1. 7.4

(a) $SSR(x_1) = 136366$, $SSR(x_3|x_1) = 2033565$, and $SSR(x_2|x_1, x_3) = 6675$. One way to get these "by hand" is from $SSR(x_3|x_1) = SSR(x_1, x_3) - SSR(x_1)$ and $SSR(x_2|x_1, x_3) = SSR(x_1, x_2, x_3) - SSR(x_1, x_2)$.

(b) To test if $x_3$ can be dropped from the model when $x_1$ and $x_2$ are there, we could use a t-test for $H_0: \beta_3 = 0$ versus $HA: \beta_3 \neq 0$. The information from that t-test would come from a fit of $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. The problem asks for an F-test though. The test statistic is:

$$F_* = \frac{\{SSE(x_1, x_3) - SSE(x_1, x_2, x_3)\}/1}{SSE(x_1, x_2, x_3)/(n - 4)} = \frac{(992204 - 985530)/(985530/48)} = 0.3250556.$$ 

We would reject $H_0$ if $F_* > F_{0.05,1.48} = 4.042652$. We do not reject $H_0$. The p-value is $Pr(F > 0.3250556) = 0.5712$ where $F \sim F_{1.48}$. This is also the p-value you get for the t-test of $H_0: \beta_3 = 0$ when you fit $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$.

(c) $SSR(x_1) = 136366$ and $SSR(x_2|x_1) = 5726$. $SSR(x_2) = 11395$ and $SSR(x_1|x_2) = 130697$. As a result, we can see that $SSR(x_1) + SSR(x_2|x_1)$ does equal $SSR(x_2) + SSR(x_1|x_2)$. That is always the case since they both equal SSR($x_1, x_2$).

2. 7.5

(a) $SSR(x_2) = 4860.3$, $SSR(x_1|x_2) = 3896.0$, and $SSR(x_3|x_1, x_2) = 364.2$.

(b) To test if $x_3$ can be dropped from the model when $x_1$ and $x_2$ are there, we could use a t-test for $H_0: \beta_3 = 0$ versus $HA: \beta_3 \neq 0$. The information from that t-test would come from a fit of $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. The problem asks for an F-test though. The test statistic is (note that we could compute the numerator in terms of SSE like we did in the previous problem. That would give an equivalent statistic:)

$$F_* = \frac{\{SSR(x_1, x_2, x_3) - SSR(x_1, x_2)\}/1}{SSE(x_1, x_2, x_3)/(n - 4)} = \frac{(9120.464 - 8756.304)/(4248.8/42)} = 3.599774.$$ 

We would reject $H_0$ if $F_* > F_{0.025,1.48} = 5.403859$. We do not reject $H_0$. The p-value is $Pr(F > 3.599774) = 0.06467673$ where $F \sim F_{1.42}$. This is also the p-value you get for the t-test of $H_0: \beta_3 = 0$ when you fit $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$.

3. 7.11
(a)  
\[ R^2_{Y1} = \frac{SSR(x1)}{SSTO} = 0.55 \]
This is the square of the correlation between x1 and y.  
\[ R^2_{Y2} = \frac{SSR(x2)}{SSTO} = 0.41 \]
This is the square of the correlation between x2 and y.  
\[ R^2_{12} = \text{corr}(x1, x2)^2 = 0 \]
This is the square of the correlation between x1 and x2.  
\[ R^2_{Y1|2} = \frac{SSR(x1|x2)}{sse(x2)} = 231.125/248.750 = 0.9291 \]
This is the fraction of variability in Y that’s left unexplained by x2 that is explained by x1.  
\[ R^2_{Y2|1} = \frac{SSR(x2|x1)}{sse(x1)} = 171.125/188.750 = 0.9066 \]
This is the fraction of variability in Y that’s left unexplained by x1 that is explained by x2.  
\[ R^2 = \frac{SSR(x1, x2)}{SSTO} = 0.9580 \]
This is the fraction of the variability in Y that’s “explained” by a linear model in x1 and x2.

Note that since x1 and x2 are uncorrelated, then SSR(x1|x2) = SSR(x1) and SSR(x2|x1) = SSR(x2) and \[ R^2_{Y1} + R^2_{Y2} = R^2. \]

4. 7.26

(a) Using only x1 and x2 as covariates, the fitted regression function is \[ E(Y) = 156.6719 -1.2677x1 -0.9208x2. \]
(b) Using x1, x2 and x3 as covariates, the fitted regression function is \[ E(Y) = 158.4913-1.1416x1-0.4420x2-13.4702x3. \] The estimates of the first 3 \( \beta \)s change.
(c) \[ SSR(x1) = 8275.4 \text{ and } SSR(x1|x3) = 3483.9, SSR(x2) = 4860.3 \text{ and } SSR(x2|x3) = 708.0. \]
(d) The differences are to be expected since the covariates are correlated.

5. 8.4a

a \( R^2 = 0.7632. \) The fit looks pretty good.

b Test \( H0 : \beta_1 = \beta_2 = 0 \) vs. \( HA : \) at least one in non-zero.  
\[ F^* = \frac{MSR(age, age2)}{MSE(age, age2)} = 91.84. \]
Since \( F^* > F_{0.05,2,57} = 3.158843, \) we reject \( H0. \)

e Testing if the quadratic term can be dropped can be done with a t-test. The regression output is

|              | Estimate | Std. Error | t value | Pr(> |t|) |
|--------------|----------|------------|---------|-------|
| (Intercept)  | 207.349608 | 29.225118  | 7.095   | 2.21e-09*** |
| x            | -2.964323 | 1.003031   | -2.955  | 0.00453**   |
| x2           | 0.014840  | 0.008357   | 1.776   | 0.08109.      |
Since the p-value for the quadratic term is greater than 0.05, we can drop it.

6. 8.9 Output is omitted. See section 8.2 for interaction plot examples.

7. 8.11

(a) Coefficients:

|          | Estimate | Std. Error | t value | Pr(>|t|) |
|----------|----------|------------|---------|----------|
| (Intercept) | 27.1500  | 6.4648     | 4.200   | 0.00123 ** |
| moist     | 5.9250   | 0.8797     | 6.735   | 2.09e-05 *** |
| sweet     | 7.8750   | 2.0444     | 3.852   | 0.00230 ** |
| moist:sweet | -0.5000  | 0.2782     | -1.797  | 0.09749 . |

(b) The model is $E(\text{like}) = \beta_0 + \beta_1 \text{moist} + \beta_2 \text{sweet} + \beta_3 (\text{moist} \times \text{sweet})$. The easiest way to test $H_0 : \beta_3 = 0$ versus $H_A : \beta_3 \neq 0$ is a t-test from regression output. From the table above, for any $\alpha > 0.09749$, we’d reject.