BIOS 6312: Exam #2 Study Guide (Spring 2021)

Below is a list of general topics to help guide your studying for Exam #2. While it is impossible to make an exhaustive list that includes every possible detail about every possible question, I suspect that focusing your efforts on these general topics will greatly serve to your benefit.

SURVIVAL ANALYSIS

1. Know how various summary measures can be gleaned from of a survival curve.
2. Know the key assumptions required in order for the Cox model to produce valid results.
3. Know the different purposes for covariate adjustment in Cox models.
4. Know how to interpret parameters from a Cox model that involves time-dependent covariates.
5. Know the two different strategies we've learned for handling competing risks and their differences.

PENALIZED REGRESSION

1. Understand why test (out-of-sample) error is the target of interest in prediction problems and *not* the training error.
2. Know the major problems associated with using training error as an estimate of test error.
3. Know the major problems associated with using p-values for variable selection purposes.
4. Understand the bias-variance tradeoff and the way that penalized regression methods choose to optimize it.
5. Understand the conceptual difference between the LASSO and the ridge penalty and its practical implications.
6. Understand the different ways that prediction error can be evaluated in linear and logistic models.

SPLINES

1. Know the major purpose of a spline.
2. Understand the bias-variance tradeoff and the way that spline methods choose to optimize it.

LONGITUDINAL DATA

1. Be able to explain the difference between marginal and conditional associations.
2. Understand why the approaches from the earlier part of the course are not adequate for longitudinal data settings.
3. Understand the differences between generalized estimating equations and generalized linear mixed effects models—what they estimate, what their assumptions are, etc.

BAYESIAN STATISTICS

1. Be able to articulate in plain language the fundamental features that distinguish the Bayesian paradigm from the frequentist paradigm.
2. On a conceptual level, understand the different reasons for and consequences of different prior distribution choices.
3. NOTE: Problems dealing with any of the highly "mathematical" aspects of Bayesian statistics are only fair game for an optional problem, but will *not* be given as a required problem.

GENERAL IDEAS

1. Know how to produce a reduced model from a full model if asked to do so (i.e., for a specific subgroup).
2. Know how to identify the right (combinations of) parameter(s) that correspond to specific scientific questions and hypothesis tests. Don't forget what you know about categorical variables, interaction terms, log-transformation, etc.
3. Be able to apply course principles to new biological/medical problems.

STATA COMMANDS OF INTEREST

```
stset
sts, graph
stcox
stcrreg
splitsample
lasso
lroc and roctab
mkspline
xtgee
regress
xtmixed
```

NOTATION, OPTIONS, POST-COMMANDS

```
gen and replace
summarize
i. and c.
## and #
robust
nolog
cluster
predict
testparm, test, and lincom
lassocoef, display(coef, penalized)
```