7.7

a)

```
f=file.choose()
CommercialProp=read.table(f)
colnames(CommercialProp)=c('Y','X1','X2','X3','X4')
> PropmodX4=lm(Y~X4,data = CommercialProp)
> summary(PropmodX4)

Call:
lm(formula = Y ~ X4, data = CommercialProp)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.1390 -0.7930  0.2890  0.9653  3.4415 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.378e+01  2.903e-01  47.482  < 2e-16 ***
X4          8.437e-06  1.498e-06   5.632 2.63e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.462 on 79 degrees of freedom
Multiple R-squared:  0.2865,  Adjusted R-squared:  0.2775 
F-statistic: 31.72 on 1 and 79 DF,  p-value: 2.628e-07

> AnovaX4=anova(PropmodX4)
> AnovaX4
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
X4         1  67.775  67.775  31.723 2.628e-07 ***
Residuals 79 168.782   2.136                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> SSRX4=AnovaX4$`Sum Sq`[1]
> SSRX4
[1] 67.7751

> PropmodX1UX4=lm(Y~X1+X4, data = CommercialProp)
> summary(PropmodX1UX4)
```

```
Call:
lm(formula = Y ~ X1 + X4, data = CommercialProp)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2032 -0.4593  0.0641  0.7730  2.5083 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.436e+01  2.771e-01  51.831  < 2e-16 ***
X1          -1.145e-01  2.242e-02  -5.105 2.27e-06 ***
X4           1.045e-05  1.363e-06   7.663 4.23e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.274 on 78 degrees of freedom
Multiple R-squared:  0.4652,  Adjusted R-squared:  0.4515 
F-statistic: 33.93 on 2 and 78 DF,  p-value: 2.506e-11

> AnovaX1UX4=anova(PropmodX1UX4)
> AnovaX1UX4
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
X1         1  14.819  14.819  9.1365  0.003389 ** 
X4         1  95.231  95.231 58.7160 4.225e-11 ***
Residuals 78 126.508   1.622                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> SSRX1gX4=AnovaX4$`Sum Sq`[2]-AnovaX1UX4$`Sum Sq`[3]
> SSRX1gX4
[1] 42.27457
> PropmodX1UX2UX4=lm(Y~X1+X2+X4,data = CommercialProp)
> summary(PropmodX1UX2UX4)

Call:
lm(formula = Y ~ X1 + X2 + X4, data = CommercialProp)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0620 -0.6437 -0.1013  0.5672  2.9583 

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.237e+01  4.928e-01  25.100  < 2e-16 ***
X1          -1.442e-01  2.092e-02  -6.891 1.33e-09 ***
X2           2.672e-01  5.729e-02   4.663 1.29e-05 ***
X4           8.178e-06  1.305e-06   6.265 1.97e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.132 on 77 degrees of freedom
Multiple R-squared:  0.583,    Adjusted R-squared:  0.5667
F-statistic: 35.88 on 3 and 77 DF,  p-value: 1.295e-14

> AnovaX1UX2UX4=anova(PropmodX1UX2UX4)
> AnovaX1UX2UX4
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 14.819  14.819  11.566  0.001067 **
X2         1 72.802  72.802  56.825 7.841e-11 ***
X4         1 50.287  50.287  39.251 1.973e-08 ***
Residuals 77 98.650   1.281
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> SSRX2gX1X4=AnovaX1UX4$`Sum Sq`[3]-AnovaX1UX2UX4$`Sum Sq`[4]
```

> SSRX2gX1X4

[1] 27.85749

```
> PropmodX1UX2UX3UX4=lm(Y~X1+X2+X3+X4,data = CommercialProp)
> summary(PropmodX1UX2UX3UX4)

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4, data = CommercialProp)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1872 -0.5911 -0.0910  0.5579  2.9441

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.220e+01  5.780e-01  21.110  < 2e-16 ***
X1          -1.420e-01  2.134e-02  -6.655 3.89e-09 ***
X2           2.820e-01  6.317e-02   4.464 2.75e-05 ***
X3           6.193e-01  1.087e+00   0.570     0.57
X4           7.924e-06  1.385e-06   5.722 1.98e-07 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.137 on 76 degrees of freedom
Multiple R-squared:  0.5847,  Adjusted R-squared:  0.5629
F-statistic: 26.76 on 4 and 76 DF,  p-value: 7.272e-14

> AnovaX1UX2UX3UX4=anova(PropmodX1UX2UX3UX4)
> AnovaX1UX2UX3UX4
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 14.819  14.819 11.4649  0.001125 **
X2         1 72.802  72.802 56.3262 9.699e-11 ***
X3         1  8.381   8.381  6.4846  0.012904 *
X4         1 42.325  42.325 32.7464 1.976e-07 ***
Residuals 76 98.231   1.293
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> SSRX3gX1X2X4=AnovaX1UX2UX4$`Sum Sq`[4]-AnovaX1UX2UX3UX4$`Sum Sq`[5]
```

> SSRX3gX1X2X4

[1] 0.4197463

b)

H0:β3=0

H1:β3$\neq$0

```
> SSE=AnovaX1UX2UX3UX4$`Sum Sq`[5]
> Fstar=(SSRX3gX1X2X4/1)/(SSE/76)
> Fstar
[1] 0.3247534
> F=qf(.99,1,76)
> F
[1] 6.980578
> pvalue=AnovaX1UX2UX3UX4$`Pr(>F)`[4]
> pvalue
[1] 1.97599e-07
```

Since Fstar is less than F, we fail to reject the null.

7.8

H0:β1=β4=0

H1:not both β1 and β4 equal 0

```
> SSE=AnovaX1UX2UX3UX4$`Sum Sq`[5]
> SSRX2X3gX1X4=AnovaX1UX4$`Sum Sq`[3]-SSE
```

```
> SSRX2X3gX1X4
[1] 28.27724
> Fstar=(SSRX2X3gX1X4/2)/(SSE/76)
> Fstar
[1] 10.9389
> F=qf(.99,2,76)
> F
[1] 4.89584
```

Since Fstar is greater than F, we reject the null.

7.10

H0:β1=-.1 and β2=.4

H1: Equalities don't hold

Yi +.1*X1-.4*X2=β0+β3*X3+β4*X4

```
>SSE=AnovaX1UX2UX3UX4$`Sum Sq`[5]
[1] 98.2306
>SSERed=AnovaRed$`Sum Sq`[3]
[1] 110.141
>F_star=((SSERed-SSE)/2)/(SSE/76)
>F_star
[1] 4.6075
>F=qf(.99,2,76)
>F
[1] 4.89584
```

Since F_star is less than F, we fail to reject the null.

7.27

a)

```
> PropmodX1UX4=lm(Y~X1+X4, data = CommercialProp)
> summary(PropmodX1UX4)

Call:
lm(formula = Y ~ X1 + X4, data = CommercialProp)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2032 -0.4593  0.0641  0.7730  2.5083
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.436e+01  2.771e-01  51.831  < 2e-16 ***
X1          -1.145e-01  2.242e-02  -5.105 2.27e-06 ***
X4           1.045e-05  1.363e-06   7.663 4.23e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.274 on 78 degrees of freedom
Multiple R-squared:  0.4652,  Adjusted R-squared:  0.4515
F-statistic: 33.93 on 2 and 78 DF,  p-value: 2.506e-11
```

Yhat=14.36-.1145X1+.1045X4

b)

Compared to the coefficients in problem 6.18c the estimated regression coefficients are larger in the first order linear regression model.

c)

```
> PropmodX3=lm(Y~X3,data = CommercialProp)
> AnovaX3=anova(PropmodX3)
> AnovaX3
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value Pr(>F)
X3         1   1.047  1.0470  0.3512 0.5551
Residuals 79 235.511  2.9811
> SSRX4gX3=AnovaX3$`Sum Sq`[2]-AnovaX3UX4$`Sum Sq`[3]
> PropmodX3=lm(Y~X3,data = CommercialProp)
> AnovaX3=anova(PropmodX3)
> AnovaX3
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value Pr(>F)
X3         1   1.047  1.0470  0.3512 0.5551
Residuals 79 235.511  2.9811
> PropmodX3UX4=lm(Y~X3+X4,data = CommercialProp)
> AnovaX3UX4=anova(PropmodX3UX4)
> AnovaX3UX4
Analysis of Variance Table
```

```
Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X3         1   1.047   1.047  0.4842    0.4886
X4         1  66.858  66.858 30.9213 3.626e-07 ***
Residuals 78 168.652   2.162
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> SSRX4gX3=AnovaX3$`Sum Sq`[2]-AnovaX3UX4$`Sum Sq`[3]
> SSRX4gX3
[1] 66.85829
> SSRX4=AnovaX4$`Sum Sq`[1]
> SSRX4
[1] 67.7751

> PropmodX1=lm(Y~X1,data = CommercialProp)
> AnovaX1=anova(PropmodX1)
> AnovaX1
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value  Pr(>F)
X1         1  14.819 14.8185  5.2795 0.02422 *
Residuals 79 221.739  2.8068
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> PropmodX1UX3=lm(Y~X1+X3,data = CommercialProp)
> AnovaX1UX3=anova(PropmodX1UX3)
> AnovaX1UX3
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value  Pr(>F)
X1         1  14.819 14.8185  5.2127 0.02515 *
X3         1   0.003  0.0027  0.0010 0.97534
Residuals 78 221.736  2.8428
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> SSRX1gX3=AnovaX3$`Sum Sq`[2]-AnovaX1UX3$`Sum Sq`[3]
> SSRX1gX3
[1] 13.7743
> SSRX1=AnovaX1$`Sum Sq`[1]
> SSRX1
[1] 14.81852
```

No, neither of the SSRs are equal to each other

d)

While they do not exactly equally each other they are relatively close, compared to variables that are highly correlated with each other. Taking a look at the correlation matrix from 6.18b, one can see that these variables are not highly correlated so the SSRs being relatively close makes sense in this circumstance.

8.3

a)

The criticism is justified and there is a possibility that the model is overfitted and is focusing too much on a certain part of the data rather than the overall trend. Something to note is that high-order polynomial models can be prone to overfitting which is where his/her concern can come from. Regardless the R^2 being .991 does not necessarily mean it is a good model, rather it is does not directly assess the ability to predict new data, if the predictions require too many adjustments to line up then it is not a good model.

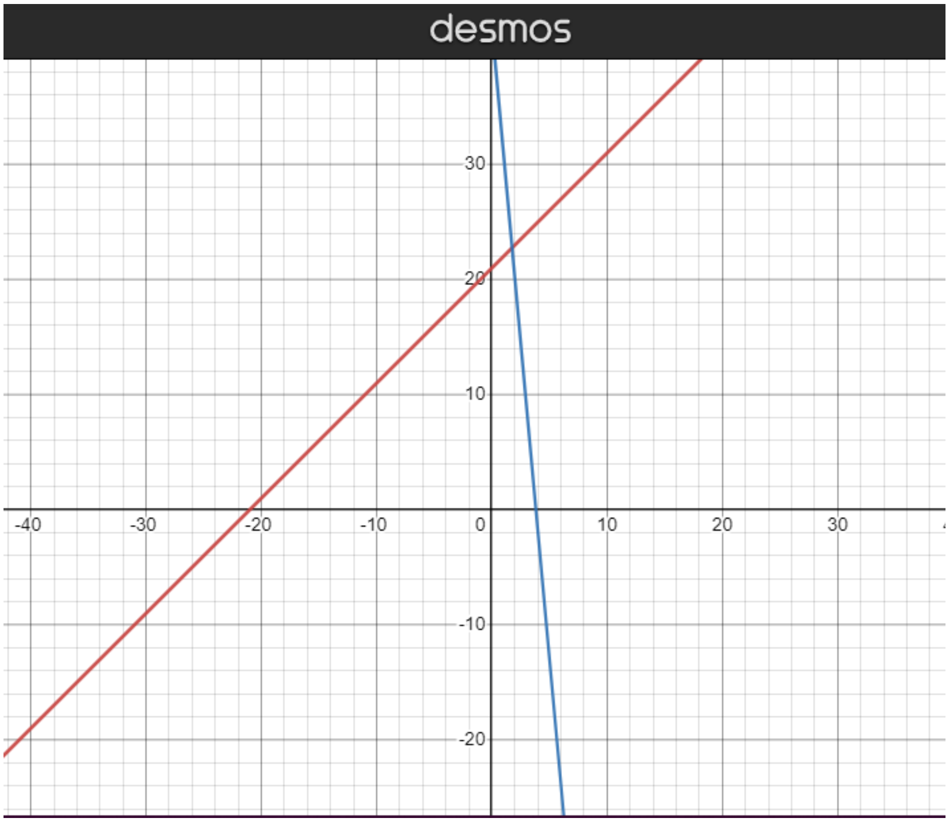
b)

The adjusted coefficient of multiple determination would be more appropriate as it would adjust for the number ot observations and parameters.Without the adjustment, adding more variables would inevitably increase R^2.

8.10

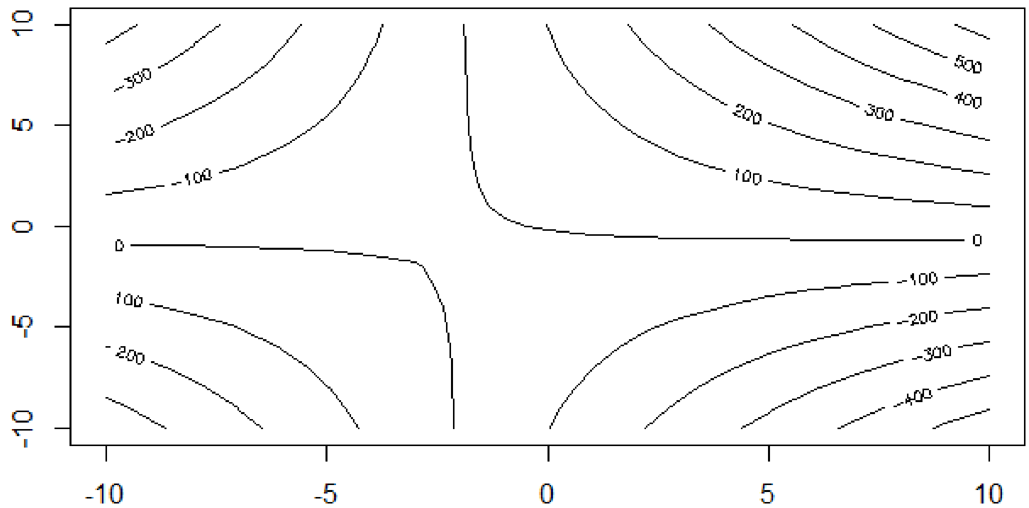
a)

When X1=1, E{Y}=21+X2 and when X1=4, E{Y}=42-11X2



b)

```
X1=-10:10
X2=-10:10
y<-matrix(nrow=length(X1),ncol=length(X2))
for(i in 1:length(X1))
{ for(j in 1:length(X2))
{ y[i,j]=2+4*X1[i]+10*X2[j]+5*X1[i]*X2[j] } }
contour(X1,X2,y)
```



8.12
A source of the difficulty could be not enough information for the model.
8.14
The coding scheme doesn't indicate that males have longer learning time, but that there is a positive relationship between the predictor and the outcome variables. Through the standard error coefficient,s(b2), it can be tested whether the relationship is significant or not.
8.16
a)
1 would be the slope, which would be the same for both types of students, major declared or not. 2 would indicate how much an effect having declared a major would implicate compared to not having declared a major.
b)

```
f=file.choose()
GPA=read.table(f)
colnames(GPA)=c('Y','X1')
f=file.choose()
Major=read.table(f)
colnames(Major)='X2'
GPA=cbind(GPA,Major)
GpaMod=lm(Y~X1+X2,data = GPA)
summary(GpaMod)
Call:
lm(formula = Y ~ X1 + X2, data = GPA)
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.70304 -0.35574  0.02541  0.45747  1.25037 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.19842    0.33886   6.488 2.18e-09 ***
X1           0.03789    0.01285   2.949  0.00385 ** 
X2          -0.09430    0.11997  -0.786  0.43341    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6241 on 117 degrees of freedom
Multiple R-squared:  0.07749,	Adjusted R-squared:  0.06172 
F-statistic: 4.914 on 2 and 117 DF,  p-value: 0.008928
```

Yhat=2.19842+.03789X1-.09430X2

c)

H0:**β**2=0

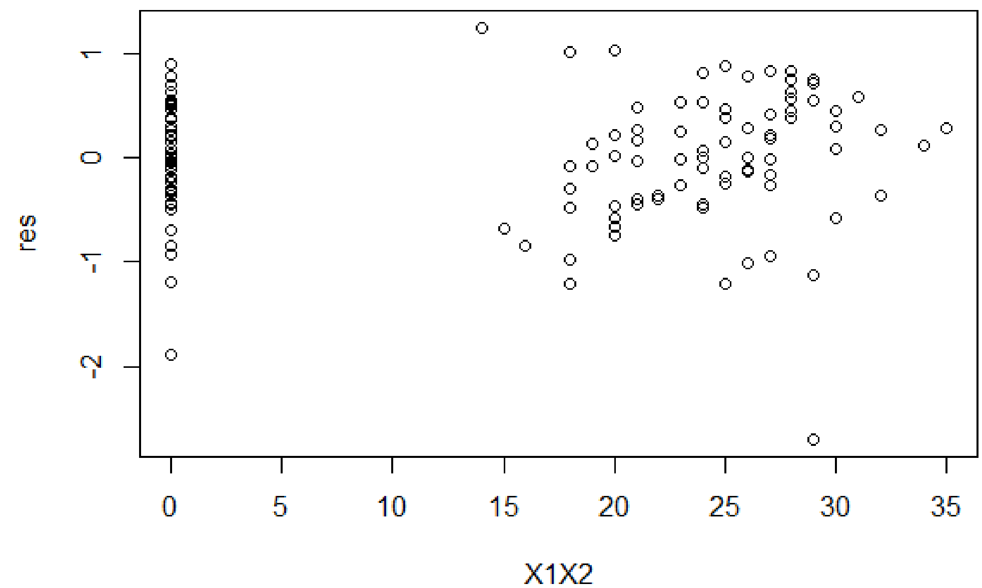
H1:**β**2=0

```
GPAanovaX1UX2=anova(GpaMod)
anova(GpaMod)
SSE=GPAanovaX1UX2$`Sum Sq`[3]
GpaModX1=lm(Y~X1,data = GPA)
summary(GpaModX1)
GPAanovaX1=anova(GpaModX1)
GPAanovaX1
SSRX1gX2=GPAanovaX1$`Sum Sq`[2]-SSE
SSRX1gX2
Fstar=(SSRX1gX2/1)/(SSE/117)
Fstar
[1] 0.6179314
F=qf(.99,1,117)
F
[1] 6.856564
```

Since Fstar is less tha F, we fail to reject the null

d)

```
res=residuals(GpaMod)
X1X2=GPA$X1*GPA$X2
plot(X1X2,res)
```



No, it seems to be fairly symmetrical.

8.20

a)

```
GPAMod=lm(Y~X1+X2+X1*X2,data = GPA)
> summary(GPAMod)

Call:
lm(formula = Y ~ X1 + X2 + X1 * X2, data = GPA)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.80187 -0.31392  0.04451  0.44337  1.47544 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.226318   0.549428   5.872 4.18e-08 ***
X1          -0.002757   0.021405  -0.129   0.8977    
X2          -1.649577   0.672197  -2.454   0.0156 *  
X1:X2        0.062245   0.026487   2.350   0.0205 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6124 on 116 degrees of freedom
```

Multiple R-squared: 0.1194,	Adjusted R-squared: 0.09664
F-statistic: 5.244 on 3 and 116 DF, p-value: 0.001982

Yhat=3.22618-.002757X1-1.649577X2+.062245X1*X2

b)
H0:β3=0
H1:β3=0

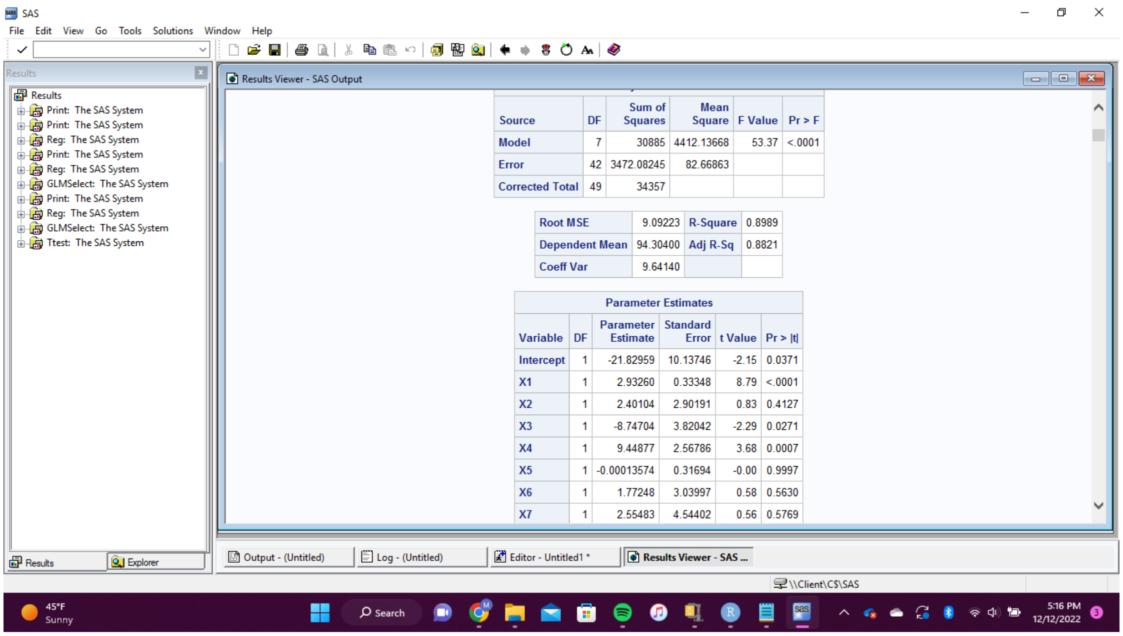
```
> AnovaX1X2=anova(GPAMod)
> AnovaX1X2
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value   Pr(>F)
X1         1  3.588  3.5878  9.5663 0.002483 **
X2         1  0.241  0.2407  0.6418 0.424691
X1:X2      1  2.071  2.0713  5.5226 0.020461 *
Residuals 116 43.506  0.3750
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> SSE=AnovaX1X2$`Sum Sq`[4]
> SSE
[1] 43.50564
> SSRX1X2gX1X2=GPAanovaX1UX2$`Sum Sq`[3]- AnovaX1X2$`Sum Sq`[4]
> SSRX1X2gX1X2
[1] 2.071257
> Fstar=(SSRX1X2gX1X2/1)/(SSE/117)
> Fstar
[1] 5.570244
> F=qf(.95,1,117)
> F
[1] 3.922173
```

Since Fstar is greater than F, we reject the null.

1)

```
data homes;
infile 'Homes1.txt';
input Y X1 X2 X3 X4 X5 X6 X7;
run;
proc print data=homes;
run;
PROC REG data=homes;
   model Y = X1 X2 X3 X4 X5 X6 X7;
```

```
run;
PROC GLMSELECT data=homes;
   model Y = X1 X2 X3 X4 X5 X6 X7/selection=stepwise;
run;
```



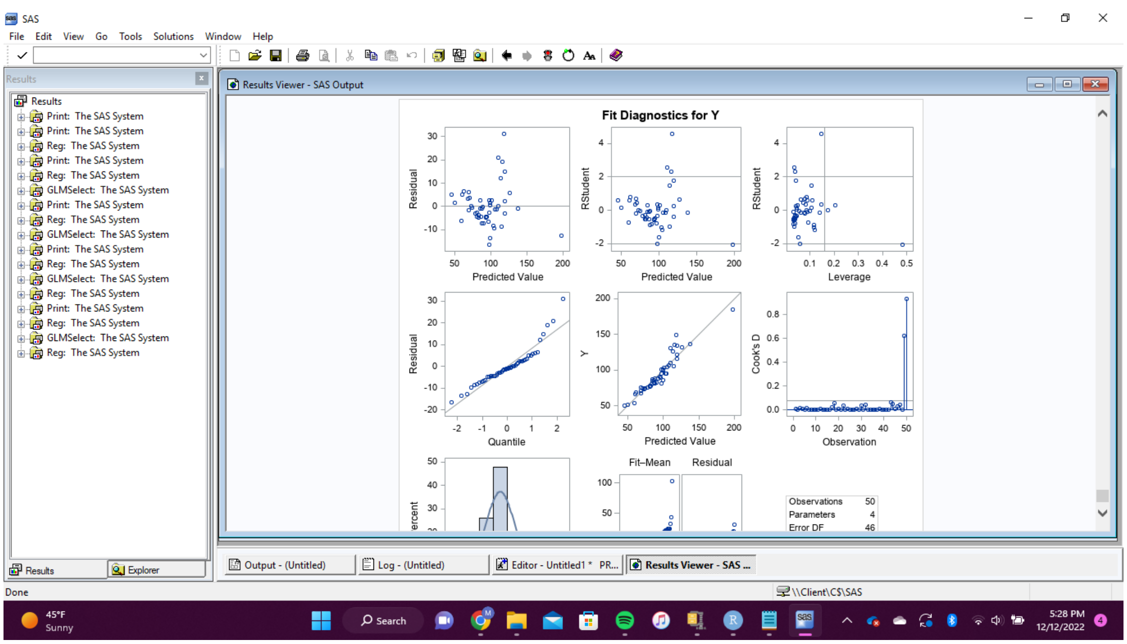
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	30885	4412.13668	53.37	<.0001
Error	42	3472.08245	82.66863		
Corrected Total	49	34357			

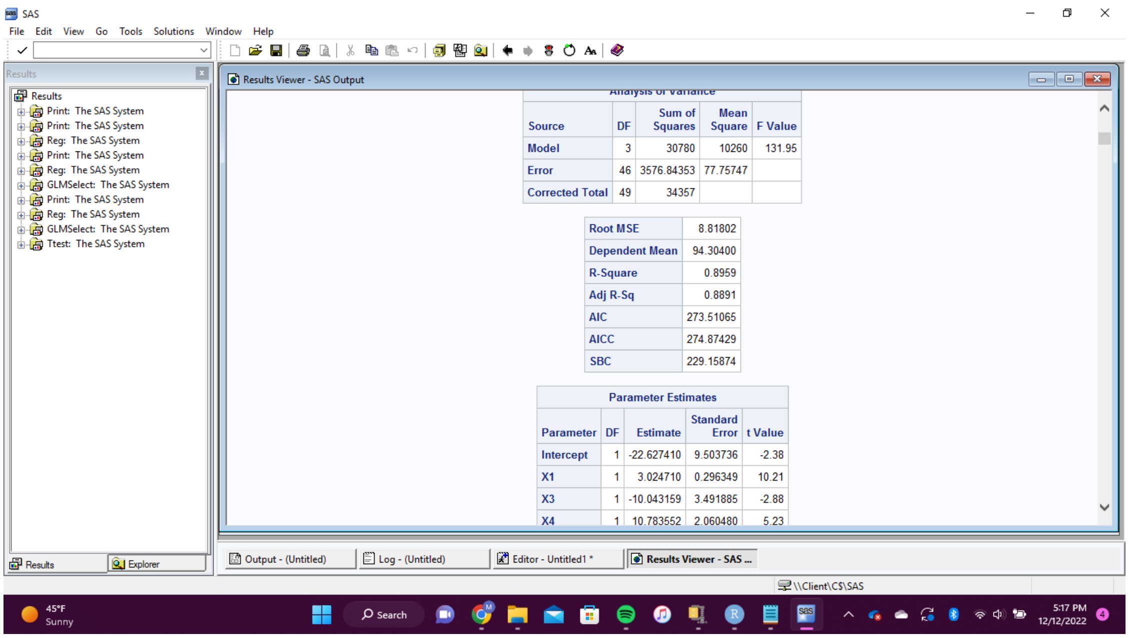
Root MSE	9.09223	R-Square	0.8989
Dependent Mean	94.30400	Adj R-Sq	0.8821
Coeff Var	9.64140		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-21.82959	10.13746	-2.15	0.0371
X1	1	2.93260	0.33348	8.79	<.0001
X2	1	2.40104	2.90191	0.83	0.4127
X3	1	-8.74704	3.82042	-2.29	0.0271
X4	1	9.44877	2.56786	3.68	0.0007
X5	1	-0.00013574	0.31694	-0.00	0.9997
X6	1	1.77248	3.03997	0.58	0.5630
X7	1	2.55483	4.54402	0.56	0.5769

,





Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	3	30780	10260	131.95
Error	46	3576.84353	77.75747	
Corrected Total	49	34357		

Root MSE	8.81802
Dependent Mean	94.30400
R-Square	0.8959
Adj R-Sq	0.8891
AIC	273.51065
AICC	274.87429
SBC	229.15874

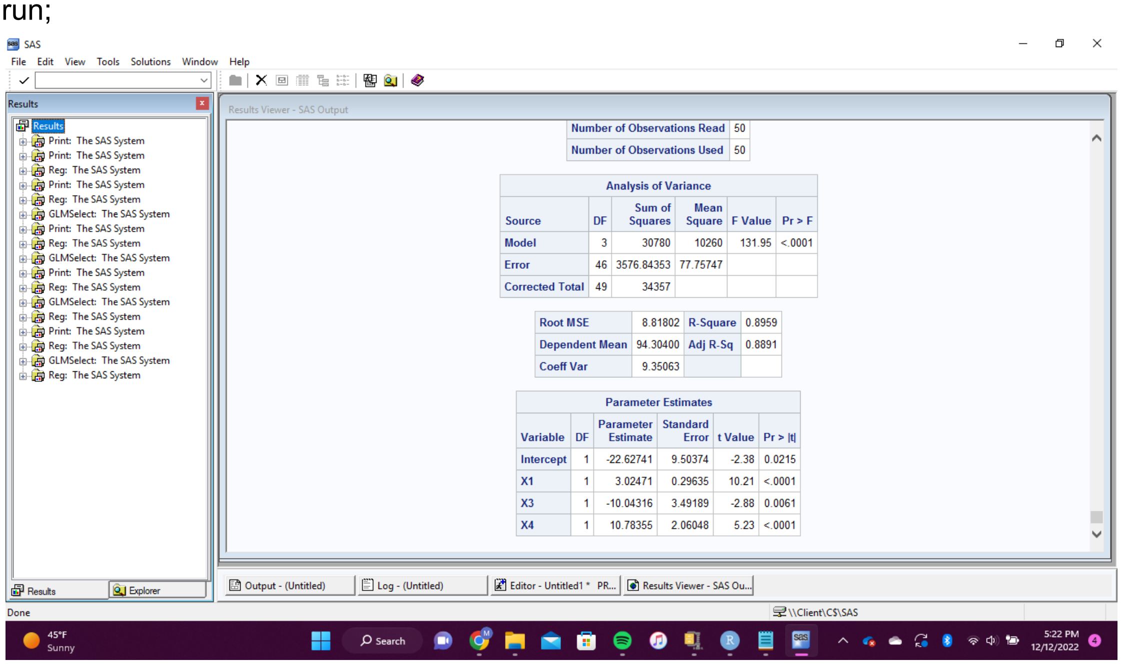
Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	-22.627410	9.503736	-2.38
X1	1	3.024710	0.296349	10.21
X3	1	-10.043159	3.491885	-2.88
X4	1	10.783552	2.060480	5.23

X1,X3, and X4 are the most important variables

2)

```
PROC REG data=homes;
        model Y = X1 X3 X4;
run;
```



Number of Observations Read	50
Number of Observations Used	50

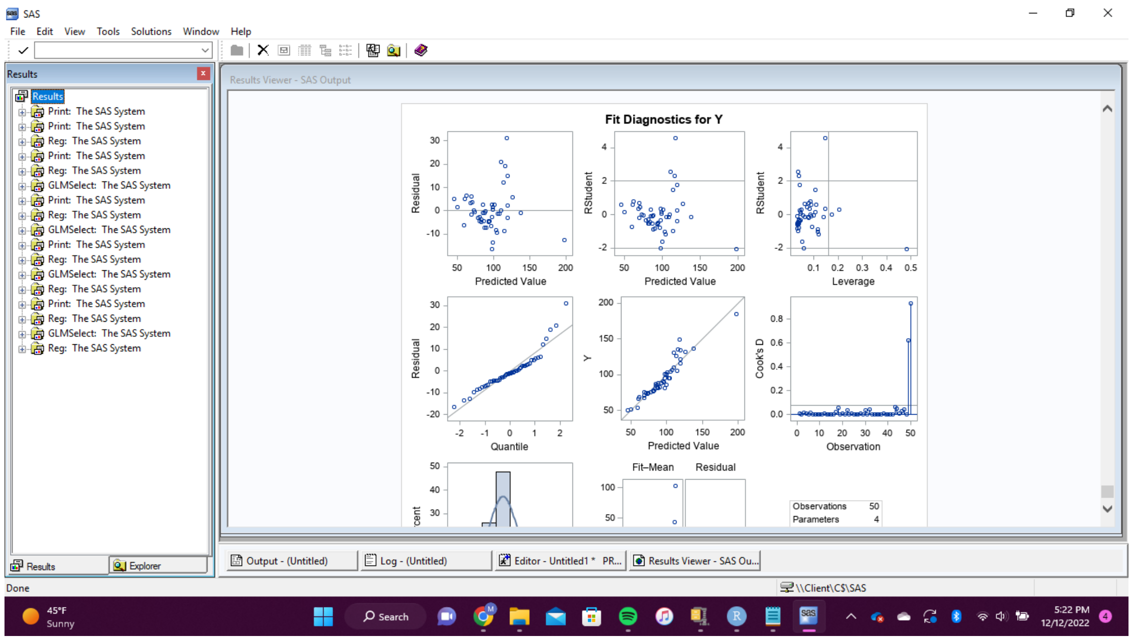
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	30780	10260	131.95	<.0001
Error	46	3576.84353	77.75747		
Corrected Total	49	34357			

Root MSE	8.81802	R-Square	0.8959
Dependent Mean	94.30400	Adj R-Sq	0.8891
Coeff Var	9.35063		

Parameter Estimates

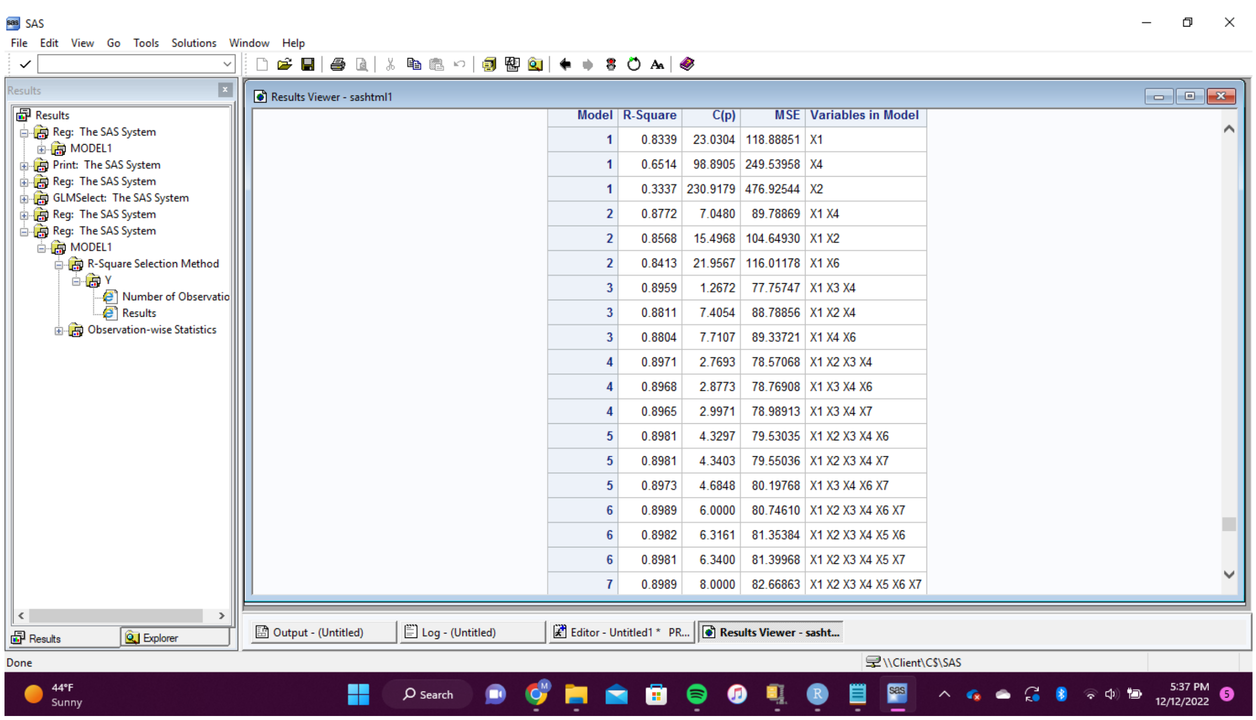
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-22.62741	9.50374	-2.38	0.0215
X1	1	3.02471	0.29635	10.21	<.0001
X3	1	-10.04316	3.49189	-2.88	0.0061
X4	1	10.78355	2.06048	5.23	<.0001



3)

a)

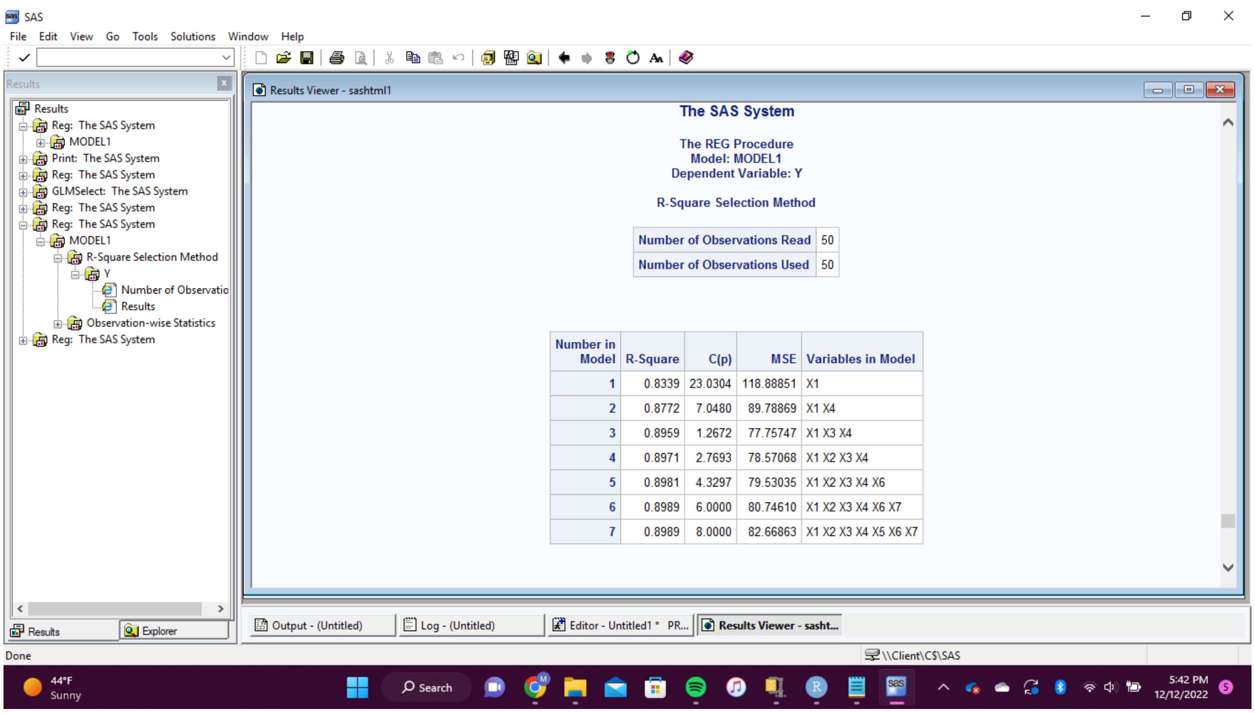
```
PROC REG data=homes;
        model Y = X1 X2 X3 X4 X5 X6 X7/selection=RSQUARE CP MSE best=3;
run;
```



Model	R-Square	C(p)	MSE	Variables in Model
1	0.8339	23.0304	118.88851	X1
1	0.6514	98.8905	249.53958	X4
1	0.3337	230.9179	476.92544	X2
2	0.8772	7.0480	89.78869	X1 X4
2	0.8568	15.4968	104.64930	X1 X2
2	0.8413	21.9567	116.01178	X1 X6
3	0.8959	1.2672	77.75747	X1 X3 X4
3	0.8811	7.4054	88.78856	X1 X2 X4
3	0.8804	7.7107	89.33721	X1 X4 X6
4	0.8971	2.7693	78.57068	X1 X2 X3 X4
4	0.8968	2.8773	78.76908	X1 X3 X4 X6
4	0.8965	2.9971	78.98913	X1 X3 X4 X7
5	0.8981	4.3297	79.53035	X1 X2 X3 X4 X6
5	0.8981	4.3403	79.55036	X1 X2 X3 X4 X7
5	0.8973	4.6848	80.19768	X1 X3 X4 X6 X7
6	0.8989	6.0000	80.74610	X1 X2 X3 X4 X6 X7
6	0.8982	6.3161	81.35384	X1 X2 X3 X4 X5 X6
6	0.8981	6.3400	81.39968	X1 X2 X3 X4 X5 X7
7	0.8989	8.0000	82.66863	X1 X2 X3 X4 X5 X6 X7

Based off this, it agrees with model X1,X3,X4 being the best model b)

```
PROC REG data=homes;
	model Y = X1 X2 X3 X4 X5 X6 X7/selection=RSQUARE CP MSE best=1;
run;
```



The SAS System

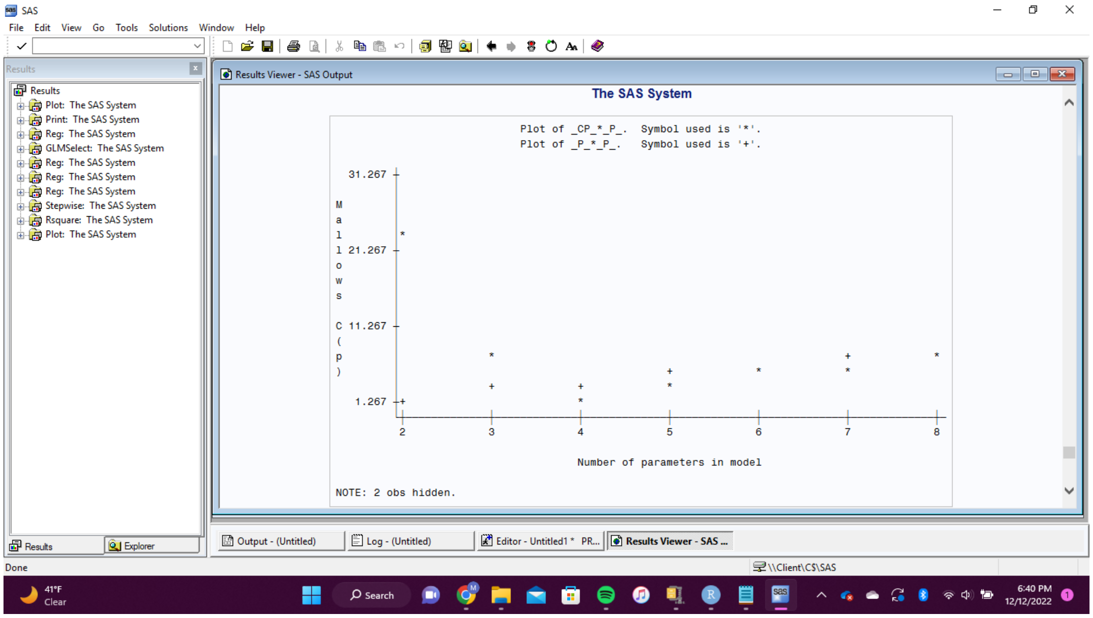
The REG Procedure
Model: MODEL1
Dependent Variable: Y

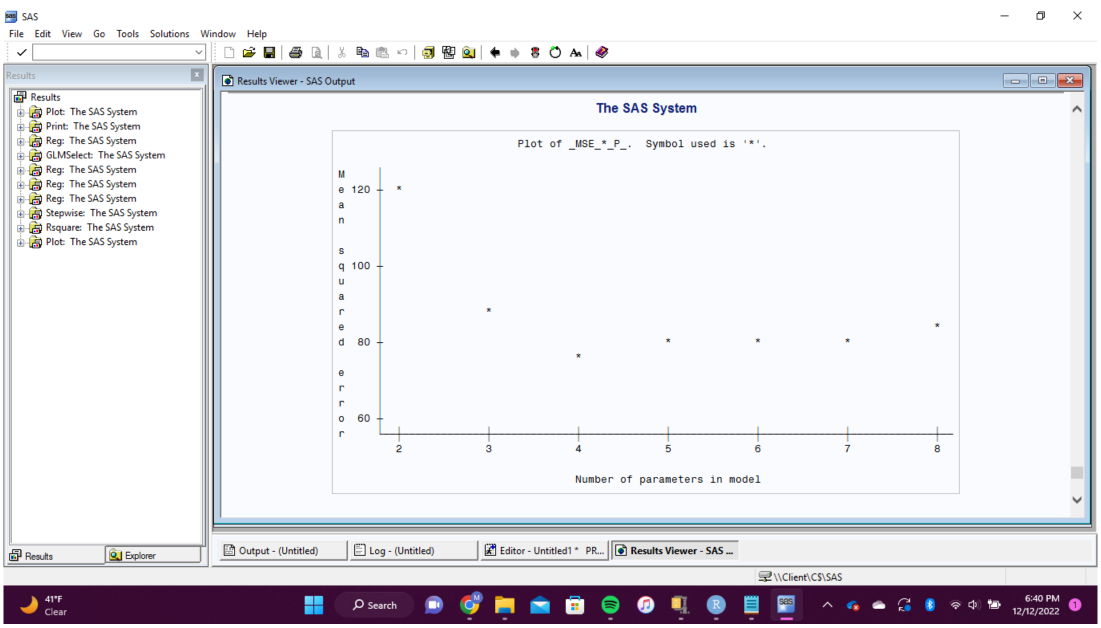
R-Square Selection Method

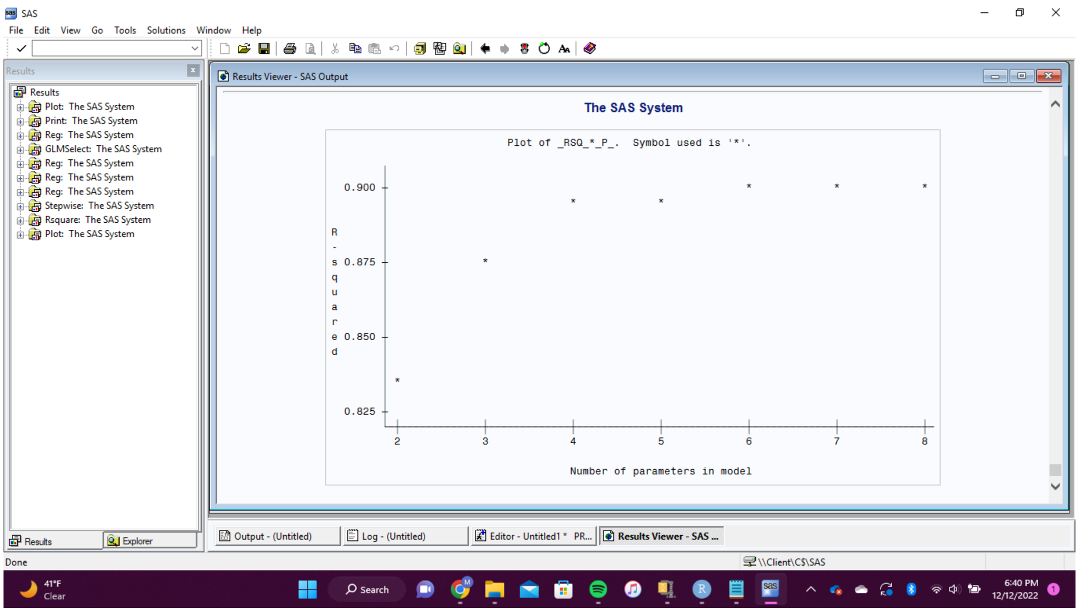
Number of Observations Read	50
Number of Observations Used	50

Number in Model	R-Square	C(p)	MSE	Variables in Model
1	0.8339	23.0304	118.88851	X1
2	0.8772	7.0480	89.78869	X1 X4
3	0.8959	1.2672	77.75747	X1 X3 X4
4	0.8971	2.7693	78.57068	X1 X2 X3 X4
5	0.8981	4.3297	79.53035	X1 X2 X3 X4 X6
6	0.8989	6.0000	80.74610	X1 X2 X3 X4 X6 X7
7	0.8989	8.0000	82.66863	X1 X2 X3 X4 X5 X6 X7

```
PROC PLOT;
PLOT _CP_ * _P_='*'  _P_*_P_='+'/overlay;
PLOT _MSE_ * _P_='*';
PLOT _RSQ_*_P_='*';
run;
```

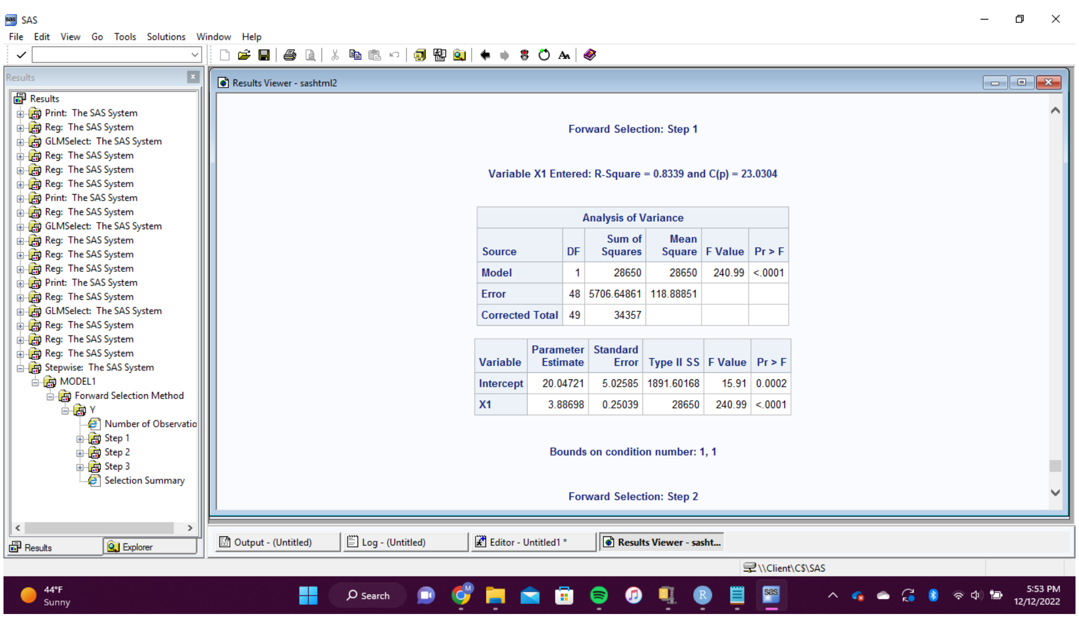

Results Viewer - SAS Output
The SAS System
Plot of _CP_*_P_. Symbol used is '*'.
Plot of _P_*_P_. Symbol used is '+'.
31.267
21.267
11.267
1.267
Mallows C(p)
2 3 4 5 6 7 8
Number of parameters in model
NOTE: 2 obs hidden.
Results
Plot: The SAS System
Print: The SAS System
Reg: The SAS System
GLMSelect: The SAS System
Reg: The SAS System
Reg: The SAS System
Reg: The SAS System
Stepwise: The SAS System
Rsquare: The SAS System
Plot: The SAS System
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1 * PR...
Results Viewer - SAS ...
Done
\\Client\C$\SAS
6:40 PM
12/12/2022


Results Viewer - SAS Output
The SAS System
Plot of _MSE_*_P_. Symbol used is '*'.
120
100
80
60
Mean squared error
2 3 4 5 6 7 8
Number of parameters in model
Results
Plot: The SAS System
Print: The SAS System
Reg: The SAS System
GLMSelect: The SAS System
Reg: The SAS System
Reg: The SAS System
Reg: The SAS System
Stepwise: The SAS System
Rsquare: The SAS System
Plot: The SAS System
Output - (Untitled)
Log - (Untitled)
Editor - Untitled1 * PR...
Results Viewer - SAS ...
\\Client\C$\SAS
6:40 PM
12/12/2022



4)

```
PROC STEPWISE data=homes;
    model Y = X1 X2 X3 X4 X5 X6 X7/FORWARD SLENTRY=.1;
run;
```



Forward Selection: Step 1

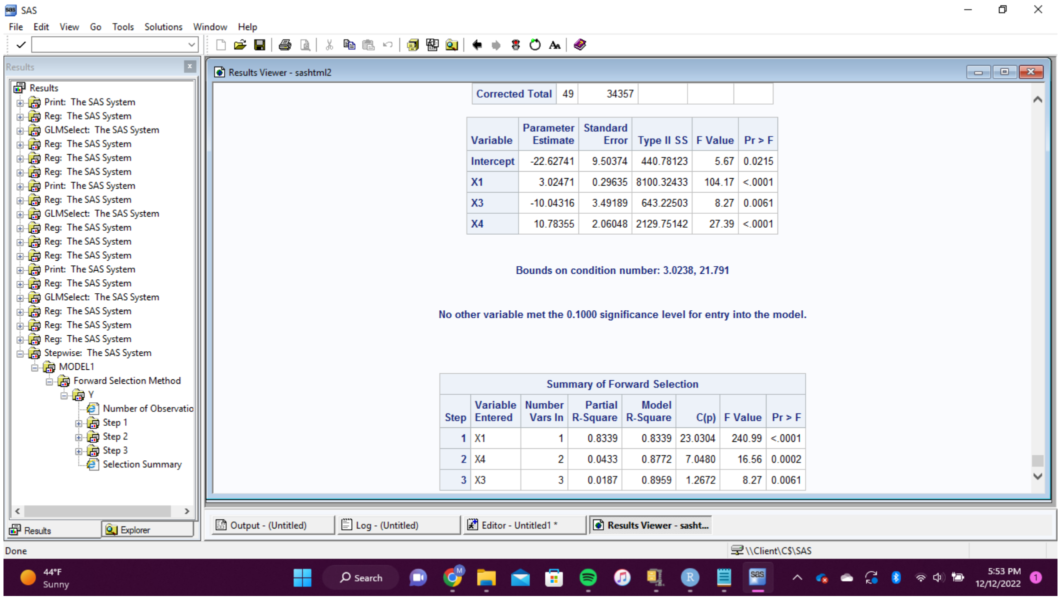
Variable X1 Entered: R-Square = 0.8339 and C(p) = 23.0304

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	28650	28650	240.99	<.0001
Error	48	5706.64861	118.88851		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	20.04721	5.02585	1891.60168	15.91	0.0002
X1	3.88698	0.25039	28650	240.99	<.0001

Bounds on condition number: 1, 1

Forward Selection: Step 2



Corrected Total	49	34357			

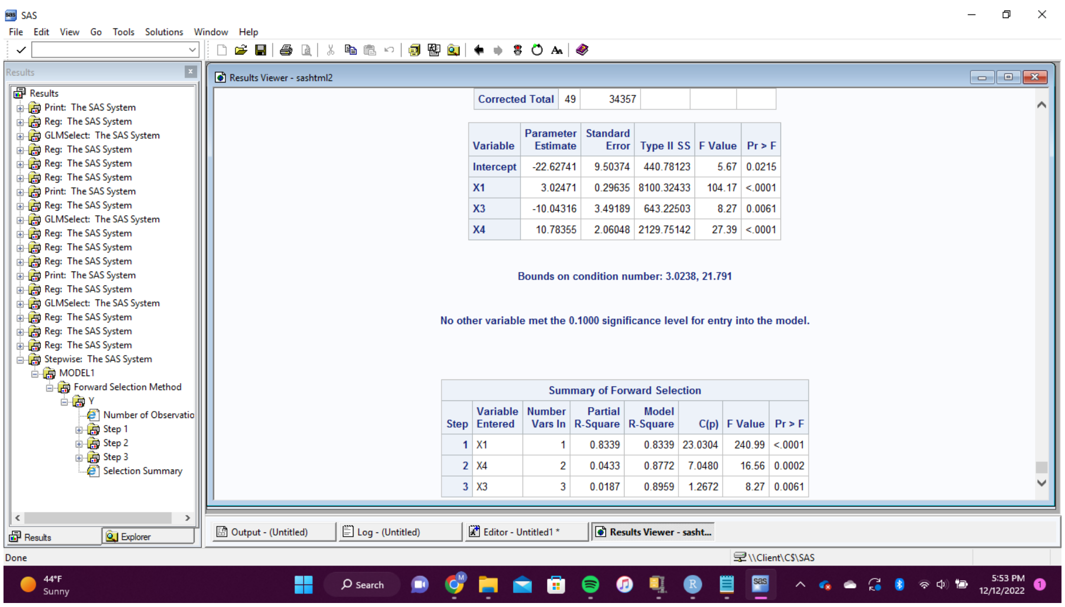
Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-22.62741	9.50374	440.78123	5.67	0.0215
X1	3.02471	0.29635	8100.32433	104.17	<.0001
X3	-10.04316	3.49189	643.22503	8.27	0.0061
X4	10.78355	2.06048	2129.75142	27.39	<.0001

Bounds on condition number: 3.0238, 21.791

No other variable met the 0.1000 significance level for entry into the model.

Summary of Forward Selection

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	X1	1	0.8339	0.8339	23.0304	240.99	<.0001
2	X4	2	0.0433	0.8772	7.0480	16.56	0.0002
3	X3	3	0.0187	0.8959	1.2672	8.27	0.0061



Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-22.62741	9.50374	440.78123	5.67	0.0215
X1	3.02471	0.29635	8100.32433	104.17	<.0001
X3	-10.04316	3.49189	643.22503	8.27	0.0061
X4	10.78355	2.06048	2129.75142	27.39	<.0001

Bounds on condition number: 3.0238, 21.791

No other variable met the 0.1000 significance level for entry into the model.

Summary of Forward Selection

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	X1	1	0.8339	0.8339	23.0304	240.99	<.0001
2	X4	2	0.0433	0.8772	7.0480	16.56	0.0002
3	X3	3	0.0187	0.8959	1.2672	8.27	0.0061


```
PROC STEPWISE data=homes;
        model Y = X1 X2 X3 X4 X5 X6 X7/BACKWARD SLSTAYY=.1;
run;
```

Backward Elimination: Step 1

Variable X5 Removed: R-Square = 0.8989 and C(p) = 6.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	30885	5147.49279	63.75	<.0001
Error	43	3472.08246	80.74610		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-21.82925	9.98721	385.75435	4.78	0.0343
X1	2.93256	0.31677	6920.45385	85.71	<.0001
X2	2.40107	2.86746	56.61541	0.70	0.4070
X3	-8.74682	3.74089	441.43997	5.47	0.0241
X4	9.44854	2.48030	1171.76701	14.51	0.0004
X6	1.77252	3.00290	28.13341	0.35	0.5581
X7	2.55523	4.39829	27.25285	0.34	0.5643

Backward Elimination: Step 2

Variable X7 Removed: R-Square = 0.8981 and C(p) = 4.3297

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	30858	6171.54078	77.60	<.0001
Error	44	3499.33531	79.53035		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-21.60587	9.90439	378.46096	4.76	0.0345
X1	2.94910	0.31310	7055.80166	88.72	<.0001
X2	2.11530	2.80361	45.27319	0.57	0.4546
X3	-8.98821	3.68965	471.96365	5.93	0.0190
X4	9.57008	2.45279	1210.72162	15.22	0.0003
X6	1.99780	2.95525	36.34519	0.46	0.5026

Bounds on condition number: 4.1893, 59.674

Backward Elimination: Step 3

Variable X6 Removed: R-Square = 0.8971 and C(p) = 2.7693

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	30821	7705.33968	98.07	<.0001
Error	45	3535.68050	78.57068		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-23.14675	9.58021	458.65973	5.84	0.0198
X1	3.00840	0.29875	7967.57341	101.41	<.0001
X2	2.01412	2.78267	41.16303	0.52	0.4729
X3	-9.49129	3.59195	548.59349	6.98	0.0113
X4	9.94029	2.37640	1374.73442	17.50	0.0001

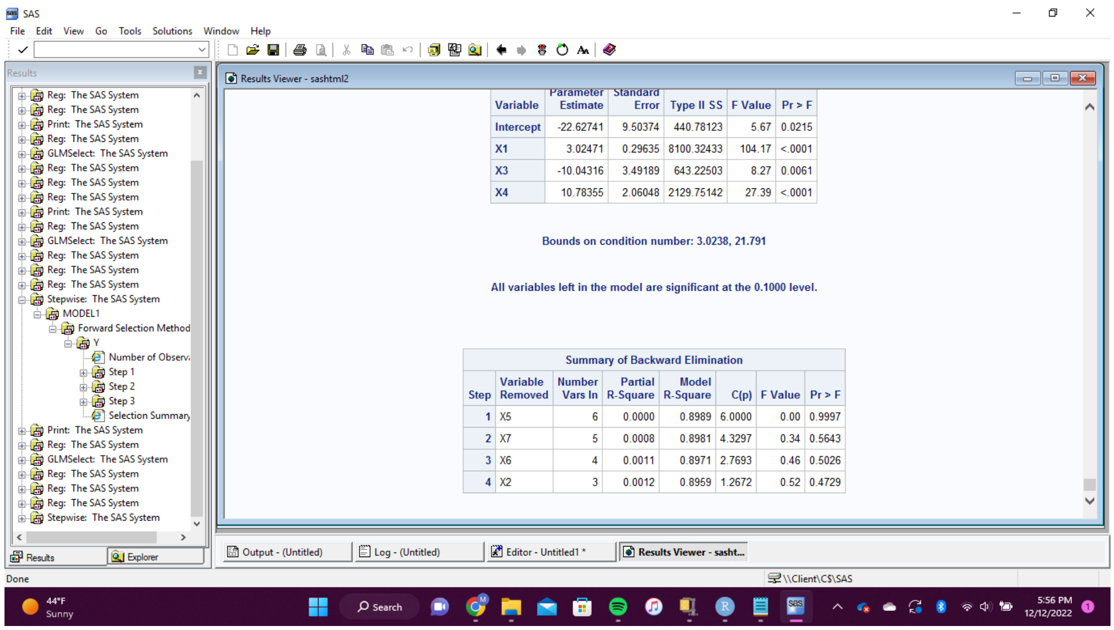
Backward Elimination: Step 4

Variable X2 Removed: R-Square = 0.8959 and C(p) = 1.2672

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	30780	10260	131.95	<.0001
Error	46	3576.84353	77.75747		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-22.62741	9.50374	440.78123	5.67	0.0215
X1	3.02471	0.29635	8100.32433	104.17	<.0001
X3	-10.04316	3.49189	643.22503	8.27	0.0061
X4	10.78355	2.06048	2129.75142	27.39	<.0001

Bounds on condition number: 3.0238, 21.791



Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-22.62741	9.50374	440.78123	5.67	0.0215
X1	3.02471	0.29635	8100.32433	104.17	<.0001
X3	-10.04316	3.49189	643.22503	8.27	0.0061
X4	10.78355	2.06048	2129.75142	27.39	<.0001

Bounds on condition number: 3.0238, 21.791

All variables left in the model are significant at the 0.1000 level.

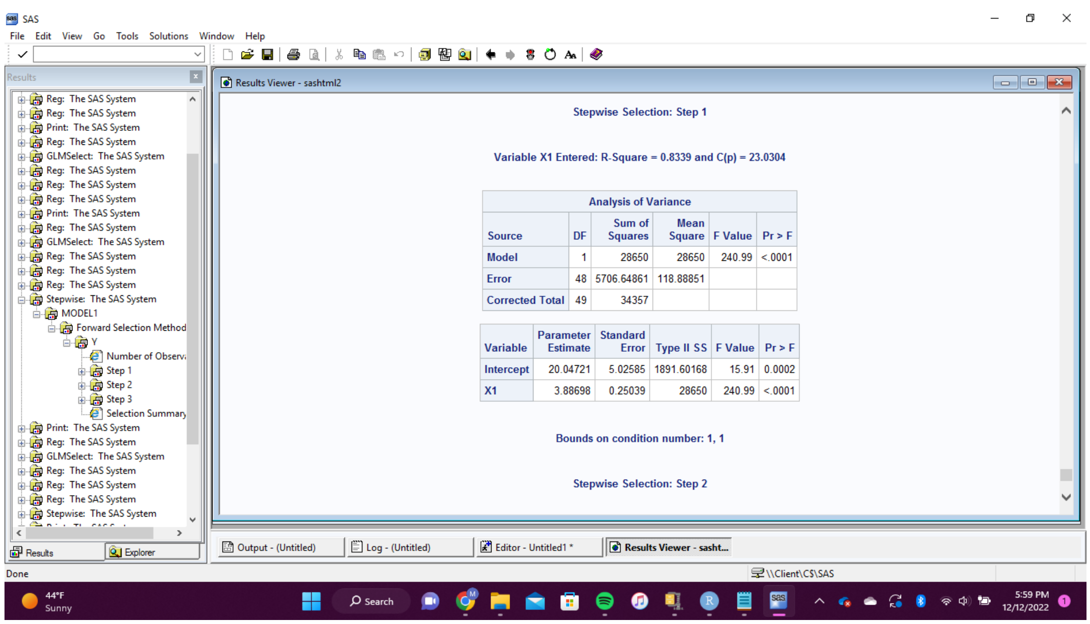
Summary of Backward Elimination

Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	X5	6	0.0000	0.8989	6.0000	0.00	0.9997
2	X7	5	0.0008	0.8981	4.3297	0.34	0.5643
3	X6	4	0.0011	0.8971	2.7693	0.46	0.5026
4	X2	3	0.0012	0.8959	1.2672	0.52	0.4729

PROC STEPWISE data=homes;

model Y = X1 X2 X3 X4 X5 X6 X7/STEPWISE SLENRTY=0.10 SLSTAY=0.10;

run;



Stepwise Selection: Step 1

Variable X1 Entered: R-Square = 0.8339 and C(p) = 23.0304

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	28650	28650	240.99	<.0001
Error	48	5706.64861	118.88851		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	20.04721	5.02585	1891.60168	15.91	0.0002
X1	3.88698	0.25039	28650	240.99	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Results Viewer - sashtml2

Stepwise Selection: Step 2

Variable X4 Entered: R-Square = 0.8772 and C(p) = 7.0480

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	30137	15068	167.82	<.0001
Error	47	4220.06856	89.78869		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-16.23717	9.92955	240.09510	2.67	0.1087
X1	2.94809	0.31716	7757.83135	86.40	<.0001
X4	7.55165	1.85592	1486.58005	16.56	0.0002

Bounds on condition number: 2.1245, 8.4978

Results Viewer - sashtml2

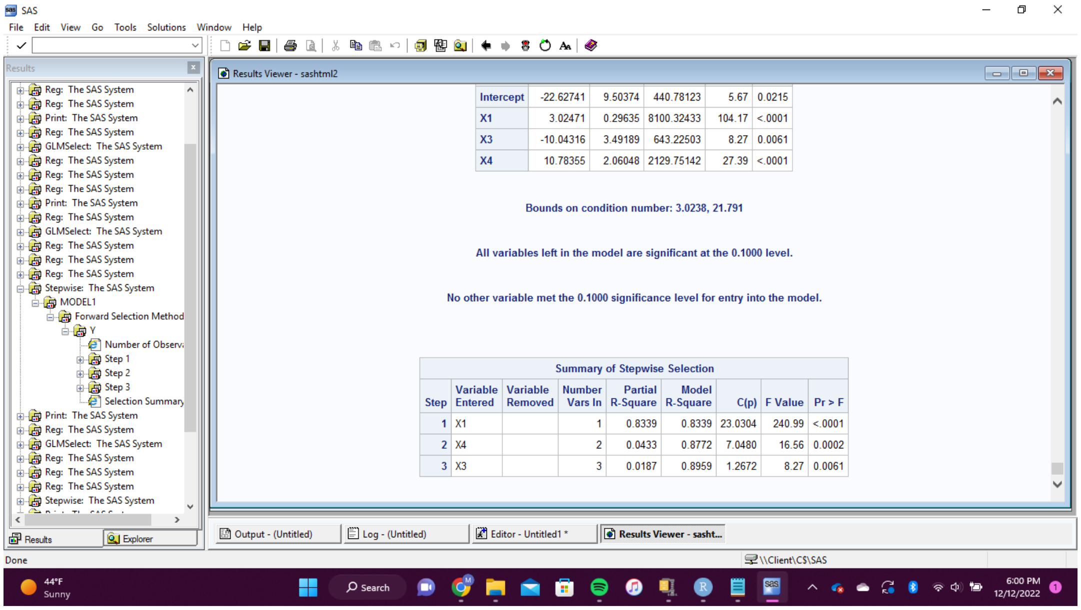
Stepwise Selection: Step 3

Variable X3 Entered: R-Square = 0.8959 and C(p) = 1.2672

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	30780	10260	131.95	<.0001
Error	46	3576.84353	77.75747		
Corrected Total	49	34357			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-22.62741	9.50374	440.78123	5.67	0.0215
X1	3.02471	0.29635	8100.32433	104.17	<.0001
X3	-10.04316	3.49189	643.22503	8.27	0.0061
X4	10.78355	2.06048	2129.75142	27.39	<.0001

Bounds on condition number: 3.0238, 21.791



Intercept	-22.62741	9.50374	440.78123	5.67	0.0215
X1	3.02471	0.29635	8100.32433	104.17	<.0001
X3	-10.04316	3.49189	643.22503	8.27	0.0061
X4	10.78355	2.06048	2129.75142	27.39	<.0001

Bounds on condition number: 3.0238, 21.791

All variables left in the model are significant at the 0.1000 level.

No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	X1		1	0.8339	0.8339	23.0304	240.99	<.0001
2	X4		2	0.0433	0.8772	7.0480	16.56	0.0002
3	X3		3	0.0187	0.8959	1.2672	8.27	0.0061