

## Final Exam: due 4:00, Tuesday, 20 Dec

- Do not discuss the exam with other students or TAs. Please send email questions to me. Office hours during the exam: 2:00–3:00 on Friday and Monday, 16 and 19 December.
- **Limit your answer to 3 pages.** Please attach a copy of your SAS code as an appendix to your answer—this does not count as part of the 3 pages.
- All SAS output should be in Consolas font.
- Please turn in a paper copy of your exam at the Biostat office, Mayo A-460, by 4:00.

In both problems, keep all observations in the analysis. Do not discard any outliers.

1. **Logistic regression.** A retrospective national survey of 1000 patients with a certain disease examined the association between age, duration of disease, and the occurrence of a disease event. The spreadsheet disease-events.xls has columns: ID number, event (1 = event, 0 = no event), patient's age (years), duration of disease.

Calculate age decade from age.

*Using odds ratios, describe the association between age decade, duration of disease, and the occurrence of a disease event. Please write sentences to describe the association, referring to tables and/or figures as needed.*

(Hint: a graph may help you describe.)

2. **Glomerular Filtration Rate (GFR).** Measuring GFR directly requires the subject to stay in bed for several hours of carefully timed blood samples, and is therefore difficult and expensive. GFR is commonly estimated from serum creatinine, which can be measured from a single blood sample. A new estimate is based on another chemical in the blood, cystatin-C.

The dataset GFR.xls contains measurements of GFR, serum creatinine, and cystatin-C from 140 adults.

- (a) Compute the reciprocals:  $SCr\_inverse = 1/\text{serum creatinine}$  and  $Cys\_inverse = 1/\text{cystatin-C}$ , and fit the model in either Proc GLM or Proc Reg:

$$GFR = SCr\_inverse \ Cys\_inverse;$$

*Report both regression coefficients, their standard errors, and their 95% confidence intervals.*

- (b) Calculate the bootstrap 95% confidence interval for each regression coefficient, using 1000 replicates.
- (c) Draw the histogram of bootstrap replicate regression coefficients for  $SCr\_inverse$  and for  $Cys\_inverse$  (two histograms).