

SW 8102 Advanced Research—Summer, 2004
Sample Data Analysis Assignment (with answers)

The following questions are based on the same data that will be used for the data analysis assignment that students will be asked to do in pairs. These questions will be answered in class as preparation for completing the regular assignment.

1. In what percentage of the cases were Intensive Family-Based Services provided?

43.4%; percentage is a direct way to report results from nominal level data (and besides that, percentage was requested)

2. Was the race of the oldest child related to the likelihood that abuse was substantiated? How likely was it that this relationship could have occurred by chance?

Not significantly; $\chi^2=2.97$, $df=2$, $p=.23$ (could have occurred by chance 23 of 100 times); chi square is an appropriate test to use when testing the relationship between two nominal level variables

3. What was the typical age of the mother or female guardian in families that were investigated? How much variation in the age was there among mothers/female guardians?

Mean=29.9; median=31; range=23 (19-42); standard deviation=6.4; mean and standard deviation can be used with ratio level data; since the mean and median are so similar, the distribution is probably not skewed; range provides additional information about the variation present

4. Was there a relationship between the age of the oldest child and the age of the mother/female guardian? What does this relationship mean, if anything? How likely was it that this relationship could have occurred by chance?

Yes, $r=.86$, $p<.001$; older mother likely to have older eldest child; $p<.001$, less than 1 chance in 1000 that this relationship occurred by chance; the Pearson's r correlation is an appropriate parametric test when examining the relationship between two ratio level variables

5. What was the typical response on the responding parent's views on spanking as a form of discipline? (Assume for the purpose of example that these responses were ordinal rather than interval.) How much variation was there in these ratings?

Median=3 (uncertain), mode=3 (uncertain); range=4 (1-strongly agree to 5-strongly disagree); with ordinal level data, the median and mode are the best measures of central tendency (the mean requires interval or ratio level data), and the range provides an appropriate measure of variation (the standard deviation also requires interval or ratio level of data)

6. Suppose you knew the number of siblings and wanted to predict length of out-of-home placement. What formula would best predict length of out-of-home placement? What would be the best prediction for length of out-of-home placement if two siblings were present?

*y (length of placement) = $1.14 * x$ (number of siblings) + 25.8; simple linear regression is used to predict one interval or ratio level (criterion) variable when you know the value of a second interval or ratio level (predictor) variable; with two siblings the best estimate of length of out-of-home placement is 28.1 days.*

7. Did sex of the oldest child affect the average length of out-of-home placement for those who were placed? What is the probability that the difference that was observed occurred by chance?

Yes, females were placed an average of 22.1 days, males were placed an average of 30.3 days; $t=3.27$, $df=31.5$, $p=.003$ (could happen by chance less than 3 times in 1000); the t -test is an appropriate parametric test to use to test a relationship when the dependent variable is interval and the independent variable is nominal and has two values

8. Was the typical age of the mothers or female guardians different based on the race of the oldest child?

The average age of mothers or guardians was different based on race of the oldest child. The average age of mothers/guardians of white children was 29.3 years; the average age of mothers/guardians of American Indian children was 28.7 years; and the average age of mothers/guardians of children of other races was 35.5 years. These differences were statistically significant ($F=4.84$; $df=2,80$; $p=.01$). One way ANOVA was used for this analysis because the criteria for a parametric test were met, the dependent variable was ratio, and the nominal independent variable had three categories.