

1.

a. Do a linear regression using PROC REG. State the estimated linear regression equation.

```
1 DATA TEMP;  
2 INFILE '/home/yeopdodo860/my_courses/tjp00/CricketChirpsvsTemperature.csv' delimiter= ',' dsd;  
3 INPUT CHIRP TEMP;  
4 RUN;  
5  
6 PROC REG DATA=TEMP;  
7 MODEL CHIRP=TEMP;  
8 RUN;  
9
```

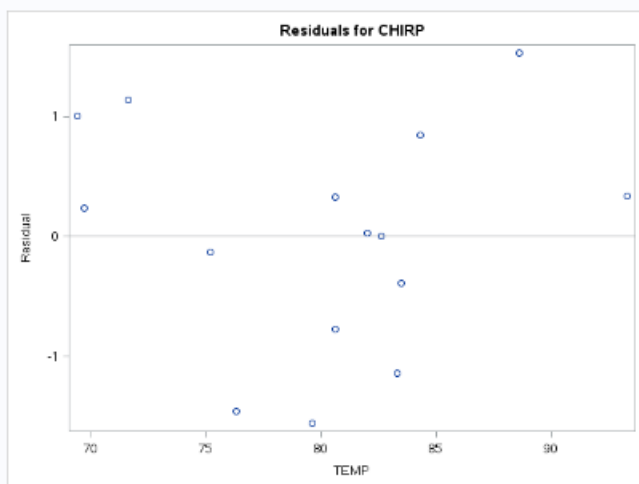
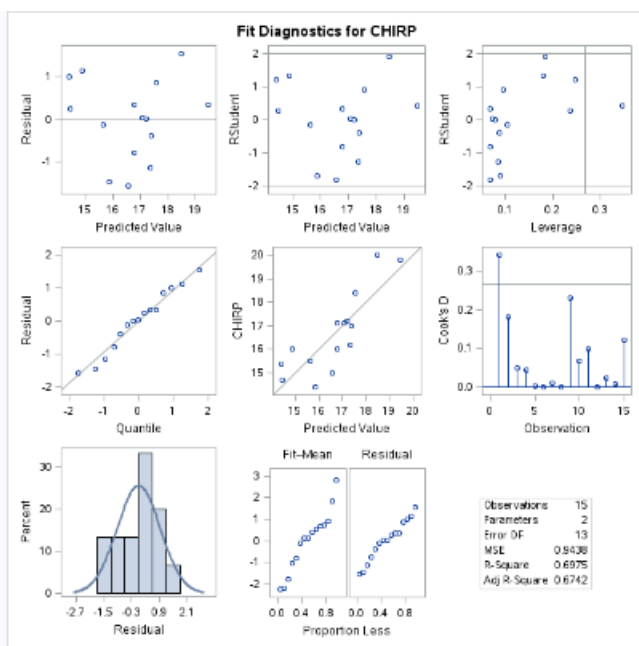
The REG Procedure
Model: MODEL1
Dependent Variable: CHIRP

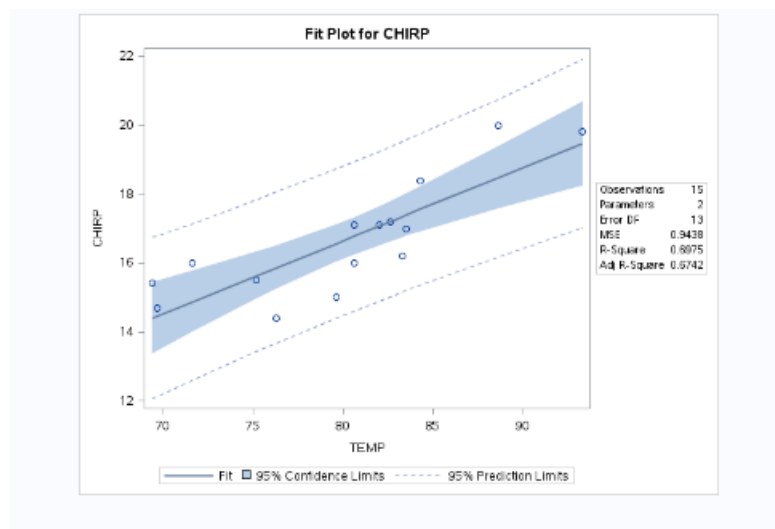
Number of Observations Read	15
Number of Observations Used	15

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	28.28733	28.28733	29.97	0.0001
Error	13	12.27001	0.94385		
Corrected Total	14	40.55733			

Root MSE	0.97152	R-Square	0.6975
Dependent Mean	16.65333	Adj R-Sq	0.6742
Coeff Var	5.83377		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-0.30914	3.10858	-0.10	0.9223
TEMP	1	0.21192	0.03871	5.47	0.0001





b. Interpret the regression output and state whether the following indicate that the regression equation is reliable and should be used.

i. The p-value for the ANOVA table.

p value is 0.83514

ii. R-Square value. Include an interpretation of what this number tells us.

R-Square	0.6975
-----------------	--------

Approximately 67.75% of the variability in chirp can be explained by or attributed to variability in temp.

iii. The p-value for parameter (β_0 and β_1) estimates. Include a conclusion about the statistical significance of the linear regression equation.

$\beta_0 = 0.9223$, changes in the predictor is not related to changes in the response since larger than alpha 0.05. less meaningful to the model

$\beta_1 = 0.0001$ changes in the predictor is related to changes in the response since less than alpha 0.05. more meaningful to the model

1.a. Do a linear regression using PROC REG. State the estimated linear regression equation.

```

1 DATA TEMP;
2 INFILE '/home/yeopdodo860/my_courses/tjp00/FireandTheftinChicago.csv' delimiter= ',' dsd;
3 INPUT FIRE THEFT;
4 RUN;
5
6 PROC REG DATA=TEMP;
7 MODEL FIRE=THEFT;
8 RUN;
9

```

The REG Procedure
Model: MODEL1
Dependent Variable: FIRE

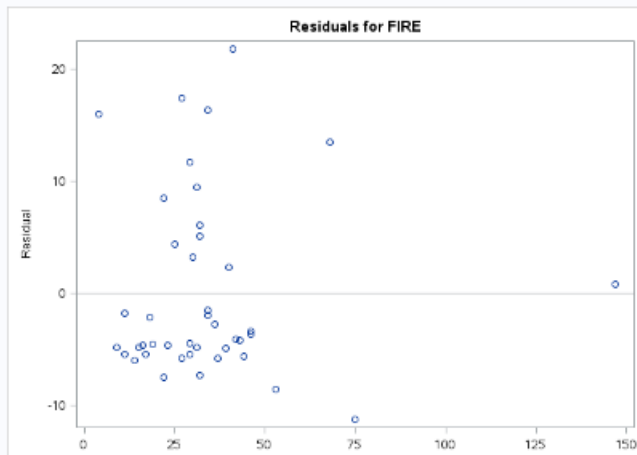
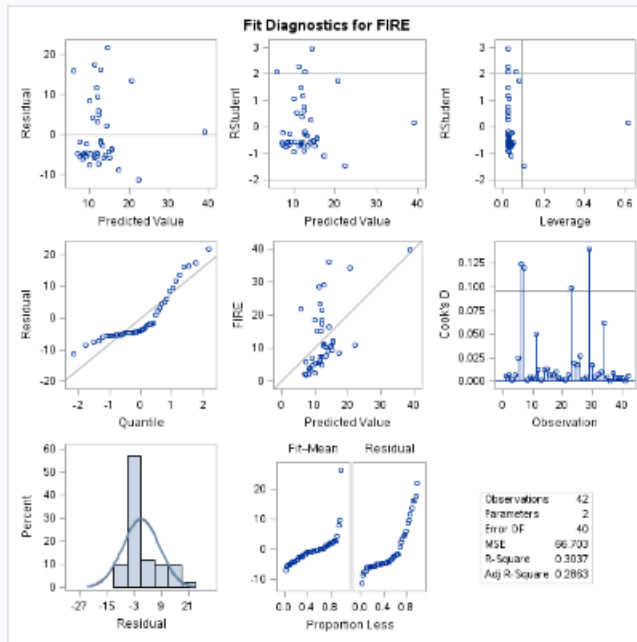
Number of Observations Read	42
Number of Observations Used	42

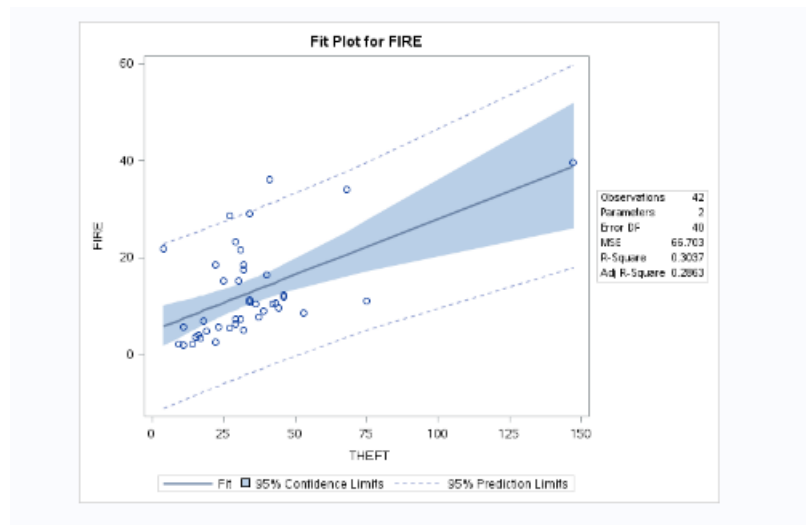
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1163.91979	1163.91979	17.45	0.0002
Error	40	2668.10807	66.70270		
Corrected Total	41	3832.02786			

Root MSE	8.16717	R-Square	0.3037
Dependent Mean	12.69286	Adj R-Sq	0.2863
Coeff Var	64.34463		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	4.90749	2.24983	2.18	0.0351
THEFT	1	0.23125	0.05536	4.18	0.0002

Exponential Smoothing: FIRE





b. Interpret the regression output and state whether the following indicate that the regression equation is reliable and should be used.

i. The p-value for the ANOVA table.

p value is 0.55

ii. R –Square value. Include an interpretation of what this number tells us.

R-Square	0.3037
-----------------	--------

Approximately 30.37% of the variability in fire can be explained by or attributed to variability in theft.

iii. The p-value for parameter (β_0 and β_1) estimates. Include a conclusion about the statistical significance of the linear regression equation.

$\beta_0 = 0.0351$, changes in the predictor is related to changes in the response since less than alpha 0.05.
more meaningful to the model

$\beta_1 = 0.0002$ changes in the predictor is related to changes in the response since less than alpha 0.05.
more meaningful to the model

1.

```
1
2 DATA FIRST;
3 FILE '/home/yeopdodo860/my_courses/session1.csv';
4 INPUT GROUP$ ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3;
5 SUM = ASSIGNMENT1 + ASSIGNMENT2 + ASSIGNMENT3;
6 PUT GROUP$ ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3 SUM;
7
8 DATALINES;
9 G 10 20 19
10 H 12 15 18
11 H 8 10 8
12 G 14 18 11
13 G 13 14 12
14 H 15 17 19
15 ;
```

2.

```
17 DATA SECOND;
18 FILE '/home/yeopdodo860/my_courses/session2.csv';
19 INPUT GROUP$ ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3;
20 SUM = ASSIGNMENT1 + ASSIGNMENT2 + ASSIGNMENT3;
21 PUT GROUP$ ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3 SUM;
22
23 DATALINES;
24 G 12 18 19
25 G 10 14
26 H 9 12
27 H 15 16 15
28 H 16 18
29 G 11 15
30
31 ;
32
```

```

41 DATA ONE;
42 INFILE '/home/yeopdodo860/my_courses/session1.csv' dsd dlm=' ';
43 INPUT GROUP$ ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3;
44 RUN;
45 DATA TWO;
46 INFILE '/home/yeopdodo860/my_courses/session2.csv' dsd dlm=' ' MISSOVER;
47 INPUT GROUP$ ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3;
48 RUN;
49
50 PROC PRINT DATA = ONE;
51 RUN;
52 PROC PRINT DATA = TWO;
53 RUN;

55 PROC SORT DATA=FIRST;
56 BY GROUP ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3 SUM;
57 RUN;
58
59 PROC SORT DATA=SECOND;
60 BY GROUP ASSIGNMENT1 ASSIGNMENT2 ASSIGNMENT3 SUM;
61 RUN;
62
63
64 DATA TRAINING;
65 MERGE FIRST SECOND;
66 BY SUM;
67 RUN;
68
69 PROC PRINT DATA = TRAINING;
70 RUN;
71
72 PROC FREQ DATA = TRAINING;
73 RUN;

```

Obs	GROUP	ASSIGNMENT1	ASSIGNMENT2	ASSIGNMENT3
1	G	10	20	19
2	H	12	15	18
3	H	8	10	8
4	G	14	18	11
5	G	13	14	12
6	H	15	17	19

Obs	GROUP	ASSIGNMENT1	ASSIGNMENT2	ASSIGNMENT3
1	G	12	18	19
2	G	10	14	.
3	H	15	16	15
4	H	16	18	.

Obs	GROUP	ASSIGNMENT1	ASSIGNMENT2	ASSIGNMENT3
1	G	10	14	.
2	G	12	18	19
3	G	14	18	11
4	H	15	16	15
5	H	16	18	.
6	H	15	17	19

The FREQ Procedure

GROUP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
G	3	50.00	3	50.00
H	3	50.00	6	100.00

ASSIGNMENT1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
10	1	16.67	1	16.67
12	1	16.67	2	33.33
14	1	16.67	3	50.00
15	2	33.33	5	83.33
16	1	16.67	6	100.00