

Biostat 510: Statistical Computing Packages

SAS Homework 3

Due Tuesday, Feb 8, 2005

Note: only problems 1 through 7 are due for this homework.

Topics:

Reading a permanent SAS data set
Recode continuous variables into categories
Contingency tables for independent samples
Contingency tables for matched samples

1. Create a new permanent SAS data set from your data set used in homework 2 having newly defined categorical variables.
 - a. Create a new variable, `ACTIVITY` that has 3 categories:
 - i. 0 to 1 hour of exercise per week (code=1)
 - ii. 2 to 4 hours of exercise per week (code=2)
 - iii. Greater than 4 hours of exercise per week (code=3)
 - b. Create a new variable `HI1` from your variable for `heartrate1`
 - i. `HRTRATE1` greater than or equal to 85 (code=1)
 - ii. `HRTRATE1` less than 85 (code=2)
 - c. Create a new variable `HI2` from your variable for `heartrate2`
 - i. `HRTRATE2` greater than or equal to 85 (code=1)
 - ii. `HRTRATE2` less than 85 (code=2)
 - d. Create a new variable `RRAN` that has values of 1 for `RAN` and 2 for `Did not Run` (if your original variable is coded this way, you do not need to make a new variable).

To create the new permanent data set, first, submit a `libname` statement, like that you used to originally define your library. Then use the two-level name for both your old SAS data set and the new one that you are creating:

```
libname b510 v9 "c:\temp\b510";
data b510.combine2;
  set b510.combine;
  /*insert codes to create new variables here*/

run;
```

2. Create user-defined formats for each of your new variables using `Proc format`. Remember that your format names can only be 8 characters long, and that they may not end with a number.
3. Check your new variables to be sure that they were correctly coded.

- a. Run Proc Means for the data set. Do the number of observations match for HRS_EXERCISE and ACTIVITY, for HRTRATE1 and HI1, for HRTRATE2 and HI2 for RAN and RRAN?
- b. Run Proc means separately for each original variable, using the categorical variable as a class variable, to check the values included in each category, as shown in the example below:

```
proc means data=b510.combine2;
  class activity;
  var hrs_excise;
run;
```

Check that the minimum and maximum values of the original continuous variable are correct within each category of your recoded variable. If not, fix the coding in your data step above, and rerun it.

Include the output from this question in your homework writeup.

4. Tabulate all of your new variables. Use the formats you created when you run the tables.
 - a. How many and what percent of the class were at activity level1, level2, level3?
 - b. How many and what percent of the class had high heartrate at time1? At time 2?
 - c. Include these tables in your homework.
5. Get a cross-tabulation of GENDER vs HI1 and ACTIVITY vs. HI1. Use your user-defined formats for these tables. Make HI1 the column variable for both of these tables. Get a chisquare test of independence for both of these cross-tabulations. Please include the chisquare test statistic, degrees of freedom and p-value when writing up your response for each table.
 - a. For GENDER*HI1: What percent of Females had high heartrate at time 1? What percent of males? Is there evidence that these two variables are related?
 - b. For ACTIVITY * HI1: What percent of students in each level of ACTIVITY had high heartrate at time 1? Is there evidence that these two variables are related?
 - i. Is there any evidence of an increasing or decreasing *trend* in the percent of students who have high heartrate at time 1, as the categories of ACTIVITY increase? (Get a Cochran-Armitage test for trend for this part of the question. Please interpret the results of this test.)
 - c. Include these tables in your homework.
6. Get a cross-tabulation of RRAN by HI2. Use RRAN as the row variable and HI2 as the column variable. Use your user-defined formats for the variables. Get a chisquare test of independence for this table, and estimates of the odds ratio and relative risk. Please include the chisquare test statistic, degrees of freedom and p-value when writing up your response.

- a. What percent of those who ran had high heart rate at time 2? What percent of those who did not run had high heart rate at time 2? Is there evidence that these two variables are related? Please explain your conclusion.
 - b. What is the odds ratio, and 95% confidence interval for the odds ratio for this table? Does the 95% CI include 1?
 - c. What is the relative risk, and 95% CI for the relative risk? Does the 95% CI for the relative risk include 1?
 - d. Are the relative risk and odds ratio similar for this table? Explain.
 - e. Include this table in your homework.
7. Get a three-way table of GENDER*RRAN*HI2. Use your user-defined formats for this problem. For this problem, please calculate the Cochran-Mantel-Haenszel statistics.
- a. What is the odds ratio and 95% CI for each table (for females and for males)?
 - b. Using the Breslow-Day test of homogeneity, is there any evidence that the odds ratio for males and for females is different? Discuss.
 - c. What is the overall odds ratio and 95% CI, adjusted for Gender?
 - d. Include these tables in your homework
8. Get a cross-tabulation of HI1*HI2. Use the user-defined formats for the table. Calculate McNemar's test of symmetry for these related proportions.
- a. What percent of students had high heartrate at time 1? What percent of students had high heartrate at time 2?
 - b. Was there a significant change in the percentage of students who had high heartrate at time 1 compared to time 2?
9. Get a three-way cross-tabulation of RRAN*HI1*HI2. Use the user-defined formats. Calculate a McNemar's test of symmetry for these related proportions.
- a. What percent of students who did not run had high heartrate at time 1? What percent of students who did not run had high heartrate at time 2?
 - b. Was there a significant change in the percentage of students who had high heartrate at time 1 compared to time 2 for those students who did not run?
 - c. What percent of students who ran had high heartrate at time 1? What percent of students who ran had high heartrate at time 2?
 - d. Was there a significant change in the percentage of students who had high heartrate at time 1 compared to time 2 for those students who ran?
 - e. Include these tables in your homework writeup.
10. Save your command file as homework3.sas. Make a printout of your commands and hand it in, along with your output from SAS. Write up brief answers to each question. Make sure that your commands can run all at once, by simply clicking on the submit button, without selecting any commands.