More Testing: *F Statistics & F Tests*

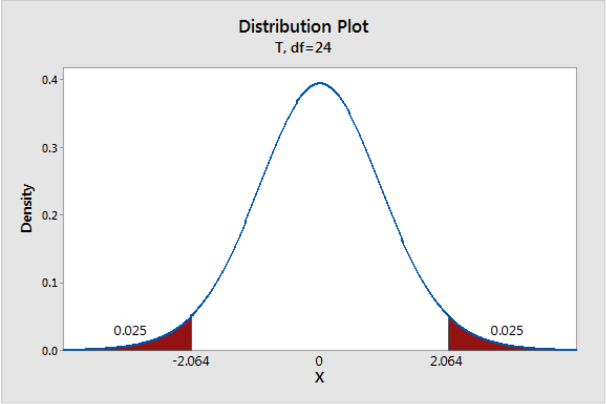
- Why an F test?
- Hypothesis Testing and t Tests: *Review*
- t Tests: Testing single parameters/restrictions
- F Tests: Testing multiple parameters & linear restrictions
- Running the F Test: *Some intuition*
- The F Statistic: F = %Δ*SSRs* / %Δ*dofs*
- The F Statistic & three *Goodness of Fit* metrics
- Running the F Test: *More Formally*

- Some F distributions
- Running F tests in Stata is a snap!
- Testing a single parameter: $F = t^2$
- … An Example
- F = t^2: *Who knew? … Well, You knew!*
- Reported F Stat's in OLS output
- *Babies and Bathwater*
- *F Stats, Adjusted* R^2 *and t Stats*
- Adding and dropping RHS variables: *F stats and Adjusted* R^2

Why bother with an F Test? Testing a *joint* Null Hypothesis

- Sometimes you want to test a joint Null Hypothesis (multiple hypotheses)… *statistical significance?*
- Examples:
  - Population effects in NFL tix prices: pop and pop^2
  - Regional fixed effects in sovereign debt models: region dummies/FEs
  - AppleMusic effects: AppleMusic dummy and trend
  - Gender differences in estimated wage equations: ftenure and female
  - Uber tipping: all those FEs (fixed effects)… day, time, pick-up and drop-off locations

Hypothesis Testing and t Tests: *Review*



- The t statistic, the *Cornerstone of Inference*: $\dfrac{B_x - \beta_x}{se(B_x)}$

- MLR.1-6: the t statistic will have a t distribution with $n-k-1$ dofs.

- Hypothesis Testing I: $H_0 : \beta_x = 0$

  - Test I: Critical value, c, defined by the significance level, α, and t_{n-k-1}: $P(|t_{n-k-1}| > c) = \alpha$

    Reject $H_0 : \beta_1 = 0$ if $|t\ stat| = \left|\dfrac{\hat{\beta}_1}{se(\hat{\beta}_1)}\right| > c$, if the t stat > the critical value, c

  - Test II: p value, defined by the *t stat* and t_{n-k-1}: $p = P\left(|t_{n-k-1}| > |t\ stat|\right)$

    Reject $H_0 : \beta_1 = 0$ if $p < \alpha$, if the p-value < the significance level, α

  - Tests I and II are equivalent: Reject under Test I if and only if you reject under Test II

t Tests: Testing single parameters/restrictions

Testing single parameter values:

- Critical value c defined by $P(|t_{n-k-1}| > c) = \alpha$

- $H_0 : \beta_x = 0 : \ t = \dfrac{B_x - \beta_x}{se(B_x)} = \dfrac{B_x}{se(B_x)} \sim t_{n-k-1}$

  - Reject if $|t| = \left|\dfrac{\hat{\beta}_1}{se(\hat{\beta}_1)}\right| > c$,

    or if $p = P\left(|t_{n-k-1}| > |t|\right) < \alpha$

- $H_0 : \beta_x = 2 : \ t = \dfrac{B_x - \beta_x}{se(B_x)} = \dfrac{B_x - 2}{se(B_x)} \sim t_{n-k-1}$

  - Reject if $|t| = \left|\dfrac{\hat{\beta}_1 - 2}{se(\hat{\beta}_1)}\right| > c$,

    or if $p = P\left(|t_{n-k-1}| > |t|\right) < \alpha$

Testing single restrictions:

- Impose the restriction and test for differences; MLR.1: $y = \beta_0 + \beta_x x + \beta_z z + U$

- Testing $H_0 : \beta_x = \beta_z$

- Impose the restriction: generate $w = x + z$

- I: … regress y on w and x: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_w w + \hat{\beta}_x x = \hat{\beta}_0 + \left(\hat{\beta}_w + \hat{\beta}_x\right)x + \hat{\beta}_w z$

  - … test for differences I: $H_0 : \beta_x = 0$

- II: … or regress y on w and z: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_w w + \hat{\beta}_z z = \hat{\beta}_0 + \hat{\beta}_w x + \left(\hat{\beta}_w + \hat{\beta}_z\right)z$

  - … test for differences II: $H_0 : \beta_z = 0$

F Tests: Testing multiple parameters & linear restrictions

- F tests: test *linear restrictions* on estimated parameters in SLR and MLR models.
- L*inear restrictions:* Linear functions (of the parameters to be estimated) is/are set equal to zero; you can have lots of restrictions (but not more than the number of parameters to be estimated); examples below
- Counting restrictions: The number of linear restrictions, q (why q? no idea!), will matter; count restrictions by counting ='s signs (drop redundant restrictions)
- Here are some examples:
  - q = 1: a) $\beta_1 = \beta_2$, and b) $\beta_1 + 2\beta_2 = 0$
  - q = 2: a) $\beta_1 = 0 \text{ and } \beta_2 = 0$, b) $\beta_1 = 1 \text{ and } \beta_2 = 2$, and c) $\beta_1 = \beta_2$, $\beta_1 = \beta_3$ and $\beta_2 = \beta_3$ (one restriction in c) is redundant)

Running the F Test: *Some intuition*

- ***Step 1***: Start with the Null hypothesis that it's A-OK to impose some linear restrictions on the estimated coefficients in our model.
- ***Step 2***: Estimate the model with and without those restrictions… and focus on the SSRs and how they change.
  - Since we've imposed a restriction (or restrictions) on the estimated coefficients, the SSRs will almost always increase: $SSR_R \geq SSR_{UR}$.
- ***Step 3***: OK, so SSRs increased. That's no surprise! But by how much? … a lot? … or maybe not so much?
  - ***Big increase in SSRs:*** If SSRs increase by a lot (whatever that is) then the restrictions severely impacted the performance of the model, and so we reject the Null Hypothesis (which was that imposing the restrictions was A-OK). ***Reject, Reject Reject!***
  - ***It's not so big***: But if not so much, then maybe those restrictions weren't so bad after all, and we might fail to reject…. Which is to say that it really was **A-OK** to impose those restrictions after all!

The F Statistic: F = %Δ*SSRs* / %Δ*dofs*

- The F statistic is defined by: $Fstat = F = \dfrac{\left(SSR_R - SSR_{UR}\right)/q}{SSR_{UR}/(n-k-1)}$,

  where q is the number of restrictions (e.g. the number of '='s), and $n-k-1$ is the number of degrees of freedom in the unrestricted (UR) model.

- By construction $F \geq 0$, assuming that F is well defined (since $SSR_R \geq SSR_{UR}$).

The F statistic is an elasticity! *Who knew?*

- The F statistic is really just an elasticity. We can rewrite the equation for F as:
  - $F = \dfrac{\left(SSR_R - SSR_{UR}\right)/SSR_{UR}}{q/(n-k-1)} = \dfrac{\%\Delta SSR}{\%\Delta dofs}$

F

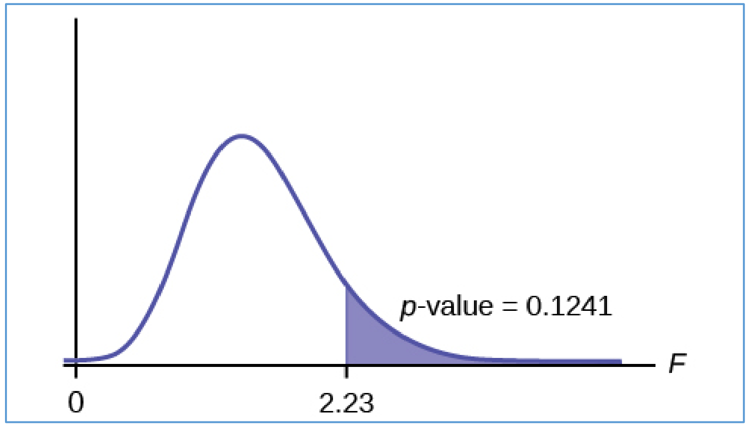
- So the F statistics tells you the %change in SSRs for a given %change in degrees of freedom (you might call this *bang per buck*).

The F Statistic & Three *Goodness of Fit* Metrics

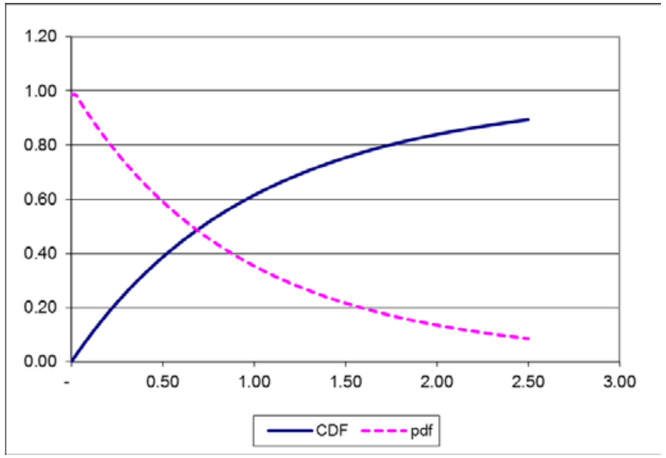
- $SSR\!: \quad F = \dfrac{\left(SSR_R - SSR_{UR}\right) / \, SSR_{UR}}{q \, / \, (n-k-1)} = \dfrac{\%\Delta SSR}{\%\Delta dofs}$
- $R^2: \quad F = \dfrac{(n-k-1)}{q} \dfrac{\left(R_{UR}^2 - R_R^2\right)}{\left(1 - R_{UR}^2\right)} = \dfrac{\Delta R^2 \; / \left(1 - R_{UR}^2\right)}{\%\Delta dofs}$
- $\text{SSE:} \quad F = \dfrac{(n-k-1)}{q} \dfrac{SSE_{UR} - SSE_R}{SSR_{UR}} = \dfrac{\Delta SSE \; / \; SSR_{UR}}{\%\Delta dofs}$

Running the F Test: *More formally*

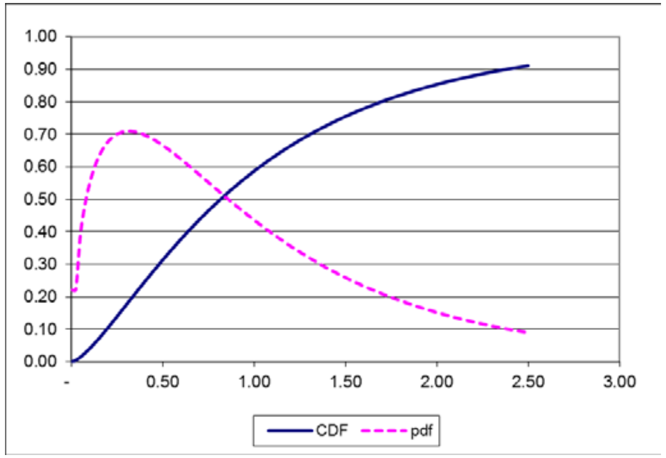
- The Null Hypothesis is that the restrictions are A-OK.
- Pick a small significance level, α , say $\alpha = .05$, the maximum acceptable probability of a False Rejection.
- Critical value: Find the critical value, c, such that $prob\left(F(q, n-k-1) > c\right) = \alpha$.
  - MLR.1-.6: The F statistic will have an F distribution with parameters *q* and *n-k-1*.
- p value: Generate the p value as the probability in the tail to the right of the Fstat: $p = prob\left(F(q, n-k-1) > Fstat\right)$
- Reject the Null Hypothesis if $Fstat > c$ or $p < \alpha$ … and fail to reject otherwise.



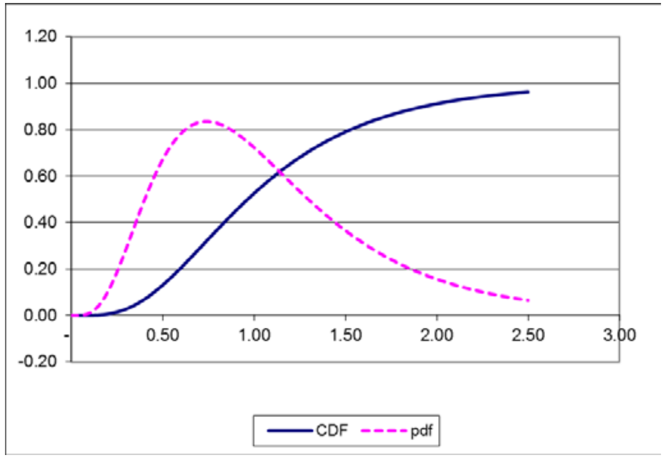
Some F Distributions



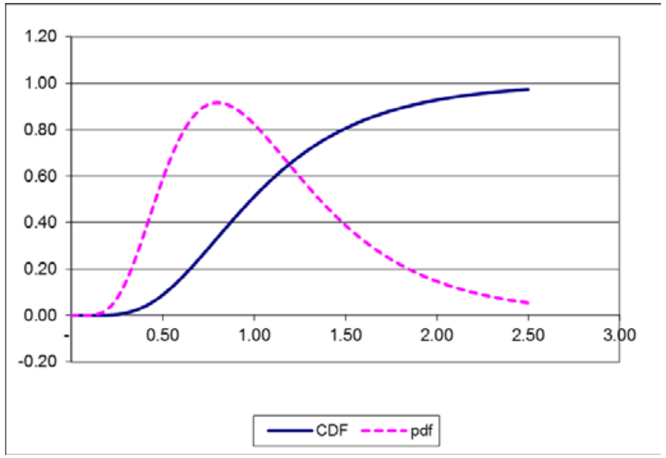
F(2,20) Distribution



F(3,20) Distribution



F(10,20) Distribution



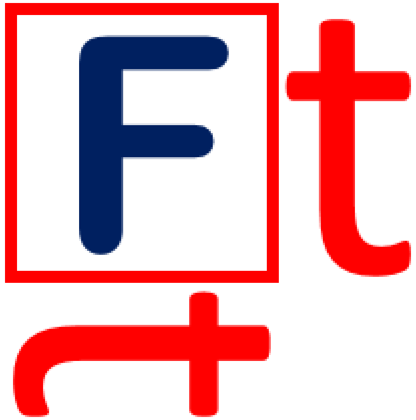
F(15,20) Distribution

Running F tests in Stata is a *snap*

- For example, to test the linear restrictions above, you would run the following Stata commands just after estimating your OLS model ($reg\ \ y\ x_1\ \ x_2\ \ x_3$):
  - $\beta_1 = \beta_2$: $test\ (x_1 = x_2)$
  - $\beta_1 + 2\beta_2 = 0$: $test\ (x_1 + 2x_2 = 0)$
  - $\beta_1 = 0\ and\ \ \beta_2 = 0$: $test\ (x_1 = 0)\ (x_2 = 0)$ or just $test\ x_1\ \ x_2$ ($'=0'$ is assumed if no value is specified)
  - $\beta_1 = 1\ and\ \ \beta_2 = 2$: $test\ (x_1 = 1)\ (x_2 = 2)$
  - $\beta_1 = \beta_2\ \ and\ \beta_1 = \beta_3$: $test\ (x_1 = x_2)\ (x_1 = x_3)$
- How to read the *test* syntax: Insert *the true parameter for the variable* in front of each variable name. So, for example: $test\ (x_1 = x_2)$ reads as *Test the Null Hypothesis that the true parameter for the variable* x_1 *equals the true parameter for the variable* x_2 .. or perhaps more concisely, *Test the Null … the true parameter for* x_1 *equals the true parameter for* x_2 .

Testing a Single Parameter: $F = t^2$

Testing $H_0: \ \beta_x = 0$



- **t test**: Reject if $\left|t_{\hat{\beta}_x}\right| > c$, where $t_{\hat{\beta}_x} = \dfrac{\hat{\beta}_x}{se_{\hat{\beta}_x}}$ and $prob\,(\left|t_{n-k-1}\right| > c) = \alpha$, or if

  $p = prob\left(\left|t_{n-k-1}\right| > \left|t_{\hat{\beta}_x}\right|\right) < \alpha$ (c is the critical value for the t test)

- **F test**: Reject if $Fstat > cc$, where $Fstat = \dfrac{\left(SSR_R - SSR_{UR}\right)}{SSR_{UR} / dofs}$ and

  $prob\left(F(1, n-k-1) > cc\right) = \alpha$, or if $p = prob\left(F(1, n-k-1) > Fstat\right) < \alpha$ (cc is the critical value for the F test)

- These tests are identical since:
  - $Fstat = t^2_{\hat{\beta}_x}$ (the F stat is the square of the t stat),
  - $cc = c^2$ (the critical value for the F test is the square of the critical value for the t test)
  - $prob\left(F(1, n-k-1) < x^2\right) = prob\left(-x < t_{n-k-1} < x\right)$ for any $x > 0$ (so you might say that the F distribution is the square of the t distribution)

Testing a Single Parameter, cont'd: *$F = t^2$ example*

Test $H_0: \beta_{hgt} = 0$ in the following SLR model:

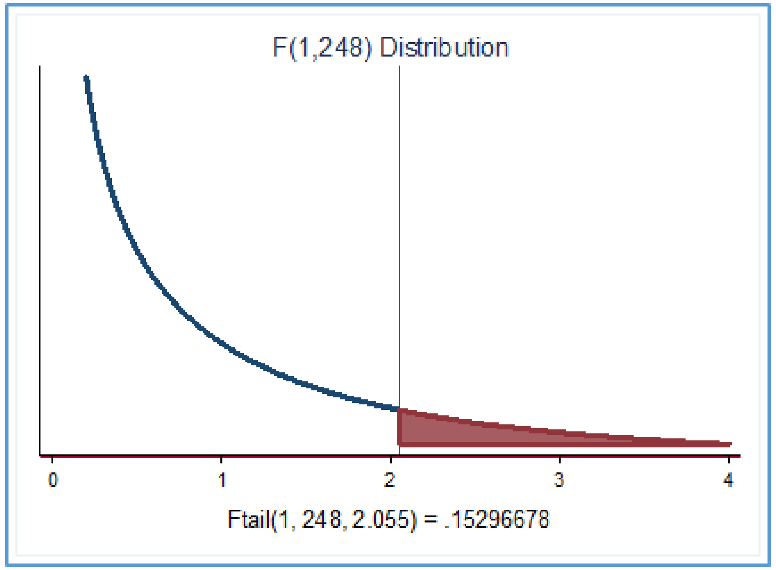
```
      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(3, 248)       =    213.67
       Model |  10872.5504         3  3624.18347   Prob > F        =    0.0000
    Residual |  4206.46623       248  16.9615574   R-squared       =    0.7210
-------------+----------------------------------   Adj R-squared   =    0.7177
       Total |  15079.0166       251  60.0757635   Root MSE        =    4.1184

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |   -.120415   .0222516    -5.41   0.000    -.1642411   -.0765888
         abd |    .879846   .0579164    15.19   0.000     .7657751    .9939168
         hgt |  -.1181607   .0824192    -1.43   0.153    -.2804915    .0441701
       _cons |  -32.66247    6.51936    -5.01   0.000    -45.50285    -19.8221
------------------------------------------------------------------------------
```

Run the F test:

```
. test hgt
 ( 1)  hgt = 0

       F(  1,   248) =    2.06
            Prob > F =    0.1529
```



Notice the equivalence of the F stat/test and the t stat/test: $1.43^2 = 2.04$ (rounding error) and $Prob > F = 0.1529 = P > |t| = 0.153$.

F = t^2: *Who knew? … Well, You knew!*

$$t_x^2 = dofs\frac{\Delta R_x^2}{1-R^2}$$

The Convergence of *Goodness of Fit* and *Inference*!

- Recall those convergence results: In SLR and MLR models, a variable's t stat reflected it's incremental contribution to R^2:
  - $t_{\hat{\beta}_x}^2 = dofs\frac{\Delta R_x^2}{1-R^2}$, where ΔR_x^2 is a RHS variable's incremental contribution to R^2.
- If you consider the full model to be unrestricted, and the restricted model to restrict the x coefficient to be zero (so effectively dropping x from the model), the F test statistic is:
  - $F = \frac{(n-k-1)}{1}\frac{\left(R_{UR}^2 - R_R^2\right)}{1-R_{UR}^2} = dofs\frac{\Delta R_x^2}{1-R^2}$.
- And since $t_{\hat{\beta}_x}^2 = F$ (we are testing just one restriction) , we have $t_{\hat{\beta}_x}^2 = dofs\frac{\Delta R_x^2}{1-R^2}$!
- So the connection between t stats and incremental R^2, which probably seemed to you to have come out of nowhere, was in fact just an example of F stats in action.

Reported F Stat's in OLS Output

(the *F stat for the regression*)

```
. reg Brozek hgt wgt abd if _n < 8

      Source |       SS           df       MS      Number of obs   =         7
-------------+----------------------------------   F(3, 3)         =     12.50
       Model |  322.483875         3  107.494625   Prob > F        =    0.0335
    Residual |  25.8046962         3   8.6015654   R-squared       =    0.9259
-------------+----------------------------------   Adj R-squared   =    0.8518
       Total |  348.288571         6  58.0480952   Root MSE        =    2.9328

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hgt |  -2.340897   1.249452    -1.87   0.158    -6.317211    1.635417
         wgt |   .1964088   .2129627     0.92   0.424    -.4813334     .874151
         abd |   1.050577   .3059068     3.43   0.041      .077045    2.024109
       _cons |   53.73654   71.80124     0.75   0.509    -174.7671    282.2401
------------------------------------------------------------------------------

. test hgt wgt abd if _n < 8

 ( 1)  hgt = 0
 ( 2)  wgt = 0
 ( 3)  abd = 0

       F(  3,     3) =   12.50
            Prob > F =    0.0335
```

```
. di (3/3)*.9259/(1-.9259)
12.495277
```

```
. di
(3/3)*(322.483875/25.8046962)
12.4971
```

- F stat/test for the regression: Testing the null hypothesis that all of the (non-intercept) true parameter values are zero.
- $$F = \frac{R^2/k}{\left[1-R^2\right]/(n-k-1)} = \frac{dofs}{k}\frac{R^2}{1-R^2} = \frac{dofs}{k}\frac{SSE}{SSR},$$
  since $R_R^2 = 0$ and $R_{UR}^2 = R^2$.
- Used to assess the overall statistical significance of the regression. In practice, the reported F stats are almost always quite sizable (in double, if not triple, digits).
- If your F stat is even close to single digits, you probably have a *crummy* model! … and should start again!

Babies and Bathwater

- Be careful about th*rowing out the baby with the bath water*… you don't want to exclude a significant explanatory variable from your model just because it happens to be associated with a set of RHS variables that are jointly insignificant.
- Or put differently: *F tests judge variables by the friends they keep!*
- **Example**: The F test does not reject at the 10% level the Null Hypothesis that the *inflation* and *deficit_gdp* parameters are zero… even though *deficit_gdp* is statistically significant at almost the 5% level (and has $p < 0.05$ when *inflation* is dropped from the model).

```
. reg NSRate corrupt gdp inflation deficit_gdp debt_gdp eurozone if _n < 30

      Source |       SS           df       MS      Number of obs   =        29
-------------+----------------------------------   F(6, 22)        =     12.99
       Model |  88.8122105         6  14.8020351   Prob > F        =    0.0000
    Residual |  25.0657205        22  1.13935093   R-squared       =    0.7799
-------------+----------------------------------   Adj R-squared   =    0.7199
       Total |  113.877931        28  4.06706897   Root MSE        =    1.0674

------------------------------------------------------------------------------
      NSRate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     corrupt |    .644807   .1078044     5.98   0.000     .4212342    .8683797
         gdp |   .0002144   .0000765     2.80   0.010     .0000557    .0003731
   inflation |   .0361479   .0846488     0.43   0.674     -.139403    .2116988
 deficit_gdp |  -.0732749    .035707    -2.05   0.052    -.1473266    .0007768
    debt_gdp |  -.0220782   .0094606    -2.33   0.029    -.0416982   -.0024581
    eurozone |   .9996265   .4721874     2.12   0.046     .0203699    1.978883
       _cons |   4.269966   1.226366     3.48   0.002     1.726638    6.813295
------------------------------------------------------------------------------
```

```
. test inflation deficit_gdp

 ( 1)  inflation = 0
 ( 2)  deficit_gdp = 0

       F(  2,    22) =    2.22
            Prob > F =    0.1320
```

F Stats, Adjusted R-squared and t Stats

- Adjusted R-square increases or decreases with changes in the RHS variables depending on the associated F statistic. Start with the unrestricted model, UR, and move to the restricted model, R… caused by dropping multiple variables from the UR model.

- The change in adjusted R-sq will be: $\Delta\bar{R}^2 = \bar{R}_R^2 - \bar{R}_{UR}^2 = \dfrac{q\left[1-\bar{R}_{UR}^2\right]}{(n-k-1)+q}\left[1-F\right]$.

- The sign of this expression will depend on whether the F statistic is greater or less than 1:
  - $Fstat < 1 \Rightarrow \Delta\bar{R}^2 = \bar{R}_R^2 - \bar{R}_{UR}^2 > 0 \Rightarrow \bar{R}_R^2 > \bar{R}_{UR}^2$,
  - $Fstat = 1 \Rightarrow \Delta\bar{R}^2 = \bar{R}_R^2 - \bar{R}_{UR}^2 = 0 \Rightarrow \bar{R}_R^2 = \bar{R}_{UR}^2$, and
  - $Fstat > 1 \Rightarrow \Delta\bar{R}^2 = \bar{R}_R^2 - \bar{R}_{UR}^2 < 0 \Rightarrow \bar{R}_R^2 < \bar{R}_{UR}^2$.

- ***You've seen this before!*** Recall that for a single restriction, the F statistic is the square of the t stat, and so, as you saw in *MLR Assessment*: $\bar{R}^2$ increases when you drop a single RHS variable having $|t| < 1$, decreases if $|t| > 1$, and is unchanged if $|t| = 1$.

Adding and Dropping RHS Variables: *F stats and Adjusted R^2*

	(1) Brozek	(2) Brozek	(3) Brozek	(4) Brozek
wgt	-0.151***	-0.129***	-0.154***	-0.132***
	(-5.21)	(-3.68)	(-4.92)	(-3.56)
abd	0.937***	0.911***	0.940***	0.914***
	(17.19)	(15.37)	(16.72)	(15.01)
hip	-0.154	-0.182	-0.153	-0.181
	(-1.22)	(-1.42)	(-1.21)	(-1.41)
thigh	0.277*	0.255*	0.277*	0.255*
	(2.43)	(2.21)	(2.42)	(2.20)
hgt		-0.0983		-0.0982
		(-1.13)		(-1.12)
ankle			0.0477	0.0459
			(0.24)	(0.23)
_cons	-41.80***	-32.33**	-42.73***	-33.24**
	(-6.45)	(-3.05)	(-5.65)	(-2.93)
N	252	252	252	252
R-sq	0.7253	0.7267	0.7254	0.7268
adj. R-sq	0.7208	0.7212	0.7198	0.7201
rmse	4.095	4.093	4.103	4.101

tstats and t tests:

- (2) to (1) (drop *hgt*): hgt |tstat| > 1, R^2 and adj R^2 decrease, RMSE increases
- (3) to (1) (drop *ankle*): *ankle* |tstat| < 1, R^2 decreases, adj R^2 increases, and RMSE decreases

F stats and F tests:

- (4) to (1) (drop *hgt* and *ankle*): Since adj R^2 increases, the F test associated with dropping *hgt* and *ankle* from (4) will have an Fstat<1 …

Here are the F test results:

```
. reg Brozek wgt abd hip thigh hgt ankle
. test hgt ankle

 ( 1)  hgt = 0
 ( 2)  ankle = 0

       F(  2,   245) =    0.66
            Prob > F =    0.5180
```

It's a Wrap!


```
. reg NSRate corrupt lngdp inflation deficit_gdp debt_gdp eurozone

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(6, 101)       =    104.46
       Model |  288.476069         6  48.0793449   Prob > F        =    0.0000
    Residual |  46.4871714       101  .460269023   R-squared       =    0.8612
-------------+----------------------------------   Adj R-squared   =    0.8530
       Total |  334.963241       107  3.13049758   Root MSE        =    .67843

------------------------------------------------------------------------------
      NSRate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     corrupt |   .5404159   .0369972    14.61   0.000     .4670235    .6138084
       lngdp |   .3366617   .0370923     9.08   0.000     .2630806    .4102428
   inflation |   -.043741    .017731    -2.47   0.015    -.0789145   -.0085674
 deficit_gdp |  -.0504287   .0129655    -3.89   0.000    -.0761487   -.0247087
    debt_gdp |  -.0092895   .0022185    -4.19   0.000    -.0136904   -.0048885
    eurozone |   .5062781   .2020622     2.51   0.014     .1054411    .9071152
       _cons |   2.661791   .2395918    11.11   0.000     2.186505    3.137077
------------------------------------------------------------------------------
```