1.1.(a)

```
1  DATA Grade;
2  INPUT ID AGE GENDER$ GPA CSCORE;
3  DATALINES;
4  1 18 M 3.7 650
5  2 18 F 2.0 490
6  3 19 F 3.3 580
7  4 23 M 2.8 530
8  5 21 M 3.5 640
9  ;
10 RUN;
```

1.1.(b)

```
12 PROC MEANS DATA = Grade MEAN MAXDEC=1;
13 Var GPA CSCORE;
14 RUN;
15
```

The MEANS Procedure

Variable	Mean
GPA	3.1
CSCORE	578.0

1.1.(c)

```
1  DATA Grade;
2  INPUT ID AGE GENDER$ GPA CSCORE;
3  INDEX = GPA + 3 * CSCORE/500;
4  DATALINES;
5  1 18 M 3.7 650
6  2 18 F 2.0 490
7  3 19 F 3.3 580
8  4 23 M 2.8 530
9  5 21 M 3.5 640
10 ;
11 RUN;
12 PROC SORT DATA = Grade;
13 BY INDEX;
14 RUN;
15
16 PROC PRINT DATA = Grade;
17 Var ID GPA CSCORE INDEX;
18 RUN;
```

Obs	ID	GPA	CSCORE	INDEX
1	2	2.0	490	4.94
2	4	2.8	530	5.98
3	3	3.3	580	6.78
4	5	3.5	640	7.34
5	1	3.7	650	7.60

1.2(a)

```
DATA DIET;
INPUT SUBJ  1-3
      HEIGHT    4-5
      WT_INIT   6-8
      WT_FINAL  9-11
      ;
      BMI_INIT = sqrt((WT_INIT/2.2)/(HEIGHT*0.0254));
      BMI_FINAL = sqrt((WT_FINAL/2.2)/(HEIGHT*0.0254));
      BMI_DIFF = BMI_FINAL - BMI_INIT
      ;
DATALINES;
00768155150
00272250240
00563240200
00170345298
;
RUN;
PROC SORT DATA = DIET;
BY SUBJ;
RUN;
PROC PRINT DATA = DIET;
VAR SUBJ HEIGHT BMI_INIT BMI_FINAL BMI_DIFF
;
RUN;
```

1.2(b)

Obs	SUBJ	HEIGHT	BMI_INIT	BMI_FINAL	BMI_DIFF
1	1	70	9.39144	8.72833	-0.66312
2	2	72	7.88271	7.72345	-0.15926
3	5	63	8.25671	7.53731	-0.71940
4	7	68	6.38680	6.28294	-0.10386

2.1

```
1 DATA Grade;
2 INPUT ID AGE GENDER$ GPA CSCORE;
3 INDEX = GPA + 3 * CSCORE/500;
4 DATALINES;
5 1 18 M 3.7 650
6 2 18 F 2.0 490
7 3 19 F 3.3 580
8 4 23 M 2.8 530
9 5 21 M 3.5 640
10 ;
11 RUN;
12
13 PROC FREQ DATA = Grade;
14 TABLES GENDER;
15 RUN;
```

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	2	40.00	2	40.00
M	3	60.00	5	100.00

2.2

```
1 DATA CLINIC;
2 INPUT ID$    1-3
3 GENDER$      4
4 RACE$    5
5 HR   6-8
6 SBP      9-11
7 DBP      12-14
8 N_PROC; 15-16
9 AVE_BP = DBP+(SBP-DBP)/3;
10 DATALINES;
11 001MW08013008010
12 002FW08811007205
13 003MB05018810002
14 004FB   10806801
15 005MW06812208204
16 006FB101    07404
17 007FW07810406603
18 008MW04811207006
19 009FB07719011009
20 010FB06616410610
21 ;
22 RUN;
23
24 PROC MEANS DATA = CLINIC N MEAN STD CLM MEDIAN;
25 var SBP DBP AVE_BP;
26 RUN;
27
```

The MEANS Procedure

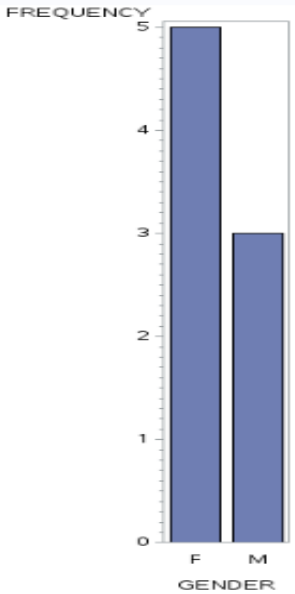
Variable	N	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean	Median
SBP	8	129.5000000	47.5424622	89.7535069	169.2464931	121.0000000
DBP	8	67.8750000	43.0628693	31.8735403	103.8764597	76.0000000
AVE_BP	8	88.4166667	43.8075698	51.7926218	125.0407116	90.6666667

2.4

```
25 PROC FREQ DATA = CLINIC ;
26 TABLES GENDER /NOCUM NOPERCENT;
27 RUN;
28
29 PROC GCHART DATA = CLINIC;
30 VBAR GENDER;
31 RUN;
```

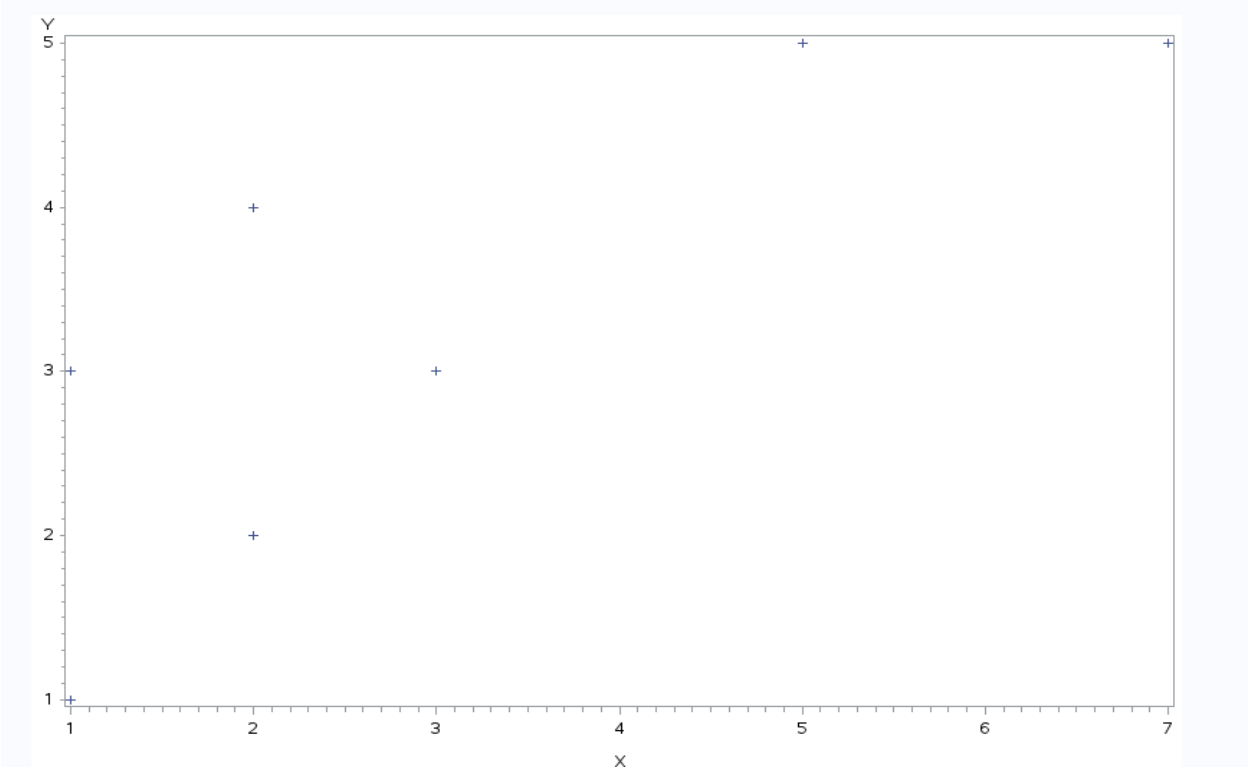
The FREQ Procedure

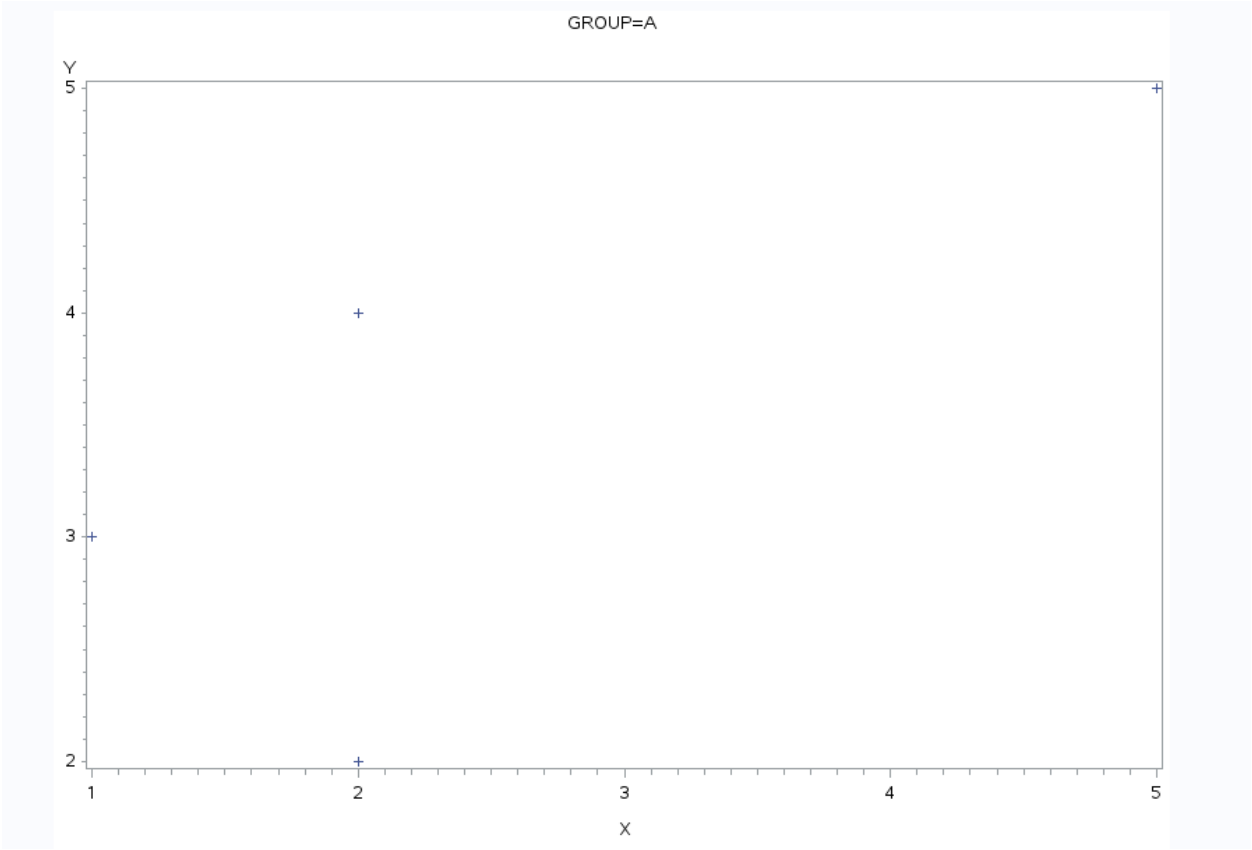
GENDER	Frequency
F	5
M	3

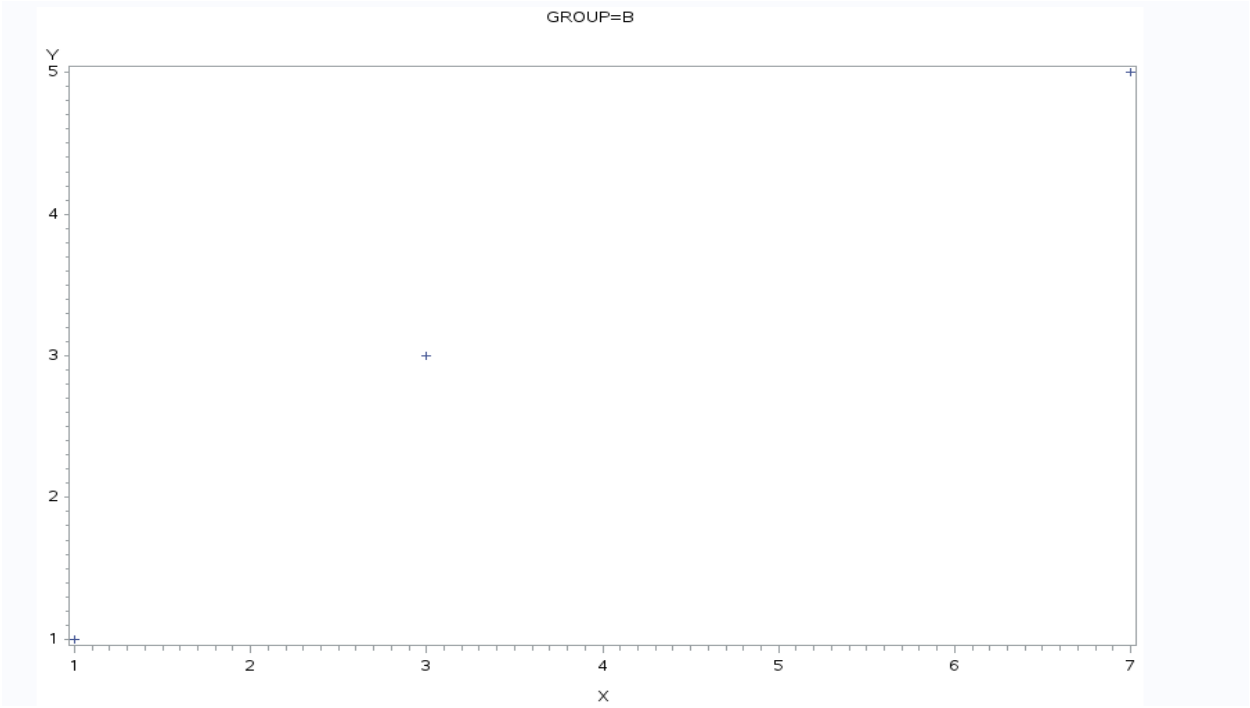


2.5

```
DATA PROB2_5;
     LENGTH GROUP$ 1;
     INPUT X Y Z GROUP $;
     DATALINES;
     2 4 6 A
     3 3 3 B
     1 3 7 A
     7 5 3 B
     1 1 5 B
     2 2 4 A
     5 5 6 A
     ;
     RUN;

PROC GCHART DATA = PROB2_5;
VBAR GROUP;
RUN;

PROC GPLOT DATA = PROB2_5;
PLOT Y * X;
RUN;

PROC SORT DATA=PROB2_5;
BY GROUP;
RUN;

PROC GPLOT DATA = PROB2_5;
BY GROUP;
PLOT Y * X;
RUN;
```

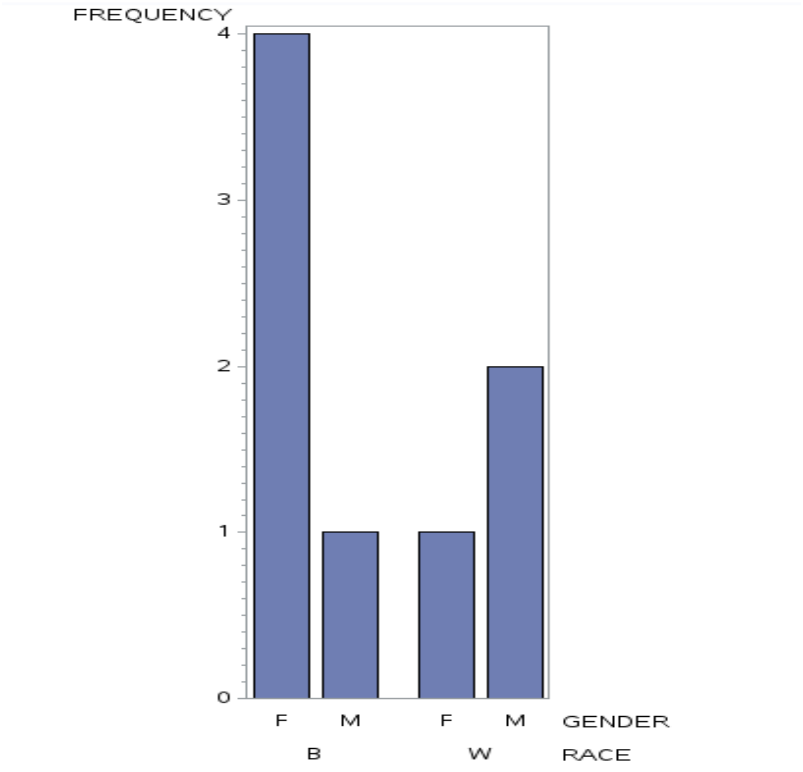


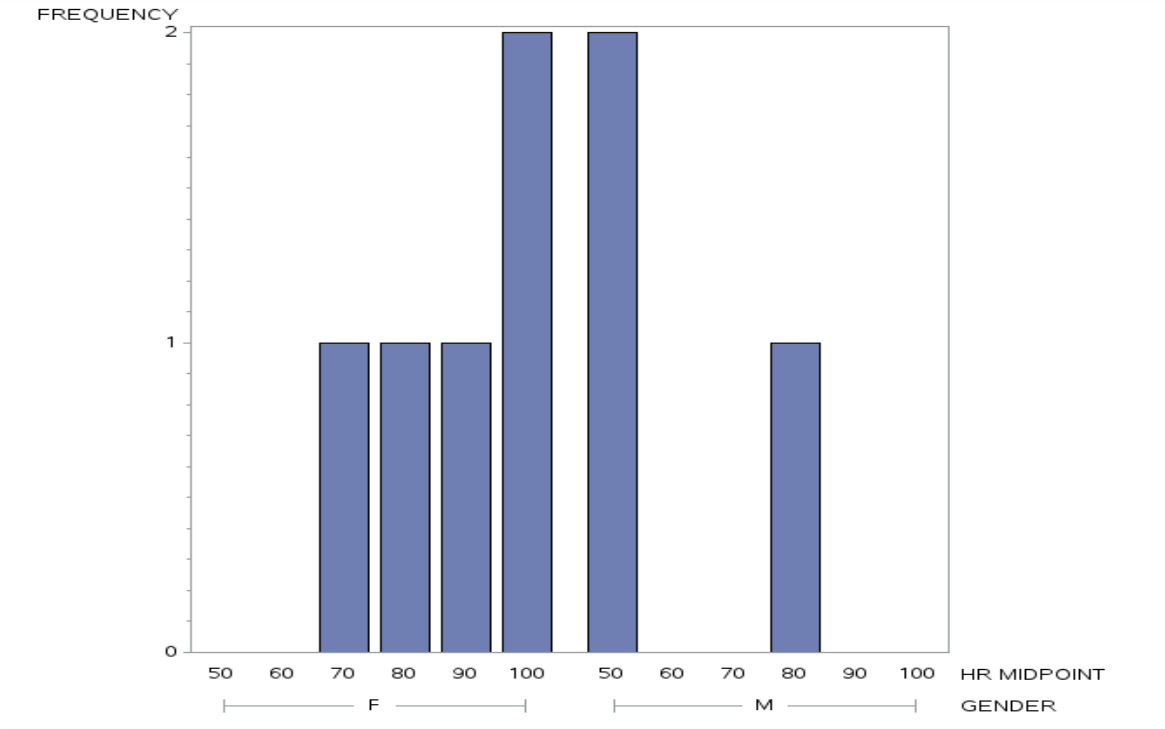

GROUP=A
Y
5
4
3
2
1
2
3
4
5
X


GROUP=B
Y
5
4
3
2
1
1
2
3
4
5
6
7
X

2.8

```
33 PROC GCHART DATA =CLINIC;
34 VBAR GENDER / GROUP = RACE;
35 RUN;
36
37 PROC GCHART DATA =CLINIC;
38 VBAR HR / GROUP = GENDER MIDPOINTS =50 TO 100 BY 10;
39 RUN;
40
```


FREQUENCY
4
3
2
1
0
F
M
F
M
GENDER
B
W
RACE


FREQUENCY
2
1
0
50
60
70
80
90
100
50
60
70
80
90
100
HR MIDPOINT
F
M
GENDER

2.9

```
1    DATA 123;
2         INPUT AGE STATUS PROGNOSIS DOCTOR GENDER STATUS2
3               STATUS3;
4       (data lines)
5         ;
6       PROC CHART DATA=123 BY GENDER;
7          VBAR STATUS
8          VBAR PROGNOSIS;
9       RUN;
10      PROC PLOT DATA=123;
11        DOCTOR BY PROGNOSIS;
12      RUN;
```

1st line : using invalid data set name by starting with a number

4th line: (data lines); is not a proper way to declare data set, DATALINES; would be valid

between 5th and 6th : there is no actual data in DATALINES; a set of data +; needed and put ; or RUN; at the end to close the PROC

6TH line: semi colon missing after DATA =123, also if you want to use BY keyword, you should use sort first and then use BY keyword.

7th line : semi colon missing after STATUS

11th line: PLOT key word missing before DOCTOR, also BY has to be replaced by *, so PLOT DOCTOR * PROGNOSIS; will be a valid line

additional points: GENDER should be marked with $, GENDER $ because of its characteristics in data, variable names gives confusion by using STATUS2 and STAUS3, skipped STATUS1.

3.1

```
1  PROC FORMAT;
2     VALUE GROUP 1 = 'ControlGroup'
3                 2 = 'Aspirin'
4                 3 = 'Iburofen';
5  RUN;
```

3.2

```
PROC FORMAT;
   VALUE GENDER 1 = 'Male'
                2 = 'Female';
   VALUE $SES 'L' = 'Low'
              'M' = 'Medium'
              'H' = 'High';
   VALUE AGEGROUP LOW-20  = '<= 20'
                  21-40   = '21 to 40'
                  41-HIGH = '41+';
RUN;
DATA DRUG;
   INPUT ID $ GENDER SES $ DRUG $ AGE;
   FORMAT GENDER GENDER. SES $SES.;
   LABEL SES = 'Socio-economic Status'
         DRUG = 'Drug Group'
         AGE  = 'Age of Subject';
  IF DRUG IN ('A' 'B' 'C' 'F') THEN
    COST = 'HIGH';
  ELSE IF DRUG NE ' ' THEN
    COST = 'LOW';

DATALINES;
001 1   L B 15
002 2 M    Z   35
003    2 H  F 76
004 1 L C 21
005 2 H . 58
;

PROC PRINT DATA=DRUG NOOBS;

RUN;

PROC FREQ DATA=DRUG;
   TABLES SES COST AGE;
   FORMAT AGE AGEGROUP.;
RUN;
```

ID	GENDER	SES	DRUG	AGE	COST
001	Male	Low	B	15	HIGH
002	Female	Medium	Z	35	LOW
003	Female	High	F	76	HIGH
004	Male	Low	C	21	HIGH
005	Female	High		58	

The FREQ Procedure

Socio-economic Status				
SES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
High	2	40.00	2	40.00
Low	2	40.00	4	80.00
Medium	1	20.00	5	100.00

COST	Frequency	Percent	Cumulative Frequency	Cumulative Percent
HIGH	3	75.00	3	75.00
LOW	1	25.00	4	100.00
Frequency Missing = 1				

Age of Subject				
AGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
<= 20	1	20.00	1	20.00
21 to 40	2	40.00	3	60.00
41+	2	40.00	5	100.00

3.3

(a)

```
PROC FORMAT;
   VALUE $GENDER 'M' = 'Male'
                 'F' = 'Female';
   VALUE $PARTY '1' = 'Republican'
                '2' = 'Democrat'
                '3' = 'Not Registered';
   VALUE YESNO 0 = 'No' 1 = 'Yes';
RUN;


DATA SURVEY;
   INPUT ID     1-3
         GENDER $ 4
         PARTY  $ 5
         VOTE     6
         FOREIGN  7
         SPEND    8;
   LABEL PARTY   =  'Political Party'
         VOTE    =  'Vote in Last Election?'
         FOREIGN = 'Agree with Government Policy?'
         SPEND   = 'Should we Increase Domestic Spending?';
   FORMAT GENDER $GENDER.
          PARTY $PARTY.
          VOTE FOREIGN SPEND YESNO.;

DATALINES;
007M1110
013F2101
137F1001
117 1111
428M3110
017F3101
037M2101
;
```

(b)

```
PROC FREQ DATA=SURVEY;
   TABLES GENDER PARTY VOTE FOREIGN SPEND;
RUN;
```

The FREQ Procedure

GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	3	50.00	3	50.00
Male	3	50.00	6	100.00
Frequency Missing = 1				

Political Party				
PARTY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Republican	3	42.86	3	42.86
Democrat	2	28.57	5	71.43
Not Registered	2	28.57	7	100.00

Vote in Last Election?				
VOTE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	1	14.29	1	14.29
Yes	6	85.71	7	100.00

Agree with Government Policy?				
FOREIGN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	4	57.14	4	57.14
Yes	3	42.86	7	100.00

Should we Increase Domestic Spending?				
SPEND	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	2	28.57	2	28.57
Yes	5	71.43	7	100.00

(c)

```
36 PROC FREQ DATA=SURVEY;
37    TABLES VOTE*(SPEND FOREIGN) / CHISQ;
38 RUN;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of VOTE by SPEND		
	SPEND(Should we Increase Domestic Spending?)		
VOTE(Vote in Last Election?)	No	Yes	Total
No	0 0.00 0.00 0.00	1 14.29 100.00 20.00	1 14.29
Yes	2 28.57 33.33 100.00	4 57.14 66.67 80.00	6 85.71
Total	2 28.57	5 71.43	7 100.00

Statistics for Table of VOTE by SPEND

Statistic	DF	Value	Prob
Chi-Square	1	0.4667	0.4945
Likelihood Ratio Chi-Square	1	0.7376	0.3904
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.4000	0.5271
Phi Coefficient		-0.2582	
Contingency Coefficient		0.2500	
Cramer's V		-0.2582	
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)	0
Left-sided Pr <= F	0.7143
Right-sided Pr >= F	1.0000
Table Probability (P)	0.7143
Two-sided Pr <= P	1.0000

Sample Size = 7

Frequency Percent Row Pct Col Pct	Table of VOTE by FOREIGN		
	FOREIGN(Agree with Government Policy?)		
VOTE(Vote in Last Election?)	No	Yes	Total
No	1 14.29 100.00 25.00	0 0.00 0.00 0.00	1 14.29
Yes	3 42.86 50.00 75.00	3 42.86 50.00 100.00	6 85.71
Total	4 57.14	3 42.86	7 100.00

Statistics for Table of VOTE by FOREIGN

Statistic	DF	Value	Prob
Chi-Square	1	0.8750	0.3496
Likelihood Ratio Chi-Square	1	1.2429	0.2649
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.7500	0.3865
Phi Coefficient		0.3536	
Contingency Coefficient		0.3333	
Cramer's V		0.3536	
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)	1
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.5714
Table Probability (P)	0.5714
Two-sided Pr <= P	1.0000

Sample Size = 7

last election and spending chi value = 0.4667, p value = 0.4945 > significance level alpha = 0.05 so it fails to reject H0, There is likely no association between two. not statistically significant

last election and foreign chi value = 0.8750 p value = 0.3496 > significance level

alpha = 0.05 so it fails to reject H0, There is likely no association between two. not statistically significant

3.4

(a)

```
DATA BLOOD;
   DO I = 1 TO 500;
      WBC = INT(RANNOR(1368)*2000 + 5000);
      X = RANUNI(0);
      IF X LT .05 THEN WBC = .;
      ELSE IF X LT .1 THEN
      WBC = WBC - 3000;
      ELSE IF X LT .15 THEN
      WBC = WBC + 4000;
      OUTPUT;
   END;
   DROP I X;
RUN;

DATA DATASTEP;
   SET BLOOD;
   LENGTH WBCGRP $ 15;

   IF WBC GE 3000 AND WBC LE 4000 THEN
    WBCGRP = 'Low';
   ELSE IF WBC GE 4001 AND WBC LE 6000 THEN
    WBCGRP = 'Medium';
   ELSE IF WBC GE 6001 AND WBC LE 12000 THEN
     WBCGRP = 'High';
   ELSE IF WBC GT 12000
    THEN WBCGRP = 'Abnormally High';
   ELSE IF WBC LT 3000 AND WBC NE . THEN
    WBCGRP = 'Abnormally Low';
   ELSE WBCGRP = 'Not Available';
RUN;

PROC FREQ DATA=DATASTEP;
   TABLES WBCGRP / MISSING NOCUM;
RUN;
```

The FREQ Procedure

WBCGRP	Frequency	Percent
Abnormally High	5	1.00
Abnormally Low	84	16.80
High	174	34.80
Low	58	11.60
Medium	157	31.40
Not Available	22	4.40

(b)

```
DATA BLOOD;
   DO I = 1 TO 500;
      WBC = INT(RANNOR(1368)*2000 + 5000);
      X = RANUNI(0);
      IF X LT .05 THEN WBC = .;
      ELSE IF X LT .1 THEN
      WBC = WBC - 3000;
      ELSE IF X LT .15 THEN
      WBC = WBC + 4000;
      OUTPUT;
   END;
   DROP I X;
RUN;

PROC FORMAT;
   VALUE WBC 3000 - 4000   = 'Low'
             4001 - 6000   = 'Medium'
             6001 - 12000  = 'High'
            LOW - <3000    = 'Abnormally Low'
            >12000 - HIGH = 'Abnormally High'
             .             = 'Not Available';
RUN;

PROC FREQ DATA=BLOOD;

   TABLES WBC / MISSING NOCUM;
   FORMAT WBC WBC.;
RUN;
```

The FREQ Procedure

WBC	Frequency	Percent
Not Available	22	4.40
Abnormally Low >	84	16.80
Low	58	11.60
Medium	157	31.40
High	174	34.80
Abnormally High	5	1.00

3.5

(1)

```
DATA DEMOG;
   INPUT WEIGHT HEIGHT GENDER $;

   IF 0 LE WEIGHT LT 101 THEN
    WeightGRP = 1;
   ELSE IF 101 LE WEIGHT LT 151 THEN
    WeightGRP = 2;
   ELSE IF 151 LE WEIGHT LE 200 THEN
    WeightGRP = 3;
   ELSE IF WEIGHT GT 200 THEN
     WeightGRP = 4;

   IF 0 LE HEIGHT LE 70 THEN
    HeightGRP = 1;
   ELSE IF HEIGHT GT 70 THEN
    HeightGRP = 2;

DATALINES;
155 68 M
98 60 F
202 72 M
280 75 M
130 63 F
.   57 F
166  . M
;

PROC FREQ DATA=DEMOG;
   TABLES WeightGRP*HeightGRP;
RUN;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of WeightGRP by HeightGRP		
	HeightGRP		
WeightGRP	1	2	Total
1	1 20.00 100.00 33.33	0 0.00 0.00 0.00	1 20.00
2	1 20.00 100.00 33.33	0 0.00 0.00 0.00	1 20.00
3	1 20.00 100.00 33.33	0 0.00 0.00 0.00	1 20.00
4	0 0.00 0.00 0.00	2 40.00 100.00 100.00	2 40.00
Total	3 60.00	2 40.00	5 100.00
Frequency Missing = 2			

(2)

```
PROC FORMAT;
   VALUE WeightFMT 0-100    = '1'
               101-150  = '2'
               151-200  = '3'
               201-HIGH = '4';
   VALUE HeightFMT 0-70     = '1'
               71-HIGH = '2';
RUN;

DATA DEMOG;
   INPUT WEIGHT HEIGHT GENDER $;

DATALINES;
155 68 M
98 60 F
202 72 M
280 75 M
130 63 F
.   57 F
166  . M
;
PROC FREQ DATA=DEMOG;
   TABLES WEIGHT*HEIGHT;
   FORMAT WEIGHT WeightFMT. HEIGHT HeightFMT.;
RUN;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of WEIGHT by HEIGHT		
	HEIGHT		
WEIGHT	1	2	Total
1	1 20.00 100.00 33.33	0 0.00 0.00 0.00	1 20.00
2	1 20.00 100.00 33.33	0 0.00 0.00 0.00	1 20.00
3	1 20.00 100.00 33.33	0 0.00 0.00 0.00	1 20.00
4	0 0.00 0.00 0.00	2 40.00 100.00 100.00	2 40.00
Total	3 60.00	2 40.00	5 100.00
Frequency Missing = 2			

3.7

```
1  DATA ASTHMA;
2     INPUT ASTHMA $ SES $ COUNT;
3
4  DATALINES;
5  YES LOW 40
6  NO LOW 100
7  YES HIGH 30
8  NO HIGH 130
9  ;
10
11 PROC FREQ DATA=ASTHMA;
12    TABLES SES*ASTHMA / CHISQ;
13    WEIGHT COUNT;
14 RUN;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of SES by ASTHMA		
	ASTHMA		
SES	NO	YES	Total
HIGH	130 43.33 81.25 56.52	30 10.00 18.75 42.86	160 53.33
LOW	100 33.33 71.43 43.48	40 13.33 28.57 57.14	140 46.67
Total	230 76.67	70 23.33	300 100.00

Statistics for Table of SES by ASTHMA

Statistic	DF	Value	Prob
Chi-Square	1	4.0262	0.0448
Likelihood Ratio Chi-Square	1	4.0234	0.0449
Continuity Adj. Chi-Square	1	3.4959	0.0615
Mantel-Haenszel Chi-Square	1	4.0128	0.0452
Phi Coefficient		0.1158	
Contingency Coefficient		0.1151	
Cramer's V		0.1158	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	130
Left-sided Pr <= F	0.9839
Right-sided Pr >= F	0.0308
Table Probability (P)	0.0148
Two-sided Pr <= P	0.0552

Sample Size = 300

Chi-square = 4.0262, p = 0.0448.

5.2

```
DATA EXAM;
INPUT (Q1-Q8)(1.);

DATALINES;
10101010
11111111
11110101
01100000
11110001
11111111
11111101
11111101
10110101
00010110
;

DATA POINT_BISERIAL;
   SET EXAM;
   RAW = SUM(OF Q1-Q8);
RUN;

PROC CORR DATA=POINT_BISERIAL NOSIMPLE;
   VAR Q1-Q8;
   WITH RAW;
RUN;
```

The CORR Procedure

1 With Variables:	RAW
8 Variables:	Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8

Pearson Correlation Coefficients, N = 10
Prob > |r| under H0: Rho=0

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
RAW	0.76447 0.0100	0.50046 0.1407	0.42471 0.2212	0.63706 0.0476	0.66254 0.0368	0.61168 0.0602	0.10403 0.7749	0.83411 0.0027

5.4

(a)

```
2  DATA SCORES;
3     DO SUBJECT = 1 TO 100;
4        IF RANUNI(1357) LT .5 THEN GROUP = 'A';
5        ELSE GROUP = 'B';
6        MATH = ROUND(RANNOR(1357)*20 + 550 + 10*(GROUP EQ 'A'));
7        SCIENCE = ROUND(RANNOR(1357)*15 + .4*MATH + 300);
8        ENGLISH = ROUND(RANNOR(1357)*20 + 500 + .05*SCIENCE +
9                  .05*MATH);
10       SPELLING = ROUND(RANNOR(1357)*15 + 500 + .1*ENGLISH);
11       VOCAB = ROUND(RANNOR(1357)*5 + 400 + .1*SPELLING +
12              .2*ENGLISH);
13       PHYSICAL = ROUND(RANNOR(1357)*20 + 550);
14       OVERALL = ROUND(MEAN(MATH, SCIENCE, ENGLISH, SPELLING, VOCAB,
15                       PHYSICAL));
16       OUTPUT;
17    END;
18 RUN;
19
20 PROC CORR DATA=SCORES NOSIMPLE;
21    VAR MATH -- OVERALL;
22 RUN;
```

The CORR Procedure

7 Variables:	MATH SCIENCE ENGLISH SPELLING VOCAB PHYSICAL OVERALL

Pearson Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	MATH	SCIENCE	ENGLISH	SPELLING	VOCAB	PHYSICAL	OVERALL
MATH	1.00000	0.32260 0.0011	0.18333 0.0679	0.10647 0.2917	0.15796 0.1165	0.03753 0.7108	0.62332 <.0001
SCIENCE	0.32260 0.0011	1.00000	0.09013 0.3725	0.10649 0.2916	0.08909 0.3781	0.05035 0.6188	0.54013 <.0001
ENGLISH	0.18333 0.0679	0.09013 0.3725	1.00000	0.09526 0.3458	0.70077 <.0001	0.01984 0.8447	0.61817 <.0001
SPELLING	0.10647 0.2917	0.10649 0.2916	0.09526 0.3458	1.00000	0.25311 0.0111	-0.17630 0.0793	0.38804 <.0001
VOCAB	0.15796 0.1165	0.08909 0.3781	0.70077 <.0001	0.25311 0.0111	1.00000	0.02422 0.8109	0.58300 <.0001
PHYSICAL	0.03753 0.7108	0.05035 0.6188	0.01984 0.8447	-0.17630 0.0793	0.02422 0.8109	1.00000	0.38172 <.0001
OVERALL	0.62332 <.0001	0.54013 <.0001	0.61817 <.0001	0.38804 <.0001	0.58300 <.0001	0.38172 <.0001	1.00000

(b) PHYSICAL is independent of all the other test scores but is correlated to OVERALL because it is one component of making up the OVERALL value. (OVERALL = ROUND(MEAN(MATH, SCIENCE, ENGLISH, SPELLING, VOCAB, PHYSICAL));)

(c)

```
20 PROC CORR DATA=SCORES NOSIMPLE;
21    VAR MATH -- PHYSICAL;
22    WITH OVERALL;
23 RUN;
24
```

The CORR Procedure

1 With Variables:	OVERALL
6 Variables:	MATH SCIENCE ENGLISH SPELLING VOCAB PHYSICAL

Pearson Correlation Coefficients, N = 100
Prob > |r| under H0: Rho=0

	MATH	SCIENCE	ENGLISH	SPELLING	VOCAB	PHYSICAL
OVERALL	0.62332 <.0001	0.54013 <.0001	0.61817 <.0001	0.38804 <.0001	0.58300 <.0001	0.38172 <.0001

5.6

(a)

```
20 SYMBOL V=DOT COLOR=BLACK;
21 PROC REG DATA=SCORES;
22    MODEL SCIENCE = MATH;
23    PLOT SCIENCE * MATH
24         RESIDUAL. * MATH;
25 RUN;
26 QUIT;
27
```

The REG Procedure
Model: MODEL1
Dependent Variable: SCIENCE

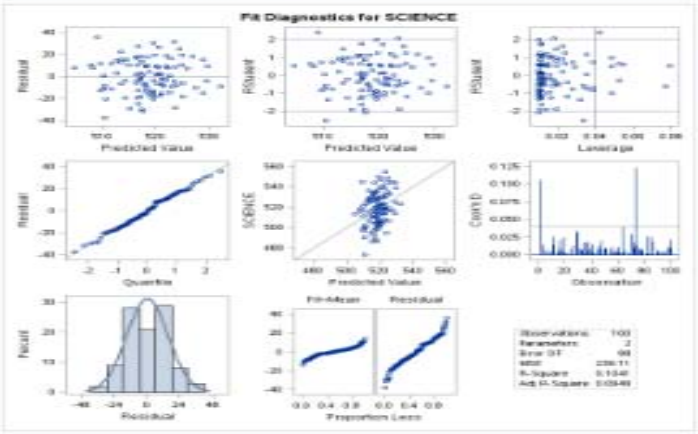
Number of Observations Read	100
Number of Observations Used	100

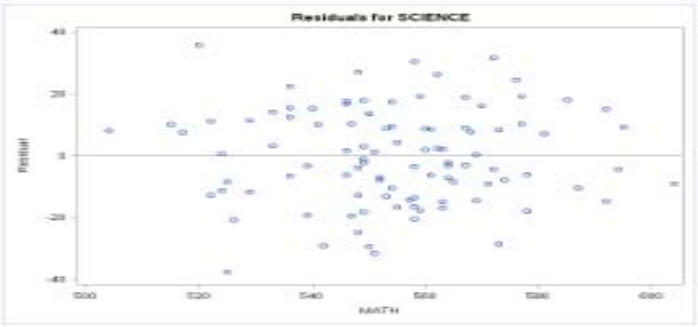
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2721.93239	2721.93239	11.38	0.0011
Error	98	23433	239.11038		
Corrected Total	99	26155			

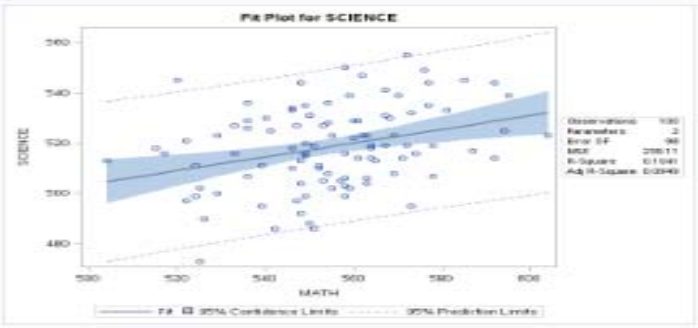
Root MSE	15.46319	R-Square	0.1041
Dependent Mean	518.55000	Adj R-Sq	0.0949
Coeff Var	2.98201		

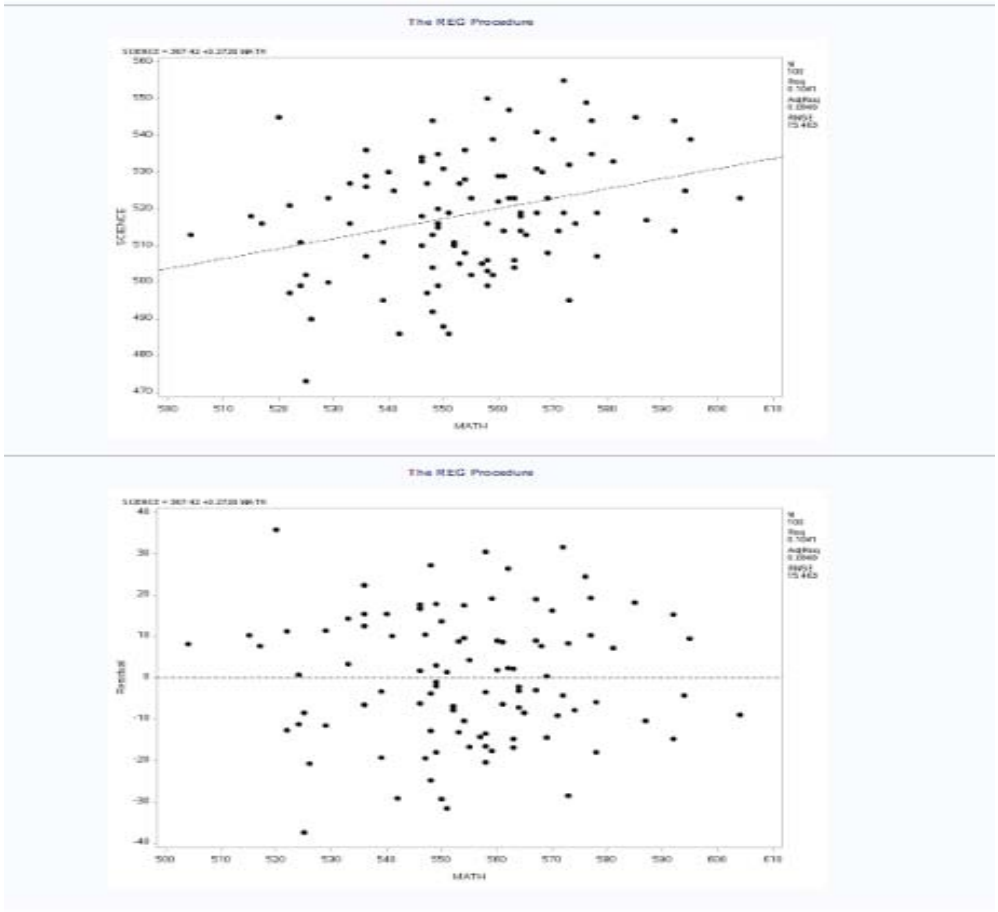
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	367.41770	44.82048	8.20	<.0001
MATH	1	0.27257	0.08079	3.37	0.0011

The REG Procedure
Model: MODEL1
Dependent Variable: SCIENCE



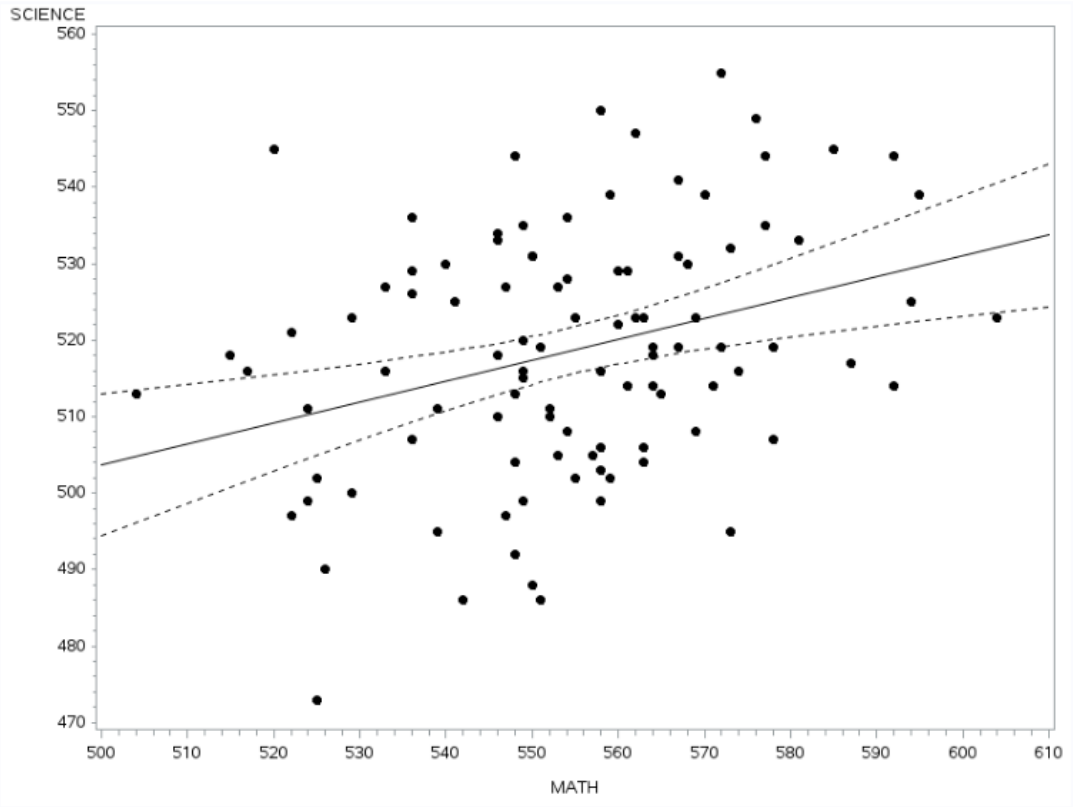


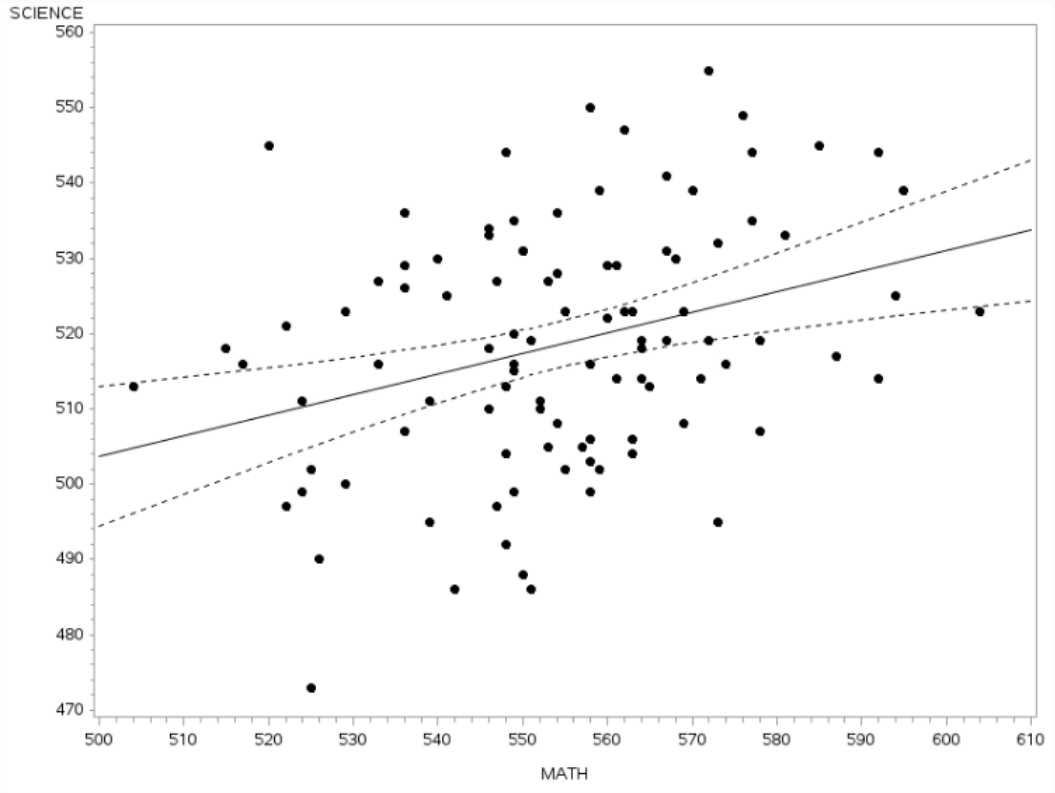




(b)

```
20 SYMBOL1 V=DOT COLOR=BLACK I=RLCLM95;
21 PROC GPLOT DATA=SCORES;
22    PLOT SCIENCE * MATH;
23 RUN;
24
25
```


SCIENCE
560
550
540
530
520
510
500
490
480
470
500
510
520
530
540
550
560
570
580
590
600
610
MATH


SCIENCE
560
550
540
530
520
510
500
490
480
470
500
510
520
530
540
550
560
570
580
590
600
610
MATH

5.8

```
DATA DOSE_RESPONSE;
   INPUT DOSE  SBP DBP;
DATALINES;
4  180 110
4  190 108
4  178 100
8  170 100
8  180  98
8  168  88
16 160  80
16 172  86
16 170  86
32 140  80
32 130  72
32 128  70
;

SYMBOL1 V=DOT I=NONE COLOR=BLACK;
PROC REG DATA=DOSE_RESPONSE;
MODEL SBP = DOSE;
PLOT SBP * DOSE
     RESIDUAL. * DOSE;
MODEL DBP = DOSE;
PLOT DBP * DOSE
     RESIDUAL. * DOSE;
RUN;
QUIT;
```

The REG Procedure
Model: MODEL1
Dependent Variable: SBP

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

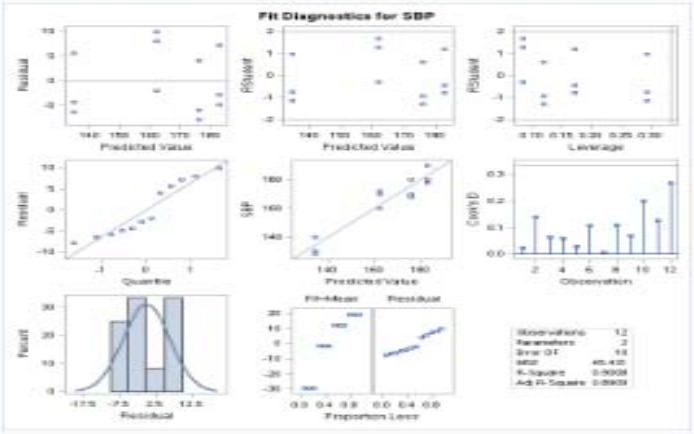
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4125.35942	4125.35942	90.81	<.0001
Error	10	454.30725	45.43072		
Corrected Total	11	4579.66667			

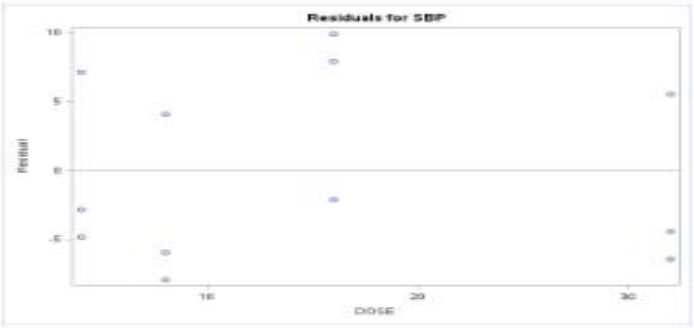
Root MSE	6.74023	R-Square	0.9008
Dependent Mean	163.83333	Adj R-Sq	0.8909
Coeff Var	4.11408		

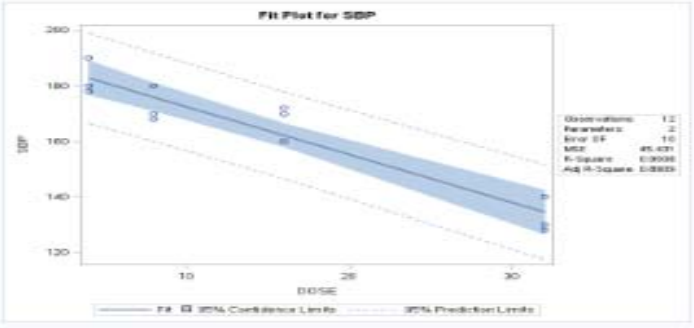
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	189.76812	3.34561	56.72	<.0001
DOSE	1	-1.72899	0.18144	-9.53	<.0001

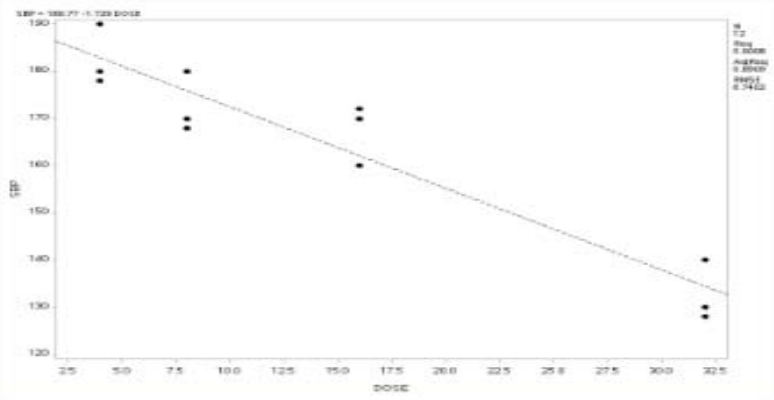
The REG Procedure
Model: MODEL1
Dependent Variable: SBP



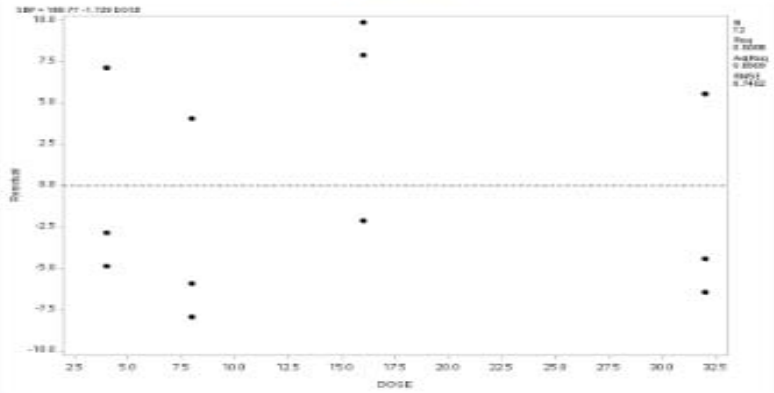




The REG Procedure



The REG Procedure



The REG Procedure
Model: MODEL2
Dependent Variable: DBP

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

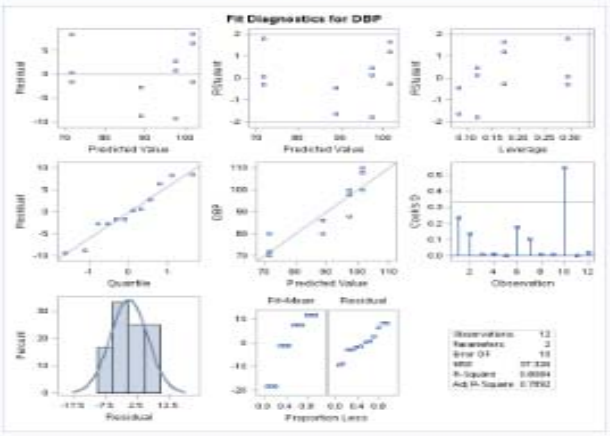
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1574.40290	1574.40290	42.18	<.0001
Error	10	373.26377	37.32638		
Corrected Total	11	1947.66667			

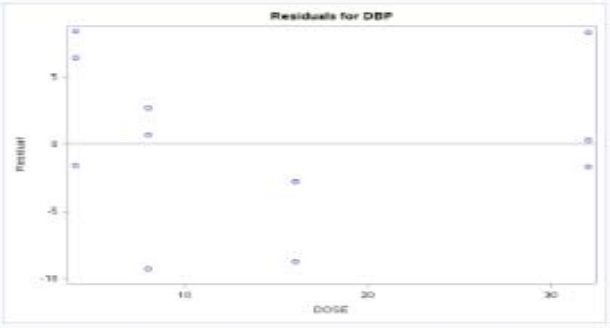
Root MSE	6.10953	R-Square	0.8084
Dependent Mean	89.83333	Adj R-Sq	0.7892
Coeff Var	6.80096		

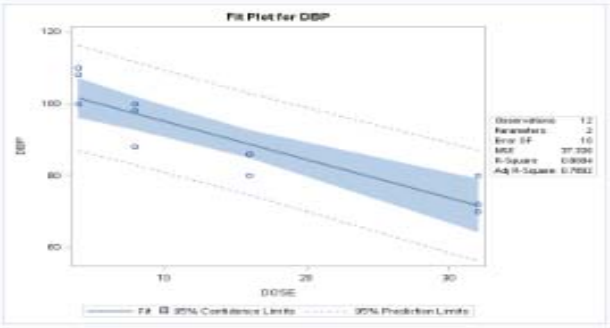
Parameter Estimates

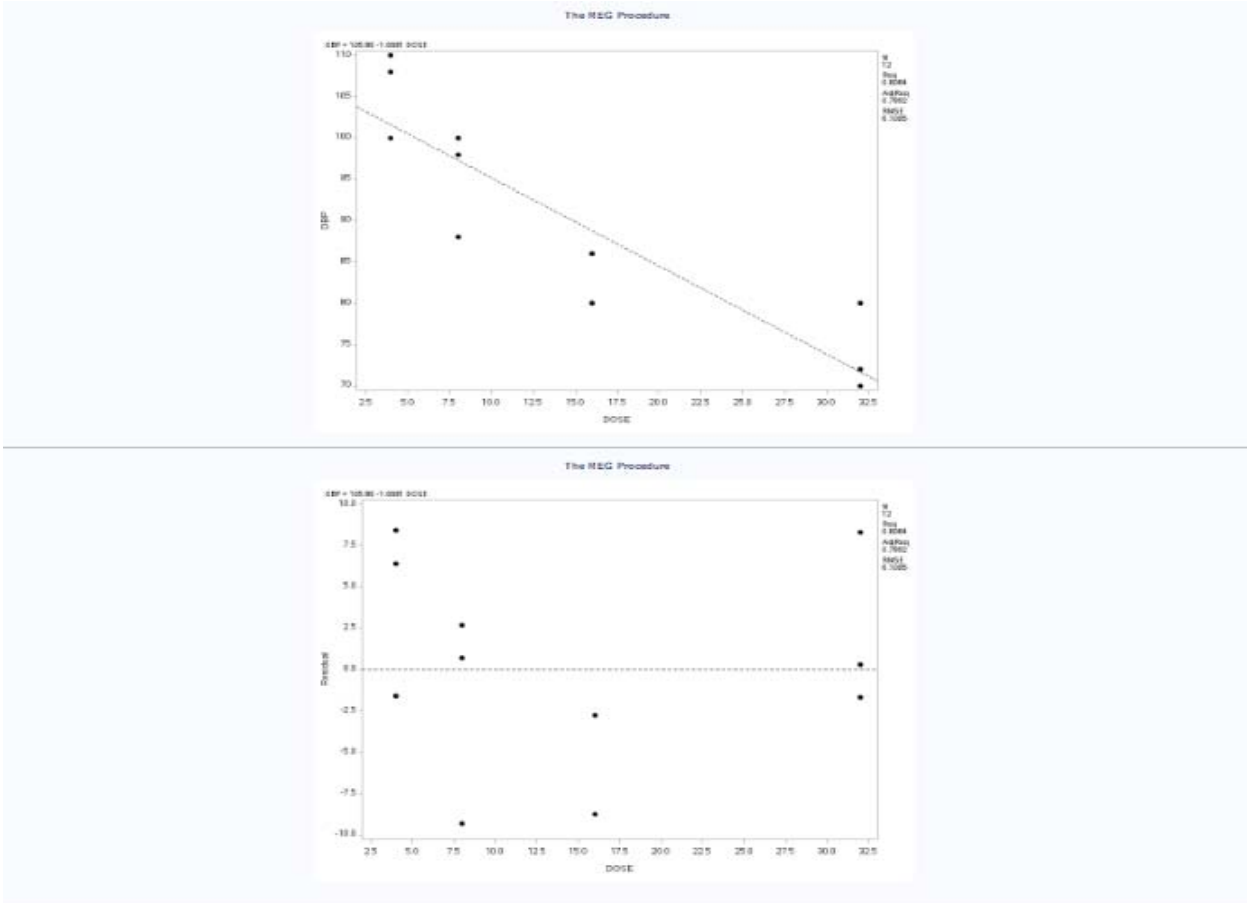
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	105.85507	3.03255	34.91	<.0001
DOSE	1	-1.06812	0.16446	-6.49	<.0001

The REG Procedure
Model: MODEL2
Dependent Variable: DBP









5.10

```
DATA LOG_DOSE;
   SET DOSE_RESPONSE;
   LOG_DOSE = LOG(DOSE);
RUN;

SYMBOL1 V=DOT I=NONE COLOR=BLACK;
PROC REG DATA=LOG_DOSE;
   MODEL SBP = LOG_DOSE;
   PLOT SBP * LOG_DOSE
        RESIDUAL. * LOG_DOSE;
   MODEL DBP = LOG_DOSE;
   PLOT DBP * LOG_DOSE
        RESIDUAL. * LOG_DOSE;
RUN;
QUIT;
```

The REG Procedure
Model: MODEL1
Dependent Variable: SBP

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

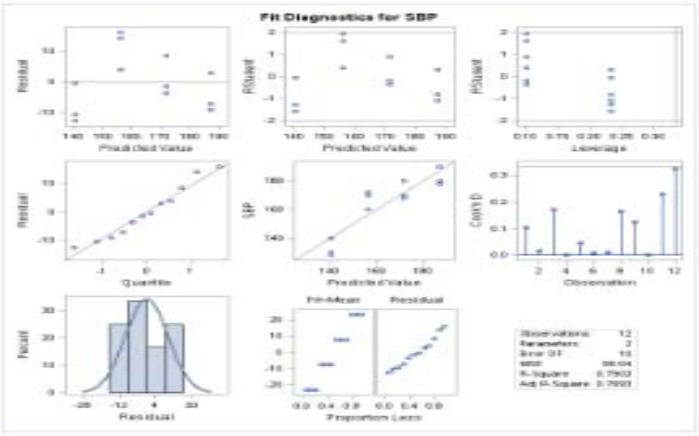
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3619.26667	3619.26667	37.68	0.0001
Error	10	960.40000	96.04000		
Corrected Total	11	4579.66667			

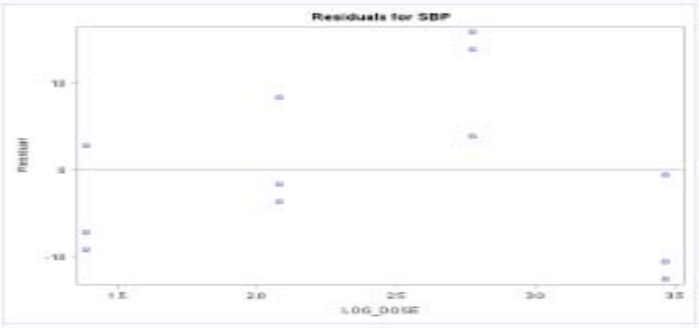
Root MSE	9.80000	R-Square	0.7903
Dependent Mean	163.83333	Adj R-Sq	0.7693
Coeff Var	5.98169		

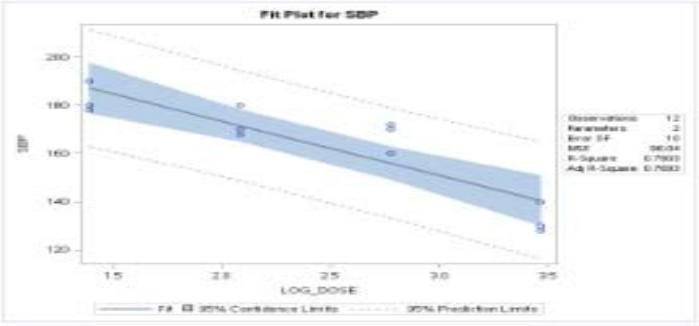
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	218.20000	9.29710	23.47	<.0001
LOG_DOSE	1	-22.40986	3.65052	-6.14	0.0001

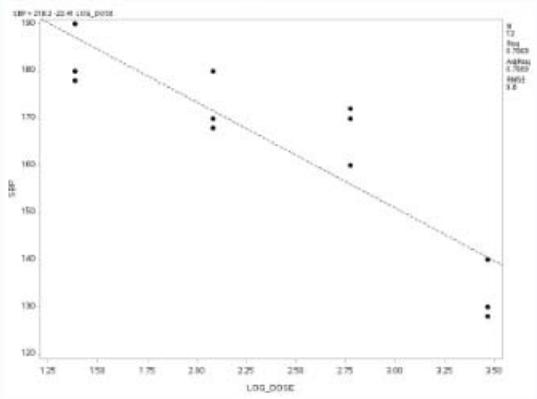
The REG Procedure
Model: MODEL1
Dependent Variable: SBP



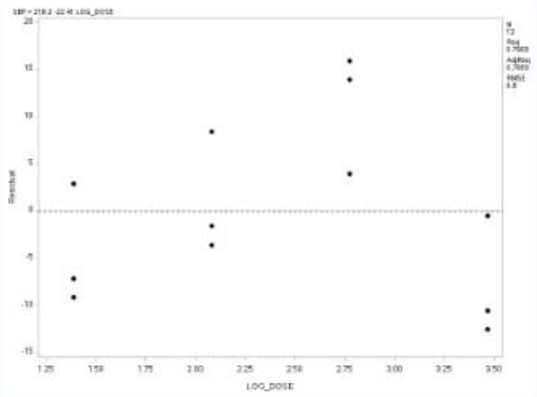




The REG Procedure



The REG Procedure



The REG Procedure
Model: MODEL2
Dependent Variable: DBP

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

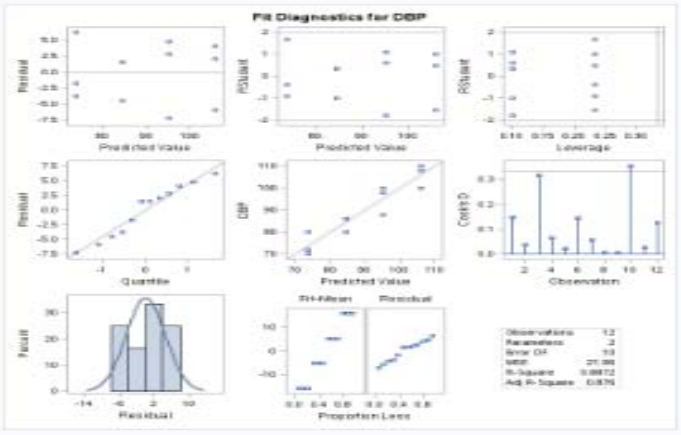
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1728.06667	1728.06667	78.69	<.0001
Error	10	219.60000	21.96000		
Corrected Total	11	1947.66667			

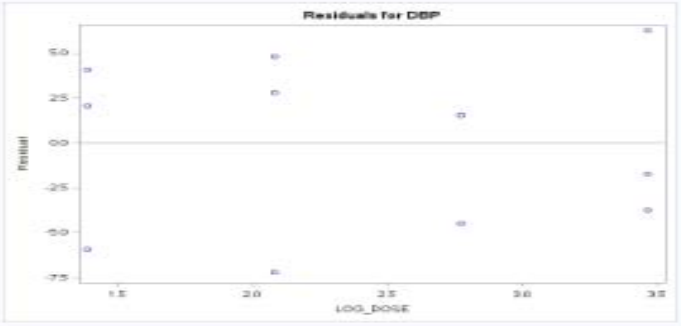
Root MSE	4.68615	R-Square	0.8872
Dependent Mean	89.83333	Adj R-Sq	0.8760
Coeff Var	5.21649		

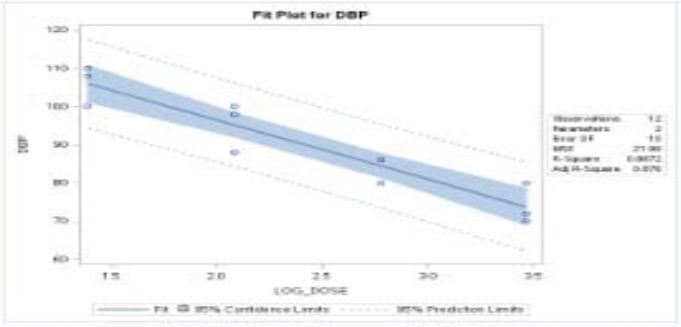
Parameter Estimates

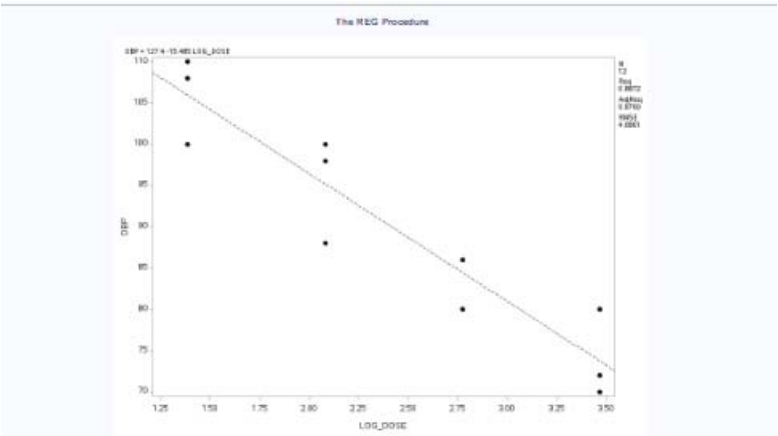
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	127.40000	4.44567	28.66	<.0001
LOG_DOSE	1	-15.48493	1.74560	-8.87	<.0001

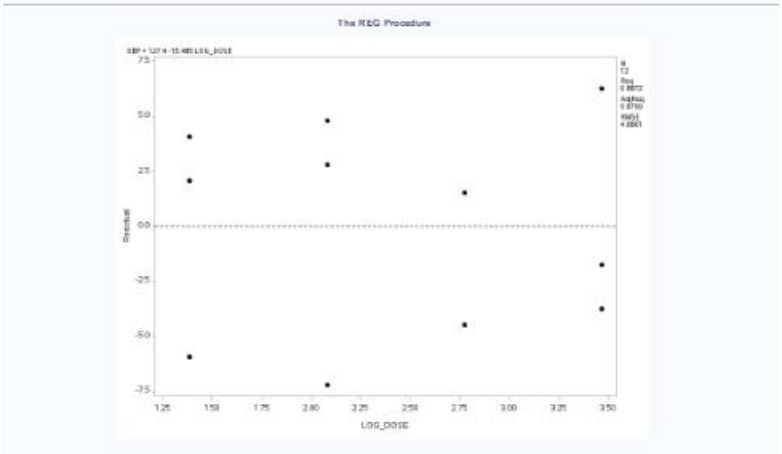
The REG Procedure
Model: MODEL2
Dependent Variable: DBP











Linear fits better for SBP because an error with linear (454) is smaller than an error with log (960).

Log fits better for DBP because an error with linear (373) is bigger than an error with log (219).

6.2

```
1  DATA PROGRAMS;
2     LENGTH i $ 5;
3     DO i = 'CODY','SMITH';
4        DO scores = 1 TO 14;
5           INPUT speed @;
6           OUTPUT;
7        END;
8     END;
9  DATALINES;
10 500 450 505 404 555 567 588 577 566 644 511 522 543 578
11 355 388 440 600 510 501 502 489 499 489 515 520 520 480
12 RUN;
13
14 PROC TTEST DATA= PROGRAMS;
15     TITLE 'READING SPEED COMPARISON';
16     CLASS i;
17     VAR speed;
18 RUN;
19 PROC NPAR1WAY DATA= PROGRAMS WILCOXON;
20     TITLE 'WILCOXON TEST';
21     CLASS i;
22     VAR speed;
23     EXACT WILCOXON;
24 RUN;
25
```

READING SPEED COMPARISON

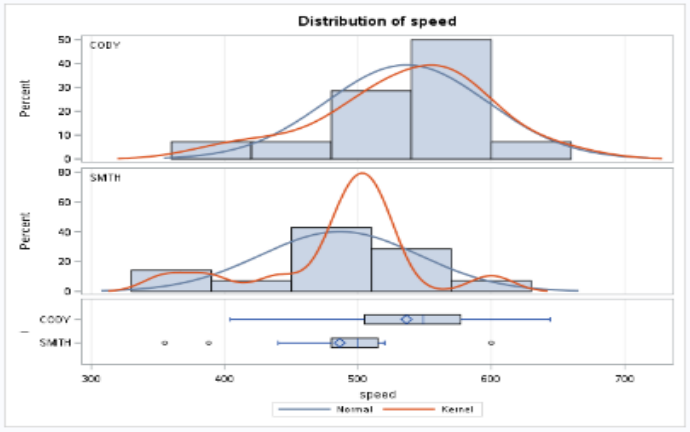
The TTEST Procedure
Variable: speed

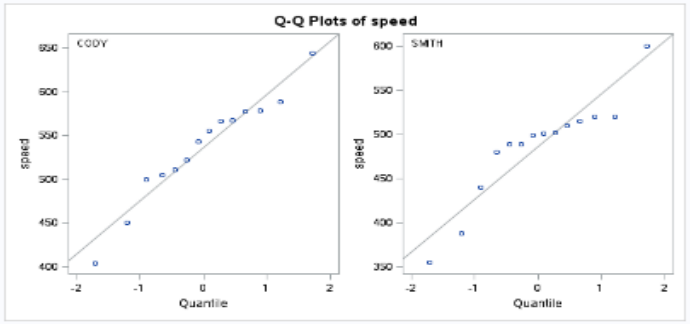
i	N	Mean	Std Dev	Std Err	Minimum	Maximum
CODY	14	536.4	60.7513	16.2365	404.0	644.0
SMITH	14	486.3	59.6843	15.9513	355.0	600.0
Diff (1-2)		50.1429	60.2202	22.7611		

i	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
CODY		536.4	501.4	571.5	60.7513	44.0419	97.8730
SMITH		486.3	451.8	520.7	59.6843	43.2684	96.1540
Diff (1-2)	Pooled	50.1429	3.3568	96.9290	60.2202	47.4244	82.5277
Diff (1-2)	Satterthwaite	50.1429	3.3560	96.9297			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	26	2.20	0.0367
Satterthwaite	Unequal	25.992	2.20	0.0367

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	13	13	1.04	0.9500


Distribution of speed
CODY
SMITH
Percent
speed
Normal
Kernel


Q-Q Plots of speed
CODY
SMITH
speed
Quantile

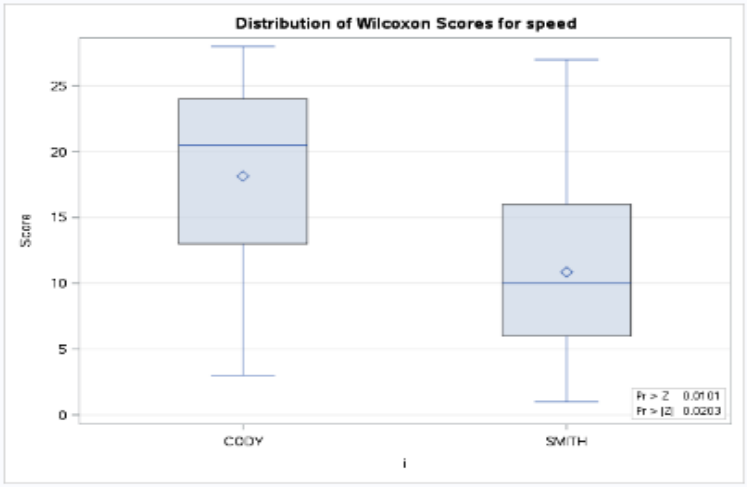
WILCOXON TEST

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable speed Classified by Variable i					
i	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
CODY	14	254.0	203.0	21.757927	18.142857
SMITH	14	152.0	203.0	21.757927	10.857143
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic (S)	254.0000
Normal Approximation	
Z	2.3210
One-Sided Pr > Z	0.0101
Two-Sided Pr > |Z|	0.0203
t Approximation	
One-Sided Pr > Z	0.0140
Two-Sided Pr > |Z|	0.0281
Exact Test	
One-Sided Pr >= S	0.0090
Two-Sided Pr >= |S - Mean|	0.0179
Z includes a continuity correction of 0.5.	

Kruskal-Wallis Test	
Chi-Square	5.4942
DF	1
Pr > Chi-Square	0.0191



6.4

```
1
2
3  DATA QUES6_4;
4     DO GROUP = 'A','B','C';
5        DO I = 1 TO 10;
6           X = ROUND(RANNOR(135)*10 + 300 + 5*(GROUP EQ 'A') - 7*(GROUP EQ 'C'));
7           Y = ROUND(RANUNI(135)*100 + X);
8           OUTPUT;
9        END;
10    END;
11    DROP I;
12 RUN;
13 PROC TTEST DATA=QUES6_4;
14    TITLE "Comparing between Group A and Group C";
15    WHERE GROUP IN ('A' 'C');
16    VAR X Y;
17    CLASS GROUP;
18 RUN;
```

Comparing between Group A and Group C

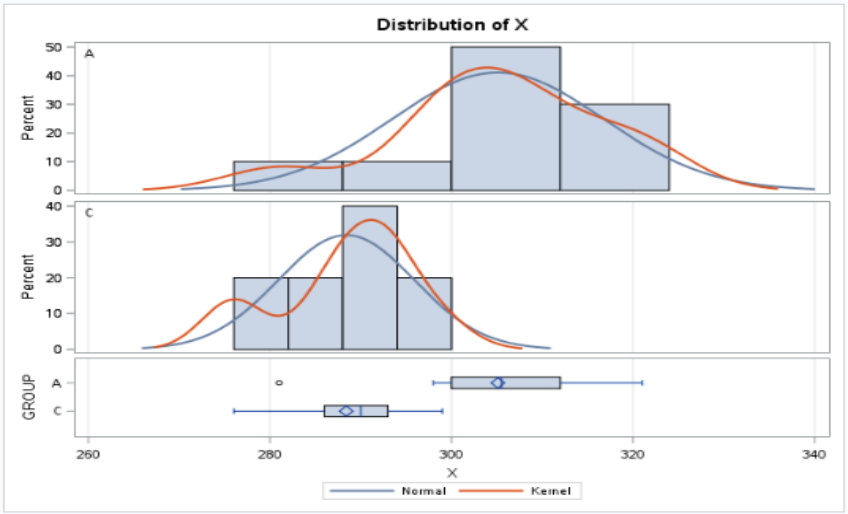
The TTEST Procedure
Variable: X

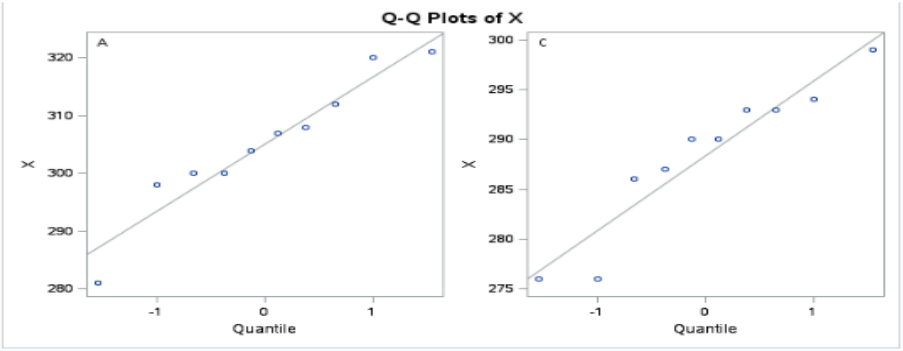
GROUP	N	Mean	Std Dev	Std Err	Minimum	Maximum
A	10	305.1	11.6376	3.6801	281.0	321.0
C	10	288.4	7.5011	2.3721	276.0	299.0
Diff (1-2)		16.7000	9.7903	4.3784		

GROUP	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		305.1	296.8	313.4	11.6376	8.0047	21.2457
C		288.4	283.0	293.8	7.5011	5.1595	13.6941
Diff (1-2)	Pooled	16.7000	7.5014	25.8986	9.7903	7.3977	14.4781
Diff (1-2)	Satterthwaite	16.7000	7.3877	26.0123			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	18	3.81	0.0013
Satterthwaite	Unequal	15.377	3.81	0.0016

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	2.41	0.2068





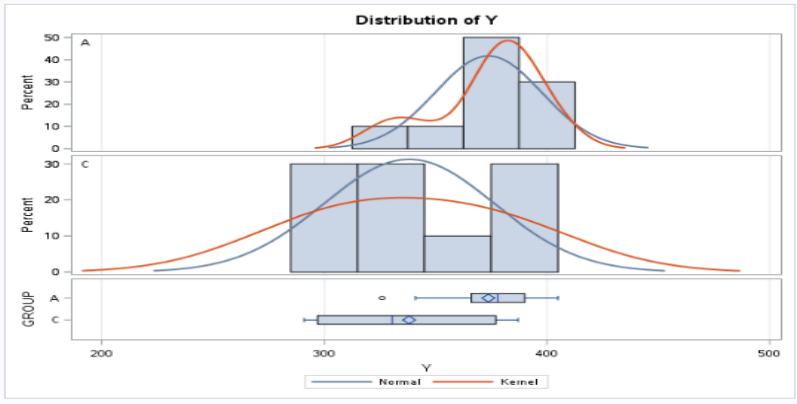
Variable: Y

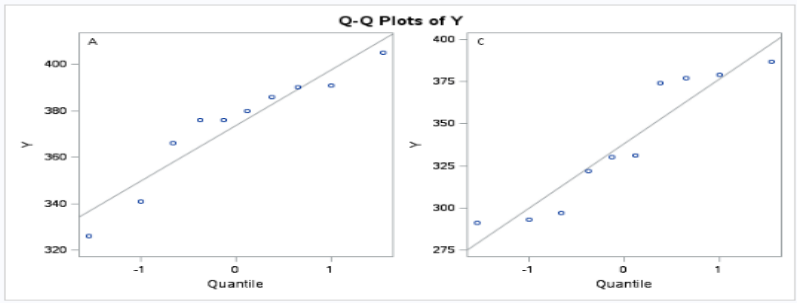
GROUP	N	Mean	Std Dev	Std Err	Minimum	Maximum
A	10	373.7	23.9214	7.5646	326.0	405.0
C	10	338.1	38.2723	12.1028	291.0	387.0
Diff (1-2)		35.6000	31.9139	14.2724		

GROUP	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		373.7	356.6	390.8	23.9214	16.4540	43.6711
C		338.1	310.7	365.5	38.2723	26.3250	69.8702
Diff (1-2)	Pooled	35.6000	5.6149	65.5851	31.9139	24.1146	47.1951
Diff (1-2)	Satterthwaite	35.6000	5.1969	66.0031			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	18	2.49	0.0226
Satterthwaite	Unequal	15.101	2.49	0.0247

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	2.56	0.1777





Procedure

7-2

```
DATA MEDICATION;
LENGTH TREAT $ 7;
DO TREAT = 'A','B','Placebo';
   DO I = 1 TO 10;
      INPUT CHOLESTEROL @;
      OUTPUT;
   END;
END;
DROP I;
DATALINES;

220  190  180  185  210  170  178  200  177  189
160  168  178  200  172  155  159  167  185  199
240  220  246  244  198  238  277  255  190  188
;

PROC ANOVA DATA=MEDICATION;
   CLASS TREAT;
   MODEL CHOLESTEROL = TREAT;
   MEANS TREAT / SNK;
RUN;
QUIT;

```

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
TREAT	3	A B Placebo

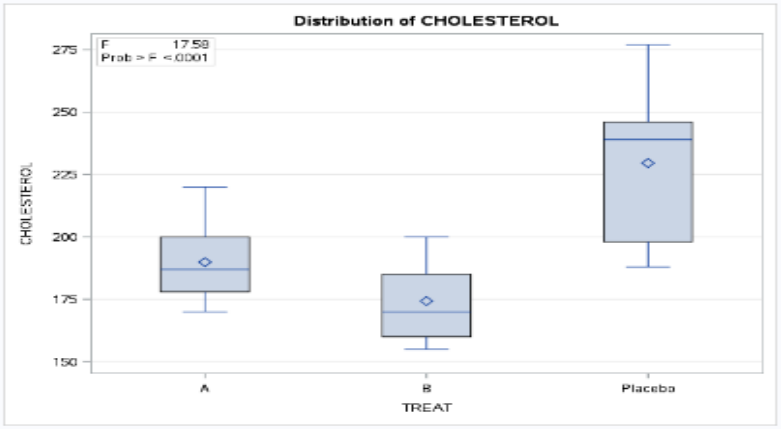
Number of Observations Read	30
Number of Observations Used	30

The ANOVA Procedure
Dependent Variable: CHOLESTEROL

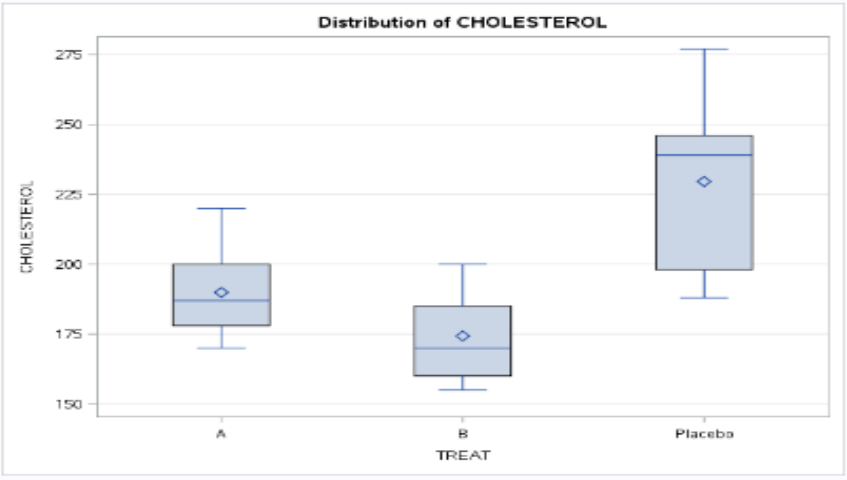
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	16258.46667	8129.23333	17.58	<.0001
Error	27	12483.40000	462.34815		
Corrected Total	29	28741.86667			

R-Square	Coeff Var	Root MSE	CHOLESTEROL Mean
0.565672	10.86340	21.50228	197.9333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
TREAT	2	16258.46667	8129.23333	17.58	<.0001



The ANOVA Procedure



The ANOVA Procedure
Student-Newman-Keuls Test for CHOLESTEROL

: 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	27
Error Mean Square	462.3481

Number of Means	2	3
Critical Range	19.730539	23.841676

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	TREAT
A	229.600	10	Placebo
B	189.900	10	A
B			
B	174.300	10	B

```
DATA PATIENTS;
DO GROUP = 'Deficiency','Normal';
   DO I = 1 TO 4;
      DO TREAT = 'Drug','Placebo';
         INPUT DEPRESSION @@;
         OUTPUT;
      END;
   END;
END;
DROP I;
DATALINES;
9  9
11 6
10 6
10 7
5  12
4  11
7  10
7  11
;
PROC ANOVA DATA=PATIENTS;
   CLASS GROUP TREAT;
   MODEL DEPRESSION = GROUP | TREAT;
   MEANS GROUP | TREAT / SNK;
RUN;
PROC MEANS DATA=PATIENTS NOPRINT NWAY;
   CLASS GROUP TREAT;
   VAR DEPRESSION;
   OUTPUT OUT=MEANS MEAN=;
RUN;
SYMBOL1 V=SQUARE C=BLACK I=JOIN;
SYMBOL2 V=CIRCLE C=BLACK I=JOIN;
PROC GPLOT DATA=MEANS;
   PLOT DEPRESSION * TREAT = GROUP;
RUN;

DATA ONEWAY;
   SET PATIENTS;
   FACTOR = TRIM(GROUP) || '-' || TREAT;
RUN;
PROC ANOVA DATA=ONEWAY;
   CLASS FACTOR;
   MODEL DEPRESSION = FACTOR;
   MEANS FACTOR / SNK;
RUN;
QUIT;
```

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
GROUP	2	Deficiency Normal
TREAT	2	Drug Plac

Number of Observations Read	16
Number of Observations Used	16

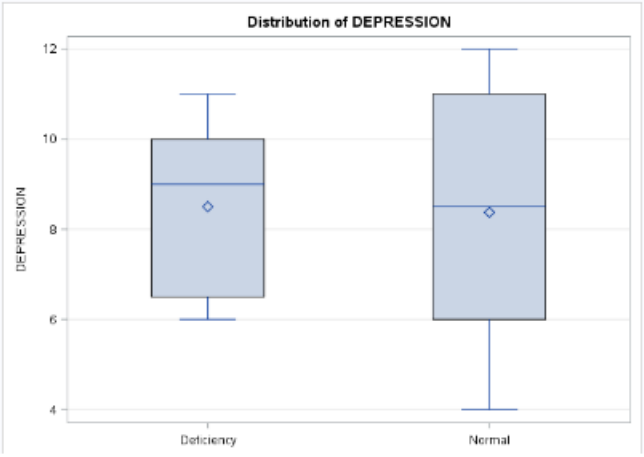
The ANOVA Procedure
Dependent Variable: DEPRESSION

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	73.18750000	24.39583333	17.48	0.0001
Error	12	16.75000000	1.39583333		
Corrected Total	15	89.93750000			

R-Square	Coeff Var	Root MSE	DEPRESSION Mean
0.813760	14.00242	1.181454	8.437500

Source	DF	Anova SS	Mean Square	F Value	Pr > F
GROUP	1	0.06250000	0.06250000	0.04	0.8360
TREAT	1	5.06250000	5.06250000	3.63	0.0811
GROUP*TREAT	1	68.06250000	68.06250000	48.76	<.0001

The ANOVA Procedure



The ANOVA Procedure
Student-Newman-Keuls Test for DEPRESSION

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

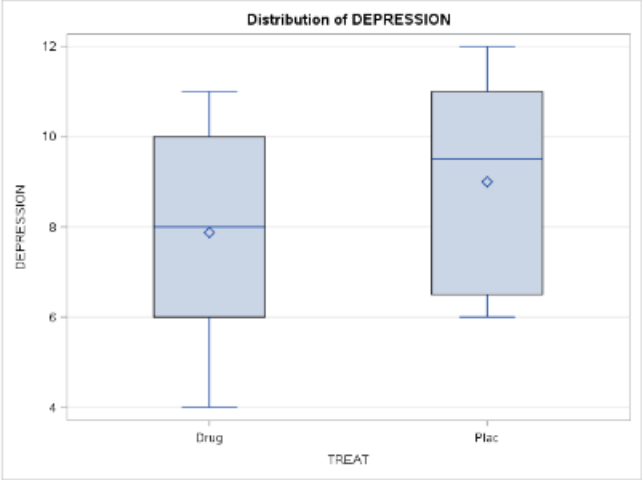
Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	1.395833

Number of Means	2
Critical Range	1.2870326

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	GROUP
A	8.5000	8	Deficiency
A			
A	8.3750	8	Normal

The ANOVA Procedure



The ANOVA Procedure
Student-Newman-Keuls Test for DEPRESSION

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

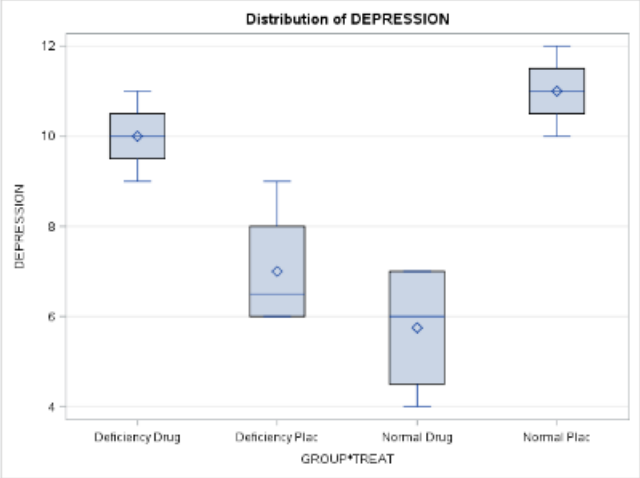
Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	1.395833

Number of Means	2
Critical Range	1.2870326

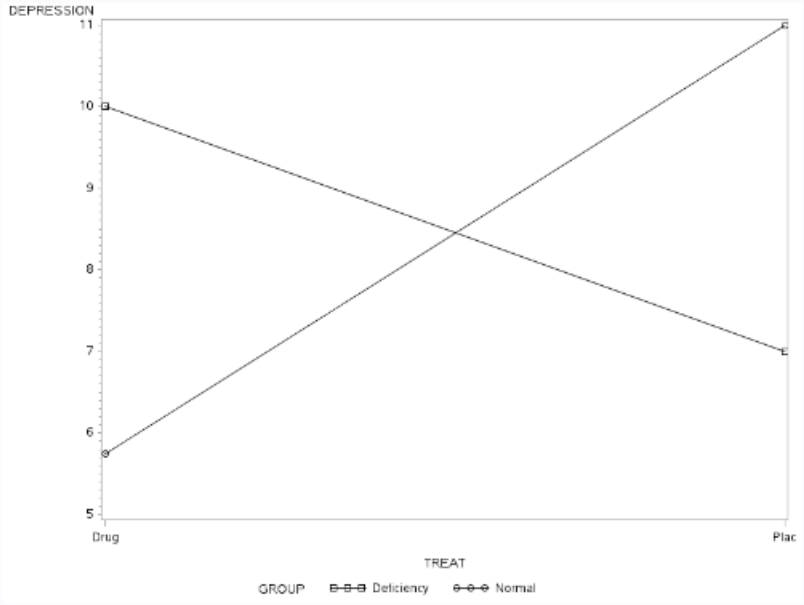
Means with the same letter are not significantly different.

SNK Grouping	Mean	N	TREAT
A	9.0000	8	Plac
A			
A	7.8750	8	Drug

The ANOVA Procedure



Level of GROUP	Level of TREAT	N	DEPRESSION Mean	DEPRESSION Std Dev
Deficiency	Drug	4	10.0000000	0.81649658
Deficiency	Plac	4	7.0000000	1.41421356
Normal	Drug	4	5.7500000	1.50000000
Normal	Plac	4	11.0000000	0.81649658



The ANOVA Procedure

Class Level Information		
Class	Levels	Values
FACTOR	4	Deficiency-Drug Deficiency-Plac Normal-Drug Normal-Plac

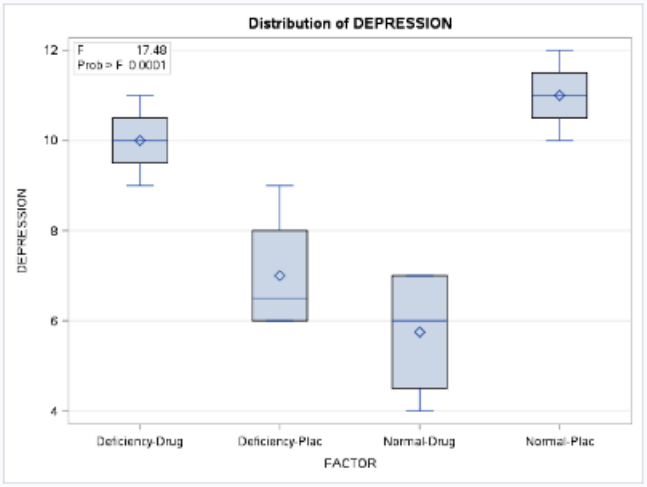
Number of Observations Read	16
Number of Observations Used	16

The ANOVA Procedure
Dependent Variable: DEPRESSION

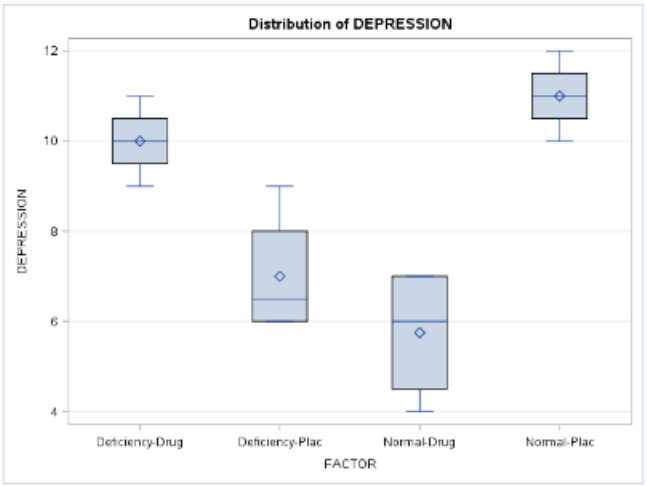
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	73.18750000	24.39583333	17.48	0.0001
Error	12	16.75000000	1.39583333		
Corrected Total	15	89.93750000			

R-Square	Coeff Var	Root MSE	DEPRESSION Mean
0.813760	14.00242	1.181454	8.437500

Source	DF	Anova SS	Mean Square	F Value	Pr > F
FACTOR	3	73.18750000	24.39583333	17.48	0.0001



The ANOVA Procedure



The ANOVA Procedure
Student-Newman-Keuls Test for DEPRESSION

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	12
Error Mean Square	1.395833

Number of Means	2	3	4
Critical Range	1.820139	2.2287451	2.4801736

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	FACTOR
A	11.0000	4	Normal-Plac
A			
A	10.0000	4	Deficiency-Drug
B	7.0000	4	Deficiency-Plac
B			
B	5.7500	4	Normal-Drug