CHAPTER 8

8.1

(a) The Minitab output of various regression models is given below. For each fitted model we list the estimated equation (with estimates, standard errors, and p-values), the coefficient of determination R^2, the root mean square error *s*, and the Durbin-Watson statistic. Minitab flags observations with unusually large standardized residuals ("R") and with unusually large leverage ("X"). The Lockerbie model is simplified by omitting insignificant variables.

Campbell (n = 13):

```
Incumbent Vote = 25.8 + 0.492 Sept Trial + 2.26 GDP Growth

Predictor        Coef     SE Coef          T        P
Constant       25.754       2.953       8.72    0.000
Sept Trial    0.49173     0.05716       8.60    0.000
GDP Growth     2.2571      0.4921       4.59    0.001

S = 1.827       R-Sq = 92.2%     R-Sq(adj) = 90.7%

Unusual Observations
Obs   Sept Tri   Incumbent        Fit      SE Fit    Residual    St Resid
  9       48.7     44.700      44.226       1.570       0.474        0.51 X

X denotes an observation whose X value gives it large influence.

Durbin-Watson statistic = 2.15
```

Abramowitz(n = 13):

```
Incumbent Vote = 45.1 - 4.69 Term + 0.179 Popularity + 2.14 GDP Growth

Predictor        Coef     SE Coef          T        P
Constant       45.059       2.865      15.73    0.000
Term           -4.691       1.337      -3.51    0.007
Popularity    0.17855     0.05567       3.21    0.011
GDP Growth     2.1389      0.6352       3.37    0.008

S = 1.984       R-Sq = 91.7%     R-Sq(adj) = 89.0%

Unusual Observations
Obs       Term   Incumbent        Fit      SE Fit    Residual    St Resid
 13       0.00     54.600      58.480       0.929      -3.880       -2.21R

R denotes an observation with a large standardized residual

Durbin-Watson statistic = 1.76
```

Holbrook (n = 13):

```
Incumbent Vote = 17.6 + 0.0998 PresPop + 0.296 PersFin - 4.00 Tenure

Predictor        Coef     SE Coef          T        P
Constant       17.606       3.865       4.56    0.001
PresPop       0.09982     0.04668       2.14    0.061
PersFin       0.29589     0.04112       7.20    0.000
Tenure         -3.995       1.002      -3.99    0.003

S = 1.505       R-Sq = 95.3%     R-Sq(adj) = 93.7%

Durbin-Watson statistic = 2.07
```

Lockerbie (n = 11):

```
The regression equation is
Incumbent Vote = 22.4 + 0.635 Inc1 - 0.184 Inc2 + 1.13 NextYearBetter
           - 1.45 Tenure

Predictor        Coef     SE Coef          T        P
Constant       22.351       7.231       3.09    0.021
Inc1           0.6352      0.5136       1.24    0.262
Inc2          -0.1836      0.4923      -0.37    0.722
NextYear       1.1251      0.2103       5.35    0.002
Tenure        -1.4488      0.2489      -5.82    0.001

S = 1.661       R-Sq = 95.4%     R-Sq(adj) = 92.3%

Durbin-Watson statistic = 1.17


The regression equation is
Incumbent Vote = 21.4 + 0.604 Inc1 + 1.13 NextYearBetter - 1.39 Tenure

Predictor        Coef     SE Coef          T        P
Constant       21.423       6.359       3.37    0.012
Inc1           0.6044      0.4747       1.27    0.244
NextYear       1.1340      0.1956       5.80    0.001
Tenure        -1.3894      0.1793      -7.75    0.000

S = 1.555       R-Sq = 95.3%     R-Sq(adj) = 93.2%

Durbin-Watson statistic = 1.32


The regression equation is
Incumbent Vote = 16.6 + 1.30 NextYearBetter - 1.37 Tenure

Predictor        Coef     SE Coef          T        P
Constant       16.646       5.329       3.12    0.014
NextYear       1.3029      0.1493       8.73    0.000
Tenure        -1.3726      0.1857      -7.39    0.000
```

```
S = 1.615        R-Sq = 94.2%     R-Sq(adj) = 92.7%

Durbin-Watson statistic = 1.26
```

(b) The sample sizes for estimating these models is extremely small (n = 13 and n = 11). Considering the extremely small sample sizes, we can not detect violations of the assumption of independent errors.

(c) The root mean square errors for most fitted models are in the range from 1.5 to 2 percentage points. They are similar to the ones in the Fair and Lewis-Beck/Tien models. The size of the root mean square error implies that the half widths of 95% prediction intervals are <u>at least</u> 3 - 4 percentage points. Incorporating the uncertainty from the estimation and considering that the sample size is very small makes the prediction intervals even wider. Furthermore, the predictions are “within-sample” predictions, which means that the case being predicted is part of the data that are used for estimation. Prediction errors for “out-of-sample” predictions (where the case being predicted is not part of the data used for the estimation) are usually larger; see (d).

(d) Leaving out case *i*, running the regression on the reduced data set, and predicting the response of the case that has been left out using the estimates from the reduced data set, leads to the PRESS residuals $e_{(i)}$ in equation (6.21) of Chapter 6. Equation (6.22) implies that the PRESS residuals can be calculated from the regular residuals and the leverages. That is,

$$e_{(i)} = y_{(i)} - \hat{y}_{(i)} = e_i/(1 - h_{ii})$$

For illustration we have calculated the residuals, leverages and PRESS residuals for the regression model considered by Campbell in the beginning of this exercise. The PRESS residuals are larger than the ordinary residuals. For example, the (out-of-sample) prediction error for 1996 is -3.76.

Year	Incumbent Vote	Sept Trial	GDP Growth	Residuals	Leverage	PRESS
1948	52.32	45.61	0.91	2.08441	0.126153	2.38533
1952	44.59	42.11	0.27	-2.48002	0.166349	-2.97488
1956	57.75	55.91	0.64	3.05900	0.093183	3.37334
1960	49.92	50.54	-0.26	-0.09906	0.134083	-0.11439
1964	61.34	69.15	0.81	-0.24520	0.361195	-0.38384
1968	49.60	41.89	1.63	-0.43144	0.280740	-0.59984
1972	61.79	62.89	1.73	1.20653	0.235919	1.57906
1976	48.95	40.00	1.17	0.88618	0.257420	1.19338
1980	44.70	48.72	-2.43	0.47371	0.738021	1.80821
1984	59.17	60.22	1.79	-0.23597	0.203862	-0.29640
1988	53.90	54.44	0.79	-0.40671	0.083538	-0.44379

1992	46.55	41.94	0.35	-0.61699	0.168430	-0.74195
1996	54.74	60.67	1.04	-3.19446	0.151107	-3.76308

(e) The four prediction models studied in this exercise are no better and no worse than the models by Fair and Lewis-Beck/Tien. While they give us some indication about the winner of presidential elections, their large uncertainty makes them only useful in the rather uninteresting situation when there is little doubt about the winner of the election.

8.2

Part 1(a): Modeling the height and the weight at referral (HeightR, WeightR) as a function of age at referral (AgeR)

Models with a linear component of Age provide an adequate representation of the relationships. Addition of Age**2 is not necessary. The models lead to an R-square of about 60 percent for height, and 45 percent for weight. Height at referral is easier to predict than weight. Birth weight is marginally significant (estimate 2.26, with p-value 0.064). Addition of birth weight to the regression of weight at referral on age at referral increases the R-square from 45.9 to 48.3 percent. Each extra pound at birth increases the weight at referral by 2.26 pounds. Average weight at referral is 73 pounds, with standard deviation 20 pounds.

Regression Analysis: HeightR versus AgeR, AgeR2**

```
The regression equation is
HeightR = 19.1 + 0.452 AgeR - 0.00120 AgeR**2

77 cases used 16 cases contain missing values

Predictor        Coef     SE Coef         T        P
Constant       19.095       9.434      2.02    0.047
AgeR           0.4523      0.1700      2.66    0.010
AgeR**2    -0.0012036   0.0007501     -1.60    0.113

S = 2.999       R-Sq = 60.4%     R-Sq(adj) = 59.3%
```

Regression Analysis: HeightR versus AgeR

```
The regression equation is
HeightR = 33.9 + 0.181 AgeR

77 cases used 16 cases contain missing values

Predictor        Coef     SE Coef         T        P
Constant       33.912       1.949     17.40    0.000
AgeR          0.18088     0.01741     10.39    0.000

S = 3.030       R-Sq = 59.0%     R-Sq(adj) = 58.5%
```

Regression Analysis: WeightR versus AgeR, AgeR2**

The regression equation is
WeightR = - 0.9 + 0.656 AgeR + 0.00009 AgeR**2

80 cases used 13 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-0.94	46.45	-0.02	0.984
AgeR	0.6555	0.8387	0.78	0.437
AgeR**2	0.000094	0.003704	0.03	0.980

S = 15.09 R-Sq = 45.9% R-Sq(adj) = 44.5%

Note: Because of the multicollinearity between AgeR and AgeR**2, both regression coefficients are (partially) insignificant. However, this does not imply that both can be omitted from the model at the same time. The results of the model given below show that AgeR is significant if it is the only variable in the model.

Regression Analysis: WeightR versus AgeR

The regression equation is
WeightR = - 2.09 + 0.677 AgeR

80 cases used 13 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-2.090	9.341	-0.22	0.824
AgeR	0.67658	0.08321	8.13	0.000

S = 14.99 R-Sq = 45.9% R-Sq(adj) = 45.2%

Regression Analysis: WeightR versus AgeR, BirthWeight

The regression equation is
WeightR = - 16.1 + 0.653 AgeR + 2.26 BirthWeight

80 cases used 13 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-16.15	11.85	-1.36	0.177
AgeR	0.65326	0.08282	7.89	0.000
BirthWeight	2.259	1.202	1.88	0.064

S = 14.75 R-Sq = 48.3% R-Sq(adj) = 46.9%

Part 1(b): Modeling the height and the weight at follow-up (HeightF, WeightF) as a function of age at follow-up (AgeF)
Similar conclusions as in 1(a). Models with a linear component of Age provide an adequate representation of the relationships. Addition of Age**2 is not needed. The

models lead to an R-square of about 40 percent for both height and weight. Birth weight is significant (estimate 4.97 with p-value 0.01). Each extra pound at birth increases the weight at follow-up by 5 pounds. Average weight at follow-up is 124 pounds, with standard deviation 32 pounds.

Regression Analysis: HeightF versus AgeF, AgeF2**

```
The regression equation is
HeightF = 10.0 + 0.458 AgeF - 0.00080 AgeF**2

81 cases used 12 cases contain missing values

Predictor        Coef     SE Coef          T         P
Constant        10.02       34.71       0.29     0.774
AgeF           0.4581      0.3937       1.16     0.248
AgeF**2     -0.000795    0.001106      -0.72     0.474

S = 4.115       R-Sq = 41.8%     R-Sq(adj) = 40.3%
```

Regression Analysis: HeightF versus AgeF

```
The regression equation is
HeightF = 34.8 + 0.176 AgeF

81 cases used 12 cases contain missing values

Predictor        Coef     SE Coef          T         P
Constant       34.801       4.090       8.51     0.000
AgeF          0.17553     0.02347       7.48     0.000

S = 4.103       R-Sq = 41.4%     R-Sq(adj) = 40.7%
```

Regression Analysis: WeightF versus AgeF, AgeF2**

```
The regression equation is
WeightF = - 158 + 2.23 AgeF - 0.00339 AgeF**2

85 cases used 8 cases contain missing values

Predictor        Coef     SE Coef          T         P
Constant       -158.2       206.4      -0.77     0.445
AgeF            2.227       2.349       0.95     0.346
AgeF**2     -0.003387    0.006620      -0.51     0.610

S = 25.24       R-Sq = 39.9%     R-Sq(adj) = 38.5%
```

Regression Analysis: WeightF versus AgeF

```
The regression equation is
WeightF = - 53.4 + 1.03 AgeF
```

85 cases used 8 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-53.37	24.17	-2.21	0.030
AgeF	1.0269	0.1388	7.40	0.000

S = 25.13   R-Sq = 39.7%   R-Sq(adj) = 39.0%

Regression Analysis: WeightF versus AgeF, BirthWeight

The regression equation is
WeightF = - 82.0 + 0.982 AgeF + 4.97 BirthWeight

85 cases used 8 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-82.04	25.84	-3.18	0.002
AgeF	0.9815	0.1353	7.25	0.000
BirthWeight	4.967	1.910	2.60	0.011

S = 24.30   R-Sq = 44.3%   R-Sq(adj) = 43.0%

Part 1(c): Modeling the combined data: HeightCo, WeightCo and AgeCo.

Models with a linear component of AgeCo provide an adequate representation of the relationship between HeightCo and AgeCo. For weight, the addition of the quadratic component AgeCo**2 becomes necessary. The scatter plot of weight against age suggests that the variability increases with the level. The scatter plot of the logarithm of weight against age indicates that the variability is stabilized by this transformation. The residuals from the regression of ln(WeightCo) on AgeCo are unremarkable. No major lack of fit can be detected.

Regression Analysis: HeightCo versus AgeCo, AgeCo2**

The regression equation is
HeightCo = 31.3 + 0.221 AgeCo -0.000144 AgeCo**2

158 cases used 28 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	31.334	4.180	7.50	0.000
AgeCo	0.22070	0.06099	3.62	0.000
AgeCo**2	-0.0001437	0.0002121	-0.68	0.499

S = 3.604   R-Sq = 77.7%   R-Sq(adj) = 77.4%

Regression Analysis: HeightCo versus AgeCo

The regression equation is
HeightCo = 34.1 + 0.180 AgeCo

158 cases used 28 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	34.060	1.136	29.97	0.000
AgeCo	0.179700	0.007717	23.29	0.000

S = 3.597 R-Sq = 77.7% R-Sq(adj) = 77.5%

Regression Analysis: WeightCo versus AgeCo, AgeCo2**

The regression equation is
WeightCo = 23.8 + 0.180 AgeCo + 0.00229 AgeCo**2

165 cases used 21 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	23.78	23.55	1.01	0.314
AgeCo	0.1799	0.3435	0.52	0.601
AgeCo**2	0.002292	0.001195	1.92	0.057

S = 20.84 R-Sq = 69.3% R-Sq(adj) = 68.9%

Regression Analysis: WeightCo versus AgeCo

The regression equation is
WeightCo = - 19.7 + 0.833 AgeCo

165 cases used 21 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	-19.659	6.507	-3.02	0.003
AgeCo	0.83340	0.04414	18.88	0.000

S = 21.01 R-Sq = 68.6% R-Sq(adj) = 68.4%

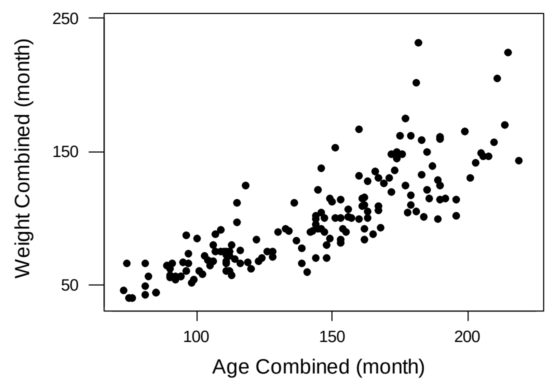
Regression Analysis: ln(WeightCo) versus AgeCo

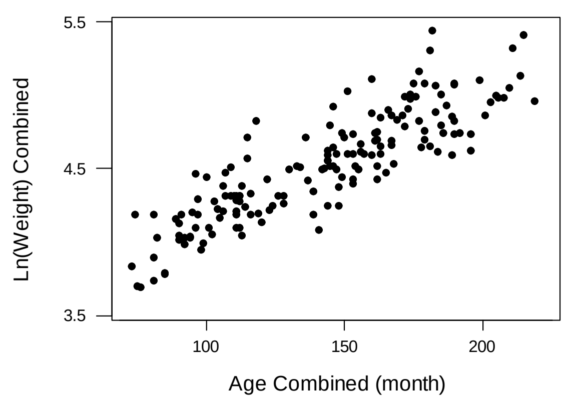
The regression equation is
ln(WeightCo) = 3.30 + 0.00864 AgeCo

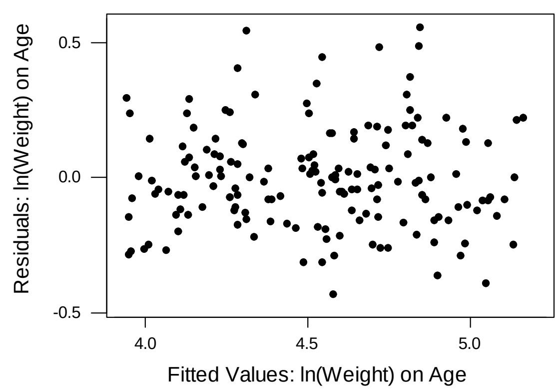
165 cases used 21 cases contain missing values

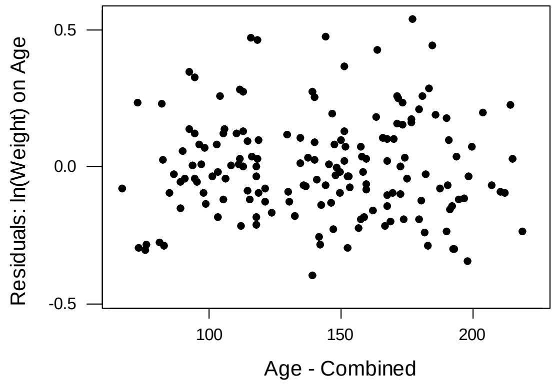
Predictor	Coef	SE Coef	T	P
Constant	3.29653	0.05685	57.99	0.000
AgeCo	0.0086442	0.0003857	22.41	0.000

S = 0.1836 R-Sq = 75.5% R-Sq(adj) = 75.4%

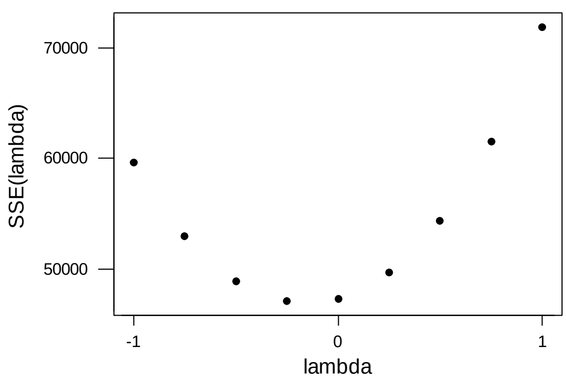

250
150
50
Weight Combined (month)
100
150
200
Age Combined (month)


5.5
4.5
3.5
Ln(Weight) Combined
100
150
200
Age Combined (month)


0.5
0.0
-0.5
Residuals: ln(Weight) on Age
4.0
4.5
5.0
Fitted Values: ln(Weight) on Age



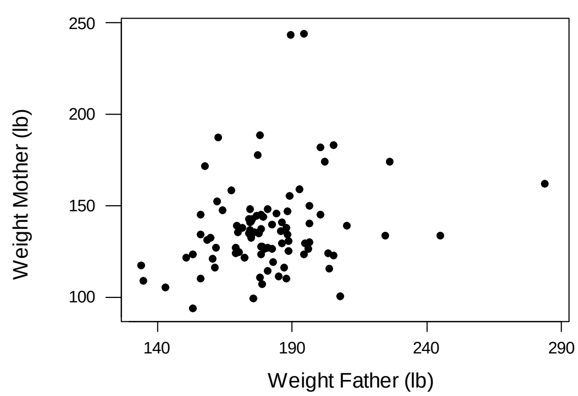
The Box-Cox transformation is applied to the response (see Section 6.5 in Chapter 6). For various values of λ we calculate the geometric mean $\bar{y}_g = (\Pi y_i)^{1/n}$ and the transformed response $(y^{\lambda} - 1)/\lambda(\bar{y}_g)^{\lambda-1}$, regress the transformed response on the explanatory variable AgeCo, and compute the error sum of squares $SSE(\lambda)$. The maximum likelihood estimate of λ minimizes $SSE(\lambda)$. The graph of $SSE(\lambda)$ against λ (given below) shows that the estimate of λ is close to 0. This confirms that the logarithmic transformation is appropriate.



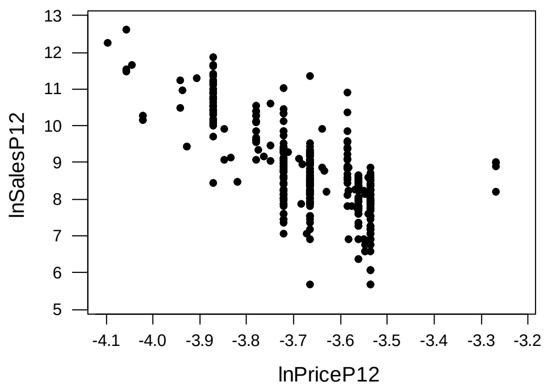
<u>Part 2(a):</u>
A plot of the weight against the height of mothers shows a relationship (correlation coefficient r = 0.336). A correlation coefficient of 0.34 implies that (only) about ten percent of the variability in weight is explained by height (because in simple linear

regression, $R^2 = r^2$). A similar conclusion can be reached for fathers. A plot of the weight against the height of fathers shows a similar-sized correlation (correlation coefficient r = 0.289).

<u>Part 2(b):</u>
The correlation between the height of mothers and the height of fathers is small (r = 0.077).
The correlation between the weight of mothers and the weight of fathers is larger (0.242). There is some (but rather weak) evidence that both partners tend to be above or below the average weight. The scatter plot shows three unusual cases. In one case the father is quite heavy, while the mother is of average weight. In the other two cases the fathers are of average weight while the mothers have weights much above average. However, the omission of these three cases does not change the correlation coefficient (r = 0.243).



8.3
(a) A scatter plot of (weekly) logarithms of sales of 12-packs of brand P (lnSalesP12) against the logs of their prices (lnPriceP12) shows an expected negative relationship. As prices increase, sales decrease.



Regression Analysis: lnSalesP12 versus lnPriceP6, lnPriceP12, lnPriceP24

```
The regression equation is
lnSalesP12 = - 3.74 + 0.921 lnPriceP6 - 7.24 lnPriceP12 + 2.92
lnPriceP24

384 cases used 15 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant       -3.740       1.598      -2.34    0.020
lnPriceP6       0.9205      0.1603      5.74    0.000
lnPriceP12     -7.2420      0.3040    -23.82    0.000
lnPriceP24      2.9233      0.2895     10.10    0.000

S = 0.7338      R-Sq = 63.0%     R-Sq(adj) = 62.7%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      347.92      115.97    215.40    0.000
Residual Error   380      204.59        0.54
Total            383      552.51
```

The results of fitting model M1 confirm a strong negative association with the product's own price. Each one percent increase in the price of 12-packs reduces the sales of 12-packs by 7.2 percent. The parameters in the model represent elasticities as the model regresses log sales on log prices; see Section 6.5.2. The elasticities of price changes in other pack-sizes of the same product (brand P) are positive and considerably smaller. Price increases in 6- and 24-packs increase the sales of 12-packs because buyers chose to buy 12-packs if the prices of other pack-sizes of their desired brand are raised. The response to price changes of 24-packs is stronger than the response to price changes of 6-packs (elasticity 2.92 as compared to 0.92).

The residuals of the regression model are stored in an additional column of the worksheet. Lagging the vector of residuals once and computing the correlation between residuals and lagged residuals results in the lag one autocorrelation of the residuals. Similar operations can be carried out to obtain higher lag autocorrelations. The lagging operation ignores missing observations in the time series. An alternative strategy is to omit all cases with missing entries, run the regression with the reduced data set (the regression estimates are unchanged), and calculate the autocorrelation function and Durbin-Watson test statistic from the resulting residuals. These latter autocorrelations are not exactly the same as the time spacing is changed by omitting missing cases. However, the differences are minor as there are relatively few missing observations. The autocorrelations shown below (calculated with the first approach) are consistently positive. In Chapter 10 we will revise the regression model by adding a time series component that takes account of this persistent positive autocorrelation.

$r_1 = 0.241$ Durbin-Watson $\approx 2(1-0.241) = 1.52$
$r_2 = 0.271$ $r_3 = 0.184$ $r_4 = 0.238$ $r_5 = 0.232$ $r_6 = 0.211$
$r_7 = 0.165$ $r_8 = 0.190$ $r_9 = 0.166$ $r_{10} = 0.114$

(b) Repeating the analysis for the other brand, brand C, leads to similar results. We find a strong negative elasticity for the price at the considered 12-pack size, and weaker and positive elasticities for prices of other pack-sizes. The response to price changes in 24-packs is stronger than the response to price changes in 6-packs (elasticity 2.08, as compared to 0.72).

Regression Analysis: lnSalesC12 versus lnPriceC6, lnPriceC12, lnPriceC24

```
The regression equation is
lnSalesC12 = - 4.32 + 0.718 lnPriceC6 - 6.31 lnPriceC12 + 2.08
lnPriceC24

384 cases used 15 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant       -4.320       1.494      -2.89    0.004
lnPriceC6       0.7176      0.1486      4.83    0.000
lnPriceC12     -6.3101      0.2606    -24.22    0.000
lnPriceC24      2.0808      0.2732      7.62    0.000

S = 0.7149      R-Sq = 64.4%     R-Sq(adj) = 64.1%


Analysis of Variance

Source            DF          SS          MS         F         P
Regression         3      351.47      117.16    229.22    0.000
Residual Error   380      194.22        0.51
Total            383      545.69
```

(c) The estimation results for model M3 show that the sales of 12-packs of brand P respond negatively to their own price changes (elasticity -6.99), and positively to price changes in other pack-sizes of brand P (elasticities 1.06 and 3.26 for 6- and 12-packs). Sales of 12-packs of brand P are not very sensitive to price changes (at all pack-sizes) of the other competing brand. Customers switch among different pack-sizes, but less among competing brands.

Regression Analysis: lnSalesP12 versus lnPriceP6, lnPriceP12, ...

```
The regression equation is
lnSalesP12 = - 5.10 + 1.06 lnPriceP6 - 6.99 lnPriceP12 + 3.26
lnPriceP24 - 0.178 lnPriceC6 - 0.349 lnPriceC12 - 0.567 lnPriceC24

383 cases used 16 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant       -5.098       1.738      -2.93    0.004
lnPriceP6       1.0578      0.2136      4.95    0.000
lnPriceP12     -6.9868      0.3606    -19.37    0.000
lnPriceP24      3.2575      0.3467      9.40    0.000
lnPriceC6      -0.1777      0.2034     -0.87    0.383
lnPriceC12     -0.3491      0.3189     -1.09    0.274
lnPriceC24     -0.5665      0.3391     -1.67    0.096

S = 0.7331      R-Sq = 63.3%     R-Sq(adj) = 62.7%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         6     348.311      58.052    108.02     0.000
Residual Error   376     202.068       0.537
Total            382     550.379
```

The results for sales of 12-packs of brand C are similar and are not shown.

(d) The estimation results for model M4 confirm that the elasticities have the expected signs. The brand P market share of 12-packs increases with decreasing 12-pack price of brand P, and increasing 12-pack price of brand C. The signs of the two price elasticities (-6.22 and 5.56) are different, but their magnitude is roughly the same. The elasticities for prices at other pack-sizes are smaller; the positive signs for brand P prices reflect a substitution effect for 12-packs when packs at other sizes of brand P become more expensive.

Regression Analysis: ln(SalesP12/SalesC12) versus lnPriceP6, lnPriceP12, ...

```
The regression equation is
ln(SalesP12/SalesC12) = 0.33 + 1.48 lnPriceP6 - 6.22 lnPriceP12
 + 2.97 lnPriceP24 - 1.19 lnPriceC6 + 5.56 lnPriceC12 - 2.54lnPriceC24
```

```
383 cases used 16 cases contain missing values

Predictor        Coef     SE Coef         T        P
Constant        0.331       1.685      0.20    0.844
lnPriceP6       1.4823      0.2071      7.16    0.000
lnPriceP12     -6.2159      0.3498    -17.77    0.000
lnPriceP24      2.9682      0.3362      8.83    0.000
lnPriceC6      -1.1860      0.1972     -6.01    0.000
lnPriceC12      5.5559      0.3092     17.97    0.000
lnPriceC24     -2.5388      0.3288     -7.72    0.000

S = 0.7109      R-Sq = 62.8%     R-Sq(adj) = 62.2%

Analysis of Variance
Source            DF          SS          MS         F        P
Regression         6     321.013      53.502    105.85    0.000
Residual Error   376     190.045       0.505
Total            382     511.058

Durbin-Watson statistic = 1.93
```

(e) The results for model M5 show that the coefficient of determination R^2 is hardly changed (0.620 versus 0.628), but the model is easier to interpret. The market share of brand P depends on the relative prices of the two brands. The market share of 12-packs increases with decreasing price ratios of 12-packs. The coefficients for other pack-sizes are considerably smaller and positive, indicating a substitution effect among the various pack-sizes.

Regression Analysis: ln(SalesP12/SalesC12) versus ln(PriceP6/PriceC6), ln(PriceP12/PriceC12), ...

```
The regression equation is
ln(SalesP12/SalesC12) = 0.126 + 1.27 ln(PriceP6/PriceC6)
           - 5.77 ln(PriceP12/PriceC12) + 2.70 ln(PriceP24/PriceC24)

383 cases used 16 cases contain missing values

Predictor                  Coef     SE Coef          T        P
Constant                 0.1258      0.0368       3.42    0.001
ln(PriceP6/PriceC6)      1.2657      0.1838       6.89    0.000
ln(PriceP12/PriceC12)   -5.7696      0.2868     -20.12    0.000
ln(PriceP24/PriceC24)    2.6998      0.2937       9.19    0.000

S = 0.7160      R-Sq = 62.0%     R-Sq(adj) = 61.7%

Durbin-Watson statistic = 1.93
```