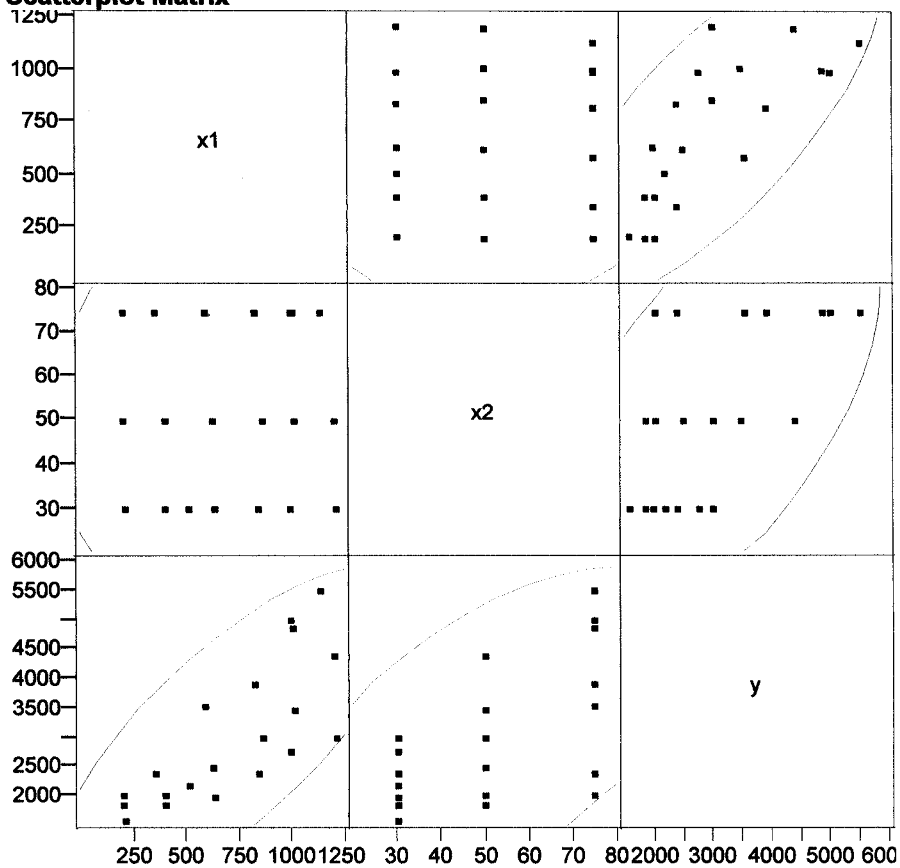
Stat 328
Summer 2004
Lab #5 Key

1. a)

Multivariate
Correlations

	x1	x2	y
x1	1.0000	0.0538	0.7667
x2	0.0538	1.0000	0.6002
y	0.7667	0.6002	1.0000

Scatterplot Matrix



Since, we see a correlation of .77 between x1 and y and .60 between x2 and y, we would be able to predict total sales on the basis of these variables.

b)

Summary of Fit

RSquare	0.901125
RSquare Adj	0.889492
Root Mean Square Error	391.5064
Mean of Response	3014.2
Observations (or Sum Wgts)	20

95% Confidence Interval

SSE = 2605713 $\chi_{(.975,17)}=30.19$ $\chi_{(.005,17)}=7.56$

(293.79, 587.09)

c)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	23747818	11873909	77.4669
Error	17	2605713	153277.25	Prob > F
C. Total	19	26353531		<.0001

This small p-value provides evidence against H_o: $\beta_1 = \beta_2 = 0$. Thus, x1 and x2 provide explanatory power for predicting sales.

d)

JMP Output

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	-558.5463	309.5127	-1.80	0.0889	-1211.561	94.4684
x1	2.548136	0.264233	9.64	<.0001	1.9906527	3.1056193
x2	34.120496	4.648559	7.34	<.0001	24.312894	43.928099

Hand Calculation

$b_1 \pm t_{(\alpha/2,df)} * S_{b1}$

$2.54 \pm 2.11 * 0.264$

2.54 ± 0.56

(1.98, 3.10) which matches the JMP Output above

e)

Effect Tests

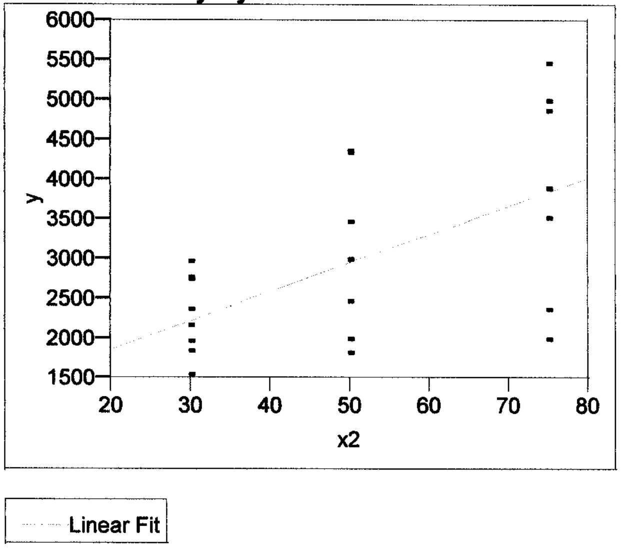
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
x1	1	1	14254376	92.9973	<.0001
x2	1	1	8257939	53.8758	<.0001

Name of Test	**t test**	**F test**
Observed Test Statistic	**t=9.64**	**F=93.00**
reference distribution	**t with 17df**	**F with 1 and 17 df**
p-value	**<0.0001**	**<0.0001**

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	23747818	11873909	77.4669
Error	17	2605713	153277.25	Prob > F
C. Total	19	26353531		<.0001

Bivariate Fit of y By x2



Linear Fit

y = 1123.7164 + 36.531084 x2

Summary of Fit

RSquare	0.360234
RSquare Adj	0.324692
Root Mean Square Error	967.818
Mean of Response	3014.2
Observations (or Sum Wgts)	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	9493442	9493442	10.1353
Error	18	16860089	936672	Prob > F
C. Total	19	26353531		0.0051

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1123.7164	632.025	1.78	0.0923
x2	36.531084	11.47478	3.18	0.0051

Partial F Test from class

H_o: $\beta_1 = 0$

F = (23747818-9493442)/(153277.25) = 92.967 $\approx$ 93.00

Both p-values for the F and t test are <0.0001 which indicates evidence against H_o. Hence, weekly newspaper advertising expenditures are useful for predicting resulting sales in the presence of shelf space allocated.

f) The observed test statistics are related in the following manner: $t^2 = F$. The denominator df for the F statistic equal the df for the t statistic.

g) H_o: $\beta_2 = 0$ vs. H_a: $\beta_2 \neq 0$
Test Statistics
$t = 7.34$ df = 17
$F = 53.9$ df = 1 and 17
p-value<0.0001 which presents evidence that $\beta_2 \neq 0$. Yes, the above test suggests that shelf space is useful for predicting sales even after accounting for advertising.

h)

From JMP

x1	x2	y	95% Lower Mean y	95% Upper Mean y	95% Lower Individual y	95% Upper Individual y
820	75	3908	3792.61766	4387.30717	3212.06706	4967.85778

i)

From JMP]

x1	x2	y	Std Error Indiv	Std Error Pred
820	75	3908	416.100521	140.934004

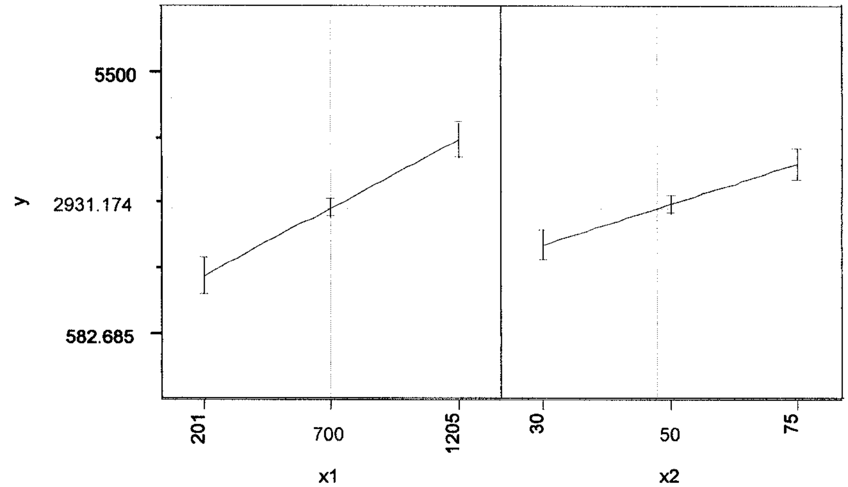
95% CI for $\mu_{y|x1=820,x22=75}$
$y \pm t_{\alpha/2,df} * S_m$
$y \pm 2.11 * 140.93$
$y = b_o + b_1 x_1 + b_2 x_2$
$= -558.55 + 2.58481*820 + 32.120*75$
$= 4089.89$
4089.89 ± 297.36
(3792.53, 4387.25) which is close to what we found from JMP in part h)

95% PI
$y \pm t_{\alpha/2,df} * S_p$
$y \pm 2.11 * 416.10$
4089.89 ± 877.97
(3211.92, 4967.86) which is close to what we found from JMP in part h)

x1	x2	95% Lower Individual y	95% Upper Individual y	95% Lower Mean y	95% Upper Mean y
500	55	1736.92305	3447.3749	2370.50732	2813.79063

j)

Prediction Profiler



y = 2931.174 when x1=700 and x2=50
95% confidence limits are 582.685 and 5500

$\mu_{y|x1,x2} = \beta_0 + \beta_1 x1 + \beta_2 x2$

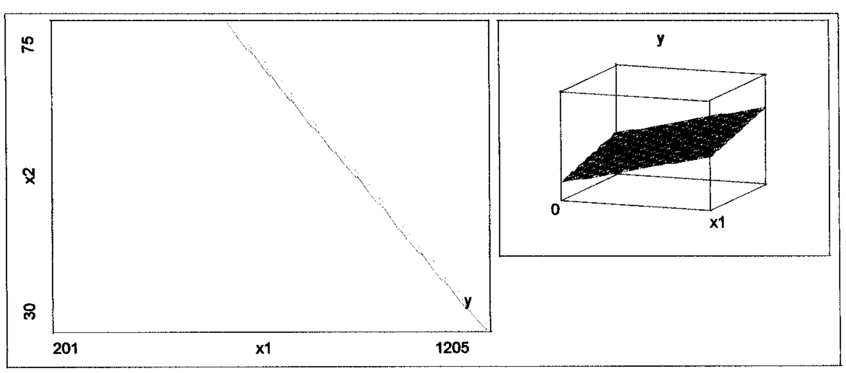
So when x1 is increased by 1 unit and x2 is held fixed, $\mu_{y|x1,x2}$ changes by β_1. Thus, holding x2 fixed and changing x1 traces out a line for mean values of y. Similarly, hold x1 fixed and vary x2; then, the mean values of y are on a line with slope β_2.

k)

Contour Profiler

Horiz	Vert	Factor	Current X
◉	○	x1	700
○	◉	x2	50

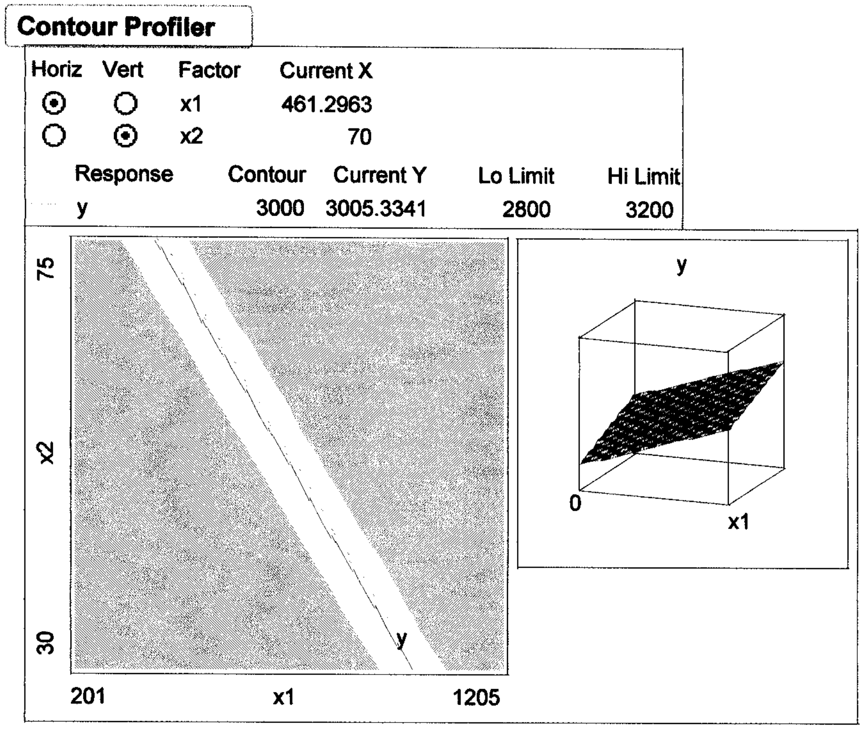
Response	Contour	Current Y	Lo Limit	Hi Limit
y	3537.5	2931.1737	.	.



A plane has been fit to these data.

l) The current value of y is exactly the same, y = 2931.174.

m)

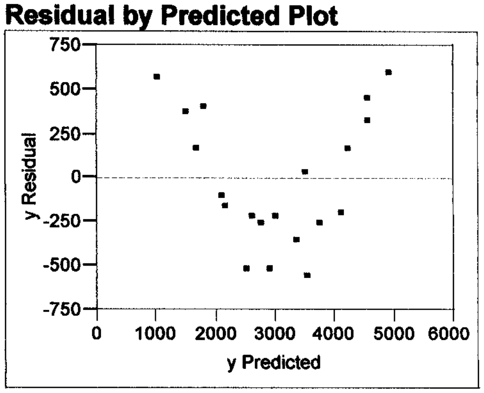


With x2=70 and y approximately 3000, we need a value of x1 of roughly 461.

n) This plot is useful because any combination of advertising expenditures and shelf space allocation it indicates whether sales will be bad (<$2800), OK (2800-3200), or good (>3200).

o)

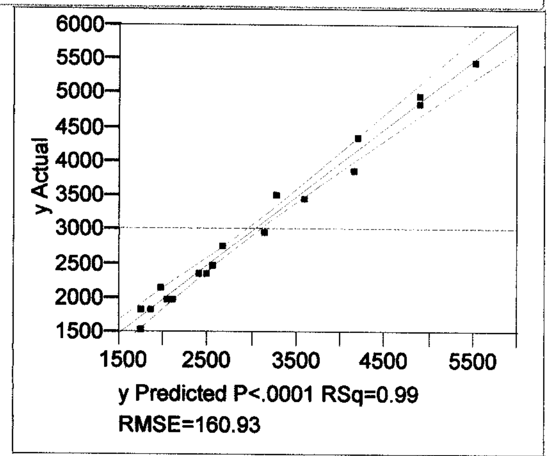
Residual by Predicted Plot



There is a systematic trend. We see that all the positive residuals correspond to either high or low sales and the negative residuals correspond to the middle sales. We do not see a random pattern of noise which suggests that perhaps this model is not appropriate for predicting sales.

p)

Actual by Predicted Plot

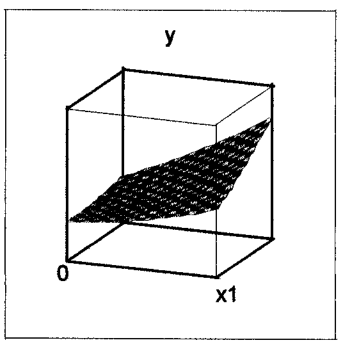


Summary of Fit

RSquare	0.986241
RSquare Adj	0.981327
Root Mean Square Error	160.9338
Mean of Response	3014.2
Observations (or Sum Wgts)	20

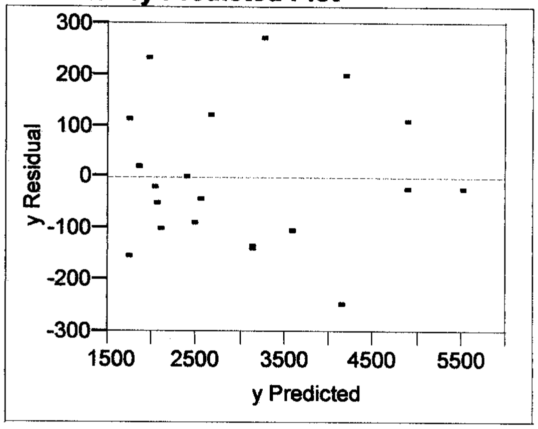
Ninety-eight point sixty-two percent of the raw variability in y is explained by this new model.

q)



We no longer see a plane.

Residual by Predicted Plot



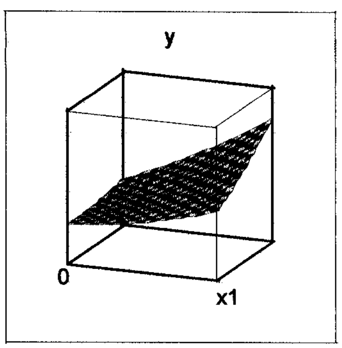
We no longer see a distance pattern of positive residuals with low and high sales and negative residuals with middle sales. Also, there are about as many positive residuals and there are negative residuals in each section of sales. This is what we should expect under the regression assumptions.

r)

$$F = \frac{(25990936 - 23747818)/3}{362596/14} = 28.9$$

This is an F with 3 and 14 df. So the p-value < 0.01. Yes, the small p-value is evidence against $\beta_3 = \beta_4 = \beta_5 = 0$. The more complicated model that allows for curvature in the mean of y is necessary.

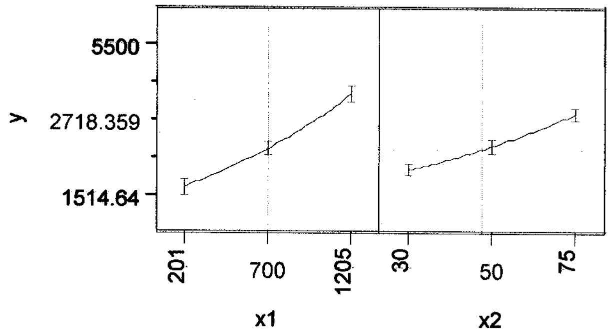
s) k)



The surface is more complicated than the plane investigated earlier. It looks like a hill will different slopes depending upon where you are and what direction you go.

s) l)

Prediction Profiler



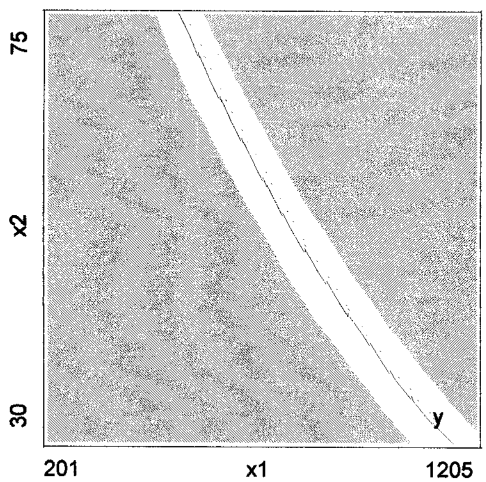
y = 2718.359 when x1 = 700 and x2 = 50. This is $212.81 less the prediction in part (j). The plane overestimates sales when x1 = 700 and x2 = 50.

s) m)

Contour Profiler

Horiz	Vert	Factor	Current X
⦿	○	x1	563.55556
○	⦿	x2	70

Response	Contour	Current Y	Lo Limit	Hi Limit
y	3000	3000.3067	2800	3200



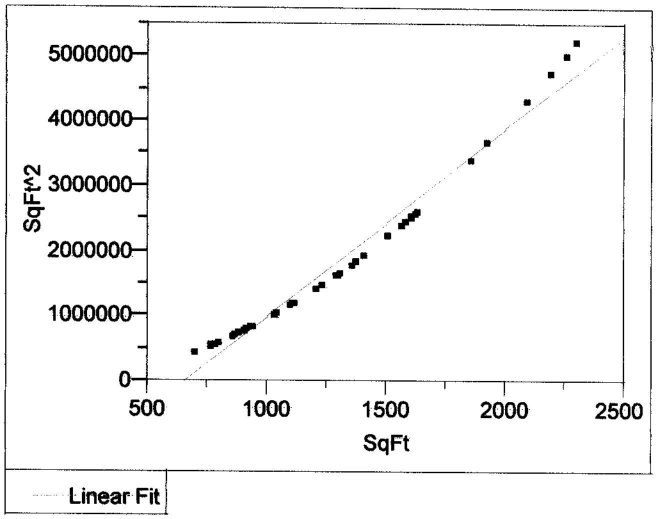
When x2 = 70 and y approximately 3000 we need a value of x1 to be approximately 564.

s) n) The above plot is useful for the same reason as provided in part n). However this is more useful to use since is corresponds with an appropriately fitting model.

Note that sales predictions differ considerably depending on which model you use. If we set x1 = 700 and x2 = 50, the plane predicts sales to be $213 more. For x2 =70, if want sales of 3000, the plane indicate we should spend $460 on advertising while the better fitting model indicates the we should spend $564.

11.51 $y = 79618$ when $x = 1000$
$y = 96778$ when $x = 1500$

11.52



Summary of Fit

RSquare	0.974047
RSquare Adj	0.973429
Root Mean Square Error	204018.6
Mean of Response	1891890
Observations (or Sum Wgts)	44

The relationship appears close to linear, but we see some curvature. The correlation is 0.987. Since these variables are highly correlated, once one is included in the model, the other variable has little explaining power to contribute.

11.52 $y = 78233$ when $x1 = 1000$
$y = 97038$ when $x1 = 1500$
The two sets of predictions are fairly similar

11.73 a) (2.108, 2.164)
b) Based on the previous confidence interval, the value 2.13 falls within the interval, so the logical conclusion is that the price is within the expected range of variability. I agree with the court ruling.