Executive Summary

The purpose of this report is to determine the *best fitting equation* to explain the level of sales at 40 Home Depot stores, and the *best store location to build* between two possible locations.

To determine the best fitting equation, we begin by examining positive/negative relationship between five independent variables (total households, advertising spending, competitors' space in square feet, poverty rating, and traffic volume) and dependent variable (sales). We assume that *households and sales are positively related, advertising and sales are positively related, competitors' space and sales are negatively related, poverty rating and sales are negatively related, and traffic volume and sales are positively related*. Observations of data listing in Appendix 1 and coefficient signs in Appendix 3 confirm our assumptions. We then look at the linearity of each independent variable.

Scatter plot in Appendix 2 shows that *all variables seem to be linear, except advertising*. To incorporate this non-linearity of advertising, we transform this variable with a square root and 1/the variable as additional variables. Before we develop any regression, we want to ensure that independent variables are not highly dependent to each other (multicollinearity problem).

Correlation matrix in Appendix 3 tells us that *the new added variables are not independent of the advertising variable* because of their high correlations. Later we will see that stepwise regression will eliminate those two newly added variables in the process. This process give us the best fitting equation by adding only significant variables, and eliminating the rest, including multicollinear variables.

Our statistical program adds advertising and households as the most significant variables to the best-fitted regression (see the step-by-step process in Appendix 4). Note that multicollinear variables are not part of the final equation. The *independent variables are significant*, as their coefficients have high t values ($t_{.05,\ 37} = 2.03$) and low p values. *None of the variables are dominant* because both have similar coefficient size (0.03 and 0.05). Both still retain positive signs, thus follow our previous assumption of *positive relationship*. At least *half/one of the independent variable, advertising, is company controlled*. The *regression is significant* with high adjusted-R square (70.5%), large F ratio (47), and very small P value. And the *forecast data fits quite well with the actual data, except for store #1* (see Appendix 5). Since the data is *not time series*, we can ignore Durbin-Watson (DW) statistic, thus serial correlation problem; and heteroscedasticity problem. Overall we are confident about the regression as the *best fitting equation* to explain sales at the 40 stores.

To select the best store location to build, we apply our best fitting regression to two alternate location data. The second location has higher sales (296.72) than the first location (294.18). This result is consistent with the coefficients and the location data. The regression indicates that household (0.05) has a higher positive impact than the advertising (0.03). The fact that the second location has higher households (941) than the first one (893) confirms our result. Thus the *second location is a better store location to build*.

Appendix 1: Listing of Variables

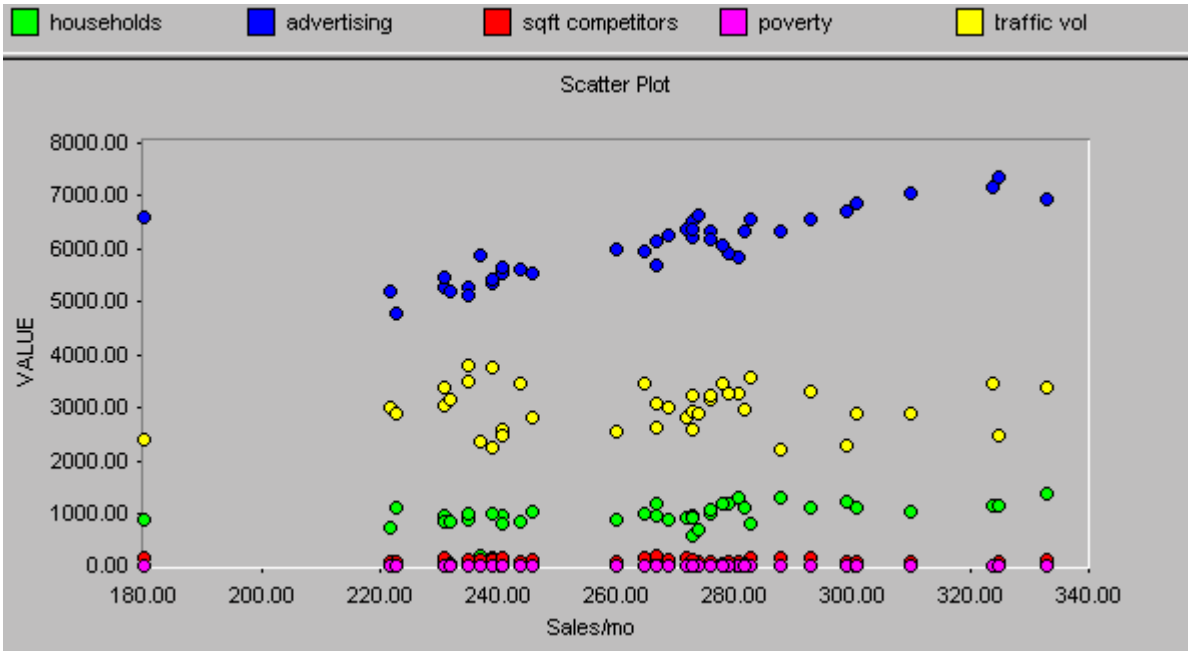
12/7/2001 19:41

FILE: StoreSales, NO. OF VARIABLES: 6, NO. OF CASES: 40
LABEL: 40storesales=F(hslds, ad, competition, poverty,traffic)

LISTING DATA MATRIX

	Sales/mo	households	advertising	sqft competitors	poverty	traffic vol
CASE 1	180	878	6575	175	7.94	2387
CASE 2	269	887	6236	134	7.59	3003
CASE 3	267	1174	5665	88	8.88	3079
CASE 4	231	957	5255	72	7.78	3030
CASE 5	265	987	5956	151	7.92	3466
CASE 6	260	871	5976	82	8.19	2563
CASE 7	310	1042	7028	82	7.77	2888
CASE 8	324	1165	7162	27	8.15	3457
CASE 9	222	745	5201	99	9.01	3004
CASE 10	283	801	6563	185	7.88	3559
CASE 11	241	946	5540	82	9.79	2592
CASE 12	333	1375	6936	129	7.93	3379
CASE 13	231	842	5446	154	8.28	3374
CASE 14	239	1006	5333	172	7.77	3743
CASE 15	281	1294	5824	94	9.06	3249
CASE 16	276	1002	6332	70	9.86	3155
CASE 17	299	1208	6716	96	7.74	2287
CASE 18	272	943	6348	152	9.66	2803
CASE 19	273	581	6500	133	6.85	3212
CASE 20	246	1044	5545	117	8.47	2796
CASE 21	239	1005	5432	144	7.95	2232
CASE 22	273	963	6215	138	10.49	2926
CASE 23	301	1104	6861	86	10.35	2888
CASE 24	267	967	6127	194	8.87	2627
CASE 25	282	1095	6335	55	9.35	2956
CASE 26	274	701	6633	100	9.11	2885
CASE 27	244	839	5616	82	8.86	3442
CASE 28	279	1200	5921	92	8.69	3252
CASE 29	241	803	5625	156	9.08	2477
CASE 30	276	1085	6160	90	8.4	3233
CASE 31	235	874	5258	137	10.24	3775
CASE 32	288	1317	6312	156	9.95	2210
CASE 33	223	1109	4774	103	9.36	2905
CASE 34	232	865	5202	68	10.41	3134
CASE 35	273	922	6364	137	9.03	2567
CASE 36	325	1142	7356	94	8.07	2476
CASE 37	235	1009	5099	144	8.09	3493
CASE 38	278	1178	6058	63	10.6	3462
CASE 39	293	1126	6559	153	9.37	3317
CASE 40	237	222	5872	115	5.84	2377

Appendix 2: Scatter plot of independent variables against dependant variable



Appendix 3: Correlation Matrix

12/7/2001 20:23

FILE: StoreSalesTransformed, NO. OF VARIABLES: 8, NO. OF(MISS. CASES: 0)
LABEL: 40storesales=F(hslds, ad, competition, poverty,traffic)

CORRELATION MATRIX

	Sales/mo	households	advertising	sqft competitors	poverty
Sales/mo	1	0.501108	0.776739	-0.268559	0.024273
households	0.501108	1	**0.215026**	**-0.19718**	**0.365073**
advertising	0.776739	0.215026	1	**-0.084159**	**-0.131506**
sqftcompetitors	-0.268559	-0.19718	-0.084159	1	**-0.18575**
poverty	0.024273	0.365073	-0.131506	-0.18575	1
traffic vol	0.070803	0.104252	-0.174262	-0.06666	0.091701
1/Advert	-0.762622	-0.192283	-0.993498	0.059165	0.125514
SqRtAdvert	0.774018	0.209627	0.999594	-0.077695	-0.129871

	traffic vol	1/Advert	SqRtAdvert
Sales/mo	0.070803	-0.762622	0.774018
households	**0.104252**	**-0.192283**	**0.209627**
advertising	**-0.174262**	***-0.993498***	***0.999594***
sqftcompetitors	**-0.06666**	**0.059165**	**-0.077695**
poverty	**0.091701**	**0.125514**	**-0.129871**
traffic vol	1	**0.181806**	**-0.176442**
1/Advert	0.181806	1	***-0.996331***
SqRtAdvert	-0.176442	-0.996331	1

Appendix 4: Stepwise Regression

```
_____________________________________________________________________12/7/2001_22:31_

  FILE: StoreSalesTransformed,  NO. OF VARIABLES: 8,  NO. OF(MISS. CASES: 0)
 LABEL:  40storesales=F(hslds, ad, competition, poverty,traffic)
_____________________________________________________________________________

                              STEPWISE REGRESSION
 DEPENDENT VARIABLE: Sales/mo
 INDEPENDENT VARIABLES:
 households      advertising     ft competitorsq poverty         raffic vol    t
 1/Advert        SqRtAdvert      CNST
 F TO ADD = 4, F TO DROP = 4, TOLERANCE = 0.001

 STEP:  1
 VARIABLE ADDED:  advertising
 R SQUARE = 0.603323 (ADJ = 0.592884),  SD. ER. EST. = 20.2985
                COEF. BETA WT.    SD. ER. F(1/38) PT. R SQ.
             --------- -------- ---------- ------- ---------
 advertising  0.039503 0.776739 0.00519615 57.7958  0.603323
        CNST 26.0148   ******** ********** *******  ********

STEP:  2
 VARIABLE ADDED:  households
 R SQUARE = 0.720349 (ADJ = 0.705233),  SD. ER. EST. = 17.2721
                 COEF. BETA WT.    SD. ER. F(1/37) PT. R SQ.
              ---------- -------- ---------- ------- ---------
 advertising  0.0356724 0.701418 0.00452733 62.0841  0.62658
  households  0.0529811 0.350285 0.0134644  15.4835  0.295016
        CNST -2.83761   ******** ********** *******  ********

VARIABLES NOT IN EQUATION:
                   PT. R SQ.   TOLERANCE   F(1/36)   P-VALUE
                  ----------- ----------- --------- ---------
 sqftcompetitors 0.0735418   0.959292    2.85766   0.0995819
         poverty 0.000563054 0.820481    0.0202814 0.887548
     traffic vol 0.0923482   0.948573    3.66279   0.063617
       1/Advert 0.000723891 0.0124844   0.026079  0.872611
      SqRtAdvert 0.00135444  0.000782058 0.0488259 0.826367

MODEL:  Sales/mo = 0.0356724advertising + 0.0529811households + -2.83761CNST

                 COEF.    SD. ER.   t(37)    P-VALUE PT. R SQ.
              ---------- ---------- ------- ---------- ---------
 advertising  0.0356724 0.00452733 7.87935 1.97544E-9  0.62658
  households  0.0529811 0.0134644  3.9349  3.52985E-4  0.295016
        CNST -2.83761   ********** ******* **********  ********

 R SQ. = 0.720349,  ADJ. R SQ. = 0.705233,  D. W. = 1.1529
 SD. ER. EST. = 17.2721,  F(2/37) = 47.6539 (P-VALUE = 5.78631E-11)
```

Appendix 5: Scatter Plot of Forecasted Sales against Actual Sales