Homework due Nov. 10

Assigned exercises: OpenIntro, Ch.9: 9.1, 9.4, 9.5, 9.6.
Supplemental exercises linked via homework web page:
Pg.764-771: # 1, 11, 15, 17, 19(a-c). (9 probs total)
Graded exercises: OpenIntro # 9.1, 9.4(a,c,d), and
other supplement, pg.764-771, # 15, 19(a-c).
Total (maximum) possible points = 20.
4 pt for each of 4 graded problems, plus 4 for completion of the rest.

Exercises from OpenIntro Ch. 9

(9.1) (a) The regression model is

$$\widehat{\text{Baby weight}} \text{ (in ounces)} = 123.05 - 8.94 \times \text{ smoke}$$

(b) The slope means: The model predicts the birth weight of babies born to mothers who smoke is on average 8.94 ounces lower than that of babies born to mothers who don't smoke.

(c) Assuming the conditions are met, the software output tests the null hypothesis H_0 : true slope of the predictor smoke is 0, vs H_A : true slope is not 0.
Since the P-value is essentially 0, there is strong evidence of a statistically significant relationship between birth weight and smoking.

(9.4) (a) The regression model is

$$\widehat{\text{days absent}} = 18.93 - 9.11 \times \text{eth} + 3.10 \times \text{sex} + 2.15 \times \text{lrn}$$

(b) **Not graded, but here are the answers:**

Slope of eth: The predicted absence for students who are not aboriginal is on average 9.11 days fewer than that of aboriginal students.

Slope of sex: The predicted absence for male students is on average 3.1 days higher than that of female students.

Slope of lrn: The predicted absence for students who are "slow learners" is on average 2.15 days higher than that of average learners.

(c) residual = observed − predicted

$$2 - (18.93 - 9.11 \times 0 + 3.10 \times 1 + 2.15 \times 1)$$
$$2 - (24.18) = \boxed{-22.18 \text{ days}}$$

(d) By defn.

$$R^2 = 1 - \frac{var(e_i)}{var(y_i)} = 1 - \frac{240.57}{264.17} = 0.0893$$

Adjusted R^2 is

$$1 - \frac{var(e_i)}{var(y_i)} \times \frac{n-1}{n-k-1} = 1 - \frac{240.57}{264.17} \times \frac{146-1}{146-3-1} = 0.0701$$

Exercises from supplement linked via homework web page

(15) (a) The regression model is

$$\widehat{\text{Asking price}} \text{ (in dollars?)} = -152037 + 9530 \times \text{ Baths} + 139.87 \times \text{ Sq ft}$$

(b) Since the R^2 is 71.1 %, about 71 % of the variability in asking price is accounted for by the model.

(c) For each additional square foot increase in area, the predicted asking price increases by \$139.87, on average, when all other variables are held constant.

(d) Not graded.

(19) (a) The regression model is

$$\widehat{\text{salary}} \text{ (in 1000 \$)} = 9.788 + 0.110\,(\text{X1}) + 0.053\,(\text{X2}) + 0.071\,(\text{X3}) + 0.004\,(\text{X4}) + 0.065\,(\text{X5})$$

(b)

$$\widehat{\text{salary}} \text{ (in 1000 \$)} = 9.788 + 0.110\,(120) + 0.053\,(9) + 0.071\,(50) + 0.004\,(60) + 0.065\,(30)$$
$$\boxed{= \$29{,}200}$$

(c) H_0 : the coefficient of $X4$ is 0 when all other variables are included.

H_A : the coefficient of $X4$ is not 0, when other variables are included.

The software output shows the t-score is 0.013, for which the P-value (with $df = 30 - 6 = 24$) is greater than 0.2. Thus we retain H_0 and conclude there is no evidence that typing speed is a statistically significant predictor of salary.