Lab #1

Biostatistics 210
(please complete and come for discussion on 26 September in Lab Review Session)

0. BACKGROUND

This lab is meant to give you practice applying methods for fitting a regression model with continuous predictors and comparing various functional forms.

1. DOWNLOAD

• Download the dataset lab1.dta. The data was collected by the United Network on Organ Sharing and contains 1 year survival outcomes on pediatric kidney transplants carried out in the US between 1990 and 2002. The variables are

•`died` (1=died in first year, 0 = survived first year)

• `age` (age of child at transplant)

•`txtype` (1=cadaveric donor, 0 = living donor)

•`prevtx` (1= previous tranplant, 0=for years 1996-2002)

Before you do anything else, explore the distribution of ages in the sample using boxplots and calculating the 5 number summary (min, max, median, 25 and 75th pctiles).

2. CATEGORICAL PREDICTOR

Try creating a binary predictor for age. To do this. you can cut at the median

```
xtile median_cut=age, nq(2)
```

Or you can chose an age (e.g. 4)

```
recode age min/4=1 4/max=0, generate(lessthan4)
```

Fit a logistic regression for each and save the results using the predict command. For the first model this would be

```
xi: logistic died i.median_cut

predict out_median
```

this saves the fit in a variable named “out_median”

You can graph the result of a categorical model using the command

```
twoway (line out_median age, sort connect(stairstep)),
ytitle(Probability of Dying in One Year) xtitle(Age at Trans-
plant) scheme(s1color)
```

You can graph the result of another fit (for instance, one saved in a variable called "out-_lt4") by adding another `(line out_lt4 age, sort connect(stairstep))` to the above command. Or you can use the overlaid plots in the dropout menus.

Q1: How do the fits for your two binary models compare graphically?

Now, attempt to fit a model with a categorical variable with 3-5 levels. You can use the recode command or the xtile command to come up with your cutpoints.

Fit a logistic regression model with your categorical variable and extract the predictions using the predict command

`xi: logistic died` *i.yourcategorical*

`predict` *p_yourcategorical*

Graph this fit using the command above

```
twoway (line p_yourcategorical age, sort connect(stairstep))
(line out_median age, sort connect(stairstep)), ytitle(Probabil-
ity of Dying in One Year) xtitle(Age at Transplant) scheme(s1-
color)
```

Q2: How does this fit compare graphically with your binary fits?

2. LINEAR FIT

Now, try the fit using age as a continuous variable and extract its prediction

```
logistic died age

predict plin
```

Compare with the categorical fit

```
twoway (line plin age, sort) (line p_yourcategorical age, sort
connect(stairstep)), ytitle(Probability of Dying in One Year)
xtitle(Age at Transplant) scheme(s1color)
```

Q3: How do the categorical and linear fits compare, do you think the linear assumption appears to be fitting the data well?

4. SPLINE

Using age, create a 3 knots spline predictor.

```
mkspline cube_age = age, cubic nknots(3) displayknots
```

Fit a logistic regression model with the splines

```
logistic died cube_age*
```

save the results

```
predict cube3_fit
```

Plot this with the results of the linear fit

```
twoway (line plin age, sort) (line cube3_fit age, sort),
ytitle(Probability of Dying in One Year) xtitle(Age at Trans-
plant) scheme(s1color) legend(order( 1 "Linear Fit" 2 "Spline
Fit"))
```

Q4: How do the spline and linear fits compare, do you think the linear assumption appears to be fitting the data well?

Test for a difference in fit between the spline and the linear assumption using the likelihood ratio test.

```
logistic died cube_age*

est store A

logistic died age

lrtest A
```

Q5: Is there evidence of non-linear effects based on the likelihood ratio tests?