Computational Problems

We have data on fish caught in the Blackfoot River by Fish, Wildlife, & Parks personnel over a number of years. They used electrofishing equipment to attract the fish to the boat, then dipped them out of the water with nets, measured length in cm and weight in grams. They are often working in cold conditions in late autumn or early spring, so some measurement error is expected.

These data are not from a random sample. The goal is to catch all fish within a reach or section of the Blackfoot River every few years to assess the health of the population. Changes over years are important to the biologists.

The data were collected by making two trips per section (Johnsrud or Scotty Brown) each sampling year. The fish caught each trip of a given year, had their weight, length, and species recorded.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| trip | length | weight | year | section | species |
| 1 | 288 | 175 | 1989 | Johnsrud | RBT |
| 1 | 288 | 190 | 1989 | Johnsrud | RBT |
| 1 | 285 | 245 | 1989 | Johnsrud | RBT |
| 1 | 322 | 275 | 1989 | Johnsrud | RBT |
| 1 | 312 | 300 | 1989 | Johnsrud | RBT |
| 1 | 363 | 380 | 1989 | Johnsrud | RBT |

1. If the researchers were only interested in Rainbow trout and Brown trout, how would you remove Bull trout and WCT (whitefish) from the data set?
2. Sometimes when sampling the fish, a technician fails to record one of the variables. How would you remove any fish from the data set with only missing weight?
3. How would you determine what seems to be the lower size limit of fish that is catchable with the equipment used by the researchers? How would you filter these "uncatchable" fish out of the data set?
4. The researchers are interested in how many fish are caught each year that weigh over 1500 grams. How would you find these numbers to report?
5. The sampling methods used by Fish, Wildlife, & Parks on the Blackfoot River has changed over the years. In the years 1989 - 1996 they used gill nets and since 1996 they have used electrofishing. How would you create a new variable **Method** to reflect these different sampling methods used over the years?
6. Which pairs of (weight, length) combinations seem difficult to believe? One way to look for unusual pairs is to use what fisheries biologists call a "condition index" where = weight and = length. If fish are highly unusual in this scale, it would be best to remove them, but you might need to compare only within species. How would you calculate and plot each Rainbow trout's condition number to assess if their measurements were recorded incorrectly?
7. The researcher want you to build a model for weight as a function of length and the available covariates. How would you select the "best" model to report? Would your model choice change if the researcher wanted to use it to predict the weight of fish next year?
8. If we consider fish bodies to be of constant density (mass per unit volume), how could we create a geometric argument for what scale to use to model weight as a function of length?
9. To what population does the model you chose apply, and why?
10. Suppose instead, that the data were entered by an undergraduate student in Excel. They chose to store the lengths and weights of the fish sampled on each trip, each year, as a list, so the first entry of each (length, weight) corresponds to the first sampled fish and so on. If you had to reformat these data from the below Excel presentation to the the long representation your were given previously, how would you reformat these data?

Table continues below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | trip | year | section | species | fish\_lengths |
| **1** | 1 | 1989 | Johnsrud | RBT | 288, 288, 285, 322, 312, 363, 269, 160, 213, 157 |
| **556** | 2 | 1989 | Johnsrud | RBT | 288, 288, 285, 322, 312, 363, 269, 160, 213, 157 |
| **1120** | 1 | 1990 | Johnsrud | RBT | 264, 368, 327, 264, 263, 290, 288, 312, 284, 263 |
| **1769** | 2 | 1990 | Johnsrud | RBT | 288, 288, 285, 322, 312, 363, 269, 160, 213, 157 |
| **2538** | 1 | 1991 | Johnsrud | RBT | 288, 288, 285, 322, 312, 363, 269, 160, 213, 157 |
| **3172** | 2 | 1991 | Johnsrud | RBT | 264, 368, 327, 264, 263, 290, 288, 312, 284, 263 |

|  |  |
| --- | --- |
|  | fish\_weights |
| **1** | 175, 190, 245, 275, 300, 380, 170, 40, 80, 35 |
| **556** | 175, 190, 245, 275, 300, 380, 170, 40, 80, 35 |
| **1120** | 185, 525, 320, 170, 165, 240, 250, 280, 230, 185 |
| **1769** | 175, 190, 245, 275, 300, 380, 170, 40, 80, 35 |
| **2538** | 175, 190, 245, 275, 300, 380, 170, 40, 80, 35 |
| **3172** | 185, 525, 320, 170, 165, 240, 250, 280, 230, 185 |

Table continues below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| trip | mark | length | weight | year |
| Min. :1.0 | Min. :0.000 | Min. : 16 | Min. : 0 | Min. :1989 |
| 1st Qu.:1.0 | 1st Qu.:0.000 | 1st Qu.:186 | 1st Qu.: 65 | 1st Qu.:1991 |
| Median :2.0 | Median :0.000 | Median :250 | Median : 150 | Median :1996 |
| Mean :1.5 | Mean :0.093 | Mean :262 | Mean : 246 | Mean :1997 |
| 3rd Qu.:2.0 | 3rd Qu.:0.000 | 3rd Qu.:330 | 3rd Qu.: 330 | 3rd Qu.:2002 |
| Max. :2.0 | Max. :1.000 | Max. :986 | Max. :4677 | Max. :2006 |
| NA | NA | NA | NA's :1796 | NA |

|  |  |
| --- | --- |
| section | species |
| Johnsrud :11648 | Brown: 3171 |
| ScottyBrown: 6704 | Bull : 553 |
| NA | RBT :12341 |
| NA | WCT : 2287 |
| NA | NA |
| NA | NA |
| NA | NA |

