

## Introduction to Regression Analysis with Stata

[stata\_intro\_regr]

1. Do graphical & numerical univariate summaries
2. Do bivariate graphical & numerical summaries, as follows:

```
mark complete
markout y x1 x2 x3 xcat1 xcat2
tab complete
```

### *Categorical vars*

```
tab xcat1 xcat2 if complete==1, cell|col|row chi2
[or: tab xcat1 xcat2, cell|col|row chi2 miss]
```

### *Quantitative vars*

```
gr y x1 x2 x3 if complete==1, box
gr y x1 x2 x3 if complete==1, con(m) band(5) xlabel ylabel
```

```
pwcorr y x1 x2 x3 if complete==1, obs bonf sig star(.05)
```

```
or without significance tests & bonferroni adjustment:
cor y x1 x2 x3 [which automatically excludes missing values]
```

### *Categorical & quantitative vars*

```
sort xcat1
by xcat1: gr7 y, hist norm
```

```
by xcat1: gr7 y, box total
```

```
tab xcat1: su(y) or: bys xcat1: su y
table xcat1, contents(mean y med y sd y min y max y)
```

```
ttest y, by(xcat1)
ttest y, by(xcat1) unequal
```

if one sd is 2+times>other sd

```
oneway y xcat1, table bonf
```

multilevel categorical var; if  
sd's test sig as highly  
unequal, categorize y by one  
of the following procedures &  
compute cross-tab

```
sort y
xtile xy=y, nq(4)
bys xy: su y
tab xy xcat1, cell|col|row chi2
```

```
sort y
gy=group(4)
bys gy: su y
tab gy xcat1, cell|col|row chi2
```

### 3. Estimate the regression model & display four diagnostic graphics

```
reg y x1 x2 x3
```

```
rvfplot, yline(0) xla yla
```

```
predict yhat
qnorm yhat, grid
```

```
avplots
```

```
sort id
lvr2plot, s([id])
```

4. Re-estimate the regression model excluding the outlying observations & display four diagnostic plots

```
reg y x1 x2 x3 if id~=91 & id~=126
```

```
rvfplot, yline(0) xla yla
```

```
predict yhat
```

```
qnorm yhat, grid
```

```
avplots
```

```
sort id
```

```
lvr2plot, s([id])
```