Assignment

Show your R work (both your input and your output) for each question below. You may find this FAQ useful.

Urban Runoff

Urbanization poses a major threat to stream and watershed ecosystems. One aspect of urbanization is the conversion of natural areas to land with impervious surfaces, thus increasing runoff of rain and, likely, pollutants. A University of Washington researcher recorded the percent of impervious land and the benthic index of biotic integrity (IBI) for 14 areas in the state of Washington. The IBI has been described as a measure of "the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a species composition and functional organization comparable to that of natural habitat in the region." In this example, the researchers were attempting to identify how the percentage of impervious surfaces in the area effects the IBI. The data for this study are below (note that much more information can be obtained at the QELP site).

```
imperv   60 43 43 34 27 25 21 18 11  8  8  8  7  7  5  4  2
IBI       9 11 13 23 31 31 21 23 27 37 39 29 31 43 33 35 37
```

Load these data into R to answer the questions below. [*Note that these are the same data that you entered in a previous class exercise.*]

1. In terms of the variables of this problem, what is the equation of the best-fit line?
2. In terms of the variables of this problem, INTERPRET the value of the slope?
3. In terms of the variables of this problem, INTERPRET the value of the y-intercept?
4. What is the predicted IBI for an impervious surface percentage of 80%?
5. What is the predicted IBI for an impervious surface percentage of 20%?
6. What is the residual if the impervious surface percentage is 30% and the IBI is 20?
7. How much will the IBI change if the percentage of impervious surface is decreased by 30%?
8. What is the correlation coefficient between IBI and the percentage of impervious surface?
9. What proportion of the variability in IBI is explained by knowing the percentage of impervious surface?
10. What aspect of this regression analysis concerns you (i.e., consider the regression assumptions)?

Sign Legibility and Age

In a study of the legibility and visibility of highway signs, a Pennsylvania research firm determined the maximum distance at which each of 30 drivers could read a newly designed sign. The participants in the study were from 18 to 82 years old. The government agency that funded the research hoped to improve highway safety for older drivers and wanted to examine the relationship between age and sign legibility distance. Use the data in vision.csv (data, meta) to construct results that can be used to answer the questions below. [*Example from Utts and Heckard, Mind on Statistics*].

1. What is the response variable? [*Hint: Read the rest of the questions before answering this question*.]
2. What is the explanatory variable?
3. In terms of the variables of this problem, what is the equation of the best-fit line?
4. In terms of the variables of this problem, interpret the value of the slope.
5. In terms of the variables of this problem, interpret the value of the intercept.
6. What is the predicted maximum distance to see the sign for a 40-year-old driver?
7. What is the predicted maximum distance to see the sign for a 90-year-old driver
8. What is the residual if the age of the driver is 50 and the maximum distance is 410 ft?
9. What is the correlation coefficient between age and maximum distance to see the sign?
10. What proportion of the variability in maximum distance to see the sign is explained by knowing the age of the driver?
11. How much will the predicted maximum distance to see the sign change if the driver's age increased by ten years?
12. What aspect of this regression analysis concerns you? [*Hint: consider the regression assumptions*]?