F Tests and F statistics

- ***Testing Linear Restrictions***
- ***F Stats and F Tests***
- ***F Distributions***
- ***F stats (w/ R^2)***
- ***F Stats and tStat^{2}s: Testing a single parameter***
- ***Example I: Bodyfat***
- ***Reported F Stat's in OLS Output***
- ***… Example I cont'd***
- ***F Stats and tStat^{2}s: Dropping multiples RHS variables***
- ***Babies and Bathwater***
- ***F Stats, Adjusted R-squared, RMSE and tStats***
- ***Adding and Dropping RHS Variables***
- ***Example II: More Bodyfat***

Testing Linear Restrictions

1. Why do we test? Because we want to know! Do we have enough evidence to the contrary to reject our Null Hypothesis? Run the test!

2. We use the F test to test linear restrictions in SLR and MLR models. *What's a linear restriction?*, you ask… well, here are some examples: [1]

   1) $\beta_1 = \beta_2$, 2) $\beta_1 + 2\beta_2 = 0$, 3) $\beta_1 = 0 \; and \; \beta_2 = 0$, and 4) $\beta_1 = 1 \; and \; \beta_2 = 2$.

   a. We'll be interested in counting the number of restrictions. To do that, just count the number of equals signs ('='s). So in the previous examples, 1) and 2) have one restriction, and 3) and 4) have two.

      i. For some unknown reason we use the letter q to count the number of restrictions.

3. Running the F test:

   a. ***Step 1***: Start with the Null hypothesis that it's A-OK to impose some linear restrictions on the estimated coefficients in our model.

   b. ***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: $SSR_R \geq SSR_{UR}$.[2]

   c. ***Step 3***: OK, so SSRs increased. That's no surprise! But by how much? … a lot? … or maybe not so much?

---

[1] Linear restrictions are linear functions of the unknown parameters.
[2] The subscripts refer to the *Restricted* and *Unrestricted* models.

i. *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 the restrictions were A-OK).

ii. *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.

4. As you'll see below, 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 ($reg\ y\ x_1\ x_2$):

   a. $\beta_1 = \beta_2$: $test\ (x_1 = x_2)$

   b. $\beta_1 + 2\beta_2 = 0$: $test\ (x_1 + 2x_2 = 0)$

   c. $\beta_1 = 0\ and\ \beta_2 = 0$: $test\ (x_1 = 0)\ (x_2 = 0)$ or just $test\ x_1\ x_2$ ('$=0$' is assumed if no value is specified)

   d. $\beta_1 = 1\ and\ \beta_2 = 2$: $test\ (x_1 = 1)\ (x_2 = 2)$

F Stats and F Tests

5. We use the F statistic (and the F distribution) to do the F test.

   a. The F statistic is defined by:

$$Fstat = F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k-1)},$$

   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.

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

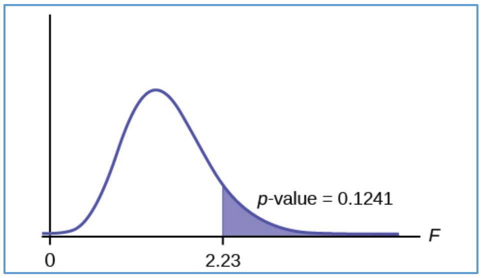
6. The F statistic as an elasticity! *Who knew?*

   a. The F statistic is really just an elasticity. We can rewrite the equation for F as:

$$F = \frac{(SSR_R - SSR_{UR})/SSR_{UR}}{q/(n-k-1)} = \frac{\%\Delta SSR}{\%\Delta dofs}$$

   b. 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*).

7. We do the F test following the classical hypothesis testing protocol:
    a. The Null Hypothesis is that the restrictions are A-OK.
    b. Focus on Type I errors and the probability of a False Rejection.
    c. Pick a small significance level, α , say $\alpha = .05$, which will be the maximum acceptable probability of a False Rejection.
    d. For the given Fstat, compute the p-value as the probability in the tail to the right of the Fstat: $p = P(F(q, n-k-1) > Fstat)$.[3]
    e. The p value is the probability that you'd observe an F statistic at least as large as Fstat when the Null Hypothesis was in fact true.
    f. Reject the Null Hypothesis if the $p < \alpha$ … and fail to reject otherwise.
        i. So reject the Null Hypothesis if the F statistic is large $\left(|\%\Delta SSR| \gg |\%\Delta dof|\right)$ and the associated probability level (*p-value*) is small (less than the significance level α). In other words: the restrictions are rejected because when they were imposed, the world got a whole lot worse.
        ii. If the Fstat (elasticity) is small $\left(|\%\Delta SSR| \ll |\%\Delta dof|\right)$ so that relatively speaking, SSRs did not increase by very much, then the associated probability level (*p-value*) is high, and the restrictions are not rejected and are deemed to be A-OK.

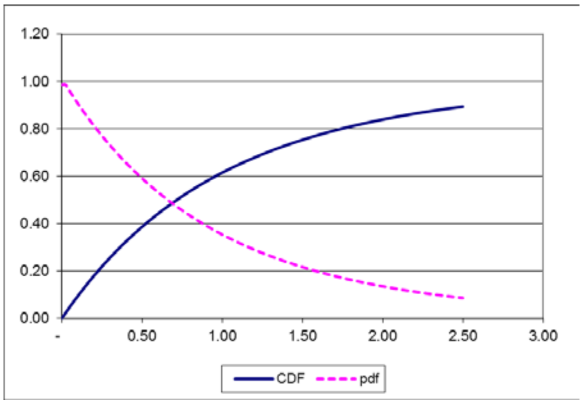


8. And so we reject the Null Hypothesis if the probability value associated with this test statistic is below the desired significance level $(p < \alpha)$ … or equivalently, if the F statistic is above the critical value c: $Fstat > c$, where $prob\{F(q, n-k-1) > c\} = \alpha$.
9. This should be very familiar sounding… as it is, after all, classical hypothesis testing.

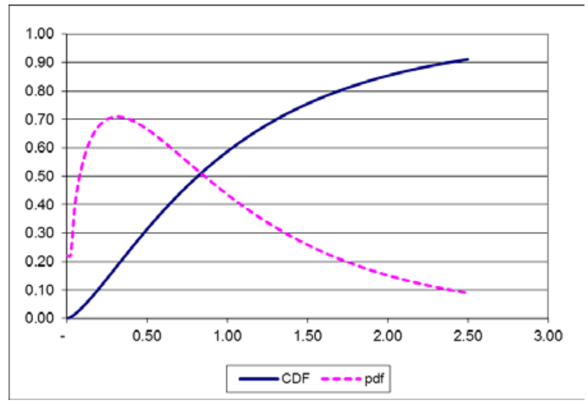
[3] Here's Stata syntax for the right tail F probability: Ftail(q, n-k-1, Fstat).

F Distributions

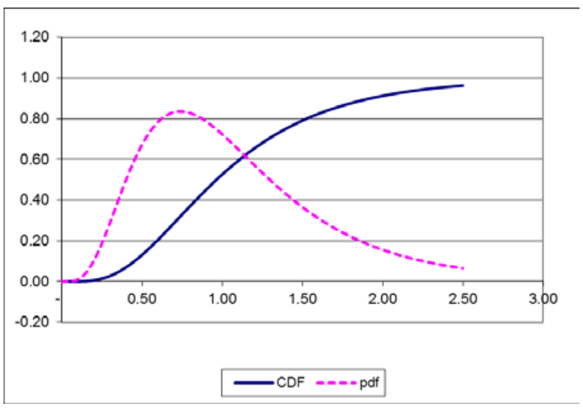
10. Under MLR.1-MLR.6, the F statistic will have a $F(q, n-k-1)$ (two parameters) distribution. Here are some F distributions for different parameter values:



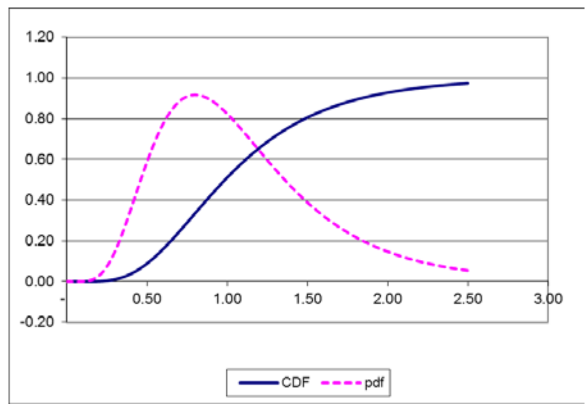
F(2,20) Distribution



F(3,20) Distribution



F(10,20) Distribution



F(15,20) Distribution

F stats (w/ R^2 and SSEs)

11. When you impose restrictions on the estimated coefficients, SSR's typically increase…. and R^2's and SSE's typically decrease. While the standard definition of the F statistic works with SSRs, you can equivalently define F stats using the other Goodness of Fit metrics:

    a. R^2's: Define $R_R^2 = 1 - \dfrac{SSR_R}{SST}$ and $R_{UR}^2 = 1 - \dfrac{SSR_{UR}}{SST}$. Then

    $$F = \frac{SST\left(\left[1-R_R^2\right]-\left[1-R_{UR}^2\right]\right)/q}{SST\left[1-R_{UR}^2\right]/(n-k-1)} = \frac{\left(R_{UR}^2-R_R^2\right)/q}{\left[1-R_{UR}^2\right]/(n-k-1)} = \frac{(n-k-1)}{q}\frac{\Delta R^2}{\left(1-R_{UR}^2\right)},$$ where

    $\Delta R^2 = \left(R_{UR}^2 - R_R^2\right)$. Alternatively, $F = \dfrac{\Delta R^2/\left(1-R_{UR}^2\right)}{\%\Delta dofs}$

b. SSE's: Since $R_R^2 = \dfrac{SSE_R}{SST}$, $R_{UR}^2 = \dfrac{SSE_{UR}}{SST}$, and $1 - R_{UR}^2 = \dfrac{SSR_{UR}}{SST}$,

$$F = \frac{\left(SSE_{UR} - SSE_R\right)/(q\ SST)}{\left[SSR_{UR}\ /\ SST\right]/(n-k-1)} = \frac{(n-k-1)}{q}\frac{\Delta SSE}{SSR_{UR}} = \frac{\Delta SSE\ /\ SSR_{UR}}{\%\Delta dofs}.$$

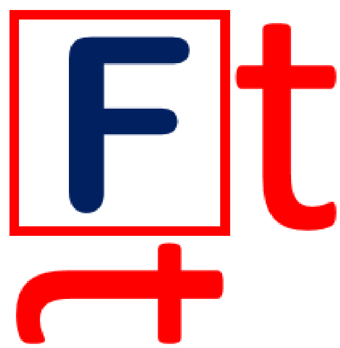
12. And so we have three equivalent ways of defining the F statistic, driven by the changes in three different *Goodness-of-Fit* metrics:

a. $F = \dfrac{\left(SSR_R - SSR_{UR}\right)/\ SSR_{UR}}{q\ /\ (n-k-1)} = \dfrac{\%\Delta SSR}{\%\Delta dofs}$,

b. $F = \dfrac{(n-k-1)}{q}\dfrac{\left(R_{UR}^2 - R_R^2\right)}{\left(1 - R_{UR}^2\right)} = \dfrac{\Delta R^2\ /\left(1 - R_{UR}^2\right)}{\%\Delta dofs}$, and

c. $F = \dfrac{(n-k-1)}{q}\dfrac{SSE_{UR} - SSE_R}{SSR_{UR}} = \dfrac{\Delta SSE\ /\ SSR_{UR}}{\%\Delta dofs}$

F Stats and tStat²s: Testing a single parameter



13. If you apply the F test to the case of testing just one parameter value (say, testing $\beta_x = 0$), the F test will effectively be the same as the t test… so there is no inconsistency between the two tests. In fact, for these tests $Fstat = tstat^2$, and the p-values for the two statistics will be the same, since the F distribution will essentially be the square of the t distribution:

$t_{\hat{\beta}_x}^2 = F_{\beta_x = 0}$ and

$prob\left\{F(1, n-k-1) < x^2\right\} = prob\left\{-x < t_{n-k-1} < x\right\}$ for $x > 0$.

(See examples below.)

14. ***The convergence of assessment and inference***. In fact, you've already seen these sorts of Fstats in action. You may recall that we previously observed that in SLR and MLR models, a variable's t stat reflected it's incremental contribution to R^2:

$t_{\hat{\beta}_x}^2 = dofs\dfrac{\Delta R_x^2}{1 - R^2}$, where ΔR_x^2 is a RHS variable's incremental contribution to R^2.

a. 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:

$$F = \frac{(n-k-1)}{1}\frac{\left(R_{UR}^2 - R_R^2\right)}{1 - R_{UR}^2} = dofs\frac{\Delta R_x^2}{1 - R^2}.$$

b. And since $t^2_{\hat{\beta}_x} = F$ (we are testing just one restriction) , we have $t^2_{\hat{\beta}_x} = dofs\frac{\Delta R^2_x}{1-R^2}$.

15. 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.

Example I: Bodyfat

16. Test $H_0: \beta_{hgt} = 0$ in the following SLR model. So we are testing the Null Hypothesis that the true *hgt* parameter is zero.

```
. reg Brozek wgt abd hgt

      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
------------------------------------------------------------------------------

. test hgt

 ( 1)  hgt = 0

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

a. Notice the equivalence of the F test and the t test: $1.43^2 = 2.04$ (rounding error) and $Prob > F = 0.1529 = P >|t| = 0.153$. Since the p value is .15 (well above .1 and .05), we cannot reject the null hypothesis that the true *hgt* parameter is 0, at any standard (and attractive) level of statistical significance.

b. Here's the *by-hand* F test for testing the null hypothesis that the *hgt* parameter is 0. To complete the calculation of the F statistic, we need to run the restricted model:

```
. reg Brozek wgt abd

      Source |       SS           df       MS      Number of obs   =       252
-------------+----------------------------------   F(2, 249)       =    318.13
       Model |  10837.6881         2  5418.84407   Prob > F        =    0.0000
    Residual |  4241.32849       249  17.0334478   R-squared       =    0.7187
-------------+----------------------------------   Adj R-squared   =    0.7165
       Total |  15079.0166       251  60.0757635   Root MSE        =    4.1272

------------------------------------------------------------------------------
      Brozek |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |  -.1364535   .0192756    -7.08   0.000    -.1744175   -.0984895
         abd |    .915138   .0525355    17.42   0.000     .8116675    1.018609
       _cons |  -41.34812   2.412986   -17.14   0.000    -46.10059   -36.59566
------------------------------------------------------------------------------
```

c. Now that we have the SSRs (SSR_R and SSR_{UR}), we can compute the F stat associated with this test:

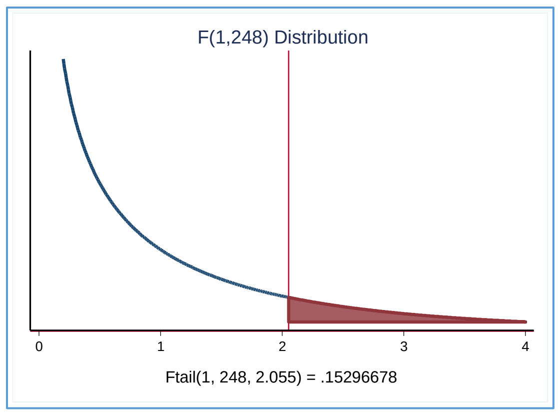
i. $$F = \frac{\left(SSR_R - SSR_{UR}\right)/SSR_{UR}}{q/(n-k-1)}$$

$$= \frac{\left(4,241.33 - 4,206.47\right)/4,206.47}{1/248}$$

$$= 2.055$$

ii. And the associated p value: Prob > F is .15

Stata: di Ftail(1,248,2.055) =.15296678



17. Other tests. Here are some examples of other F tests, which could as well be t tests, testing for equality of parameters:

```
. test wgt = hgt

 ( 1)  wgt - hgt = 0

       F(  1,   248) =    0.00
            Prob > F =    0.9812


. test wgt = abd

 ( 1)  wgt - abd = 0

       F(  1,   248) =  161.61
            Prob > F =    0.0000
```

Reported F Stat's in OLS Output

18. Standard regression packages typically report by default the F statistic (and related p value) associated with a particular F test… of the Null Hypothesis that all of the (non-intercept) parameter values are zero. The reported ***F statistic***, or ***F stat*** for short, is useful in seeing whether the whole lot of RHS variables have explanatory power… so it has something to say about the collective precision of estimation of the RHS variables as a group.

19. In conducting this F test, the restricted model has $R_R^2 = 0$, since the dependent variable is regressed on just the constant term. And since $R_{UR}^2 = R^2$ (the reported R-squared for the regression), we have:

$$F = \frac{R^2 / k}{\left[1 - R^2\right] / (n-k-1)} = \frac{dofs}{k}\frac{R^2}{1-R^2} = \frac{dofs}{k}\frac{SSE}{SSR}.$$

This is the reported F statistic, used to assess the overall statistical significance of the regression..

20. 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 need to find a better group of explanatory variables… or better data… or maybe both!

21. So now you know. That F test can be used to test the Null hypothesis that all of the (non-intercept) parameter values are zero. And as I said before: If you cannot reject that hypothesis, you have a really crummy model.

Example I …continued

22. Test $H_0: \ \beta_{wgt} = 0, \ \beta_{abd} = 0 \ \ and \ \ \beta_{hgt} = 0$ in the SLR model above. So we are testing the Null Hypothesis that the true parameters for the RHS variables are all zero. Run the F test after running the regression:

```
. test wgt abd hgt

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

       F(  3,   248) =  213.67
            Prob > F =    0.0000
```

a. Notice the agreement with the reported F(3, 248) for the regression… as well as Prob > F.

b. And here's a replication of the F stat for the regression, F(3.248):

$$F = \frac{R^2 / k}{\left[1 - R^2\right] / (n-k-1)} = \frac{.7210/3}{(1-.7210)/248} = 213.63 \ \ldots \ \text{close enough}$$

F Stats and tStat^{2}s: Dropping multiples RHS variables

23. Sometimes we wonder whether the inference gods will bless the dropping of multiple RHS variables, or equivalently, the constraining of multiple RHS OLS/MLR estimated coefficients to be 0. Bring on the F test!

    a. Perhaps surprisingly, there can be an approximate relation between the test F statistic and the reported t stats (for the dropped variables) in the original (full) MLR model.

24. To fix expectations: Start with an MLR model with the usual k RHS variables. We want to test the Null Hypothesis that the true parameters for q of the RHS variables are all 0… to see if we might perhaps have justification for dropping those RHS variables from the model.

25. The F statistic for the associated F Test is:

$$F = \frac{(n-k-1)}{q}\frac{\left(R_{UR}^2 - R_R^2\right)}{1-R_{UR}^2} = dofs\frac{\left(\Delta R^2 / q\right)}{1-R^2} = dofs\frac{\left(\Delta SSE / q\right)}{SSR}.$$ [4]

And so the F stat will be driven by $\Delta R^2 / q$ ($\Delta SSE / q$), the average (per dropped variable) change in R^2 (SSE).

26. We know that if we are thinking about dropping just one RHS variable, $F = \left(t\ stat\right)^2$, and so you might wonder:

    Is there a relationship between the F statistic for the new test, and the reported t stats in the full model for the variables under consideration.

27. The answer is, perhaps somewhat surprisingly, yes, maybe, sort of, …it all depends, since in many cases, $F \sim \frac{1}{q}\sum\left(t\ stat_i\right)^2$.

    a. If the dropped RHS variables have zero correlations amongst themselves, then: $F = \frac{1}{q}\sum\left(t\ stat_i\right)^2$ … and so if you want to do the eyeball test in thinking about dropping multiple RHS variables under these conditions, just average the squares of the reported t stats. And so if the individual t stats are all quite healthy, then the test's F stat will be as well. And if not, then definitely not!

    b. Without those zero correlations, the estimate (using $\frac{1}{q}\sum\left(t\ stat_i\right)^2$ from the full model) will be more or less close, and may be smaller or larger than the F stat, depending on circumstances. But it's an easy place to start, especially if you are not in a position to run the actual F test.

28. So as a practical matter, the F test is likely to reject Null Hypotheses when the reported t stats in the full model are large… but may be more accommodating, when the reported t stats are more modest.

---

[4] This corrects a typo in an earlier version of this material.

29. Here's a set of examples using the bodyfat dataset. In these examples, the full model has two RHS variables, and so the F statistic, in testing whether you can drop the two RHS variables, is just the reported F stat for the regression.

```
. corr Brozek age thigh chest wgt
(obs=252)

         |  Brozek      age    thigh    chest      wgt
---------+---------------------------------------------
  Brozek |  1.0000
     age |  0.2892   1.0000
   thigh |  0.5613  -0.2001   1.0000
   chest |  0.7029   0.1764   0.7299   1.0000
     wgt |  0.6132  -0.0127   0.8687   0.8942   1.00
```

	Models		
	(1)	**(2)**	**(3)**
age	0.18	0.10	0.26
	6.40	***3.83***	***8.99***
wgt	0.16		
	13.30		
chest		0.62	
		15.12	
thigh			0.95
			13.87
_cons	(18.37)	(48.13)	(49.17)
	(7.13)	(11.72)	(10.86)
F	**107.85**	**136.07**	**116.29**
avg t^2	***108.92***	***121.61***	***136.59***
diff	***-1.07***	***14.45***	***-20.30***

30. Since *age* and *wgt* are highly uncorrelated, the average of the t stat^2 provides a reasonable estimate of the F stat. In the other two cases, the estimates are more or less close to the true F stat. Importantly, however, note that the estimates may be above or below the true F stat, depending on circumstances.

Babies and Bathwater

31. 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 variables that are jointly insignificant.

    Or put differently: *F tests judge variables by the friends they keep!*

32. **Example**: Here's an example using a sample from a sovereign debt dataset. 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). But *inflation* is so statistically insignificant, it pulls down *deficit_gdp* in the process when the F test (of the Null Hypothesis that $\beta_{inflation} = \beta_{deficit\_gdp} = 0$) is conducted.

```
. 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, RMSE and t Stats

33. We previously considered ***adjusted R²*** ($\bar{R}^2$) as one of several *Goodness-of-Fit* metrics in MLR models. At that time we focused on the impacts of adding and subtracting explanatory variables one-by-one, and the relationship between changes in $\bar{R}^2$ and the magnitudes of the t stats of the added or subtracted RHS variables.

34. Recall that $\bar{R}^2$ was developed in an attempt to adjust the ***coefficient of determination*** for the fact that when you add RHS variables to a model, R^2 cannot decline since SSR cannot increase. Small decreases in SSRs will not generate a higher adj R^2; larger decreases will… and what is *small* or *large* will also depend on how many changes were made.

35. From *MLR Assessment*:

    a. Adjusted R^2 is defined: $\bar{R}^2 = 1 - \dfrac{SSR}{SST}\dfrac{(n-1)}{(n-k-1)} = 1 - \dfrac{MSE}{S_{YY}}$.

    i. Since $\dfrac{(n-1)}{(n-k-1)} > 0$, $\bar{R}^2 \le R^2$, with the difference inversely related to k.

    ii. For given S_{YY}, adjusted R^2 increases if and only if MSE decreases. So if you are adding or subtracting RHS variables from a MLR model, $\bar{R}^2$ and *MSE* will move in exactly opposite directions.

36. We now move from the world of one-by-one inclusions/exclusions to the more general cases of adding or subtracting multiple explanatory variables.

37. It turns out that the change in adj R^2, in say, adding the variables to the model) is closely related to the F-statistic associated with testing whether those new RHS variables should be in the model.

38. The change in adjusted R^2 in going from one model to the other is defined by:

    $\bar{R}^2_{UR} - \bar{R}^2_{R} = \dfrac{1}{S_{YY}}\left(MSE_R - MSE_{UR}\right) = \dfrac{1}{S_{YY}}\left(\dfrac{SSR_R}{(n-k-1)+q} - \dfrac{SSR_{UR}}{n-k-1}\right)$. After much algebra,

    this is: $\bar{R}^2_{UR} - \bar{R}^2_{R} = \dfrac{q\left[1-\bar{R}^2_{UR}\right]}{(n-k-1)+q}\left[F-1\right]$.

    a. The sign of this expression will depend on whether the F statistic is greater or less than 1:

    $\bar{R}^2$ increases if and only if the F statistic associated with the restrictions' F test exceeds 1.

39. You've in fact 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 add a RHS variable if and only if the magnitude of the t stat of the added RHS variable exceeds 1. Now you see that that insight extends to the case of multiple RHS explanatory variables… just use the F stat rather than the t stat!

Adding and Dropping RHS Variables

40. This last fact is useful when considering dropping RHS variables from a model.

    a. **One RHS variable:**

    i. |t stat| > 1: If a RHS variable has a |t stat| > 1 then dropping that variable from the model will cause the adjusted R-squared to decline.

    1. Put differently: Adding RHS variables with |t stat| > 1 will lead to higher $\bar{R}^2$.

    ii. |t stat| < 1: And if |t stat| < 1, then adjusted R-squared will increase when that variable is dropped from the model.[5]

    b. **Multiple RHS variables:**

    i. Fstat > 1: If the Fstat associated with dropping multiple RHS variables is > 1 then dropping those variables from the model will cause adjusted R-squared to decline.

    1. Put differently: Adding RHS variables with Fstat > 1 will lead to higher $\bar{R}^2$.

    ii. Fstat < 1: And if Fstat < 1, then adjusted R-squared will increase when those variables are dropped from the model.

41. **Relation to statistical significance**: It could be that additional RHS variables that lead to a higher $\bar{R}^2$ are not statistically significant. Adding a RHS variable with |t stat| > 1 will be associated with higher $\bar{R}^2$; but statistical significance will generally want to see a |t stat| > 2 or so.

    a. If the added variable is statistically significant (so its t stat is above 1), then $\bar{R}^2$ increased when that variable was added to the model.

    b. But it could be that the new variable is not statistically significant, even though adjusted R-squared increased when the variable was added to the model.

    c. **An Example:** Spoze that you have a RHS variable with a t stat of 1.5. We would normally say that that variable was not statistically significant at standard significance levels. And so you might be tempted to toss that variable from your MLR model. But if you do that, $\bar{R}^2$ will decline, since the t stat > 1.

    d. This highlights a tension between two approaches to MLR model building. You could be focused on maximizing $\bar{R}^2$ … or perhaps you are more focused on statistical significance. These two objectives can conflict at times. *Deal with it!*

[5] And yes, if t stat = 0, then there will be no change to adjusted R-squared when the variable is dropped from the model.

	prob (F(q, n-k-1) > 1)						prob (F(q, n-k-1) > 2)						prob (F(q, n-k-1) > 4)						
	q: #restrictions						q: #restrictions						q: #restrictions						
dofs	1	2	3	4	5	10	1	2	3	4	5	10	1	2	3	4	5	10	***dofs***
10	34%	40%	43%	45%	47%		19%	19%	18%	17%	16%		7%	5%	4%	3%	3%		10
20	33%	39%	41%	43%	44%	48%	17%	16%	15%	13%	12%	9%	6%	3%	2%	2%	1%	0%	20
50	32%	38%	40%	42%	43%	46%	16%	15%	13%	11%	9%	5%	5%	2%	1%	1%	0%	0%	50
100	32%	37%	40%	41%	42%	45%	16%	14%	12%	10%	9%	4%	5%	2%	1%	0%	0%	0%	100
250	32%	37%	39%	41%	42%	44%	16%	14%	11%	10%	8%	3%	5%	2%	1%	0%	0%	0%	250
500	32%	37%	39%	41%	42%	44%	16%	14%	11%	9%	8%	3%	5%	2%	1%	0%	0%	0%	500

Example II: More Bodyfat

	(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:

(1) to (2) (add *hgt*): hgt |tstat| > 1, R^2 and adj R^2 increase, RMSE decreases

(1) to (3) (add *ankle*): *ankle* |tstat| < 1, R^2 increases, adj R^2 decreases, and RMSE increases

F stats and F tests:
(4) to (1) (drop *hgt* and *ankle*): Since adj R^2 **increases**,[6] 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
```

[6] In a previous version of this handout, the incorrectly said *decreases*.