

Part 1: Getting Started, Data, Packages
Answers to Selected Exercises

3. Regressions of SAT math scores on teachers' salary and on teachers' salary and percentage of students taking the exam:

```
> library(car)
> mod.1 <- lm(SATM ~ pay, data=States)
> summary(mod.1)
```

```
Call:
lm(formula = SATM ~ pay, data = States)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-69.688 -15.903   2.918  16.258  70.312
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 595.1869   25.5289  23.314 < 2e-16 ***
pay          -3.1607    0.8134  -3.886 0.000306 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 30.53 on 49 degrees of freedom
Multiple R-Squared: 0.2355, Adjusted R-squared: 0.2199
F-statistic: 15.1 on 1 and 49 DF, p-value: 0.0003065
```

```
> mod.2 <- lm(SATM ~ pay + percent, data=States)
> summary(mod.2)
```

```
Call:
lm(formula = SATM ~ pay + percent, data = States)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-29.6969 -13.7879   0.7006   9.9880  42.9610
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 513.6990   16.9174  30.365 < 2e-16 ***
pay          0.9718    0.6292   1.545  0.129
percent     -1.3743    0.1387  -9.906 3.45e-13 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 17.68 on 48 degrees of freedom
 Multiple R-Squared: 0.7489, Adjusted R-squared: 0.7384
 F-statistic: 71.58 on 2 and 48 DF, p-value: 3.947e-15

Notice that the coefficient of teachers' salary is negative in the simple regression ($B = -3.1607$) but positive in the multiple regression ($B_1 = 0.9718$). The difference is due to “controlling” for percentage of students taking the exam in the multiple regression.

A scatterplot matrix for the data reveals that there are two groups of states – those with relatively low and relatively high percentages of students taking the SAT exam (and slight nonlinearity in some of the marginal relationships):

```
> scatterplot.matrix(~ SATM + pay + percent, span=.75,
+ cex=2, data=States)
```

