Mercury levels in fish tissue for large mouth bass in the Wacamaw and Lumber Rivers

Rivers in North Carolina contain small concentrations of mercury which can accumulate in fish over their lifetimes. Because mercury cannot be excreted from the body, it builds up in the tissues. The concentration of mercury in fish tissue can be obtained at considerable expense by catching fish and sending samples to a lab for analysis. Directly measuring the mercury concentration in the water is impossible since it is almost always below detectable limits.

A study was recently conducted in the Wacamaw and Lumber Rivers to investigate mercury levels in tissues of large mouth bass. At several stations along each river, a group of fish were caught, weighed, and measured. In addition a tissue sample from each fish caught was sent to the lab so that the tissue concentration of mercury could be determined for each fish.

Questions:

1. Is there a relationship between mercury concentration and size (weight and/or length) of a fish?

2. A concentration over 1 part per million (0 on the log scale) is considered unsafe for human consumption. In light of this, what recommendations can you make for fish caught from these rivers?

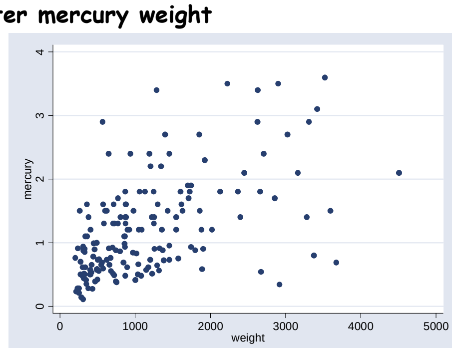
RELATIONSHIPS BETWEEN VARIABLES….

corr mercury weight

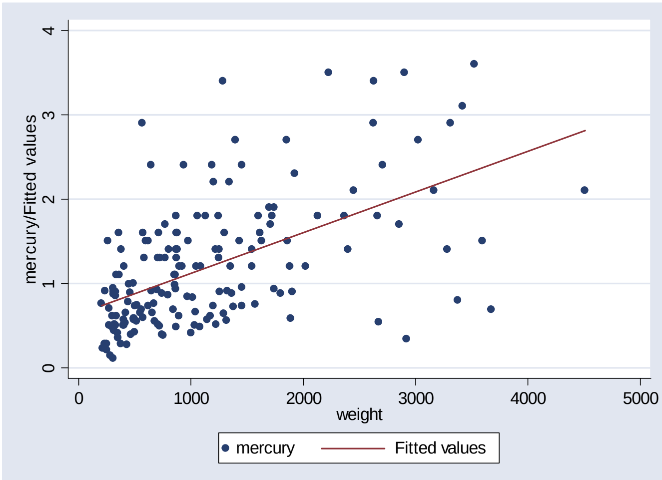
```
           |  mercury   weight
-----------+------------------
   mercury |   1.0000
    weight |   0.5538   1.0000
```

Correlation tells us whether two variables are **linearly related**, and if they are what direction their relationship is in. Positive correlation = positive relationship (one variable goes up, the other tends to go up as well; one variable goes down, the other variable tends to go down as well). Negative correlation = negative relationship (the variables tend to move in the opposite directions). Similar information can be obtained from a scatterplot diagram:

scatter mercury weight



To get the quantification of the relationship we employ MODELS. The first simple model we have is a simple linear regression. This fits a line through the scatterplot of two variables. The line is estimated using the least squares procedure that estimates the intercept and the slope.



. regress mercury weight

```
      Source |       SS       df       MS              Number of obs =     171
-------------+------------------------------           F(  1,   169) =   74.77
       Model |  30.2510497     1  30.2510497           Prob > F      =  0.0000
    Residual |  68.3712259   169  .404563467           R-squared     =  0.3067
-------------+------------------------------           Adj R-squared =  0.3026
       Total |  98.6222756   170  .580131033           Root MSE      =  .63605

------------------------------------------------------------------------------
     mercury |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |   .0004818   .0000557     8.65   0.000      .0003718    .0005918
       _cons |   .6386813   .0803536     7.95   0.000      .4800552    .7973073
```

The estimated intercept is 0.64 and the estimated slope is 0.0005.

Intercept denotes the average mercury level for a fish that has weight of 0. It is meaningless in many situations as you can see.

Slope denotes the average change in mercury when weight goes up by 1 unit (in this case gram). So if you catch two fishes, one weighing x and the other weighing (x+1) grams you'd expect their mercury levels to differ by 0.0005 ppm.

So the estimated model (line) is:

$$\text{Expected Mercury} = 0.64 + 0.0005 \text{ weight}$$

This model lets us predict how much mercury a fish of any given weight will have on average. (Out of all fishes of that given weight, the average observed mercury level will be equal to the fitted value from the model above). For example, we can predict (and give CI for) the average mercury level for a fish of 1kg. (This is inference about the predicted values.)

We can also say something about the actual mercury level of the fish of 1kg based on our model – we can find not only its mean (and the CI for it based on the above), but also its forecast interval: the middle 95% of all mercury weights for all fishes of 1kg. This prediction interval takes into account the uncertainty in the prediction of the mean as well as the actual randomness of mercury weight around the mean.

Ok, so we have postulated a linear relationship between these two variables, and STATA has estimated one for us. STATA has in fact estimated that there is a positive linear relationship between these two variables. But is that relationship STATISTICALLY SIGNIFICANT?

We can answer that question by looking at the p-value of the weight in the above table: it reads 0.000.
This pretty much means that the relationship will be significant at any significance level (recall that we usually use 0.05 or 0.1 level). Alternatively – look at the 95% confidence interval: it does not include 0, which tells us that at 5% level we would reject the null hypothesis of no relationship between mercury and weight. So the mercury-weight relationship IS statistically significant.

```
. reg  mercury weight

      Source |       SS       df       MS              Number of obs =     171
-------------+------------------------------           F(  1,   169) =   74.77
       Model |  30.2510497     1  30.2510497           Prob > F      =  0.0000
    Residual |  68.3712259   169  .404563467           R-squared     =  0.3067
-------------+------------------------------           Adj R-squared =  0.3026
       Total |  98.6222756   170  .580131033           Root MSE      =  .63605

------------------------------------------------------------------------------
     mercury |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |   .0004818   .0000557     8.65   0.000     .0003718    .0005918
       _cons |   .6386813   .0803536     7.95   0.000     .4800552    .7973073
------------------------------------------------------------------------------
```


TESTING SINGLE HYPOTHESIS IN STATA

```
. test weight
 ( 1)  weight = 0
       F(  1,   169) =   74.77
            Prob > F =    0.0000
```

NOTE: F is the square of t above (in SLR only)!
NOTE: you could also use weight coefficient's
t-stat, p-value, CI, F-stat and its p-value

```
.  test weight=1

 ( 1)  weight = 1

       F(  1,   169) = 3.2e+08
            Prob > F =    0.0000
```

NOTE: you could test this using the CI for the weight coefficient in the table as well.

TESTING MULTIPLE HYPOTHESIS SIMULTANEOUSLY

```
. test weight _cons
 ( 1)  weight = 0
 ( 2)  _cons = 0
       F(  2,   169) =  337.55
            Prob > F =    0.0000
```

```
. test weight=_cons=1

 ( 1)  weight - _cons = 0
 ( 2)  weight = 1

       F(  2,   169) = 4.4e+08
            Prob > F =    0.0000
```

```
. test weight=_cons+1=1

 ( 1)  weight - _cons = 1
 ( 2)  weight = 1

       F(  2,   169) = 4.4e+08
            Prob > F =    0.0000
```

NOTE: Testing both hypotheses at once requires adjustment to the signif level of each in order to preserve the overall significance level of the entire test.

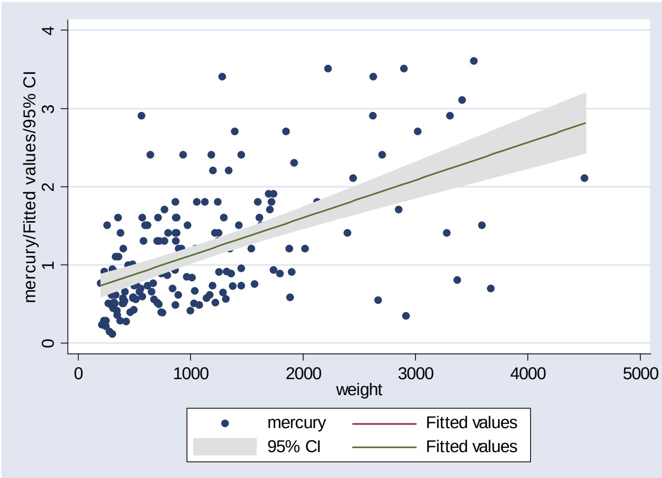
Bonferroni adjustment

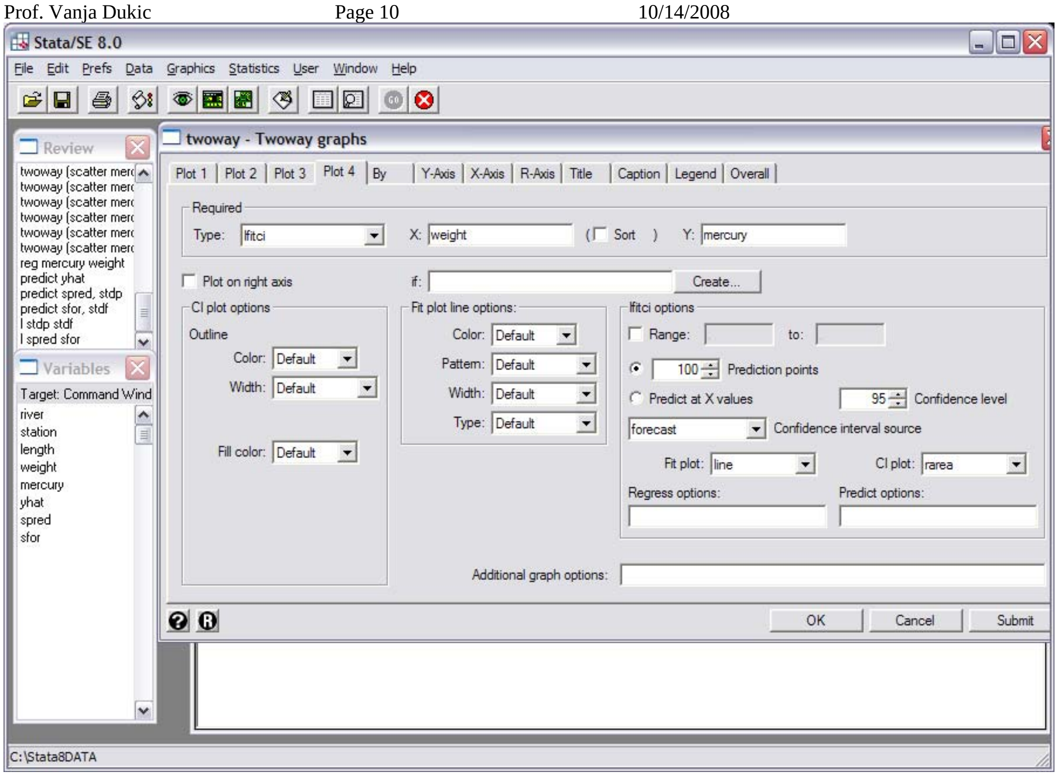
```
. test weight

 ( 1)  weight = 0

       F(  1,   169) =   74.77
            Prob > F =    0.0000
```

```
. test _cons

 ( 1)  _cons = 0

       F(  1,   169) =   63.18
            Prob > F =    0.0000
```

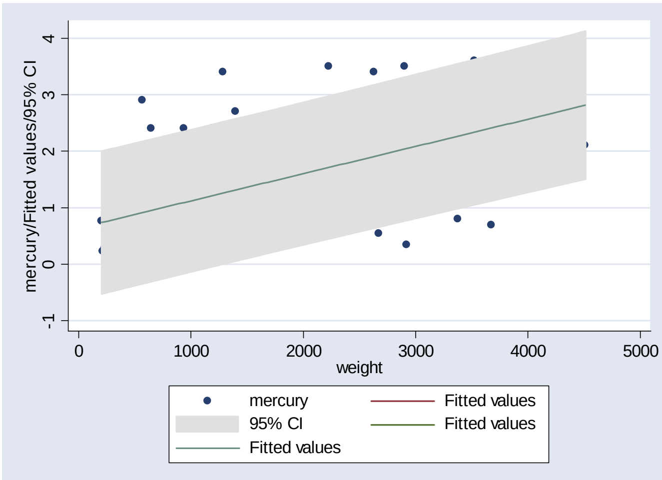
INFERENCE FOR FITS AND FORECASTS

```
twoway (scatter mercury weight) (lfit mercury weight) (lfitci mercury weight)
```





twoway (scatter mercury weight) (lfit mercury weight) (lfitci mercury weight) (lfitci mercury weight, stdf)




```
. predict yhat
(option xb assumed; fitted values)

. predict spred, stdp

. predict sfor, stdf
```



```
. l spred sfor

     +---------------------+
     |    spred       sfor |
     |---------------------|
  1. | .0551914   .6384431 |
  2. | .0628405   .6391497 |
  3. | .1068311   .6449623 |
  4. | .0487234   .6379165 |
  5. | .0495763   .6379822 |
     |---------------------|
  6. | .0488295   .6379246 |
  7. | .0510455   .6380981 |
  8. | .0515619   .6381395 |
  9. | .0488058   .6379228 |
 10. | .0490598   .6379423 |
     |---------------------|
 11. | .1334374   .6498992 |
 12. |  .110068   .6455063 |
 13. | .0979597   .6435523 |
 14. | .1489693   .6532651 |
 15. | .0643477   .6392997 |
….
```