Homework 3

STAT3530 Applied Linear Models

Due Thursday, September 17, in class.

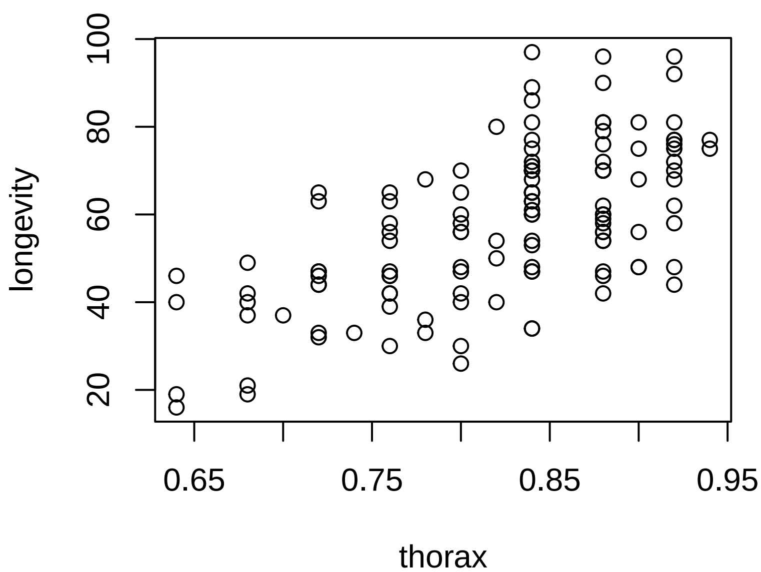
You may collaborate, but write up your own solutions. Submit a paper copy of this document with handwritten solutions below each prompt. Please attach any plots you are asked to make and label them by problem number.

Each dataset loads from the course site; the line is given at the start of each problem.

Fruit flies

1. The `flies` data give `thorax` length in millimeters and `longevity` in days for 124 male fruit flies. This problem considers the hypothesis that larger flies live longer on average.

```
load(url("https://tdruiz-stat3530.share.connect.posit.cloud/_data/flies.RData"))
```



   a. Write the linear model that facilitates testing the hypothesis of interest.

b. Fit the model and report the estimates and standard errors.

c. Check the residual diagnostics and comment on model assumptions. Attach your plots separately.

d. Test the hypothesis that larger flies live longer on average. State the hypotheses, report the test statistic and p-value, and interpret the result.

e. Report and interpret a 95% confidence interval for the change in mean longevity per 0.1mm increase in thorax length (a millimeter is a large change on this scale).

f. Compute and interpret a 95% interval for the expected lifespan of a fly with a 0.8mm thorax.

Spot the issue

The `sim` data contain simulated data consisting of a single predictor `x` and six different responses `y1` through `y6` generated from different models.

```r
load(url("https://tdruiz-stat3530.share.connect.posit.cloud/_data/sim.RData"))
```

Regress each response on `x` separately, for example:

```r
fit <- lm(y1 ~ x, data = sim)          # and so on for y2, ..., y6
```

2. For each response, fit the simple linear model and check the assumptions.

    a. Make residual diagnostic plots for each of the six fits. You do NOT need to attach them.

    b. Complete the table. Write "none" when nothing is wrong.

response	assumption that fails	which plot showed it	what you saw
y1			
y2			
y3			
y4			
y5			
y6			

    c. Does the independence assumption fail for any of the models? If so, how did you detect it?

Transformations

Kleiber's law

3. The `kleiber` data give body mass (kg) and resting metabolic rate (kJ/day) for 95 animal species, ranging from a 3.6 gram shrew to a 3 ton elephant.

   ```
   load(url("https://tdruiz-stat3530.share.connect.posit.cloud/_data/kleiber.RData"))
   ```

   Two competing theories predict how metabolic rate should scale with mass. If heat loss through the body surface is what limits metabolism, then rate should scale as mass$^{2/3}$, since surface area grows as the 2/3 power of volume. Kleiber's law instead claims an exponent of 3/4.

   Both theories say $\text{metab} = c \cdot \text{mass}^{\beta_1}$ for some constant c.

   a. Report the correlation between metabolism and mass on the original scale. Does this suggest a linear relationship?

   b. Now fit the model `metab ~ mass` and make the residual-versus-fitted plot. Attach your plot separately. Does this suggest a linear relationship? Explain.

   c. Write the model implied by the theory as a linear model and identify the parameters corresponding to β_1 and c.

   d. Fit the model in (b) and check the residual diagnostics. Attach your plots separately. Does the model seem sufficiently well-specified for inference on β_1? Why or why not?

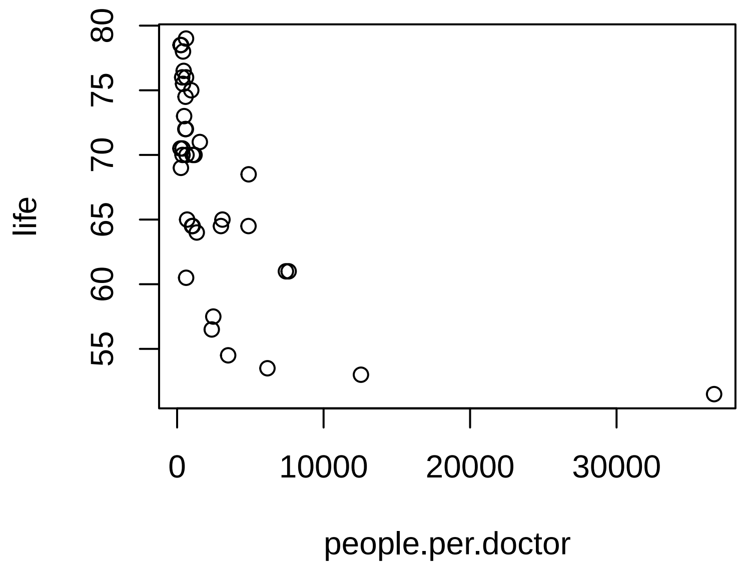
e. Report and interpret a 95% confidence interval for β_1. Which (if either) theory does this support?

f. Test $H_0 : \beta_1 = 3/4$ and, separately, $H_0 : \beta_1 = 2/3$. Note that `summary()` tests neither of these, so you will need to build the test statistic yourself from the estimate and its standard error. For simplicity, compare with the critical value 1.96 for a two-sided test at level $\alpha = 0.05$. Is either theory inconsistent with the data?

Life expectancy and access to doctors

4. The `doctors` data record life expectancy in years and the number of people per doctor for 38 countries in 1993.

```
load(url("https://tdruiz-stat3530.share.connect.posit.cloud/_data/doctors.RData"))
```



a. Identify and fit an appropriately specified model for the relationship between life expectancy and the number of people per doctor. Attach residual diagnostics for your final model; write the model and report parameter estimates and standard errors.

b. Report in context a 95% confidence interval for the parameter that measures the relationship.

c. Would you conclude from this fit that training more doctors raises life expectancy? Explain briefly.