Stat 328 Final Exam

Summer 2000 Prof. Vardeman

1. Attached to this exam is a yellow JMP-IN printout useful in the analysis of some data taken from *Statistics for Business and Economics* by McClave, Benson and Sincich. The data concern gasoline consumption from 1970 through 1993. For these years, in the JMP data table there are the year (numbered 1 through 24, 1 being 1970) and:

$y =$ Auto Fuel Consumption (billion gallons)
$y' =$ Per Capita Auto Fuel Consumption (1000 gallons)
$x_1 =$ US Population Size (millions)
$x_2 =$ Average Gross Weekly Earnings (1982 dollars)
$x_3 =$ Price of Regular ($ per gallon)
$x_4 =$ Relative Price of Auto Fuel (x_3/CPI)
$x_5 =$ Available Dollars $= x_2 \cdot x_1$

Initially consider trying to model y in terms of the x's. Some relevant correlations and scatterplots are on page 2 of the printout.

(a) Which single variable (of x_1 through x_5) would produce the largest R^2 in a SLR analysis **of y**? What sign would be on its fitted regression coefficient, b? Explain both.

variable: explanations:

sign on its b:

Now consider analysis of y' (not y). Page 3 of the printout concerns a SLR analysis **of y'** as a function of x_2, the single predictor of y' with the highest R^2 in SLR.

(b) There are, on the plot on page 3 of the printout, two clusters of points. The one on the right comes from the first 10 data points and the one on the left from the last 14. These 2 groups correspond to years before and after the "Mideast oil embargo." A sensible thing to do in this context would be to ask whether there should be 2 SLR's instead of just one here. Defining a dummy variable

$$d = \begin{cases} 1 & \text{for years 1-10} \\ 0 & \text{for years 11-24} \end{cases}$$

and fitting the model $y' = \beta_0 + \beta_1 d + \beta_2 x_2 + \beta_3 d \cdot x_2 + \epsilon$ to these data produces $SSE = .00238427$. Give some quantitative assessment of whether single SLR is adequate here, or whether 2 are needed. (Explain what you calculate, why you do so and what it indicates. A full blown test isn't required.)

Regardless of your answer in (b), now consider a single SLR analysis of these (x_2, y') data.

(c) Using the simple linear regression model, give 95% confidence limits for the increase in mean yearly per capita fuel consumption that seems to accompany a \$1 increase in weekly earnings. (No need to simplify.)

(d) Suppose your job is to make an early guess (before figures on fuel use are available) at 2000 per capita auto fuel consumption, based on reports that say $x_2 = 270$ for this year. Give 95% prediction limits. (You may read these from the printout instead of doing calculations.)

Vardeman used JMP-IN's stepwise regression facility and found that the $k = 2$ variable regression equation with the largest R^2 (that can be built from these predictors) is

$$y' = \beta_0 + \beta_1 x_1 + \beta_5 x_5 + \epsilon \quad (= \beta_0 + \beta_1 x_1 + \beta_5 x_2 \cdot x_1 + \epsilon) \tag{*}$$

Pages 4 through 8 of the printout concern the fitting and use of this model.

(e) What is a p-value for testing whether population size (x_1) could be safely dropped from this model (leaving x_5)? Where did you get it and what does it indicate here?

p-value: origin and interpretation:

(f) What "L" is the difference in mean fuel consumptions for two years with (respectively) x_1 of 270 (million) and 273 (million) and x_2 of 300 and 310? (Give the linear combination of β's of interest here.)

(g) What, for $x_1 = 270$, are 95% confidence limits for the increase in mean per capita fuel consumption this model says accompanies a $1 increase in average gross weekly earnings? (No need to simplify.)

(h) What do you see in the plot of residuals for this model against $\widehat{y}'$ on page 5 of the printout? (If it makes any difference, the 10 points on the right are the 10 "pre-embargo" points.)

(i) From looking at the JMP data table on pages 7 (to which some things have been added subsequent to fitting the model (*)) identify a case that you think is probably highly important in the fitting of model (*). What draws your attention to this case?

(j) Page 8 of the printout is a plot of e_i versus e_{i-1} for this data set, where i is both the case number and the year number ($i = 1$ being 1970). What does this plot indicate about the MLR regression analysis under model (*)? Note that 1993 (the $i = 24$) case in the data set, had $e_{24} \approx .01$. If I told you the values for x_1 and x_5 for 1994 (for example, say they are $x_1 = 261$ and $x_5 = 66,000$) and asked you to adjust your prediction of y_{25} derived from the fitted version of model (*), how (in qualitative terms) would you make use of this information about e_{24}? (Would you adjust up or down, and why?)

2. The pink printout attached to this printout is from a JMP analysis of some data again from *Statistics for Business and Economics* by McClave, Benson and Sincich, this time fashioned after a study from the *Journal of Marketing Research* meant to evaluate the effectiveness of short-run supermarket strategies. 3 different Display levels ($1 =$ "normal space," $2 =$ "normal plus end-of-aisle" and $3 =$ "twice normal") and 3 different Price levels ($1 =$ "regular," $2 =$ "reduced" and $3 =$ "cost to store") were used. The response variable

$$y = \text{weekly sales } (\$100)$$

was observed for a single product 3 different times for each of the 3×3 different combinations of levels of the two factors.

(a) On page 2 of the printout is a plot of y versus a "cell number" variable that names the combinations of Display and Price $1 = (1,1)$, $2 = (1,2)$, $3 = 1,3)$, $4 = (2,1)$, $\ldots$, $9 = (3,3)$. Do you see anything in that plot to cause you to worry about the constant variance assumption? There is on pages 3&4 a JMP-IN "Fit Model" MLR analysis using "cell" as a nominal variable. On it is a pooled estimate of σ. What is that estimate, and what in the context of this problem is it estimating?

comment on plot:

$s_{\text{Pooled}} =$ _______________

σ here is:

(b) Give and interpret a p-value for testing whether there is any difference at all among the 9 different cell means (9 different short run marketing strategies) in terms of mean sales.

p-value:

origin and interpretation:

(c) Give a 95% prediction interval for the next sales figure for the "twice normal space"&"cost to store" marketing strategy for this item. (No need to simplify.)

Pages 5 through 10 of the printout refer to a JMP analysis of these data using "nominal" variables Display and Price. First there is a "no-interactions" analysis. Then there is a "full model"/"with interactions" analysis.

(d) Does it look from the printout that one really needs to include interactions in a two-factor analysis of these data? Explain. (Offer some quantitative support for your position.)

Regardless of your answer to (d), for the rest of this exam, consider a "with interactions" analysis.

(e) Give 95% two-sided confidence limits for the main effect of the "normal space" level of Display. What is being measured here in terms of the 9 different mean sales figures? (No need to simplify the limits.)

limits:

interpretation:

(f) Below is a 3×3 table. Write in it estimated interactions for the 9 different combinations of levels of Display and Price. (Some of these can be gotten directly from the printout, others you'll have to do a bit of arithmetic to get.)

Price

		1	2	3
	1			
Display	2			
	3			

(g) Suppose that in fact the MLR model with (with dummies) is a good one and that estimates of the model parameters here are in fact exactly the true values (that's too much to hope for, but pretend they are perfect). What do you assess as the fraction of weeks that the "normal space"&"regular price" marketing strategy will produce sales of at least $1000?

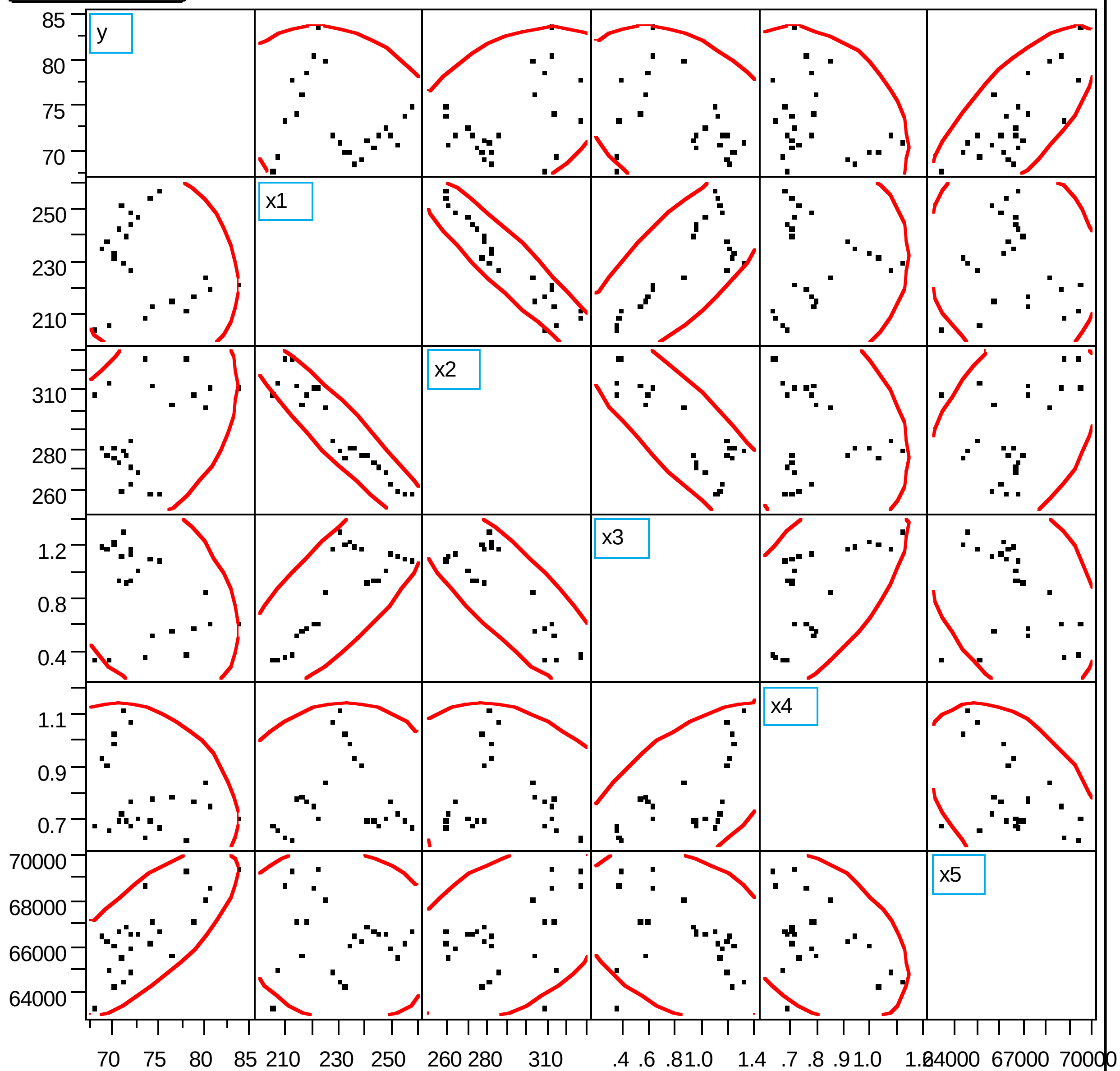
AutoFuel

Rows	Year	y	y'	x1	x2	x3	x4	x5
1	1	67.8	0.33057	205.1	308.84	0.36	0.689	63343.08
2	2	69.51	0.335635	207.1	314.35	0.36	0.672	65101.89
3	3	73.5	0.350167	209.9	327.51	0.37	0.641	68744.35
4	4	78	0.368098	211.9	327.45	0.4	0.633	69386.65
5	5	74.2	0.346891	213.9	313.91	0.53	0.786	67145.35
6	6	76.5	0.354167	216	303.96	0.57	0.794	65655.36
7	7	78.8	0.361468	218	308.35	0.59	0.775	67220.3
8	8	80.7	0.366485	220.2	311.88	0.62	0.763	68675.98
9	9	83.8	0.37646	222.6	312.42	0.63	0.715	69544.69
10	10	80.2	0.356286	225.1	302.91	0.86	0.852	68185.04
11	11	71.9	0.315766	227.7	285.32	1.19	1.074	64967.36
12	12	71	0.308696	230	280.75	1.31	1.125	64572.5
13	13	70.1	0.301765	232.3	276.95	1.22	1.033	64335.49
14	14	69.9	0.298081	234.5	281.83	1.24	0.998	66089.14
15	15	68.7	0.290732	236.3	281.87	1.21	0.942	66605.88
16	16	69.3	0.290566	238.5	277.96	1.2	0.917	66293.46
17	17	71.4	0.296635	240.7	278.15	0.93	0.703	66950.7
18	18	70.6	0.290774	242.8	275.09	0.95	0.706	66791.85
19	19	71.9	0.293469	245	272.21	0.95	0.683	66691.45
20	20	72.7	0.293975	247.3	269.55	1.02	0.713	66659.72
21	21	72	0.288115	249.9	264.23	1.16	0.774	66031.08
22	22	70.7	0.279889	252.6	259.9	1.14	0.729	65650.74
23	23	73.9	0.28935	255.4	259.17	1.12	0.705	66192.02
24	24	75.1	0.290972	258.1	258.57	1.1	0.678	66736.92

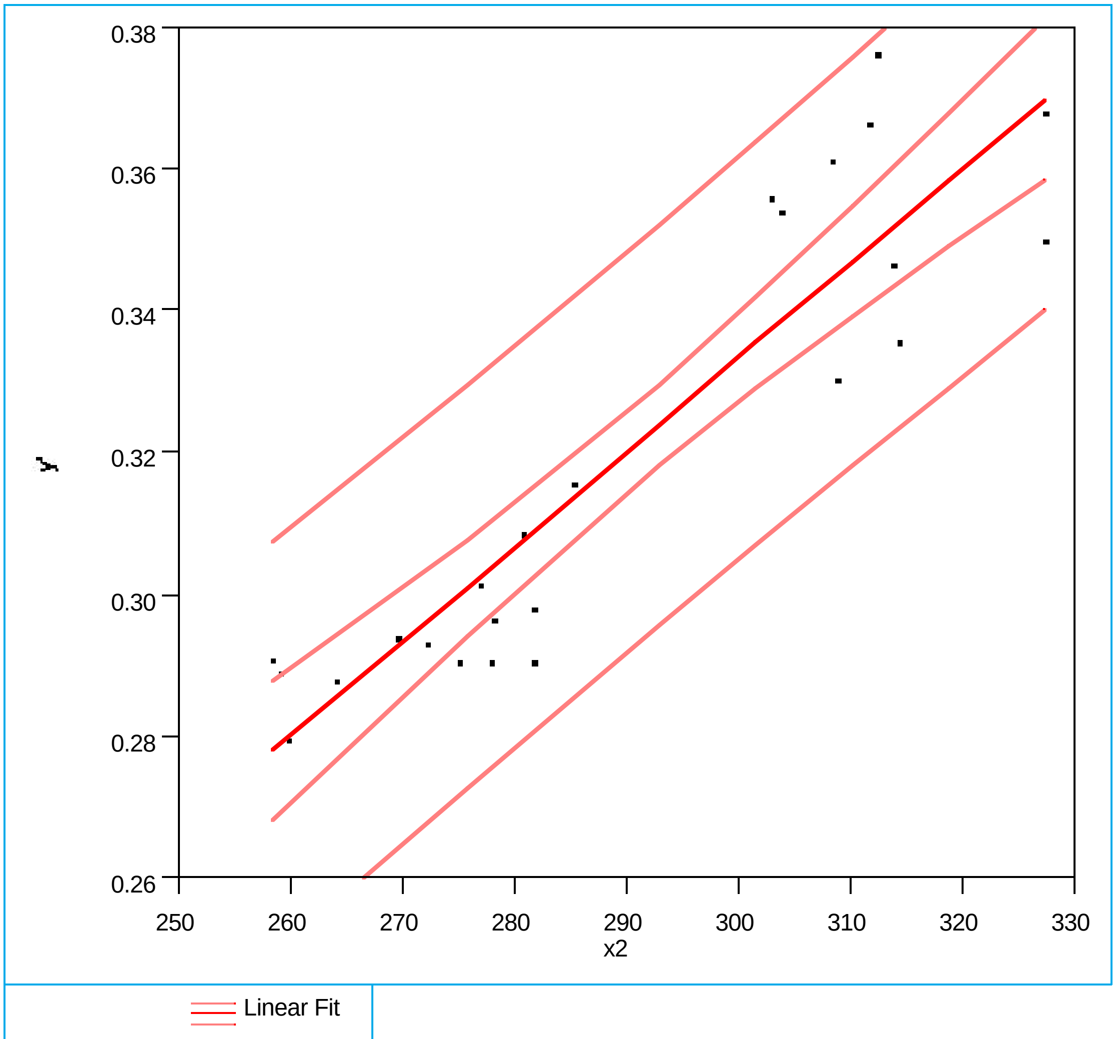
Correlations

Variable	y	x1	x2	x3	x4	x5
y	1.0000	-0.2337	0.4484	-0.3940	-0.2701	0.7521
x1	-0.2337	1.0000	-0.9476	0.7992	0.0487	-0.1168
x2	0.4484	-0.9476	1.0000	-0.8619	-0.2193	0.4236
x3	-0.3940	0.7992	-0.8619	1.0000	0.6298	-0.3560
x4	-0.2701	0.0487	-0.2193	0.6298	1.0000	-0.4383
x5	0.7521	-0.1168	0.4236	-0.3560	-0.4383	1.0000

Scatterplot Matrix



y' By x2



Linear Fit

y' = -0.0662 + 0.00133 x2

Summary of Fit

RSquare	0.835699
RSquare Adj	0.828231
Root Mean Square Error	0.013311
Mean of Response	0.319792
Observations (or Sum Wgts)	24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.01982645	0.019826	111.9007
Error	22	0.00389794	0.000177	Prob>F
C Total	23	0.02372439		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.066219	0.036592	-1.81	0.0840
x2	0.0013324	0.000126	10.58	<.0001

Response: y'

Summary of Fit

RSquare	0.86164
RSquare Adj	0.848463
Root Mean Square Error	0.012502
Mean of Response	0.319792
Observations (or Sum Wgts)	24

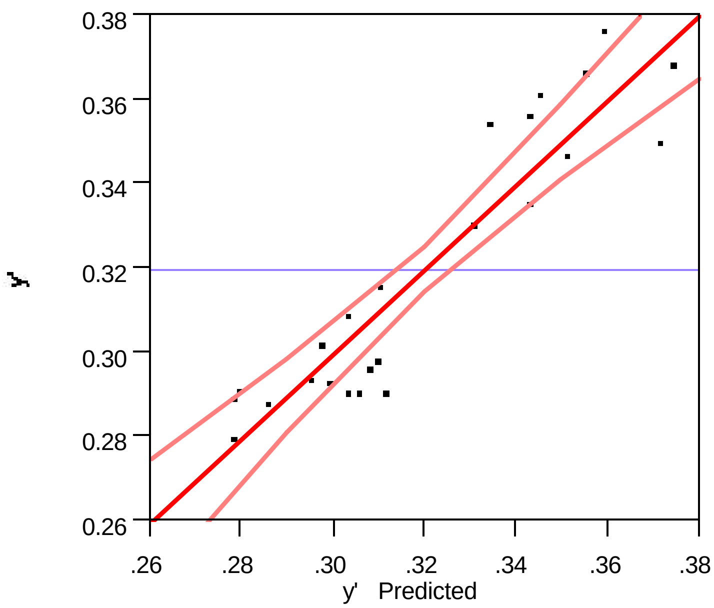
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.07819	0.122224	0.64	0.5293
x1	-0.001541	0.000164	-9.42	<.0001
x5	0.000009	0.000002	5.34	<.0001

Effect Test

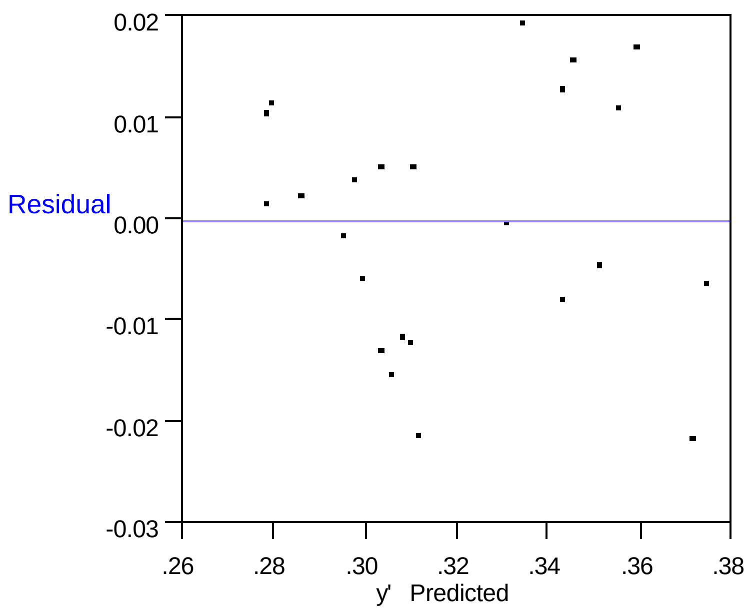
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
x1	1	1	0.01386884	88.7268	<.0001
x5	1	1	0.00445715	28.5149	<.0001

Whole-Model Test

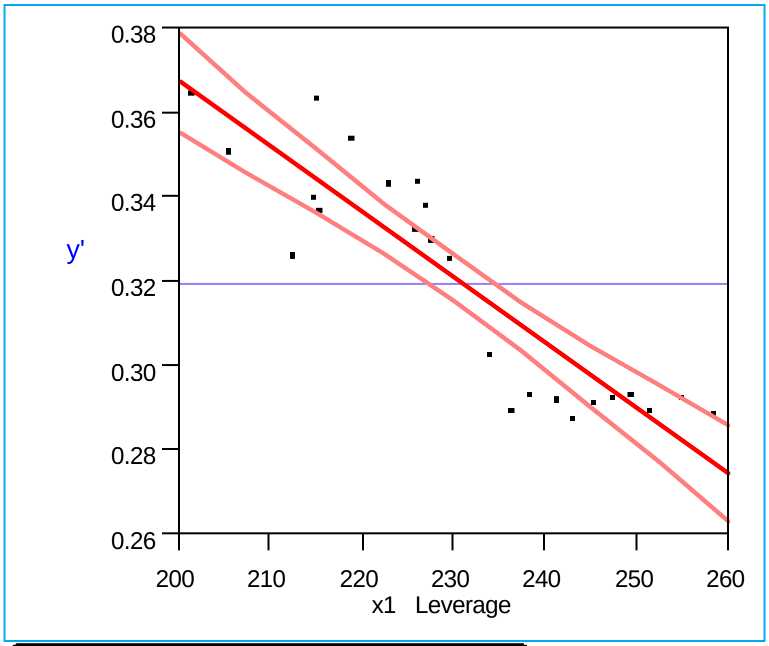


Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.02044189	0.010221	65.3892
Error	21	0.00328250	0.000156	Prob>F
C Total	23	0.02372439		<.0001



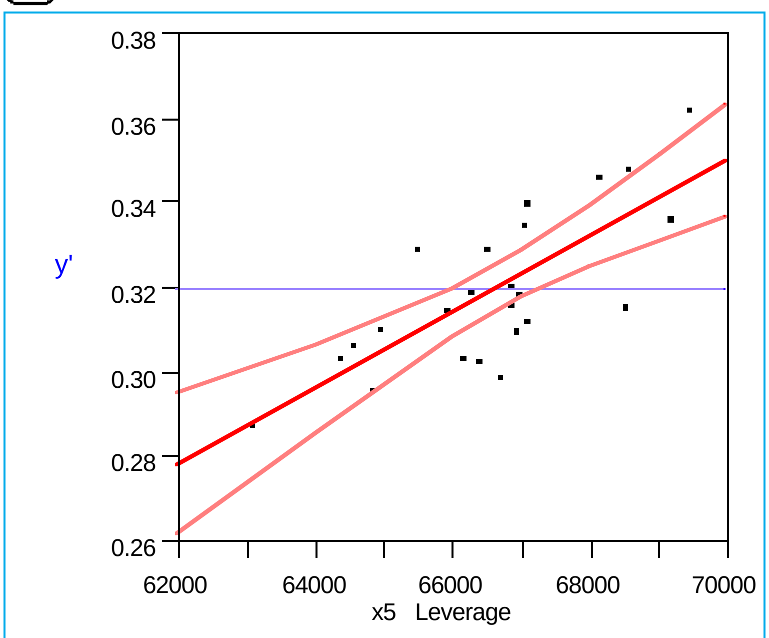
x1



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
0.01386884	88.7268	1	<.0001

x5



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
0.00445715	28.5149	1	<.0001

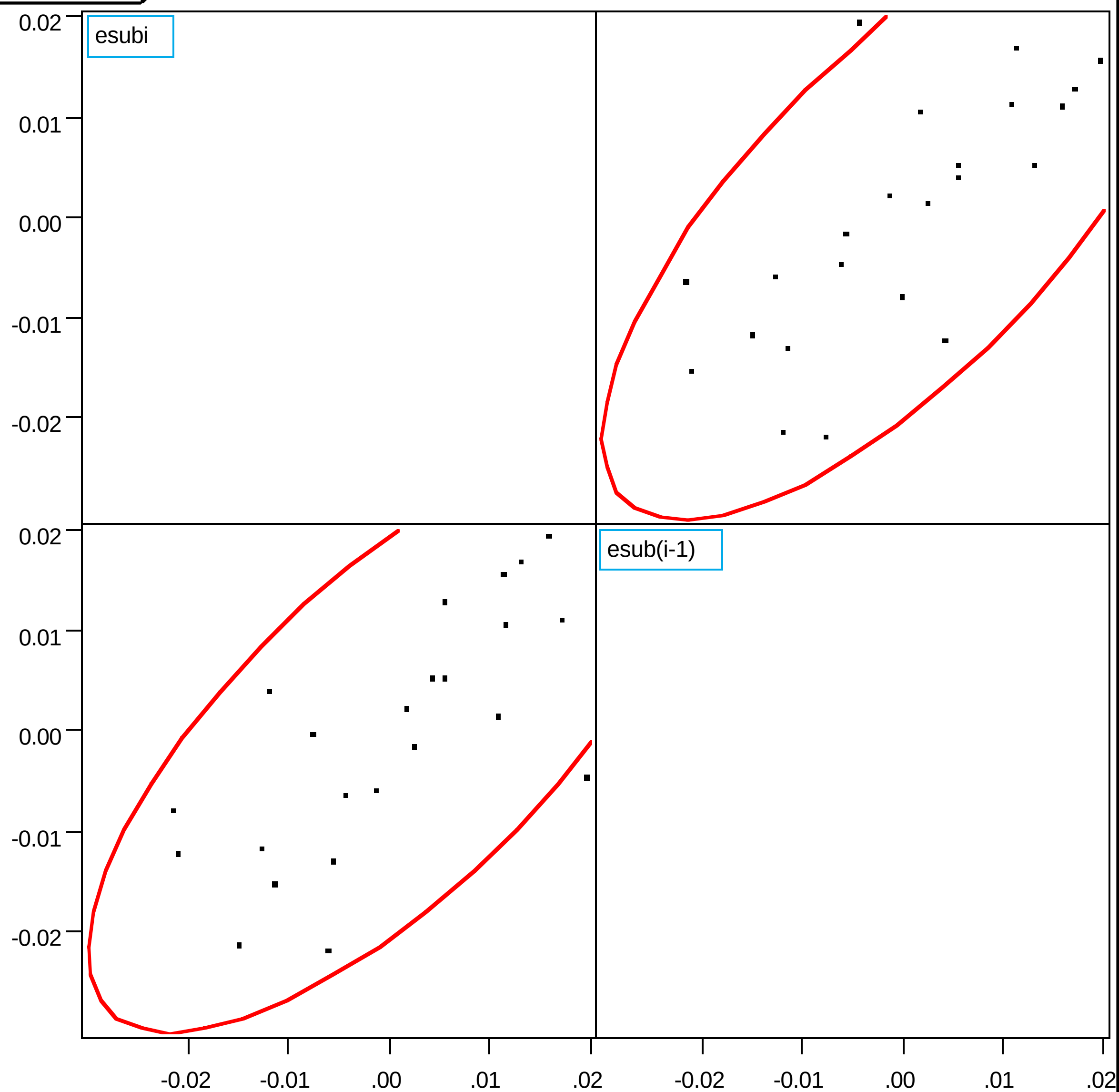
Rows	Year	y'	x1	x5	Predicted y'	Residual y'	hy'	Cook's D Influencey'
1	1	0.33057	205.1	63343.08	0.330591	-0.00002	0.377233	8.865e-7
2	2	0.335635	207.1	65101.89	0.343294	-0.00766	0.191471	0.036642
3	3	0.350167	209.9	68744.35	0.371671	-0.0215	0.184017	0.272538
4	4	0.368098	211.9	69386.65	0.374353	-0.00625	0.22516	0.031289
5	5	0.346891	213.9	67145.35	0.351154	-0.00426	0.093032	0.004383
6	6	0.354167	216	65655.36	0.334544	0.019622	0.10008	0.101469
7	7	0.361468	218	67220.3	0.345507	0.01596	0.074328	0.047122
8	8	0.366485	220.2	68675.98	0.355182	0.011303	0.132417	0.047932
9	9	0.37646	222.6	69544.69	0.35928	0.01718	0.203671	0.202166
10	10	0.356286	225.1	68185.04	0.343223	0.013063	0.090936	0.040043
11	11	0.315766	227.7	64967.36	0.310336	0.00543	0.09163	0.006983
12	12	0.308696	230	64572.5	0.303247	0.005449	0.114295	0.009224
13	13	0.301765	232.3	64335.49	0.297575	0.00419	0.130581	0.006468
14	14	0.298081	234.5	66089.14	0.309923	-0.01184	0.047313	0.01559
15	15	0.290732	236.3	66605.88	0.311787	-0.02105	0.046835	0.048735
16	16	0.290566	238.5	66293.46	0.305592	-0.01503	0.052121	0.027931
17	17	0.296635	240.7	66950.7	0.3081	-0.01147	0.062455	0.019918
18	18	0.290774	242.8	66791.85	0.303438	-0.01266	0.068079	0.026807
19	19	0.293469	245	66691.45	0.299146	-0.00568	0.076881	0.006199
20	20	0.293975	247.3	66659.72	0.295316	-0.00134	0.088698	0.00041
21	21	0.288115	249.9	66031.08	0.285666	0.002449	0.104674	0.00167
22	22	0.279889	252.6	65650.74	0.278091	0.001798	0.129494	0.001178
23	23	0.28935	255.4	66192.02	0.278634	0.010716	0.143484	0.047898
24	24	0.290972	258.1	66736.92	0.279363	0.01161	0.171117	0.071588

Correlations

Variable	esubi	esub(i-1)
esubi	1.0000	0.7274
esub(i-1)	0.7274	1.0000

1 rows not used due to missing values.

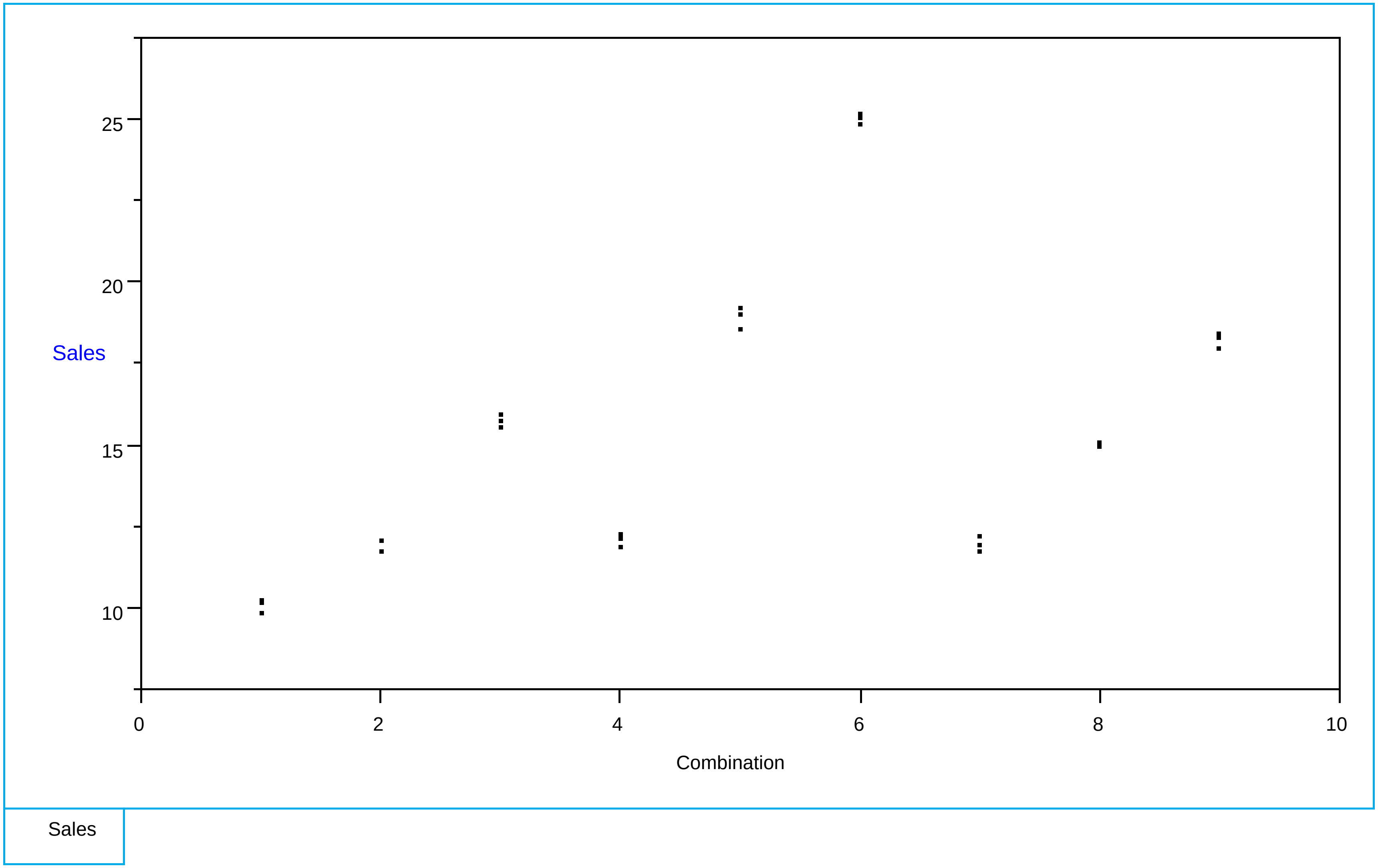
Scatterplot Matrix



supermarket

Rows	Combination	Display	Price	Sales
1	1	1	1	9.89
2	1	1	1	10.25
3	1	1	1	10.3
4	2	1	2	12.11
5	2	1	2	12.15
6	2	1	2	11.82
7	3	1	3	15.77
8	3	1	3	15.59
9	3	1	3	15.98
10	4	2	1	11.91
11	4	2	1	12.33
12	4	2	1	12.21
13	5	2	2	18.6
14	5	2	2	19.1
15	5	2	2	19.26
16	6	2	3	24.92
17	6	2	3	25.27
18	6	2	3	25.11
19	7	3	1	12.26
20	7	3	1	12.02
21	7	3	1	11.8
22	8	3	2	15.16
23	8	3	2	15.01
24	8	3	2	14.98
25	9	3	3	18.01
26	9	3	3	18.33
27	9	3	3	18.52

Plot



Response: Sales

Summary of Fit

RSquare	0.99832
RSquare Adj	0.997573
Root Mean Square Error	0.222428
Mean of Response	15.50593
Observations (or Sum Wgts)	27

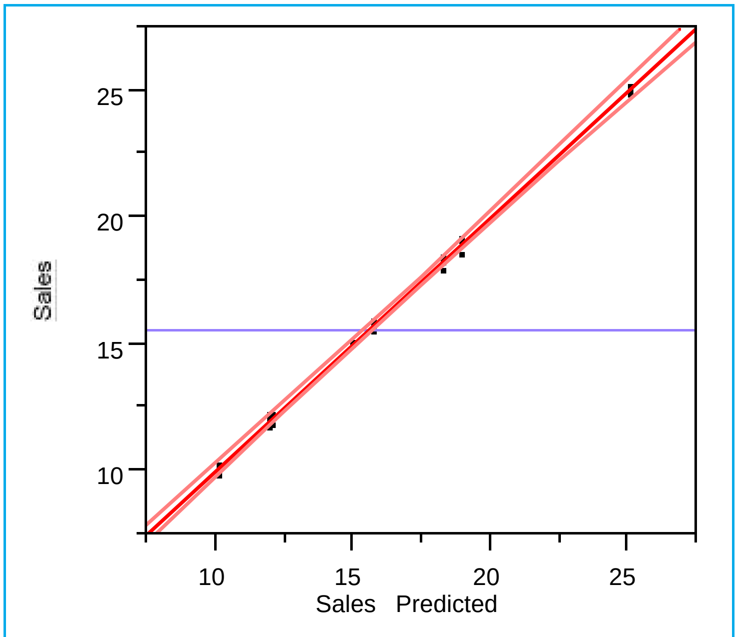
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	15.505926	0.042806	362.24	<.0001
Combinat[1-9]	-5.359259	0.121074	-44.26	<.0001
Combinat[2-9]	-3.479259	0.121074	-28.74	<.0001
Combinat[3-9]	0.2740741	0.121074	2.26	0.0362
Combinat[4-9]	-3.355926	0.121074	-27.72	<.0001
Combinat[5-9]	3.4807407	0.121074	28.75	<.0001
Combinat[6-9]	9.5940741	0.121074	79.24	<.0001
Combinat[7-9]	-3.479259	0.121074	-28.74	<.0001
Combinat[8-9]	-0.455926	0.121074	-3.77	0.0014

Effect Test

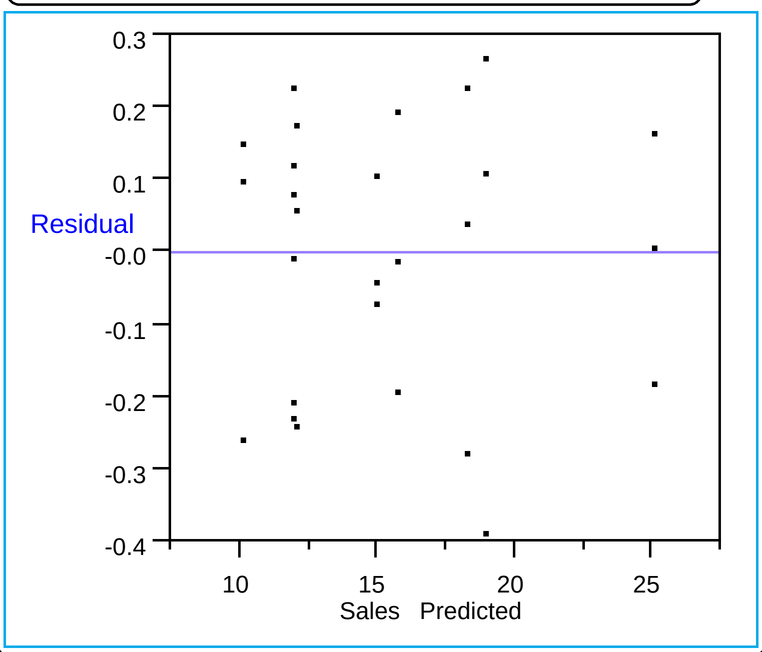
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Combination	8	8	529.11512	1336.849	<.0001

Whole-Model Test

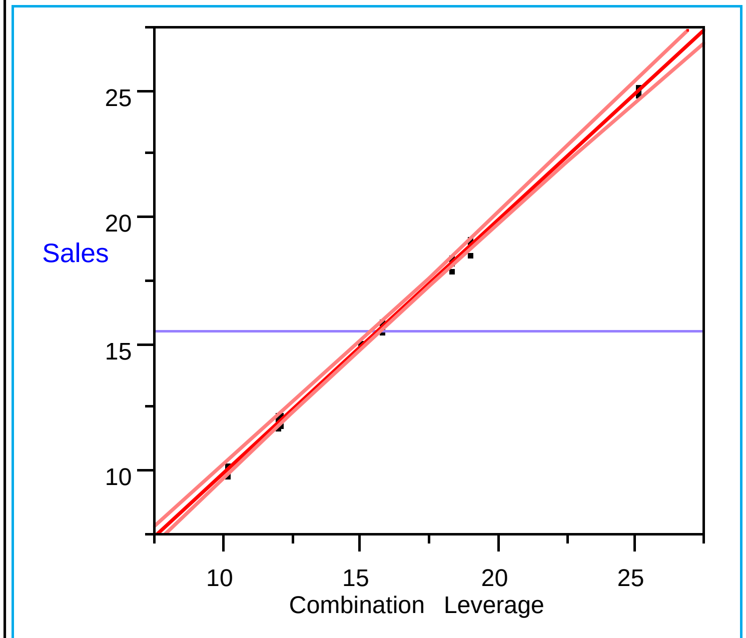


Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	529.11512	66.1394	1336.849
Error	18	0.89053	0.0495	Prob>F
C Total	26	530.00565		<.0001



Combination



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
529.11512	1336.849	8	<.0001

Least Squares Means

Level	Least Sq Mean	Std Error	Mean
1	10.14666667	0.1284186825	10.1467
2	12.02666667	0.1284186825	12.0267
3	15.78000000	0.1284186825	15.7800
4	12.15000000	0.1284186825	12.1500
5	18.98666667	0.1284186825	18.9867
6	25.10000000	0.1284186825	25.1000
7	12.02666667	0.1284186825	12.0267
8	15.05000000	0.1284186825	15.0500
9	18.28666667	0.1284186825	18.2867

Response: Sales

Summary of Fit

RSquare	0.901961
RSquare Adj	0.884136
Root Mean Square Error	1.536836
Mean of Response	15.50593
Observations (or Sum Wgts)	27

Lack of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack of Fit	4	51.070481	12.7676	258.0669
Pure Error	18	0.890533	0.0495	Prob>F
Total Error	22	51.961015		<.0001
				Max RSq
				0.9983

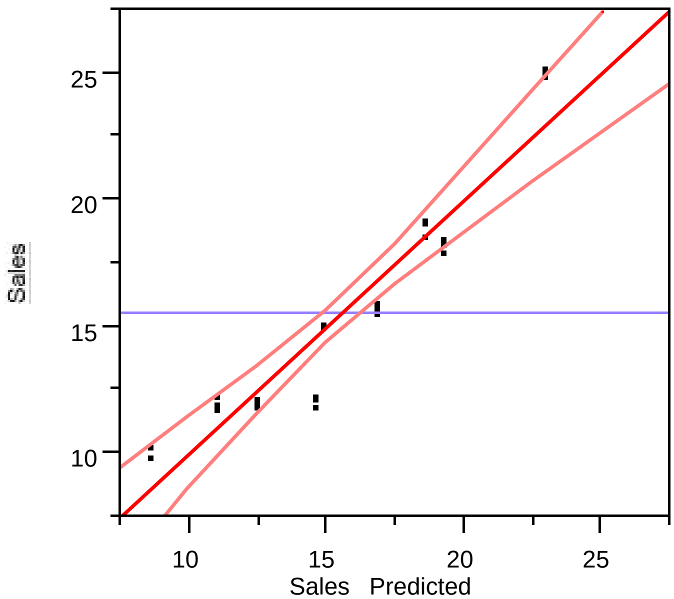
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	15.505926	0.295764	52.43	<.0001
Display[1-3]	-2.854815	0.418274	-6.83	<.0001
Display[2-3]	3.2396296	0.418274	7.75	<.0001
Price[1-3]	-4.064815	0.418274	-9.72	<.0001
Price[2-3]	-0.151481	0.418274	-0.36	0.7207

Effect Test

Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Display	2	2	169.13925	35.8063	<.0001
Price	2	2	308.90539	65.3944	<.0001

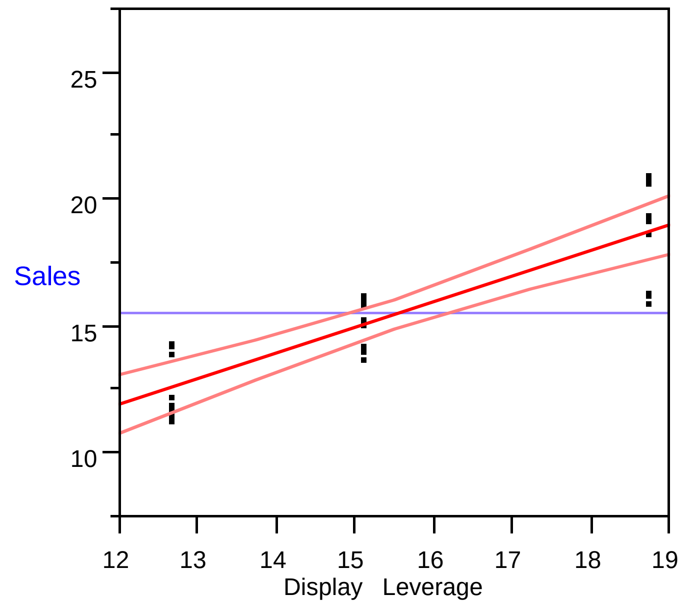
Whole-Model Test



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	478.04464	119.511	50.6003
Error	22	51.96101	2.362	Prob>F
C Total	26	530.00565		<.0001

Display



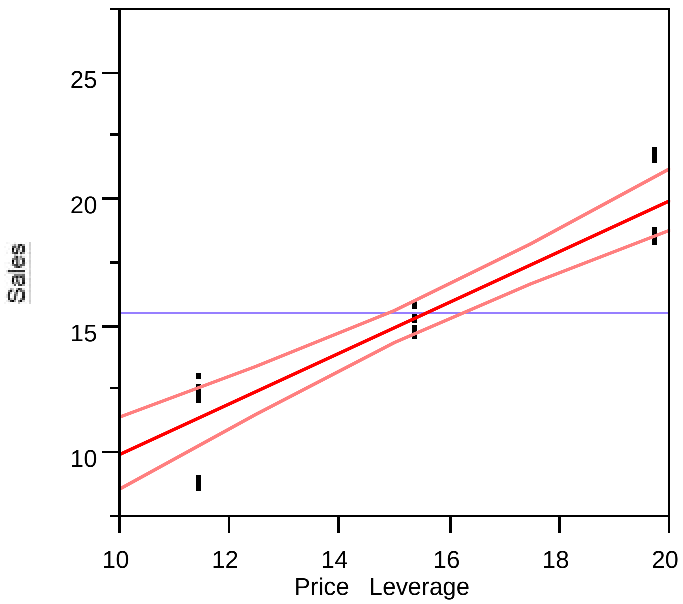
Effect Test

Sum of Squares	F Ratio	DF	Prob>F
169.13925	35.8063	2	<.0001

Least Squares Means

Level	Least Sq Mean	Std Error	Mean
1	12.65111111	0.5122786036	12.6511
2	18.74555556	0.5122786036	18.7456
3	15.12111111	0.5122786036	15.1211

Price



Effect Test

Sum of Squares	F Ratio	DF	Prob>F
308.90539	65.3944	2	<.0001

Least Squares Means

Level	Least Sq Mean	Std Error	Mean
1	11.44111111	0.5122786036	11.4411
2	15.35444444	0.5122786036	15.3544
3	19.72222222	0.5122786036	19.7222

Response: Sales

Summary of Fit

RSquare	0.99832
RSquare Adj	0.997573
Root Mean Square Error	0.222428
Mean of Response	15.50593
Observations (or Sum Wgts)	27

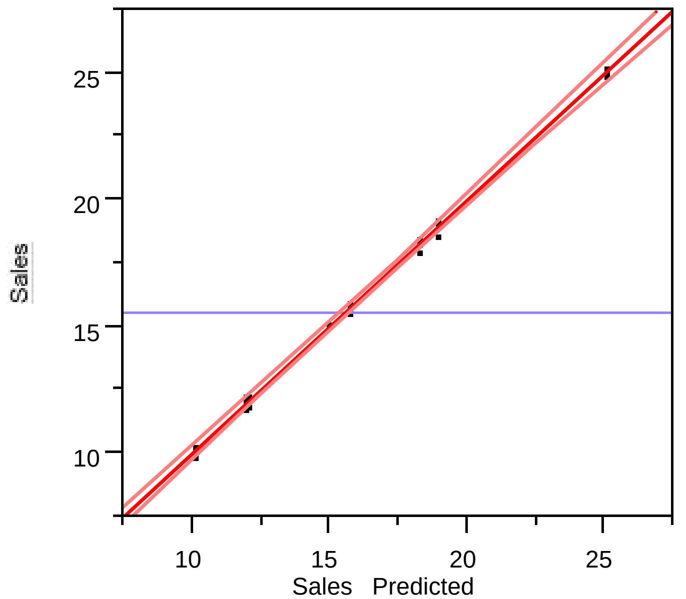
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	15.505926	0.042806	362.24	<.0001
Display[1-3]	-2.854815	0.060537	-47.16	<.0001
Display[2-3]	3.2396296	0.060537	53.51	<.0001
Price[1-3]	-4.064815	0.060537	-67.15	<.0001
Price[2-3]	-0.151481	0.060537	-2.50	0.0222
Display[1-3]*Price[1-3]	1.5603704	0.085612	18.23	<.0001
Display[1-3]*Price[2-3]	-0.472963	0.085612	-5.52	<.0001
Display[2-3]*Price[1-3]	-2.530741	0.085612	-29.56	<.0001
Display[2-3]*Price[2-3]	0.3925926	0.085612	4.59	0.0002

Effect Test

Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Display	2	2	169.13925	1709.373	<.0001
Price	2	2	308.90539	3121.892	<.0001
Display*Price	4	4	51.07048	258.0669	<.0001

Whole-Model Test

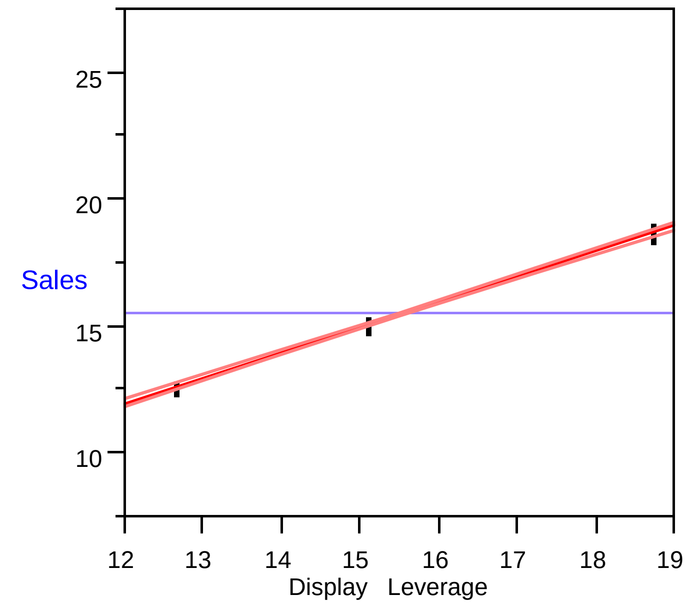


Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	529.11512	66.1394	1336.849
Error	18	0.89053	0.0495	Prob>F
C Total	26	530.00565		<.0001

Press	2.0037

Display



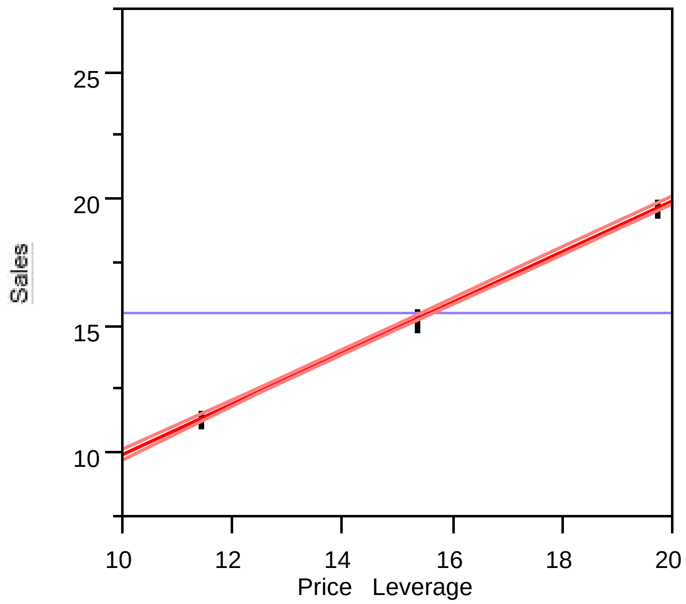
Effect Test

Sum of Squares	F Ratio	DF	Prob>F
169.13925	1709.373	2	<.0001

Least Squares Means

Level	Least Sq Mean	Std Error	Mean
1	12.65111111	0.0741425609	12.6511
2	18.74555556	0.0741425609	18.7456
3	15.12111111	0.0741425609	15.1211

Price



Effect Test

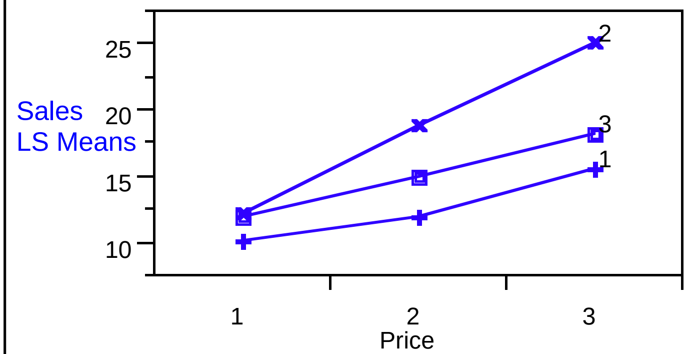
Sum of Squares	F Ratio	DF	Prob>F
308.90539	3121.892	2	<.0001

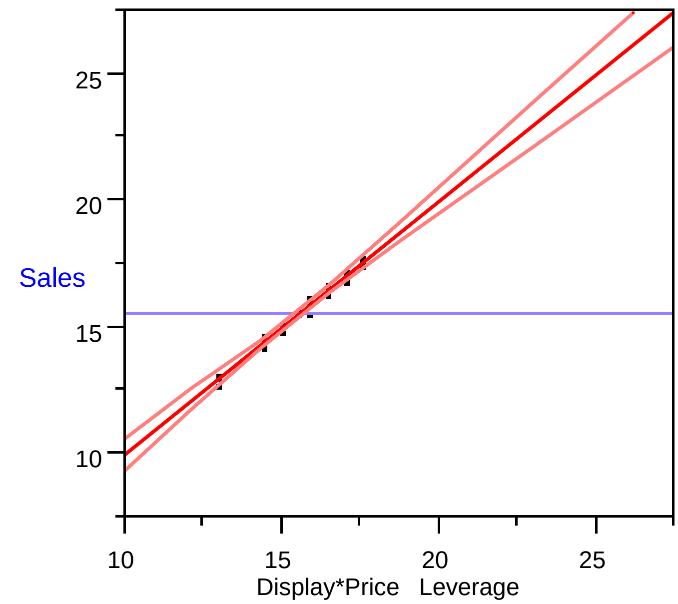
Least Squares Means

Level	Least Sq Mean	Std Error	Mean
1	11.44111111	0.0741425609	11.4411
2	15.35444444	0.0741425609	15.3544
3	19.72222222	0.0741425609	19.7222

Display*Price

Profile Plot





Effect Test

Sum of Squares	F Ratio	DF	Prob>F
51.070481	258.0669	4	<.0001

Least Squares Means

Level	Least Sq Mean	Std Error
1,1	10.14666667	0.1284186825
1,2	12.02666667	0.1284186825
1,3	15.78000000	0.1284186825
2,1	12.15000000	0.1284186825
2,2	18.98666667	0.1284186825
2,3	25.10000000	0.1284186825
3,1	12.02666667	0.1284186825
3,2	15.05000000	0.1284186825
3,3	18.28666667	0.1284186825