Biostat 200C Homework 3

Due May 14 @ 11:59PM

Q1.

The **log-logistic** distribution with the probability density function

$$f(y) = \frac{e^\theta \lambda y^{\lambda-1}}{(1 + e^\theta y^\lambda)^2}$$

is sometimes used for modelling survival times.

- (a) Find the survivor function $S(y)$, the hazard function $h(y)$ and the cumulative hazard function $H(y)$.

- (b) Show that the median survival time is $\exp(-\theta/\lambda)$.

- (c) Use R to plot the hazard function for $\lambda = 1$ and $\lambda = 5$ with $\theta = -5$, $\theta = -2$, and $\theta = 1/2$, in one figure.

Q2. ELMR Exercise 7.5

The data arise from a large postal survey on the psychology of debt. The frequency of credit card use `ccarduse` is a three-level factor ranging from never, occasionally to regularly.

```
data(debt)
help(debt)
```

- (a) Declare the response as an ordered factor and make a plot showing the relationship to `prodebt`. Comment on the plot. Use a table or plot to display the relationship between the response and the income group.

- (b) Fit a proportional odds model for credit card use with all the other variables as predictors. What are the two most significant predictors and what is their qualitative effect on the response? What is the least significant predictor?

- (c) Use stepwise AIC to select a smaller model than the full set of predictors. You will need to handle the missing values carefully. Report on the qualitative effect of the predictors in your chosen model. Can we conclude that the predictors that were dropped from the model have no relation to the response?

- (d) Compute the median values of the predictors in your selected model. At these median values, compare the predicted outcome probabilities for both smokers and nonsmokers.

- (e) Fit a proportional hazards model to the same set of predictors and recompute the two sets of probabilities from the previous question. Does it make a difference to use this type of model?