The Real Estate Data

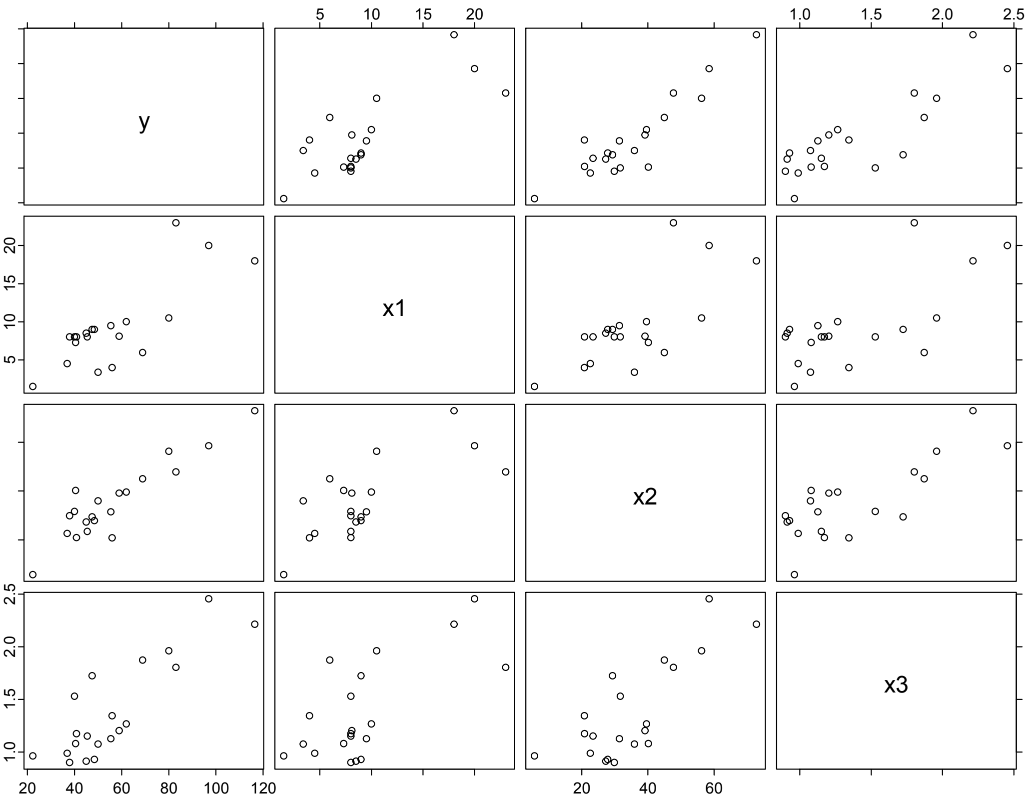
A property appraiser wants to model the relation between the sale price (Y) of a residential property in a mid-sized city and the following three predictors:

1. X_1 - Appraied Land Value (in thousand dollars)
2. X_2 - Appraied Improvements (in thousand dollars)
3. X_3 - Area (in thousand square feet).

	y	x1	x2	x3
1	68.9	5.96	44.967	1.873
2	48.5	9.00	27.860	0.928
3	55.5	9.50	31.439	1.126
4	62.0	10.00	39.592	1.265
5	116.5	18.00	72.827	2.214
6	45.0	8.50	27.317	0.912
7	38.0	8.00	29.856	0.899
8	83.0	23.00	47.752	1.803
9	59.0	8.10	39.117	1.204
10	47.5	9.00	29.349	1.725
11	40.5	7.30	40.166	1.080
12	40.0	8.00	31.679	1.529
13	97.0	20.00	58.510	2.455
14	45.5	8.00	23.454	1.151
15	40.9	8.00	20.897	1.173
16	80.0	10.50	56.248	1.960
17	56.0	4.00	20.859	1.344
18	37.0	4.50	22.610	0.988
19	50.0	3.40	35.948	1.076
20	22.4	1.50	5.779	0.962

Exploratory Data Analysis (EDA)

Paired Scatterplots



Correlation Matrix:

	y	x1	x2	x3
y	1.0000000	0.7897767	0.9156607	0.8489815
x1	0.7897767	1.0000000	0.7289003	0.6889384
x2	0.9156607	0.7289003	1.0000000	0.7881460
x3	0.8489815	0.6889384	0.7881460	1.0000000

Multiple Linear Regression Models

Model I: $y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

Table of Parameter Estimates

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	25.9903	6.4612	4.0225	0.0008
x1	3.3290	0.6094	5.4626	0.0000

Analysis of Variance Table

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
x1	1	6102.224	6102.224	29.84018	0.00003449665
Residuals	18	3680.944	204.497		
Total	19	9783.168			

Model II: $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

Table of Parameter Estimates

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	8.7235	4.9444	1.7643	0.0956
x1	1.1003	0.5376	2.0466	0.0565
x2	1.0705	0.1882	5.6872	0.0000

Analysis of Variance Table

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Model	2	8515.009	4257.505	57.07264	2.87024e-08
Residuals	17	1268.159	74.598		
Total	19	9783.168			

Model III: $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$

Table of Parameter Estimates

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	1.4703	5.7463	0.2559	0.8013
x1	0.8145	0.5122	1.5901	0.1314
x2	0.8204	0.2112	3.8850	0.0013
x3	13.5286	6.5857	2.0543	0.0567

Analysis of Variance Table

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Model	3	8779.677	2926.559	46.66219	3.8998e-08
Residuals	16	1003.491	62.718		
Total	19	9783.168			

Model IV: $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \beta_4 X_{i2}^2 + \beta_5 X_{i3} + \varepsilon_i$

(Polynomial Regression)

Table of Parameter Estimates

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.647940	11.075654	1.593	0.133
x1	0.309235	1.601071	0.193	0.850
x1.sq	0.020606	0.057947	0.356	0.727
x2	0.210999	0.512521	0.412	0.687
x2.sq	0.008899	0.006042	1.473	0.163
x3	9.615461	6.811813	1.412	0.180

Analysis of Variance Table

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Model	5	8956.068	1791.214	30.308	4.9e-07
Residuals	14	827.1	59.1		
Total	19	9783.168			

Model V: $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$ (Interaction)

Table of Parameter Estimates

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	25.5807	8.6580	2.9546	0.0093
x1	-1.0523	1.0650	-0.9881	0.3378
x2	0.6470	0.2518	2.5699	0.0206
x1:x2	0.0469	0.0207	2.2669	0.0376

Analysis of Variance Table

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Model	3	8823.289	2941.096	49.02481	2.739e-08
Residuals	16	959.879	59.992		
Total	19	9783.168			

Confounding Effect of X_1 and X_2 on X_3

	$\hat{\beta}$	**s.e.($\hat{\beta}$)**
Crude	41.3636	6.0683
X_1	28.2735	7.2787
X_2	16.3732	6.6169
X_1, X_2	13.5286	6.5857
X_1, X_2, X_1^2	11.9626	6.8759

Model Selection: All Possible Regression with the Real-Estate Example

model	SSE	df	p	MSE	R2	R2.a	AIC	BIC
1	9783.168	19	0	514.9036	0	0	187.7684	189.7598
x1	3680.944	18	1	204.4969	0.62375	0.60284	170.2185	173.2057
x2	1580.622	18	1	87.81232	0.83843	0.82946	153.3115	156.2987
x3	2731.757	18	1	151.7643	0.72077	0.70526	164.254	167.2412
x1 + x2	1268.159	17	2	74.5976	0.87037	0.85512	150.9064	154.8894
x1 + x3	1950.091	17	2	114.7112	0.80067	0.77722	159.5126	163.4956
x2 + x3	1162.073	17	2	68.35726	0.88122	0.86724	149.1592	153.1422
x1 + x2 + x3	1003.491	16	3	62.7182	0.89743	0.87819	148.2248	153.2035
x1 + x2 + x3 + x1:x2	839.9916	15	4	55.9994	0.91414	0.89124	146.6678	152.6422
x1 + x2 + x3 + x1:x3	971.0894	15	4	64.7393	0.90074	0.87427	149.5684	155.5428
x1 + x2 + x3 + x2:x3	856.9413	15	4	57.1294	0.91241	0.88905	147.0674	153.0418

Model Diagnostics: the Real-Estate Example

	x1	x2	x3	y	yhat	h	e	z	r	r.jack	cook.d
1	5.96	44.97	1.873	68.9	68.6	0.2972	0.343	0.04334	0.05170	0.0501	2.83e-04
2	9.00	27.86	0.928	48.5	44.2	0.1412	4.287	0.54134	0.58416	0.5717	1.40e-02
3	9.50	31.44	1.126	55.5	50.2	0.0851	5.265	0.66480	0.69502	0.6833	1.12e-02
4	10.00	39.59	1.265	62.0	59.2	0.0861	2.788	0.35205	0.36826	0.3581	3.20e-03
5	18.00	72.83	2.214	116.5	105.8	0.3673	10.666	1.34680	1.69319	1.8095	4.16e-01
6	8.50	27.32	0.912	45.0	43.1	0.1369	1.856	0.23440	0.25231	0.2448	2.52e-03
7	8.00	29.86	0.899	38.0	44.6	0.1453	-6.644	-0.83890	-0.90743	-0.9021	3.50e-02
8	23.00	47.75	1.803	83.0	83.8	0.5279	-0.774	-0.09768	-0.14217	-0.1377	5.65e-03
9	8.10	39.12	1.204	59.0	56.4	0.1054	2.551	0.32206	0.34051	0.3309	3.42e-03
10	9.00	29.35	1.725	47.5	56.2	0.2086	-8.717	-1.10068	-1.23723	-1.2597	1.01e-01
11	7.30	40.17	1.080	40.5	55.0	0.1797	-14.481	-1.82853	-2.01895	-2.2644	2.23e-01
12	8.00	31.68	1.529	40.0	54.7	0.0926	-14.662	-1.85143	-1.94356	-2.1531	9.63e-02
13	20.00	58.51	2.455	97.0	99.0	0.3807	-1.977	-0.24965	-0.31724	-0.3081	1.55e-02
14	8.00	23.45	1.151	45.5	42.8	0.0941	2.700	0.34088	0.35816	0.3482	3.33e-03
15	8.00	20.90	1.173	40.9	41.0	0.1244	-0.100	-0.01264	-0.01351	-0.0131	6.49e-06
16	10.50	56.25	1.960	80.0	82.7	0.2296	-2.687	-0.33928	-0.38654	-0.3760	1.11e-02
17	4.00	20.86	1.344	56.0	40.0	0.1844	15.976	2.01725	2.23375	2.6072	2.82e-01
18	4.50	22.61	0.988	37.0	37.1	0.0969	-0.052	-0.00657	-0.00691	-0.0067	1.28e-06
19	3.40	35.95	1.076	50.0	48.3	0.2151	1.710	0.21596	0.24376	0.2365	4.07e-03
20	1.50	5.78	0.962	22.4	20.4	0.3014	1.952	0.24649	0.29490	0.2863	9.38e-03