

# Biostat 510

## Homework 3

### Due Thursday, February 7, 2008

State the alpha level that you are using for any statistical tests carried out for this homework. When reporting the results of a statistical test, include the test statistic, the degrees of freedom and the p-value of the test. **Include your user-defined formats for each variable in all oneway and twoway tables.**

To carry out this homework you will need to download the two data sets from my web page: autism\_demog.sas7bdat, and voter.sas7bdat.

The autism\_demog data set is part of a prospective longitudinal study of 213 children reported by researchers at the University of Michigan (Oti, Anderson, and Lord, 2004). The children were placed into three diagnostic groups when they were 2 years old: Autism, Pervasive Developmental Disorder (PDD), and non-spectrum. Study participants were children who had had consecutive referrals to one of two autism clinics before the age of 3 years. Social development was assessed at each age (2 years, 5 years, 9 years, 13 years, and 14 years) using the Vineland Adaptive Behavior Interview survey form, a parent-reported measure of socialization. The variable VSAE (Vineland Socialization Age Equivalent) was a combined score that included assessments of interpersonal relationships, play/leisure time activities, and coping skills. Initial language development was assessed using the Sequenced Inventory of Communication Development (SICD) scale; children were placed into one of three groups (SICDEGP) based on their initial SICD scores on the expressive language subscale at age 2. Their diagnostic group was assessed at age 2 and at age 9 (BESTEST2 and BESTEST9, respectively).

Variable	Description	Codes
newinit	ID	
agemom	age of mother when child was born	Age in years
gender		1=Male 2=Female
bestest2	diagnosis at age 2	1=Autism 2=PDD 3=Non-Spectrum
bestest9	diagnosis at age 9	1=Autism 2=PDD 3=Non-Spectrum
bestviq	verbal IQ at age 2	
bestnviq	non-verbal IQ at age 2	
sicdegp	expressive language group	1=Low 2=Medium 3=High

1. Create a new permanent SAS data set called b510.autism\_recode and add some new variables to your data set.
  - a) Create two new variables,
    - i. AUTISM\_AGE2, based on diagnosis at age 2. For your new variable 1=Autism, 2=Other (PDD or Non-Spectrum).
    - ii. AUTISM\_AGE9, based on diagnosis at age 9. For your new variable 1=Autism, 2=Other (PDD or Non-Spectrum).
  - b) Get a proc contents for your new data set (include the output in your writeup)
    - i. How many observations are there in your data set?
    - ii. What are the types of the variables?
  - c) Get descriptive statistics for your variables (include the output in your writeup).
    - i. How many observations are there for each variable?
  
2. Create user-defined formats for variables listed below (these can be temporary).
  - a) BESTEST2 and BESTEST9 (Remember, you can use the same format for more than one variable).
  - b) AUTISM\_AGE2 and AUTISM\_AGE9.
  - c) SICDEGP
  - d) GENDER

Get oneway frequency tables for each of these variables, assigning the user-defined formats to them. **Include your oneway frequencies in your homework writeup.**

3. Examine the Gender of the children included in this study.
  - a) What proportion of the children are male? Female?
  - b) Carry out a chi-square goodness of fit test to test the null hypothesis that there are equal proportions of males and females. What do you conclude? **(Include the chi-square goodness of fit test output in your writeup).**
  
4. Examine the relationship of Gender to Diagnosis. **(Include the output from this question in your writeup).**
  - a) Create a cross-tabulation with Gender as the row variable and diagnosis at age 2 as the Column variable.
  - b) What percent of Males are Autistic? PDD? Non-Spectrum? What percent of females fall into each of these diagnostic categories?
  - c) Carry out a chi-square test to test whether gender is independent of diagnosis at age 2? What do you conclude?
  - d) Is the Mantel-Haenszel test appropriate here? Why or why not?
  
5. Examine the relationship of Diagnosis at Age 2 to expressive language group. **(Include output from this question in your writeup).**
  - a) Get a cross-tabulation with diagnosis at age 2 as the row variable (the exposure) and expressive language group as the column variable (the outcome). Get the expected values, and the row percentage, along with the frequencies (don't print the column percents, or the overall table percents). Get a chi-square test of independence.

- b) What proportion of Autistic children are in the highest category of expressive language? What proportion of PDD children? What proportion of Non-Spectrum children?
  - c) Carry out a test to test the null hypothesis that diagnosis at age 2 is independent of expressive language group. What do you conclude?
6. Compare the proportion of children diagnosed with autism at age 2 to the proportion of children diagnosed with autism at age 9. **(Include the output in your writeup).**
- a) Create a cross-tab of AUTISM\_AGE2 vs. AUTISM\_AGE9. What proportion of children were diagnosed with autism at age 2? At age 9?
  - b) How many children were included in this table?
  - c) Was there a change in the proportion of children included in the autism diagnostic category at age 9 compared to age 2? Use a McNemar's test here, because you are looking at paired data. Give the results of your statistical test and explain them.
  - d) Why do you think the diagnosis of some of these children changed from age 2 to age 9?

The second part of the homework uses the Voter.sas7bdat permanent SAS data set. This data set was originally supplied with SPSS, and was transferred into a SAS data set. Download this data set from my web page. A description of the variables included in the voter data set is included below:

Variable	Description	Codes
Pres92	how the respondent voted for president in 1992	1=Bush 2=Perot 3=Clinton
Age	Age of respondent in years	Age in years
Agecat	age categories	1=less than 35 2=35-44 3=45-64 4=65+
Educ	highest year of school completed	Numeric
Degree	highest degree completed	0=less than high school 1=high school 2=junior college 3=bachelor 4=graduate degree
Sex	respondent's sex	1=male 2=female

7. Create user-defined formats for PRES92, AGECAT, DEGREE, and SEX. **(Include the output from this question in your writeup).**
- a) Get oneway tabulations for all of these variables, using the formats.
  - b) Get a Proc Contents for this data set
    - i. How many observations are there? What types are the variables?