**Codebook for Head Start data used in Chapter 11, #4**

See Table 11.6 for variable descriptions.

Part (h) asks you to create a binned graph. Here is code to help you do this.

**R:**

# Load the Rdata data set called “dta”

# Each observation has been assigned to a bin via the dta$Bin variable.

# Each observation also has a variable indicating the average value of all observations

# in that bin. This variable is called dta$BinMean.

# We can plot the observations by bin and bin mean using the code below.

# There are multiple observations (lines in the data set) at each point, but this does not

# affect our plot.

plot(dta$Bin, dta$BinMean, type = "n", ylab = "Mortality", xlab = "Poverty")

points(dta$Bin, dta$BinMean, cex=1.5, pch = 19, col="Blue")

abline(v= 0)

Aux.reg.1 = lm(BinMean ~ Bin, data = dta[dta$Poverty < 0,])

lines(c(-6, 0), c(Aux.reg.1$coef[1] - Aux.reg.1$coef[2]\*6, Aux.reg.1$coef[1]), col= "Blue", lwd = 2.0)

Aux.reg.2 = lm(BinMean ~ Bin, data = dta[dta$Poverty > 0,])

lines(c(0, 3), c(Aux.reg.1$coef[1] - Aux.reg.1$coef[2]\*6, Aux.reg.1$coef[1]), col= "Red", lwd = 2.0)

**Stata:**

The trick to creating a binned graph is associating each observation with a bin label that is in the middle of the bin. Here is Stata code to create a *Bin* variable is

scalar BinNum = 50

scalar BinMin = -6

scalar BinMax = 3

scalar BinLength = (BinMax-BinMin)/BinNum

gen Bin = BinMin + BinLength\*(0.5+(floor((Poverty-BinMin)/BinLength)))

The *Bin* variable here sets the value for each observation to the middle of the bin; there are likely to be other ways to do it.

egen BinMean = mean(Mortality), by(Bin)

graph twoway (scatter BinMean Bin, ytitle("Mortality") /\*

\*/ xtitle("Poverty") msize(large) xline(0.0) )/\*

\*/ (lfit BinMean Bin if HeadStart == 0, clcolor(blue)) /\*

\*/ (lfit BinMean Bin if HeadStart == 1, clcolor(red))