Exercises

1. A regression model is to be developed for predicting the ability of soil to absorb chemical contaminants. Ten observations have been taken on a soil absorption index (y) and two regressors: x_1 = amount of extractable iron ore and x_2 = amount of bauxite. We wish to fit the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1.17991 & -7.30982\text{ E-3} & 7.3006\text{ E-4} \\ -7.30982\text{ E-3} & 7.9799\text{ E-5} & -1.23713\text{ E-4} \\ 7.3006\text{ E-4} & 1.23713\text{ E-4} & 4.6576\text{ E-4} \end{bmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 220 \\ 36{,}768 \\ 9{,}965 \end{bmatrix}$$

(a) Estimate the regression coefficients in the model specified above.

(b) What is the predicted value of the absorption index y when $x_1 = 200$ and $x_2 = 50$?

2. The data in the following Table are the 1976 team performance statistics for the teams in the National Football League (Source: The Sporting News).

(a) Fit a multiple regression model relating the number of games won to the teams' passing yardage (x_2), the percent rushing plays (x_7), and the opponents' yards rushing (x_8).

(b) Estimate σ^2

(c) What are the standard errors of the regression coefficients?

(d) Use the model to predict the number of games won when x_2= 2000 yards, x_7= 60%, and x_8 = 1800.

(e) Test for significance of regression using $\alpha = 0.05$. What is the *P*-value for this test?

(f) Conduct the *t*-test for each regression coefficient β_2, β_7, β_8 and . Using $\alpha = 0.05$, what conclusions can you draw about the variables in this model?

(g) What proportion of total variability is explained by this model?

(h) Construct a normal probability plot of the residuals. What conclusion can you draw from this plot?

(i) Plot the residuals versus and versus each regressor, and comment on model adequacy.

(j) Are there any influential points in these data?

Team	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Washington	10	2113	1985	38.9	64.7	+4	868	59.7	2205	1917
Minnesota	11	2003	2855	38.8	61.3	+3	615	55.0	2096	1575
New England	11	2957	1737	40.1	60.0	+14	914	65.6	1847	2175
Oakland	13	2285	2905	41.6	45.3	−4	957	61.4	1903	2476
Pittsburgh	10	2971	1666	39.2	53.8	+15	836	66.1	1457	1866
Baltimore	11	2309	2927	39.7	74.1	+8	786	61.0	1848	2339
Los Angeles	10	2528	2341	38.1	65.4	+12	754	66.1	1564	2092
Dallas	11	2147	2737	37.0	78.3	−1	797	58.9	2476	2254
Atlanta	4	1689	1414	42.1	47.6	−3	714	57.0	2577	2001
Buffalo	2	2566	1838	42.3	54.2	−1	797	58.9	2476	2254
Chicago	7	2363	1480	37.3	48.0	+19	984	68.5	1984	2217
Cincinnati	10	2109	2191	39.5	51.9	+6	819	59.2	1901	1686
Cleveland	9	2295	2229	37.4	53.6	−5	1037	58.8	1761	2032
Denver	9	1932	2204	35.1	71.4	+3	986	58.6	1709	2025
Detroit	6	2213	2140	38.8	58.3	+6	819	59.2	1901	1686
Green Bay	5	1722	1730	36.6	52.6	−19	791	54.4	2288	1835
Houston	5	1498	2072	35.3	59.3	−5	776	49.6	2072	1914
Kansas City	5	1873	2929	41.1	55.3	+10	789	54.3	2861	2496
Miami	6	2118	2268	38.6	69.6	+6	582	58.7	2411	2670
New Orleans	4	1775	1983	39.3	78.3	+7	901	51.7	2289	2202
New York Giants	3	1904	1792	39.7	38.1	−9	734	61.9	2203	1988
New York Jets	3	1929	1606	39.7	68.8	−21	627	52.7	2592	2324
Philadelphia	4	2080	1492	35.5	68.8	−8	722	57.8	2053	2550
St. Louis	10	2301	2835	35.3	74.1	+2	683	59.7	1979	2110
San Diego	6	2040	2416	38.7	50.0	0	576	54.9	2048	2628
San Francisco	8	2447	1638	39.9	57.1	−8	848	65.3	1786	1776
Seattle	2	1416	2649	37.4	56.3	−22	684	43.8	2876	2524
Tampa Bay	0	1503	1503	39.3	47.0	−9	875	53.5	2560	2241

3. The pull strength of a wire bond is an important characteristic. The following table gives information on pull strength (y), die height (x_1), post height (x_2), loop height (x_3), wire length (x_4), bond width on the die (x_5), and bond width on the post (x_6).

(a) Fit a multiple linear regression model using x_2, x_3, x_4, and as the regressors.

(b) Estimate σ^2.

(c) Find the $Se(\widehat{\beta}_j)$. How precisely are the regression coefficients estimated, in your opinion?

(d) Use the model from part (a) to predict pull strength when $x_2 = 20$, $x_3 = 30$, $x_4 = 90$, and $x_5 = 2.0$.

y	x_1	x_2	x_3	x_4	x_5	x_6
8.0	5.2	19.6	29.6	94.9	2.1	2.3
8.3	5.2	19.8	32.4	89.7	2.1	1.8
8.5	5.8	19.6	31.0	96.2	2.0	2.0
8.8	6.4	19.4	32.4	95.6	2.2	2.1
9.0	5.8	18.6	28.6	86.5	2.0	1.8
9.3	5.2	18.8	30.6	84.5	2.1	2.1
9.3	5.6	20.4	32.4	88.8	2.2	1.9
9.5	6.0	19.0	32.6	85.7	2.1	1.9
9.8	5.2	20.8	32.2	93.6	2.3	2.1
10.0	5.8	19.9	31.8	86.0	2.1	1.8
10.3	6.4	18.0	32.6	87.1	2.0	1.6
10.5	6.0	20.6	33.4	93.1	2.1	2.1
10.8	6.2	20.2	31.8	83.4	2.2	2.1
11.0	6.2	20.2	32.4	94.5	2.1	1.9
11.3	6.2	19.2	31.4	83.4	1.9	1.8
11.5	5.6	17.0	33.2	85.2	2.1	2.1
11.8	6.0	19.8	35.4	84.1	2.0	1.8
12.3	5.8	18.8	34.0	86.9	2.1	1.8
12.5	5.6	18.6	34.2	83.0	1.9	2.0

4. The compressive strength of concrete is being studied, and four different mixing techniques are being investigated. The following data have been collected.

Mixing Technique	Compressive Strength (psi)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

(a) Test the hypothesis that mixing techniques affect the strength of the concrete. Use $\alpha = 0.05$

(b) Find the *P*-value for the *F*-statistic computed in part (a).

(c) Analyze the residuals from this experiment.

(d) Use Fisher's LSD method with $\alpha = 0.05$ to analyse the mean compressive strength of the four mixing techniques.

5. An article in Lubrication Engineering (December 1990) describes the results of an experiment designed to investigate the effects of carbon material properties on the progression of blisters on carbon face seals. The carbon face seals are used extensively in equipment such as air turbine starters. Five different carbon materials were tested, and the surface roughness was measured. The data are as follows:

Carbon Material Type	Surface Roughness					
EC10	0.50	0.55	0.55	0.36		
EC10A	0.31	0.07	0.25	0.18	0.56	0.20
EC4	0.20	0.28	0.12			
EC1	0.10	0.16				

(a) Does carbon material type have an effect on mean surface roughness? Use $\alpha = 0.05$.

(b) Find the residuals for this experiment. Does a normal probability plot of the residuals indicate any problem with the normality assumption?

(c) Plot the residuals versus $\hat{y}_{ij}$. Comment on the plot.

(d) Find a 95% confidence interval on the difference between the mean surface roughness between the EC10 and the EC1 carbon grades.

6. An article in the Journal of Agricultural Engineering Research (Vol. 52, 1992, pp. 53–76) describes an experiment to investigate the effect of drying temperature of wheat grain on the baking quality of bread. Three temperature levels were used, and the response variable measured was the volume of the loaf of bread produced. The data are as follows:

Temperature (°C)	Volume (CC)				
70.0	1245	1235	1285	1245	1235
75.0	1235	1240	1200	1220	1210
80.0	1225	1200	1170	1155	1095

(a) Does drying temperature affect mean bread volume? Use $\alpha = 0.01$

(b) Find the P-value for this test.

(c) Use the Fisher's LSD method to determine which means are different.

(d) Analyze the residuals from this experiment and comment on model adequacy.

7. An article in Industrial Quality Control (1956, pp. 5–8) describes an experiment to investigate the effect of two factors (glass type and phosphor type) on the brightness of a television tube. The response variable measured is the current (in microamps) necessary to obtain a specified brightness level. The data are shown in the following table:

Glass Type	**Phosphor Type**		
	1	2	3
1	280	300	290
	290	310	285
	285	295	290
2	230	260	220
	235	240	225
	240	235	230

(a) State the hypotheses of interest in this experiment.

(b) Test the above hypotheses and draw conclusions using the analysis of variance with $\alpha = 0.05$.

(c) Analyze the residuals from this experiment.

8. An engineer who suspects that the surface finish of metal parts is influenced by the type of paint used and the drying time. He selected three drying times—20, 25, and 30 minutes— and used two types of paint. Three parts are tested with each combination of paint type and drying time. The data are as follows:

	Drying Time (min)		
Paint	20	25	30
1	74	73	78
	64	61	85
	50	44	92
2	92	98	66
	86	73	45
	68	88	85

(a) State and test the appropriate hypotheses using the analysis of variance with $\alpha = 0.05$.

(b) Analyze the residuals from this experiment.

9. The percentage of hardwood concentration in raw pulp, the freeness, and the cooking time of the pulp are being investigated for their effects on the strength of paper. The data from a three-factor factorial experiment are shown in the following table.

Percentage of Hardwood Concentration	Cooking Time 1.5 hours			Cooking Time 2.0 hours		
	Freeness			Freeness		
	350	500	650	350	500	650
10	96.6	97.7	99.4	98.4	99.6	100.6
	96.0	96.0	99.8	98.6	100.4	100.9
15	98.5	96.0	98.4	97.5	98.7	99.6
	97.2	96.9	97.6	98.1	96.0	99.0
20	97.5	95.6	97.4	97.6	97.0	98.5
	96.6	96.2	98.1	98.4	97.8	99.8

(a) Analyze the data using the analysis of variance assuming that all factors are fixed. Use $\alpha = 0.05$.

(b) Find *P*-values for the F-ratios in part (a).

(c) The residuals are found by $e_{ijkl} = y_{ijkl} - \bar{y}_{ijk\cdot}$. Graphically analyze the residuals from this experiment.

10. An engineer is interested in the effect of cutting speed (*A*), metal hardness (*B*), and cutting angle (*C*) on the life of a cutting tool. Two levels of each factor are chosen, and two replicates of a 2^3 factorial design are run. The tool life data (in hours) are shown in the following table:

Treatment Combination	Replicate	
	I	II
(1)	221	311
a	325	435
b	354	348
ab	552	472
c	440	453
ac	406	377
bc	605	500
abc	392	419

(a) Analyze the data from this experiment.

(b) Find an appropriate regression model that explains tool life in terms of the variables used in the experiment.

(c) Analyze the residuals from this experiment.

11. The data shown here represent a single replicate of a 2^5 design that is used in an experiment to study the compressive strength of concrete. The factors are mix (A), time (B), laboratory (C), temperature (D), and drying time (E).

$$
\begin{array}{llrllr}
(1) & = & 700 & e & = & 800 \\
a & = & 900 & ae & = & 1200 \\
b & = & 3400 & be & = & 3500 \\
ab & = & 5500 & abe & = & 6200 \\
c & = & 600 & ce & = & 600 \\
ac & = & 1000 & ace & = & 1200 \\
bc & = & 3000 & bce & = & 3006 \\
abc & = & 5300 & abce & = & 5500 \\
d & = & 1000 & de & = & 1900 \\
ad & = & 1100 & ade & = & 1500 \\
bd & = & 3000 & bde & = & 4000 \\
abd & = & 6100 & abde & = & 6500 \\
cd & = & 800 & cde & = & 1500 \\
acd & = & 1100 & acde & = & 2000 \\
bcd & = & 3300 & bcde & = & 3400 \\
abcd & = & 6000 & abcde & = & 6800
\end{array}
$$

(a) Estimate the factor effects.

(b) Which effects appear important? Use a normal probability plot.

(c) If it is desirable to maximize the strength, in which direction would you adjust the process variables?

(d) Analyze the residuals from this experiment.

12. An experiment has run a single replicate of a 2^4 design and calculated the following factor effects:

$$
\begin{array}{lll}
A = 80.25 & AB = 53.25 & ABC = -2.95 \\
B = -65.50 & AC = 11.00 & ABD = -8.00 \\
C = -9.25 & AD = 9.75 & ACD = 10.25 \\
D = -20.50 & BC = 18.36 & BCD = -7.95 \\
 & BD = 15.10 & ABCD = -6.25 \\
 & CD = -1.25 &
\end{array}
$$

(a) Construct a normal probability plot of the effects.

(b) Identify a tentative model, based on the plot of effects in part (a).

(c) Estimate the regression coefficients in this model, assuming that $\bar{y} = 400$