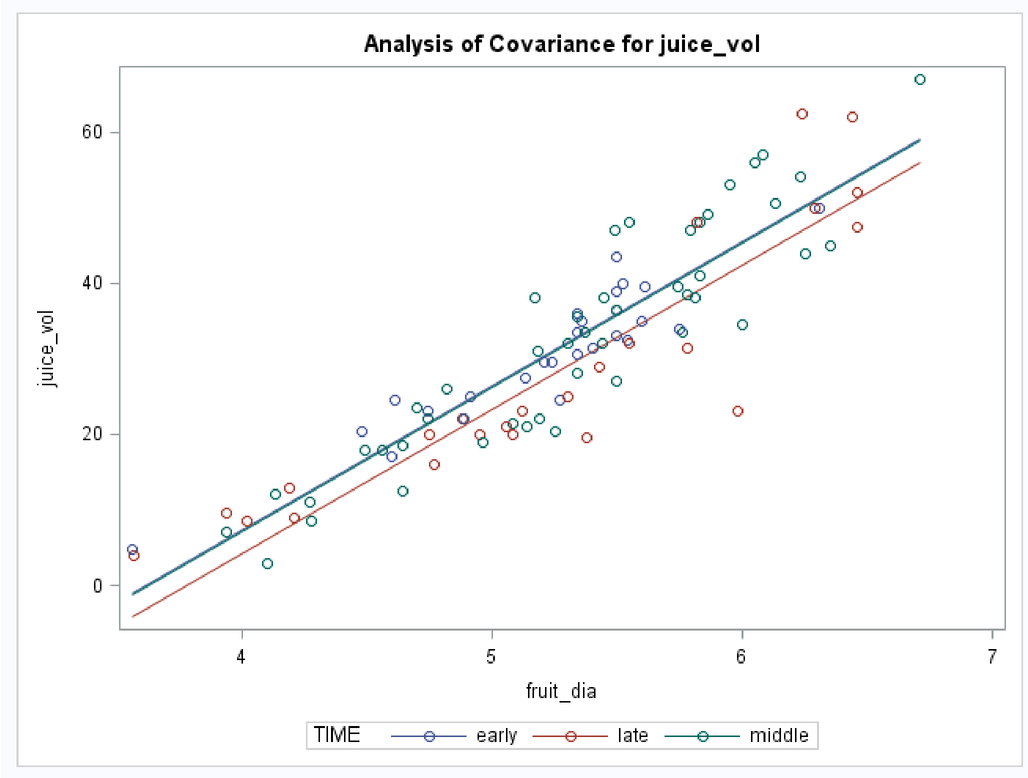
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	16199.16253	5399.72084	154.49	<.0001
Error	92	3215.53747	34.95149		
Corrected Total	95	19414.70000			

R-Square	Coeff Var	Root MSE	juice_vol Mean
0.834376	19.10171	5.911979	30.95000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TIME	2	382.53556	191.26778	5.47	0.0057
fruit_dia	1	15816.62696	15816.62696	452.53	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TIME	2	159.95756	79.97878	2.29	0.1072
fruit_dia	1	15816.62696	15816.62696	452.53	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-69.27213804	B	4.87397133	-14.21	<.0001
TIME early	0.22116895	B	1.45457229	0.15	0.8795
TIME late	-2.89533938	B	1.49166324	-1.94	0.0553
TIME middle	0.00000000	B	.	.	.
fruit_dia	19.09779150		0.89775759	21.27	<.0001



4. Chapter 8: 8.2, 8.4, 8.6, 8.8, 8.10

8.2

```
DATA tan.STATIN;
 DO SUBJ = 1 TO 20;
      IF RANUNI(1557) LT .5 THEN GENDER = 'FEMALE';
      ELSE GENDER = 'MALE';
      IF RANUNI(0) LT .3 THEN DIET = 'HIGH FAT';
      ELSE DIET = 'LOW FAT';
      DO DRUG = 'A' , 'B' , 'C';
            LDL = ROUND(RANNOR(1557)*20 + 110 + 5*(DRUG EQ 'A') - 10*(DRUG EQ 
'B') - 5*(GENDER EQ 'FEMALE') +10*(DIET EQ 'HIGH FAT'));
            HDL = ROUND(RANNOR(1557)*10 + 20 + 0.2*LDL + 12*(DRUG EQ 'B'));
            TOTAL = ROUND(RANNOR(1557)*20 + LDL + HDL + 50 - 10*(GENDER EQ 
'FEMALE') + 10*(DIET EQ 'HIGH FAT'));
            OUTPUT;
      END;
 END;
RUN;
PROC ANOVA DATA = tan.STATIN;
CLASS SUBJ DRUG;
MODEL LDL HDL TOTAL = SUBJ DRUG /NOUNI;
TITLE "Patient Cholestrol-LDL-HDL for the drugs";
MEANS DRUG /SNK;
RUN;
```

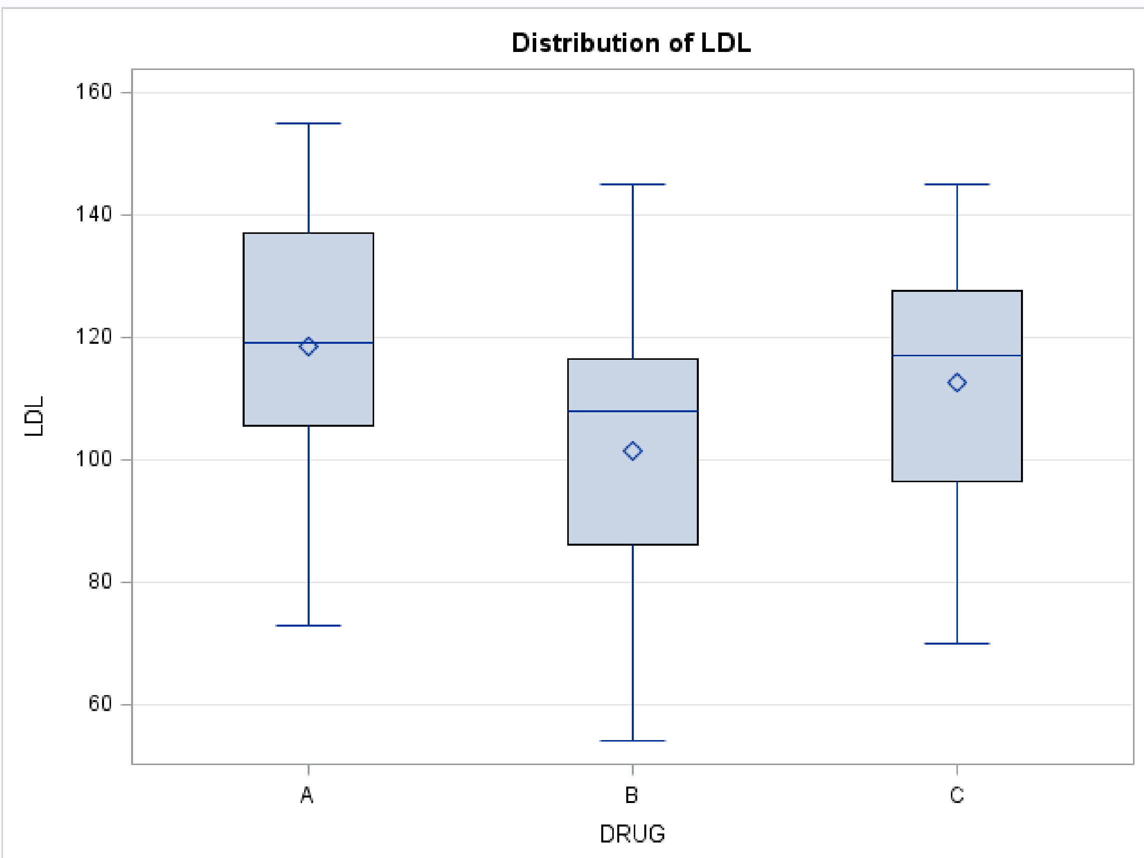
Patient Cholestrol-LDL-HDL for the drugs

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
SUBJ	20	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
DRUG	3	A B C

Number of Observations Read	60
Number of Observations Used	60

The ANOVA Procedure



Student-Newman-Keuls Test for LDL

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

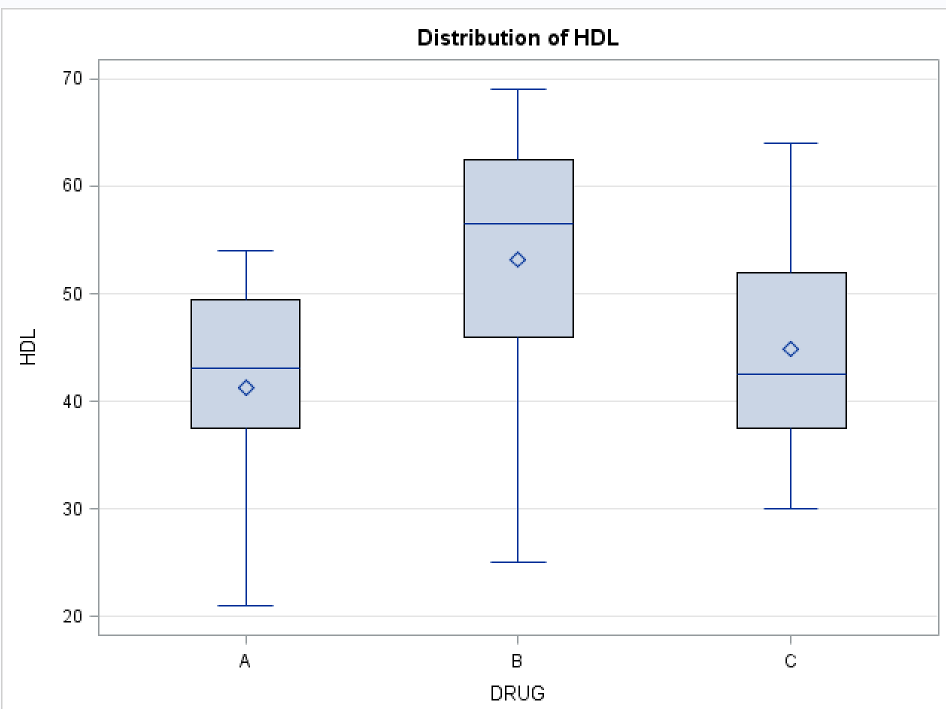
Alpha	0.05
Error Degrees of Freedom	38
Error Mean Square	341.8781

Number of Means	2	3
Critical Range	11.836443	14.259841

Means with the same letter are not significantly different.				
SNK Grouping		Mean	N	DRUG
	A	118.600	20	A
	A			
B	A	112.700	20	C
B				
B		101.350	20	B

Patient Cholestrol-LDL-HDL for the drugs

The ANOVA Procedure



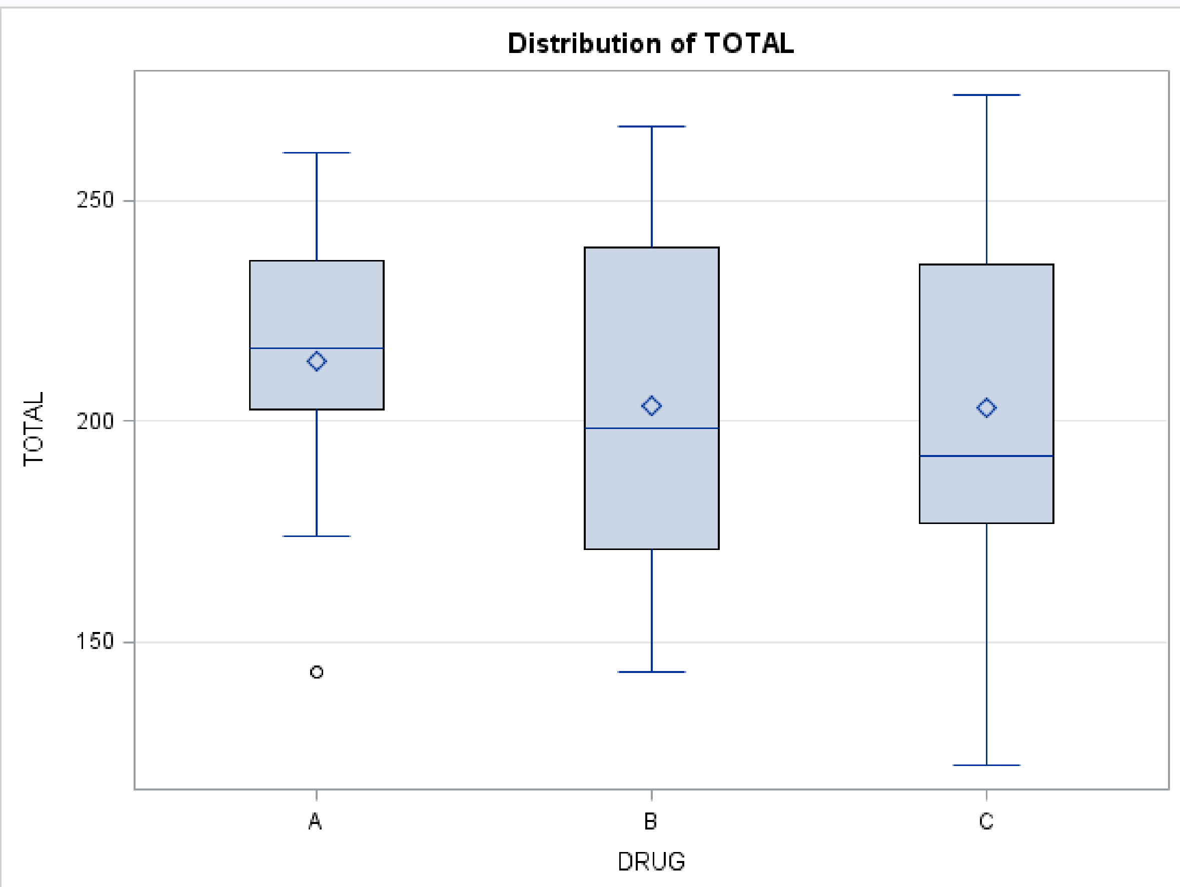
Student-Newman-Keuls Test for HDL

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.05
Error Degrees of Freedom	38
Error Mean Square	110.3535

Number of Means	2	3
Critical Range	6.7247888	8.1016249

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	DRUG
A	53.150	20	B
B	44.800	20	C
B			
B	41.200	20	A



Student-Newman-Keuls Test for TOTAL

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

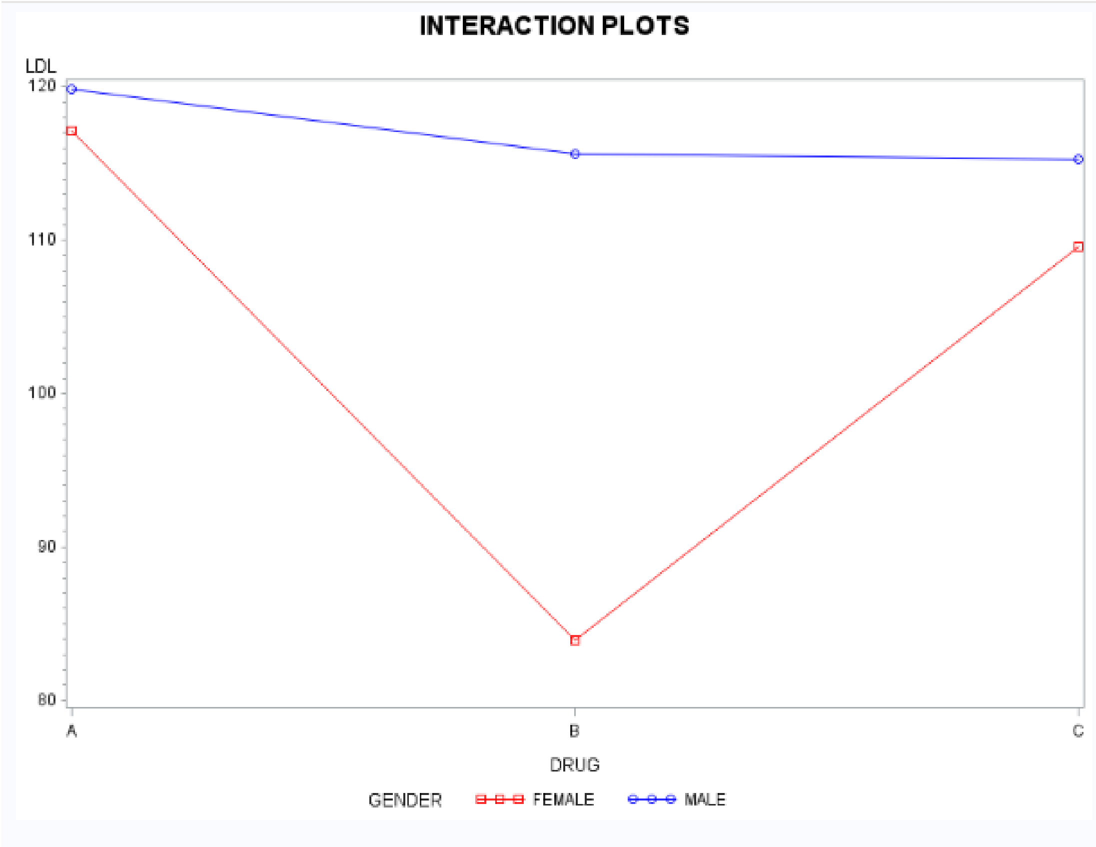
Alpha	0.05
Error Degrees of Freedom	38
Error Mean Square	1069.716

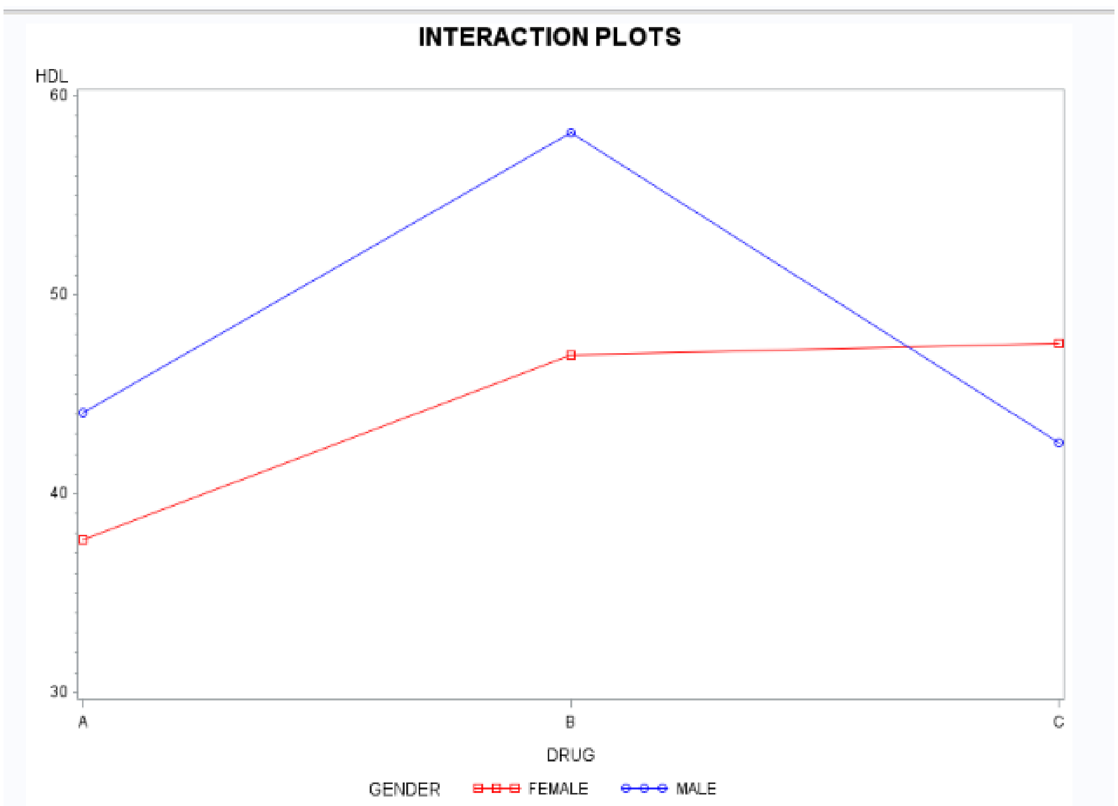
Number of Means	2	3
Critical Range	20.937257	25.22396

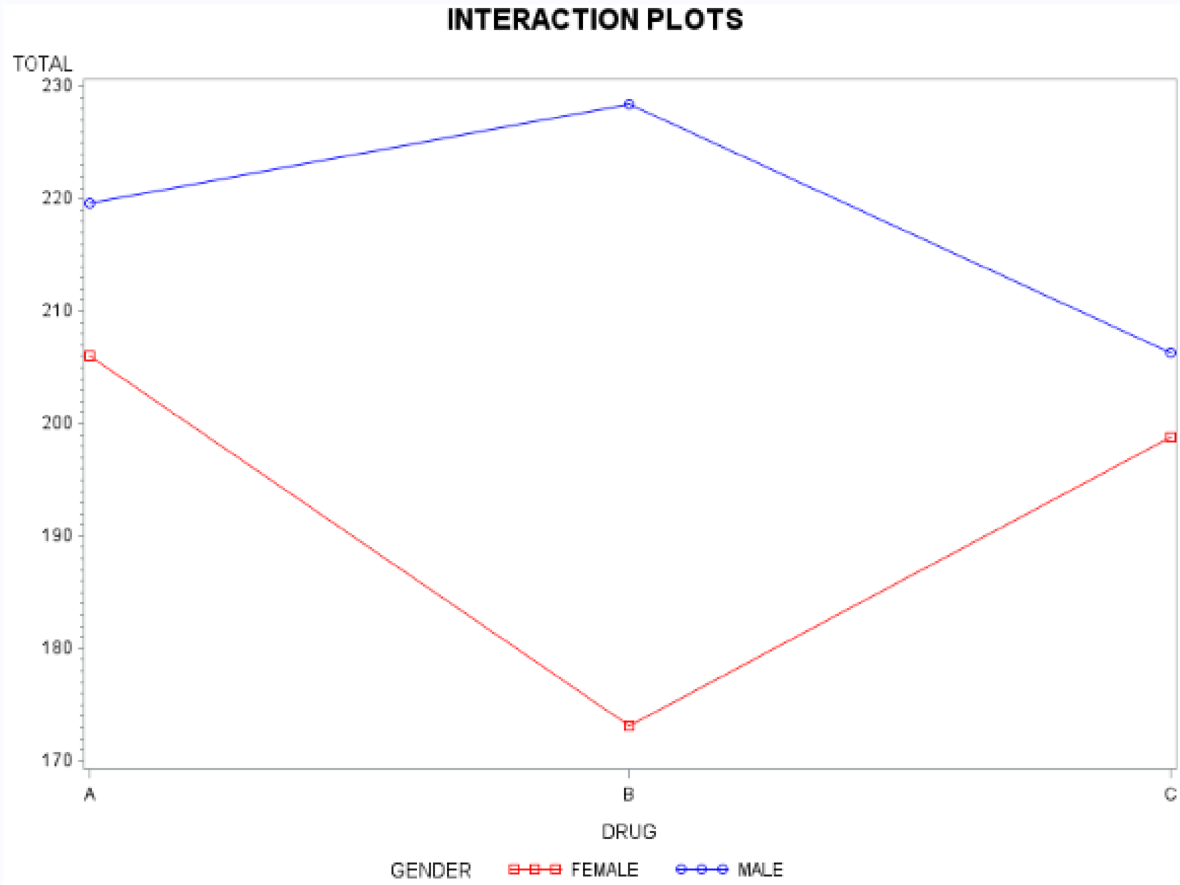
Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	DRUG
A	213.55	20	A
A			
A	203.55	20	B
A			
A	202.95	20	C

8.4

```
PROC ANOVA DATA = tan.STATIN;
CLASS SUBJ GENDER DRUG;
MODEL TOTAL HDL LDL  = GENDER SUBJ(GENDER) DRUG GENDER*DRUG
DRUG*SUBJ(GENDER);
MEANS GENDER|DRUG;
TITLE "Gender Add"
TEST H = GENDER E = SUBJ(GENDER);
TEST H = DRUG GENDER*DRUG E = DRUG*SUBJ(GENDER);
RUN;
PROC MEANS DATA = tan.STATIN NOPRINT NWAY;
CLASS GENDER DRUG;
VAR TOTAL HDL LDL;
OUTPUT OUT = G_D MEAN=;
RUN;
OPTIONS LINESIZE  = 60 PAGESIZE = 35;
SYMBOL1 VALUE = SQUARE COLOR = RED INTERPOL = JOIN;
SYMBOL2 VALUE = CIRCLE COLOR = BLUE INTERPOL = JOIN;
PROC GPLOT DATA = G_D;
TITLE "INTERACTION PLOTS";
PLOT TOTAL*DRUG=GENDER HDL*DRUG=GENDER LDL*DRUG=GENDER;
RUN;
```


INTERACTION PLOTS
LDL
120
110
100
90
80
A
B
C
DRUG
GENDER
FEMALE
MALE


INTERACTION PLOTS
HDL
60
50
40
30
A
B
C
DRUG
GENDER
FEMALE
MALE



Gender Add TEST E = SUBJ(GENDER)

The ANOVA Procedure

Dependent Variable: TOTAL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	59	78282.98333	1326.83023	.	.
Error	0	0.00000	.		
Corrected Total	59	78282.98333			

R-Square	Coeff Var	Root MSE	TOTAL Mean
1.000000	.	.	206.6833

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GENDER	1	9644.74091	9644.74091	.	.
SUBJ(GENDER)	18	26570.90909	1476.16162	.	.
DRUG	2	1418.13333	709.06667	.	.
GENDER*DRUG	2	6706.95758	3353.47879	.	.
SUBJ*DRUG(GENDER)	36	33942.24242	942.84007	.	.

Gender Add TEST E = SUBJ(GENDER)

The ANOVA Procedure

Dependent Variable: HDL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	59	8236.183333	139.596328	.	.
Error	0	0.000000	.		
Corrected Total	59	8236.183333			

R-Square	Coeff Var	Root MSE	HDL Mean
1.000000	.	.	46.38333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GENDER	1	261.786027	261.786027	.	.
SUBJ(GENDER)	18	2277.730640	126.540591	.	.
DRUG	2	1503.233333	751.616667	.	.
GENDER*DRUG	2	685.669024	342.834512	.	.
SUBJ*DRUG(GENDER)	36	3507.764310	97.437897	.	.

Gender Add TEST E = SUBJ(GENDER)

The ANOVA Procedure

Dependent Variable: LDL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	59	29810.18333	505.25734	.	.
Error	0	0.00000	.		
Corrected Total	59	29810.18333			

R-Square	Coeff Var	Root MSE	LDL Mean
1.000000	.	.	110.8833

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GENDER	1	2662.71532	2662.71532	.	.
SUBJ(GENDER)	18	11081.46801	615.63711	.	.
DRUG	2	3074.63333	1537.31667	.	.
GENDER*DRUG	2	2524.47104	1262.23552	.	.
SUBJ*DRUG(GENDER)	36	10466.89562	290.74710	.	.

Level of GENDER	N	TOTAL		HDL		LDL	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
FEMALE	27	192.666667	34.3208840	44.0740741	12.1177744	103.518519	25.2987844
MALE	33	218.151515	34.4656724	48.2727273	11.3970191	116.909091	18.1200228

Level of DRUG	N	TOTAL		HDL		LDL	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
A	20	213.550000	28.5463151	41.2000000	10.3649206	118.600000	21.3008525
B	20	203.550000	38.2202413	53.1500000	12.0842438	101.350000	22.7879816
C	20	202.950000	42.0694601	44.8000000	10.0451612	112.700000	20.8354455

Level of GENDER	Level of DRUG	N	TOTAL		HDL		LDL	
			Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
FEMALE	A	9	206.111111	32.6934414	37.6666667	10.3440804	117.111111	24.0283397
FEMALE	B	9	173.111111	20.7511713	47.0000000	13.5092561	83.888889	18.5704365
FEMALE	C	9	198.777778	40.6594939	47.5555556	10.8755587	109.555556	21.8065994
MALE	A	11	219.636364	24.5490233	44.0909091	9.9040855	119.818182	19.9139056
MALE	B	11	228.454545	30.2468631	58.1818182	8.3404818	115.636364	14.5552240
MALE	C	11	206.363636	44.8514721	42.5454545	9.2017785	115.272727	20.6934333

8.6

```
PROC ANOVA DATA = tan.STATIN;
TITLE "Add DIET";
CLASS GENDER DIET SUBJ DRUG;
MODEL TOTAL HDL LDL = GENDER DIET GENDER*DIET SUBJ(GENDER DIET) DRUG
DIET*DRUG GENDER*DIET*DRUG GENDER*DRUG DRUG*SUBJ(GENDER DIET)/NOUNI;
MEANS GENDER|DIET / SNK E=SUBJ(GENDER DIET);
MEANS DRUG DIET*DRUG GENDER*DRUG  GENDER*DIET*DRUG;
TEST H = GENDER DIET GENDER*DIET E = SUBJ(GENDER DIET);
TEST H = DRUG GENDER*DRUG GENDER*DRUG*DIET DIET*DRUG  E = DRUG*SUBJ(GENDER
DIET);
RUN;
```

Add DIET

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
SUBJ	20	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
GENDER	2	FEMALE MALE
DIET	2	HIGH FAT LOW FAT
DRUG	3	A B C

Number of Observations Read	60
Number of Observations Used	60

Add DIET

The ANOVA Procedure

Student-Newman-Keuls Test for TOTAL

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.05
Error Degrees of Freedom	16
Error Mean Square	984.7991
Harmonic Mean of Cell Sizes	29.7

Note: Cell sizes are not equal.

Number of Means	2
Critical Range	17.262629

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	GENDER
A	218.152	33	MALE
B	192.667	27	FEMALE

Add DIET

The ANOVA Procedure

Student-Newman-Keuls Test for HDL

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.05
Error Degrees of Freedom	16
Error Mean Square	86.98748
Harmonic Mean of Cell Sizes	29.7

Note: Cell sizes are not equal.

Number of Means	2
Critical Range	5.1305211

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	GENDER
A	48.273	33	MALE
A			
A	44.074	27	FEMALE

Add DIET

The ANOVA Procedure

Student-Newman-Keuls Test for LDL

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.05
Error Degrees of Freedom	16
Error Mean Square	349.1548
Harmonic Mean of Cell Sizes	29.7

Note: Cell sizes are not equal.

Number of Means	2
Critical Range	10.278792

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	GENDER
A	116.909	33	MALE
B	103.519	27	FEMALE

Level of GENDER	Level of DIET	N	TOTAL		HDL		LDL	
			Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
FEMALE	HIGH FAT	6	224.000000	34.4209239	54.6666667	5.9217115	128.333333	20.7042669
FEMALE	LOW FAT	21	183.714286	29.2645568	41.0476190	11.7833620	96.428571	22.0761669
MALE	HIGH FAT	9	234.333333	28.7836759	49.5555556	11.9280249	124.666667	17.5997159
MALE	LOW FAT	24	212.083333	34.9806365	47.7916667	11.4169179	114.000000	17.7959448

Level of DRUG	N	TOTAL		HDL		LDL	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
A	20	213.550000	28.5463151	41.2000000	10.3649206	118.600000	21.3008525
B	20	203.550000	38.2202413	53.1500000	12.0842438	101.350000	22.7879816
C	20	202.950000	42.0694601	44.8000000	10.0451612	112.700000	20.8354455

Level of DIET	Level of DRUG	N	TOTAL		HDL		LDL	
			Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
HIGH FAT	A	5	237.600000	18.1052479	45.0000000	9.4868330	140.400000	13.6857590
HIGH FAT	B	5	227.000000	27.9374301	57.6000000	3.9115214	110.000000	8.0622577
HIGH FAT	C	5	226.000000	44.9722136	52.2000000	12.1119775	128.000000	18.1934054
LOW FAT	A	15	205.533333	27.1447618	39.9333333	10.6399964	111.333333	18.3290038
LOW FAT	B	15	195.733333	38.6993294	51.6666667	13.5786948	98.466667	25.5059283
LOW FAT	C	15	195.266667	39.6366833	42.3333333	8.3037570	107.600000	19.5733054

Level of GENDER	Level of DRUG	N	TOTAL		HDL		LDL	
			Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
FEMALE	A	9	206.111111	32.6934414	37.6666667	10.3440804	117.111111	24.0283397
FEMALE	B	9	173.111111	20.7511713	47.0000000	13.5092561	83.888889	18.5704365
FEMALE	C	9	198.777778	40.6594939	47.5555556	10.8755587	109.555556	21.8065994
MALE	A	11	219.636364	24.5490233	44.0909091	9.9040855	119.818182	19.9139056
MALE	B	11	228.454545	30.2468631	58.1818182	8.3404818	115.636364	14.5552240
MALE	C	11	206.363636	44.8514721	42.5454545	9.2017785	115.272727	20.6934333

Level of GENDER	Level of DIET	Level of DRUG	N	TOTAL		HDL		LDL	
				Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
FEMALE	HIGH FAT	A	2	236.000000	35.3553391	49.0000000	4.2426407	146.500000	12.0208153
FEMALE	HIGH FAT	B	2	198.000000	7.0710678	54.5000000	0.7071068	106.000000	12.7279221
FEMALE	HIGH FAT	C	2	238.000000	50.9116882	60.5000000	4.9497475	132.500000	12.0208153
FEMALE	LOW FAT	A	7	197.571429	28.8782337	34.4285714	9.1988612	108.714286	19.3796653
FEMALE	LOW FAT	B	7	166.000000	17.3301279	44.8571429	14.8034745	77.571429	14.9427479
FEMALE	LOW FAT	C	7	187.571429	33.3609409	43.8571429	9.0448617	103.000000	19.6044213
MALE	HIGH FAT	A	3	238.666667	5.1316014	42.3333333	12.0138809	136.333333	15.5026879
MALE	HIGH FAT	B	3	246.333333	11.5902258	59.6666667	3.7859389	112.666667	4.7258156
MALE	HIGH FAT	C	3	218.000000	50.0899191	46.6666667	12.8970281	125.000000	23.5796522
MALE	LOW FAT	A	8	212.500000	25.2982213	44.7500000	9.8524834	113.625000	18.3609640
MALE	LOW FAT	B	8	221.750000	32.8666135	57.6250000	9.6944388	116.750000	17.0608156
MALE	LOW FAT	C	8	202.000000	45.5756827	41.0000000	7.9642056	111.625000	19.9279954

8.8

```
DATA tan.STATIN;
 DO SUBJ = 1 TO 20;
      IF RANUNI(1557) LT .5 THEN GENDER = 'FEMALE';
      ELSE GENDER = 'MALE';
      IF RANUNI(0) LT .3 THEN DIET = 'HIGH FAT';
      ELSE DIET = 'LOW FAT';
      DO DRUG = 'A' , 'B' , 'C';
            LDL = ROUND(RANNOR(1557)*20 + 110 + 5*(DRUG EQ 'A') - 10*(DRUG EQ
'B') - 5*(GENDER EQ 'FEMALE') +10*(DIET EQ 'HIGH FAT'));
            HDL = ROUND(RANNOR(1557)*10 + 20 + 0.2*LDL + 12*(DRUG EQ 'B'));
            TOTAL = ROUND(RANNOR(1557)*20 + LDL + HDL + 50 - 10*(GENDER EQ
'FEMALE') + 10*(DIET EQ 'HIGH FAT'));
            OUTPUT;
      END;
 END;
RUN;

PROC PRINT DATA=tan.STATIN;
RUN;
```

```
DATA NEW_STAT;
SET tan.STATIN;
DO I = 1 TO 60;
RETAIN LDL_A 0;
RETAIN LDL_B 0;
RETAIN LDL_C 0;
IF DRUG = 'A' THEN DO;
      LDL_A = LDL;
      LDL_B=0;
      LDL_C=0;
      END;
ELSE IF DRUG = 'B' THEN DO;
      LDL_B = LDL;
      LDL_A=0;
      LDL_C=0;
      END;
ELSE IF DRUG = 'C' THEN DO;
      LDL_C = LDL;
      LDL_A=0;
      LDL_B=0;
      END;
END;
DROP I DIET HDL LDL;
RUN;

PROC MEANS DATA = NEW_STAT NOPRINT NWAY;
CLASS SUBJ GENDER;
VAR LDL_A LDL_B LDL_C;
OUTPUT OUT = STAT1 MAX =;
TITLE "";
RUN;
DATA STAT1;
SET FINAL;
DROP _TYPE_ _FREQ_;
RUN;
PROC PRINT DATA = FINAL;
RUN;
PROC ANOVA DATA = FINAL;
CLASS GENDER;
MODEL LDL_A LDL_B LDL_C = GENDER/NOUNI;
REPEATED DRUG 3 (1 2 3);
MEANS GENDER;
RUN;
```

Obs	SUBJ	GENDER	DIET	DRUG	LDL	HDL	TOTAL
1	1	MALE	LOW FAT	A	132	48	245
2	1	MALE	LOW FAT	B	86	45	169
3	1	MALE	LOW FAT	C	105	43	192
4	2	MALE	LOW FAT	A	119	40	211
5	2	MALE	LOW FAT	B	145	66	255
6	2	MALE	LOW FAT	C	117	38	192
7	3	FEMALE	LOW FAT	A	113	21	204
8	3	FEMALE	LOW FAT	B	54	25	146
9	3	FEMALE	LOW FAT	C	92	48	189
10	4	FEMALE	LOW FAT	A	138	22	228
11	4	FEMALE	LOW FAT	B	94	47	164
12	4	FEMALE	LOW FAT	C	94	31	152
13	5	FEMALE	HIGH FAT	A	138	46	211
14	5	FEMALE	HIGH FAT	B	97	55	193
15	5	FEMALE	HIGH FAT	C	141	64	274
16	6	MALE	HIGH FAT	A	152	54	243
17	6	MALE	HIGH FAT	B	118	57	234
18	6	MALE	HIGH FAT	C	99	36	161
19	7	FEMALE	LOW FAT	A	109	39	179
20	7	FEMALE	LOW FAT	B	74	55	178
21	7	FEMALE	LOW FAT	C	94	41	190
22	8	FEMALE	LOW FAT	A	117	37	201
23	8	FEMALE	LOW FAT	B	94	57	192
24	8	FEMALE	LOW FAT	C	122	51	230
25	9	FEMALE	LOW FAT	A	108	45	223
26	9	FEMALE	LOW FAT	B	76	64	173
27	9	FEMALE	LOW FAT	C	74	38	147

TRANSPOSED:

Obs	SUBJ	GENDER	_TYPE_	_FREQ_	LDL_A	LDL_B	LDL_C
1	1	MALE	3	3	132	86	105
2	2	MALE	3	3	119	145	117
3	3	FEMALE	3	3	113	54	92
4	4	FEMALE	3	3	138	94	94
5	5	FEMALE	3	3	138	97	141
6	6	MALE	3	3	152	118	99
7	7	FEMALE	3	3	109	74	94
8	8	FEMALE	3	3	117	94	122
9	9	FEMALE	3	3	108	76	74
10	10	MALE	3	3	136	111	131
11	11	MALE	3	3	139	121	110
12	12	MALE	3	3	120	127	127
13	13	MALE	3	3	97	123	120
14	14	FEMALE	3	3	73	86	117
15	15	MALE	3	3	88	107	70
16	16	MALE	3	3	95	113	107
17	17	FEMALE	3	3	155	115	124
18	18	MALE	3	3	119	112	137
19	19	MALE	3	3	121	109	145
20	20	FEMALE	3	3	103	65	128

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
GENDER	2	FEMALE MALE

Number of Observations Read	20
Number of Observations Used	20

The ANOVA Procedure
Repeated Measures Analysis of Variance

Repeated Measures Level Information			
Dependent Variable	LDL_A	LDL_B	LDL_C
Level of DRUG	1	2	3

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no DRUG Effect
H = Anova SSCP Matrix for DRUG
E = Error SSCP Matrix

S=1 M=0 N=7.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.57952640	6.17	2	17	0.0097
Pillai's Trace	0.42047360	6.17	2	17	0.0097
Hotelling-Lawley Trace	0.72554694	6.17	2	17	0.0097
Roy's Greatest Root	0.72554694	6.17	2	17	0.0097

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no DRUG*GENDER Effect
H = Anova SSCP Matrix for DRUG*GENDER
E = Error SSCP Matrix

S=1 M=0 N=7.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.61482936	5.32	2	17	0.0160
Pillai's Trace	0.38517064	5.32	2	17	0.0160
Hotelling-Lawley Trace	0.62646755	5.32	2	17	0.0160
Roy's Greatest Root	0.62646755	5.32	2	17	0.0160

The ANOVA Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GENDER	1	2662.71532	2662.71532	4.33	0.0521
Error	18	11081.46801	615.63711		

The ANOVA Procedure
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

						Adj Pr > F	
Source	DF	Anova SS	Mean Square	F Value	Pr > F	G - G	H-F-L
DRUG	2	3074.63333	1537.31667	5.29	0.0097	0.0111	0.0097
DRUG*GENDER	2	2524.47104	1262.23552	4.34	0.0205	0.0226	0.0205
Error(DRUG)	36	10466.89562	290.74710				

Greenhouse-Geisser Epsilon	0.9471
Huynh-Feldt-Lecoutre Epsilon	1.0552

Level of GENDER	N	LDL_A		LDL_B		LDL_C	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
FEMALE	9	117.111111	24.0283397	83.888889	18.5704365	109.555556	21.8065994
MALE	11	119.818182	19.9139056	115.636364	14.5552240	115.272727	20.6834333

Compared to 8.4 the result is similar.

8.10

```
PROC ANOVA DATA = FINAL;
TITLE "Diet 8.10";
CLASS GENDER DIET;
MODEL LDL_A LDL_B LDL_C = GENDER|DIET GENDER*DIET/NOUNI;
REPEATED DRUG/NOM;
RUN;
```