# Lab #3

*Biostatistics 210*
(OPTIONAL: NOT FOR HAND-IN)

## 0. BACKGROUND

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

## 1. DOWNLOAD

• Download the model building dataset lab3.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.*

## 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

logistic died median_cut

predict 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 Transplant) scheme(s1color)

You can graph the result of another fit (say 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 do come up with your cutpoints.

Fit a logistic regression model with your category 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)), ytitle(Probability of Dying in One Year) xtitle(Age at Transplant) scheme(s1color)

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

## 2. LINEAR FIT

Now, try the fit of using age as a continuous variable and extract it's 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

Create a 3 knots spline basis for age

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 *a*ge, sort), ytitle(Probability of Dying in One Year) xtitle(Age at Transplant) 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?**

## 5. ADJUSTED SPLINE

We can also consider if the effect of age is modified by other variable. Do fit the spline model we simply add those other variables to the model. Let’s try previous transplant and whether the transplant is from a living or cadaveric donor.

logistic died prevtx txtype cube_age*

To graph this can’t use the predict command but have to do a little calculations. Here we use the fact that 14% of the over sample has a prevtx and 48% get their organ from a cadaveric donor. The command on the next page  can calculate the linear predictor for different ages each of which have 14% previous transplant and 48% from cadaveric donor.

```
gen
out=_b[_cons]+_b[prevtx]*0.14+_b[txtype]*0.48+_b[cube_age1]*cube_age1+_b[cube_a
ge2]*cube_age2
```

The command above allows age to vary by fixed prevtx to a population with 14% previous transplant and 48% cadaveric donors. The command below changes that linear predictor into a predicted probability.

```
gen pr = exp(out)/(1+exp(out))
```

This can then be plotted

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

**Q5: Does this change your conclusions?**

I