STAT 451

Hand-in Homework 3

Problems 1-2 due Friday, Sep. 25 by midnight; problem 3 (grad) due Sun, Sep. 27 by midnight.

Problems 1 and 2 are to be done in a single R Markdown document. Problem 3 is for graduate students and is to be done in a separate R Markdown document and submitted separately. Use sentences to report results; don't just give a value. Remember context (your answers should not have "x" or "y" in them). Use inline R code to report values. Remember to include units where appropriate. Submit only the Rmd document.

1. Honeybees use a "waggle dance" to communicate the location of a food source (direction and distance) to the rest of the swarm. The orientation of the dance indicates the direction of the source and the duration of the dance indicates distance. The table below shows the distance in meters and the duration, in seconds, for seven waggle dances (from Seeley, Honeybee Democracy, Princeton University Press, 2010). You can enter these data by hand into R (see item 5 in SomeRelevantRCommands). The goal is to build a model for predicting distance from duration.

   (a) Create an appropriate scatterplot (with appropriately labeled axes) and succinctly describe the relationship. Remember context.

   (b) Report the correlation coefficient between duration and distance using inline code, e.g. "The correlation is..."

   (c) Find the equation of the least squares line and write the equation using variable names. Put it on a separate line by leaving a blank line before and after the equation. Use inline code for the intercept and slope (intercept is `coef(m)[1]` and slope is `coef(m)[2]` where `m` is the model). Remember, the left side of the equation has the word "Predicted".

   (d) Create a residual plot and a histogram of the residuals (side-by-side). Does it appear there are any problems using a linear model?

   (e) Interpret the slope of the regression line (`coef(m)[2]`) in the context of the problem.

   (f) Report and interpret the values of R^2 and s_e using inline code (see 11.d on p. 8 of SomeRelevantRCommands.pdf for extracting these values).

   (g) Predict the distance to the food source if a honeybee does a waggle dance lasting 1 second. Repeat for 3 seconds. Use inline code.

Distance (m)	Duration (sec)
200	0.40
250	0.45
500	0.95
950	1.30
1950	2.00
3500	3.10
4300	4.10

2. The file `Cranes.csv` contains the winter count of whooping cranes at Aransas National Wildlife Refuge in Texas for every year from 1938 through 2010 (they changed the counting method starting in about 2011).

   (a) Plot count vs. year (include the plot). With regularly spaced data over time, it is usual to connect the points, see 10.c on p. 6 in SomeRelevantRCommands.

   (b) The relationship between count and year is obviously not linear. So plot log10 of count vs. year (include the plot). This does a pretty good job of linearizing the relationship.

   (c) Fit the linear model for predicting log10(Count) from year. Report only the intercept and slope of the model (no residual plots or anything else) using inline R code. Interpret the slope coefficient in terms of percent growth, as we did in class for the car price data.

   (d) What's the predicted number of cranes from your model for 2020 and 2050? Use inline R code.

3. (Grad only) The file `TrotSpeeds.csv` contains some of the data from a study in which animals were trained to run on a treadmill to study how speed, stride and gait compared across different species. One part of the study recorded the speeds at which the animals progressed from walking to trotting, looking for a relationship with body mass. This file contains the body mass (in kg.) and trot speed (in meters per second) for 20 individual animals of 15 species. The goal is to develop a model for predicting trot speed from body mass.

   (a) Plot trot speed vs. body mass (include the plot). Succinctly describe the relationship between the two variables remembering the guidelines in chapter 6.

   (b) Find a transformation of either variable (or both) which both linearizes the relationship as well as making the variability roughly constant. Use the Circle of Transformations (p. 256 of the text) as a guideline. Stick with the standard transformations in Table 8.1. Include only the final scatterplot. You don't need to discuss which transformations you tried.

   (c) Report the regression model for the transformation you decided on in part b) from the summary command.

(d) Include a residual plot and histogram of residuals (side-by-side) for your model in part c). Do they appear OK?

(e) Report and interpret the value of the slope coefficient for your model.

(f) Calculate the predicted trot speed for a 3 kg. cat and a 70 kg. human. Does the prediction for the human seem reasonable? (All the animals in the study were 4-legged.)