Homework due Oct. 27/29

Assigned exercises: Ch.8, OpenIntro Stats, exercises: 8.2, 8.6, 8.12, 8.15, 8.20, 8.21, 8.24, 8.25, 8.31, 8.33, 8.35, 8.44. (12 exercises total)
Graded exercises: 8.12, 8.20, 8.24(a,c), 8.25(a-c), 8.35.
Total (maximum) possible points = 20.
3 pt for each of 5 graded problems, plus 5 for completion of the rest.

Exercises from Ch.8, OpenIntro Stats

(8.12) (a) The relationship between average temperature and crawling age of babies is approximately linear, moderately strong and negative. There appears to be an outlier at a point where the average temperature is about 52°F and the average crawling age around 28 weeks.

(b) The relationship would not change in form, strength, or direction if we change the units.

(c) If we change the units, the correlation would remain the same, $r = -0.70$.

(8.20) Since the residual is positive, we under-estimated the incidence of skin cancer. This is because residual = observed − predicted. A positive residual corresponds to larger observed value than predicted.

(8.24) (a) The conditions are assumed to have been checked previously, when this exercise was first introduced.

The regression equation has the form: $\hat{y} = b_0 + b_1 x$
where the x-variable is shoulder girth and the y-variable is height.
Given summary statistics are:
$\bar{x} = 107.2,\ s_x = 10.37$ cm, $r = 0.67$
$\bar{y} = 171.14,\ s_y = 9.41$ cm

We have: $b_1 = r\dfrac{s_y}{s_x} = 0.67\left(\dfrac{9.41}{10.37}\right) = 0.604$ cm/cm
Then we have: $\hat{y} = b_0 + 0.604x$.

Find b_0 by plugging in $(\bar{x}, \bar{y})$: $171.14 = b_0 + 0.604(107.2) \;\Rightarrow\; b_0 = 106.39$

Equation of regression line is:

$\hat{y} = 106.39 + 0.604\,x$ OR $\boxed{\widehat{\text{height}} = 106.39 + 0.604\ (\text{shoulder girth})}$

(c) $R^2 = 0.67^2 = 0.44$ or 44%.
Interpretation: The R^2 value of 44% means that about 44% of the variability in height is explained by the variability in shoulder girth.

(b), (d)-(f) are not graded.

(8.25) (a) The regression equation is:

$\widehat{\text{Murder rate}} = -29.901 + 2.559\ (\text{Poverty }\%)$

(b) Interpretation of intercept: In metropolitan areas where there is no poverty, the model predicts a murder rate of -29.901 per million, on average. This is clearly meaningless in reality, but it is what the intercept says.

(c) Interpretation of slope: For each additional percent increase in poverty, the model predicts an increase of 2.559 murders per million, on average.

(d)-(e) are not graded, but here are the answers:

(d) Interpretation of R^2: Approximately 70.52% of the variability in murder rates is explained by variability in the percentage living in poverty.

(e) The correlation coefficient $= \sqrt{R^2/100} = 0.8398$.

(8.35) (a) Let β_1 denote the true slope of a possible linear relationship between poverty percentage and murder rates. The hypotheses are:

Null hypothesis $\quad H_0 : \beta_1 = 0$

Alt hypothesis $\quad H_A : \beta_1 \neq 0$

(b) From the given regression output we see that the test statistic (t-score) is 6.562, with P-value ≈ 0. Thus, we reject the null hypothesis and conclude that poverty percentage is a statistically significant predictor of murder rates. Of course, this assumes all necessary conditions for inference are met, which we have not checked!

(c) Confidence interval $=$ statistic $\pm t^*_{df} \times SE$

Here we have: statistic $= 2.559$, $n = 20 \Rightarrow df = 18$, $t^*_{18} = 2.10, SE = 0.39$.

Confidence interval $= 2.559 \pm 2.10 \times 0.39 = (1.74, 3.378)$

Interpretation of CI: For each additional percent increase in poverty, the model predicts the murder rate on average will increase by 1.74 to 3.378 per million

(d) is not graded, but here is the answer:

(d) Yes, the entire confidence interval is above 0, which leads to the same conclusion: poverty percentage is a statistically significant predictor of murder rates.