

Biostatistics 208 HW #1, Due 1/31/17

INSTRUCTIONS

The first part of the assignment is to complete Problems 2.2-2.6 at the end of chapter 2 in the course textbook, *Regression Methods in Biostatistics*. The problems are also posted on the course website. For Problems 2.2 and 2.3, you can control the number of observations using the `set obs` command before you generate the pseudo-random data, as shown in the first lab. Also, note that the `rnormal()` function used in the first lab is a simpler option to the `invnorm(uniform())` function specified in problem 2.2. For example, to generate a variable `x` containing random observations drawn from the standard normal distribution: `generate x=rnormal()`

The datasets for Problems 2.4-2.6 can be found on the course website.

The second part of the assignment involves a linear regression model with a categorical predictor, the four-level behavioral pattern variable `behpat` in the WCGS dataset used in Problem 2.6.

1. Using the `tab, sum()` command, compute average systolic blood pressure (variable `sbp`) according to behavioral pattern. (Note that the `tabstat` command can also be used here.)
2. Using the `i.` prefix (to be discussed in the second lecture and lab) for the categorical predictor variable `behpat`, run an unadjusted linear regression with outcome `sbp`.
3. Write down the values of the indicator variables required to represent `behpat` in the model, using a table similar to table 4.3 in the textbook.
4. Using the `lincom` command, and the appropriate `i.` prefixes to specify the various behavior patterns, obtain the average SBP in the four behavior pattern groups.
5. Use the `testparm` command to obtain an F -test of the null hypothesis that mean SBP is the same in all four behavioral pattern groups. Write a brief description of the results.
6. Fit the same model as in part 2, but use the `ib` operator with the prefix “`ibn.`” rather than “`i.`” for the behavior pattern variable, and exclude the intercept using the `noconstant` option in the `regress` command. Explain differences between resulting coefficient estimates compared to those from the previously fitted model. Finally use `testparm` again to obtain an F -test of the null hypothesis that mean SBP is the same in all four behavioral pattern groups. (Note that you may need to consult the Stata help file and manual for `testparm` and section 11.4.3.2 of the manual on base levels for factor variables to complete this problem.)

Please email your completed assignment to biostat208hw@gmail.com by 10:30 AM on Tuesday, Jan 31, 2017.