Chapter 8 Textbook exercises

Solutions to even-numbered questions
Statistics and statistical programming
Northwestern University
MTS 525

Aaron Shaw

November 11, 2020

All exercises taken from the *OpenIntro Statistics* textbook, 4th edition, Chapter 8.

8.6 British married straight couples

a) Husband and wife ages are positively, linearly correlated, with a few possible outliers.
b) Husband and wife heights also appear to be positively and linearly correlated, but very weakly.
c) The age plot shows a much more evident linear trend. The data points also align more tightly.
d) No. Correlation is not affected by rescaling. Revisit the formulas for calculating correlation for more depth on this..

8.36 The heads of babies

a) Substituting into the regression equation:
$\bar{y} = 3.91 + .78 * 28$

```
y_bar = 3.91 + .78 * 28

y_bar
```

```
## [1] 25.75
```

b) Let's write out the hypotheses first:

$$H_0 : \ \beta_{gestational\_age} = 0$$

$$H_A : \ \beta_{gestational\_age} \neq 0$$

The test statistic for this can be calculated as a t-score as follows:

$$T = \frac{0.78 - 0}{.35} = 2.23$$

We can look up a p-value for this in a t-table for $df = 23$ or we can figure it out in R. The `pt(q,df)` function gives the proportion of the T-distribution which is less than `q` with `df` degrees of freedom.

```
(1 - pt(q = 2.23, df = 23)) * 2 # To get 2-tailed p-value, we multiply this by 2
```

```
## [1] 0.03579437
```

```r
pt(q = 2.23, df = 23, lower.tail = FALSE) * 2 # Alternatively, we can get the area of the upper tail an
```

```
## [1] 0.03579437
```

The p-value is less than 0.05, so with a traditional $\alpha = 0.05$ we would reject H_0 and conclude that there is substantial evidence of an association between gestational age and head circumference for low birth-weight babies (the true slope parameter is unlikely to be zero).

8.40 Cats

(a) Here are the hypotheses:

$$H_0 : \beta_{body\_weight} = 0$$
$$H_A : \beta_{body\_weight} \neq 0$$

(b) The estimate is large (and positive) with a small standard error, resulting in a large test-statistic and a correspondingly small p-value. This is likely lower than whatever critical value was selected for the significance level, so the data provide compelling evidence in support of the alternative hypothesis that the true slope parameter for body weight is not equal to zero, and is positively associated with heart weight in cats.

(c) The confidence interval for the parameter can be calculated like this:

$$b_{body\_weight} \pm t^* \times SE$$

Plugging in values from the model estimates and 1.98 for the value of t^*:

$$4.034 \pm 1.98 \times 0.250$$

$$(3.539, 4.529)$$

In words, this model indicates that for each additional kilogram in a cat's weight we can be 95% confident that the wight of it's heart will increase by 3.539 to 4.529 grams on average.

(d) Yes. The alternative hypothesis from parts (a) and (b) corresponds to the 95% CI above zero calculated in part (d).

8.44 Rating professors

a) Here's how to calculate the slope:

$$\bar{y} = b_0 + b_1 \bar{x}$$
$$3.9983 = 4.010 + b_1 \times -0.0883$$

I'll let R handle the arithmetic:

```r
b_1 = (3.9983-4.010) / -0.0883
b_1
```

```
## [1] 0.1325028
```

b) The null hypothesis is that $H_0 : \beta_1 = 0$. With a t-score of 4.13 and a p-value of ~0.00 there is strong evidence of a positive relationship between beauty and teaching scores (we would reject the null hypothesis).

c)

1. Nearly normal residuals: The distribution of residuals looks a little bit (left) skewed but fairly normal.

2. Homoscedasticity (constant variance) of residuals: The residuals have approximately equal variance across the range of beauty around zero.

3. Independent observations/residuals: The question does not really shed much light on this. Looking at the original paper, it is actually a bit unclear how the instructors were chosen and whether or not they are a random sample. The residuals do not seem to have any relationship with the order of data collection, so that's good at least.

4. Linear relationship between independent/dependent variables: From the scatterplot, the relationship appears to be linear (or at least not obviously non-linear).