Using Y to represent per capita expenditure on schools, the model takes the form

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \gamma_1 T_1 + \gamma_2 T_2 + \delta_1 T_1 \cdot X_1 \\ &\quad + \delta_2 T_1 \cdot X_2 + \delta_3 T_1 \cdot X_3 + \alpha_1 T_2 \cdot X_1 + \alpha_2 T_2 \cdot X_2 \\ &\quad + \alpha_3 T_2 \cdot X_3 + \varepsilon. \end{aligned}$$

From the definitions of T_1 and T_2, the above model is equivalent to

$$\begin{aligned} \text{For 1960}: \quad & Y = (\beta_0 + \gamma_1) + (\beta_1 + \delta_1)X_1 + (\beta_2 + \delta_2)X_2 \\ & \qquad + (\beta_3 + \delta_3)X_3 + \varepsilon, \\ \text{For 1970}: \quad & Y = (\beta_0 + \gamma_2) + (\beta_1 + \alpha_1)X_1 + (\beta_2 + \alpha_2)X_2 \\ & \qquad + (\beta_3 + \alpha_3)X_3 + \varepsilon, \\ \text{For 1975}: \quad & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon. \end{aligned}$$

As noted earlier, this method of analysis necessarily implies that the variability about the regression function is assumed to be equal for all three years. One formal hypothesis of interest is

$$H_0 : \gamma_1 = \gamma_2 = \delta_1 = \delta_2 = \delta_3 = \alpha_1 = \alpha_2 = \alpha_3 = 0,$$

which implies that the regression system has remained unchanged throughout the period of investigation (1960–1975).

The data for this example, which we refer to as the Education Expenditures data, appear in Tables 5.12, 5.13, and 5.14 and can be obtained from the book's Web Site. The reader is invited to perform the analysis described above as an exercise.

EXERCISES

5.1 Using the model defined in (5.6):

(a) Check to see if the usual least squares assumptions hold.

(b) Test $H_0 : \gamma = 0$ using the F-test.

(c) Test $H_0 : \gamma = 0$ using the t-test.

(d) Verify the equivalence of the two tests above.

5.2 Using the model defined in (5.8):

(a) Check to see if the usual least squares assumptions hold.

(b) Test $H_0 : \delta = 0$ using the F-test.

(c) Test $H_0 : \delta = 0$ using the t-test.

(d) Verify the equivalence of the two tests above.

5.3 Perform a thorough analysis of the Ski Sales data in Table 5.11 using the ideas presented in Section 5.6.

5.4 Perform a thorough analysis of the Education Expenditures data in Tables 5.12, 5.13, and 5.14 using the ideas presented in Section 5.7.

Table 5.12 Education Expenditures Data (1960)

Row	STATE	Y	X_1	X_2	X_3	Region
1	ME	61	1704	388	399	1
2	NH	68	1885	372	598	1
3	VT	72	1745	397	370	1
4	MA	72	2394	358	868	1
5	RI	62	1966	357	899	1
6	CT	91	2817	362	690	1
7	NY	104	2685	341	728	1
8	NJ	99	2521	353	826	1
9	PA	70	2127	352	656	1
10	OH	82	2184	387	674	2
11	IN	84	1990	392	568	2
12	IL	84	2435	366	759	2
13	MI	104	2099	403	650	2
14	WI	84	1936	393	621	2
15	MN	103	1916	402	610	2
16	IA	86	1863	385	522	2
17	MO	69	2037	364	613	2
18	ND	94	1697	429	351	2
19	SD	79	1644	411	390	2
20	NB	80	1894	379	520	2
21	KS	98	2001	380	564	2
22	DE	124	2760	388	326	3
23	MD	92	2221	393	562	3
24	VA	67	1674	402	487	3
25	WV	66	1509	405	358	3
26	NC	65	1384	423	362	3
27	SC	57	1218	453	343	3
28	GA	60	1487	420	498	3
29	FL	74	1876	334	628	3
30	KY	49	1397	594	377	3
31	TN	60	1439	346	457	3
32	AL	59	1359	637	517	3
33	MS	68	1053	448	362	3
34	AR	56	1225	403	416	3
35	LA	72	1576	433	562	3
36	OK	80	1740	378	610	3
37	TX	79	1814	409	727	3
38	MT	95	1920	412	463	4
39	ID	79	1701	418	414	4
40	WY	142	2088	415	568	4
41	CO	108	2047	399	621	4
42	NM	94	1838	458	618	4
43	AZ	107	1932	425	699	4
44	UT	109	1753	494	665	4
45	NV	114	2569	372	663	4
46	WA	112	2160	386	584	4
47	OR	105	2006	382	534	4
48	CA	129	2557	373	717	4
49	AK	107	1900	434	379	4
50	HI	77	1852	431	693	4

Table 5.13 Education Expenditures Data (1970)

Row	STATE	Y	X_1	X_2	X_3	Region
1	ME	189	2828	351	508	1
2	NH	169	3259	346	564	1
3	VT	230	3072	348	322	1
4	MA	168	3835	335	846	1
5	RI	180	3549	327	871	1
6	CT	193	4256	341	774	1
7	NY	261	4151	326	856	1
8	NJ	214	3954	333	889	1
9	PA	201	3419	326	715	1
10	OH	172	3509	354	753	2
11	IN	194	3412	359	649	2
12	IL	189	3981	349	830	2
13	MI	233	3675	369	738	2
14	WI	209	3363	361	659	2
15	MN	262	3341	365	664	2
16	IA	234	3265	344	572	2
17	MO	177	3257	336	701	2
18	ND	177	2730	369	443	2
19	SD	187	2876	369	446	2
20	NB	148	3239	350	615	2
21	KS	196	3303	340	661	2
22	DE	248	3795	376	722	3
23	MD	247	3742	364	766	3
24	VA	180	3068	353	631	3
25	WV	149	2470	329	390	3
26	NC	155	2664	354	450	3
27	SC	149	2380	377	476	3
28	GA	156	2781	371	603	3
29	FL	191	3191	336	805	3
30	KY	140	2645	349	523	3
31	TN	137	2579	343	588	3
32	AL	112	2337	362	584	3
33	MS	130	2081	385	445	3
34	AR	134	2322	352	500	3
35	LA	162	2634	390	661	3
36	OK	135	2880	330	680	3
37	TX	155	3029	369	797	3
38	MT	238	2942	369	534	4
39	ID	170	2668	368	541	4
40	WY	238	3190	366	605	4
41	CO	192	3340	358	785	4
42	NM	227	2651	421	698	4
43	AZ	207	3027	387	796	4
44	UT	201	2790	412	804	4
45	NV	225	3957	385	809	4
46	WA	215	3688	342	726	4
47	OR	233	3317	333	671	4
48	CA	273	3968	348	909	4
49	AK	372	4146	440	484	4
50	HI	212	3513	383	831	4

Table 5.14 Education Expenditures Data (1975)

Row	STATE	Y	X_1	X_2	X_3	Region
1	ME	235	3944	325	508	1
2	NH	231	4578	323	564	1
3	VT	270	4011	328	322	1
4	MA	261	5233	305	846	1
5	RI	300	4780	303	871	1
6	CT	317	5889	307	774	1
7	NY	387	5663	301	856	1
8	NJ	285	5759	310	889	1
9	PA	300	4894	300	715	1
10	OH	221	5012	324	753	2
11	IN	264	4908	329	649	2
12	IL	308	5753	320	830	2
13	MI	379	5439	337	738	2
14	WI	342	4634	328	659	2
15	MN	378	4921	330	664	2
16	IA	232	4869	318	572	2
17	MO	231	4672	309	701	2
18	ND	246	4782	333	443	2
19	SD	230	4296	330	446	2
20	NB	268	4827	318	615	2
21	KS	337	5057	304	661	2
22	DE	344	5540	328	722	3
23	MD	330	5331	323	766	3
24	VA	261	4715	317	631	3
25	WV	214	3828	310	390	3
26	NC	245	4120	321	450	3
27	SC	233	3817	342	476	3
28	GA	250	4243	339	603	3
29	FL	243	4647	287	805	3
30	KY	216	3967	325	523	3
31	TN	212	3946	315	588	3
32	AL	208	3724	332	584	3
33	MS	215	3448	358	445	3
34	AR	221	3680	320	500	3
35	LA	244	3825	355	661	3
36	OK	234	4189	306	680	3
37	TX	269	4336	335	797	3
38	MT	302	4418	335	534	4
39	ID	268	4323	344	541	4
40	WY	323	4813	331	605	4
41	CO	304	5046	324	785	4
42	NM	317	3764	366	698	4
43	AZ	332	4504	340	796	4
44	UT	315	4005	378	804	4
45	NV	291	5560	330	809	4
46	WA	312	4989	313	726	4
47	OR	316	4697	305	671	4
48	CA	332	5438	307	909	4
49	AK	546	5613	386	484	4
50	HI	311	5309	333	831	4

Table 5.15 Corn Yields by Fertilizer Group

Fertilizer 1	Fertilizer 2	Fertilizer 3	Control Group
31	27	36	33
34	27	37	27
34	25	37	35
34	34	34	25
43	21	37	29
35	36	28	20
38	34	33	25
36	30	29	40
36	32	36	35
45	33	42	29

5.5 Three types of fertilizer are to be tested to see which one yields more corn crop. Forty similar plots of land were available for testing purposes. The 40 plots are divided at random into four groups, ten plots in each group. Fertilizer 1 was applied to each of the ten corn plots in Group 1. Similarly, Fertilizers 2 and 3 were applied to the plots in Groups 2 and 3, respectively. The corn plants in Group 4 were not given any fertilizer; it will serve as the control group. Table 5.15 gives the corn yield y_{ij} for each of the forty plots.

(a) Create three indicator variables F_1, F_2, F_3, one for each of the three fertilizer groups.

(b) Fit the model $y_{ij} = \mu_0 + \mu_1 F_{i1} + \mu_2 F_{i2} + \mu_3 F_{i3} + \varepsilon_{ij}$.

(c) Test the hypothesis that, on the average, none of the three types of fertilizer has an effect on corn crops. Specify the hypothesis to be tested, the test used, and your conclusions at the 5% significance level.

(d) Test the hypothesis that, on the average, the three types of fertilizer have equal effects on corn crop but different from that of the control group. Specify the hypothesis to be tested, the test used, and your conclusions at the 5% significance level.

(e) Which of the three fertilizers has the greatest effects on corn yield?

5.6 In a statistics course personal information was collected on all the students for class analysis. Data on age (in years), height (in inches), and weight (in pounds) of the students are given in Table 5.16 and can be obtained from the book's Web Site. The sex of each student is also noted and coded as 1 for women and 0 for men. We want to study the relationship between the height and weight of students. Weight is taken as the response variable, and the height as the predictor variable.

(a) Do you agree or do you think the roles of the variables should be reversed?

(b) Is a single equation adequate to describe the relationship between height and weight for the two groups of students? Examine the standardized

Table 5.16 Class Data on Age (in Years), Height (in Inches), Weight (in Pounds), and Sex (1 = Female, 0 = Male)

Age	Height	Weight	Sex	Age	Height	Weight	Sex
19	61	180	0	19	65	135	1
19	70	160	0	19	70	120	0
19	70	135	0	21	69	142	0
19	71	195	0	20	63	108	1
19	64	130	1	19	63	118	1
19	64	120	1	20	72	135	0
21	69	135	1	19	73	169	0
19	67	125	0	19	69	145	0
19	62	120	1	27	69	130	1
20	66	145	0	18	64	135	0
19	65	155	0	20	61	115	1
19	69	135	1	19	68	140	0
19	66	140	0	21	70	152	0
19	63	120	1	19	64	118	1
19	69	140	0	19	62	112	1
18	66	113	1	19	64	100	1
18	68	180	0	20	67	135	1
19	72	175	0	20	63	110	1
19	70	169	0	20	68	135	0
19	74	210	0	18	63	115	1
20	66	104	1	19	68	145	0
20	64	105	1	19	65	115	1
20	65	125	1	19	63	128	1
20	71	120	1	20	68	140	1
19	69	119	1	19	69	130	0
20	64	140	1	19	69	165	0
20	67	185	1	19	69	130	0
19	60	110	1	20	70	180	0
20	66	120	1	28	65	110	1
19	71	175	0	19	55	155	0

residual plot from the model fitted to the pooled data, distinguishing between the male and female students.

(c) Find the best model that describes the relationship between the weight and the height of students. Use interaction variables and the methodology described in this chapter.

(d) Do you think we should include age as a variable to predict weight? Give an intuitive justification for your answer.

5.7 Presidential Election Data (1916–1996): The data in Table 5.17 were kindly provided by Professor Ray Fair of Yale University, who has found that the proportion of votes obtained by a presidential candidate in a United States

Table 5.17 Presidential Election Data (1916–1996)

Year	V	I	D	W	G	P	N
1916	0.5168	1	1	0	2.229	4.252	3
1920	0.3612	1	0	1	−11.463	16.535	5
1924	0.4176	−1	−1	0	−3.872	5.161	10
1928	0.4118	−1	0	0	4.623	0.183	7
1932	0.5916	−1	−1	0	−14.901	7.069	4
1936	0.6246	1	1	0	11.921	2.362	9
1940	0.5500	1	1	0	3.708	0.028	8
1944	0.5377	1	1	1	4.119	5.678	14
1948	0.5237	1	1	1	1.849	8.722	5
1952	0.4460	1	0	0	0.627	2.288	6
1956	0.4224	−1	−1	0	−1.527	1.936	5
1960	0.5009	−1	0	0	0.114	1.932	5
1964	0.6134	1	1	0	5.054	1.247	10
1968	0.4960	1	0	0	4.836	3.215	7
1972	0.3821	−1	−1	0	6.278	4.766	4
1976	0.5105	−1	0	0	3.663	7.657	4
1980	0.4470	1	1	0	−3.789	8.093	5
1984	0.4083	−1	−1	0	5.387	5.403	7
1988	0.4610	−1	0	0	2.068	3.272	6
1992	0.5345	−1	−1	0	2.293	3.692	1
1996	0.5474	1	1	0	2.918	2.268	3

presidential election can be predicted accurately by three macroeconomic variables, incumbency, and a variable which indicates whether the election was held during or just after a war. The variables considered are given in Table 5.18. All growth rates are annual rates in percentage points. Consider fitting the initial model

$$\begin{aligned} V &= \beta_0 + \beta_1 \cdot I + \beta_2 \cdot D + \beta_3 \cdot W + \beta_4 \cdot (G \cdot I) \\ &\quad + \beta_5 \cdot P + \beta_6 \cdot N + \varepsilon \end{aligned} \tag{5.11}$$

to the data.

(a) Do we need to keep the variable I in the above model?

(b) Do we need to keep the interaction variable $(G \cdot I)$ in the above model?

(c) Examine different models to produce the model or models that might be expected to perform best in predicting future presidential elections. Include interaction terms if needed.

Table 5.18 Variables for the Presidential Election Data (1916–1996) in Table 5.17

Variable	Definition
YEAR	Election year
V	Democratic share of the two-party presidential vote
I	Indicator variable (1 if there is a Democratic incumbent at the time of the election and -1 if there is a Republican incumbent)
D	Indicator variable (1 if a Democratic incumbent is running for election, -1 if a Republican incumbent is running for election, and 0 otherwise)
W	Indicator variable (1 for the elections of 1920, 1944, and 1948, and 0 otherwise)
G	Growth rate of real per capita GDP in the first three quarters of the election year
P	Absolute value of the growth rate of the GDP deflator in the first 15 quarters of the administration
N	Number of quarters in the first 15 quarters of the administration in which the growth rate of real per capita GDP is greater than 3.2%