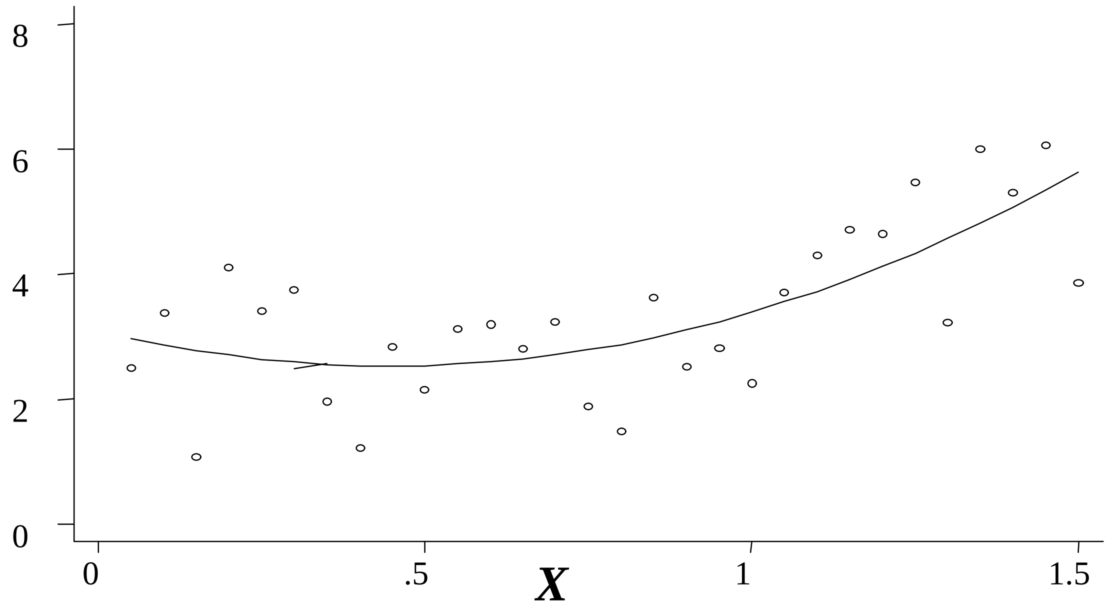
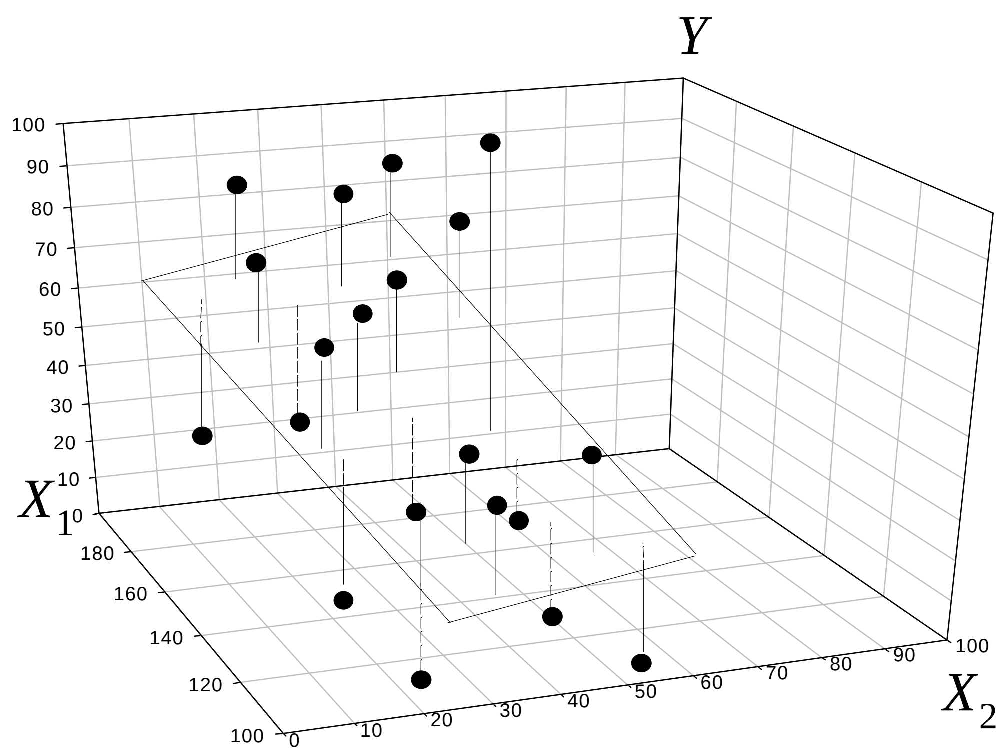
Multiple regression

Multiple regression is an extension of the simple regression situation. We are still trying to describe Y as (now) a *linear* combination of several predictors (X's). The predictors can be powers of one another $Y = \beta_o + \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon$ or $Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ (where $X_2 = X_1^2$), or they can be distinct such as $Y=\beta_o+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k+\varepsilon$. In the first case, the graphical representation of the problem is as follows:



In the second case, the model is harder to visualize, and impossible to do so beyond the two-predictor situation (when the dimension of the problem rises above three).

In all cases, the regression *surface* (notice we have departed from the simple line) is going to be a *hyperplane* (a plane in three dimensions). The figure below shows the two-predictor situation.



The least-squares regression surface

The idea for finding the "best" regression *surface* is identical as the simple linear case. That is, the best surface is the one that minimizes the squared deviations of the estimated values from the observations. That is, the least-squares surface is the one that minimizes

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_o - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \cdots \hat{\beta}_k X_{ki}\right)^2$$

As with simple linear regression, $\hat{Y}_i = \hat{\beta}_o + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$

Assumptions of multiple regression

1. **Independence**: The Y observations are statistically independent of each other. Usually this is not the case when multiple measurements are taken on the same subject. Other techniques must then be used that account for this dependency.

2. **Linearity**: The mean value of Y for each combination of $X_1, X_2, \ldots, X_k$ is a linear combination of them. That is, $E(Y_i) = \mu_{Y \mid X_1, X_2, \cdots X_k} = \beta_o + \beta_1 X_{1i} + \cdots \beta_k X_{ki}$.

3. **Homoskedacity**: The variance of Y is the same for any fixed combination of $X_1, X_2, \ldots, X_k$. That is $\sigma^2_{Y|X_1,X_2,\cdots X_k} = V(Y|X_1,X_2,\cdots,X_k) \equiv \sigma^2$ or alternatively, that $\sigma^2_{\varepsilon|X_1,X_2,\cdots X_k} \equiv \sigma^2$.

4. **Normality**: For any fixed combination of $X_1, X_2, \ldots, X_k$ the variable Y is normally distributed. That is, $\varepsilon \sim N(0, \sigma^2)$.

Explaining variability

Our task is to explain the variability in the data. Using similar methods as before, we have

$$\underbrace{\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2}_{\text{Total sum of squares}} = \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i-\overline{Y}\right)^2}_{\text{Regression sum of squares}} + \underbrace{\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2}_{\text{Residual sum of squares}}$$

The multiple regression ANOVA table

Source of variability	Sums of squares (SS)	df	Mean squares (MS)	F	Prob > F
Model	$SSR\begin{cases} SS(\beta_1) \\ SS(\beta_2 \mid \beta_1) \\ \vdots \\ SS(\beta_k \mid \beta_1,\beta_2,\cdots,\beta_{k-1}) \end{cases}$	k	$MSR=SSR/k$	$F=\dfrac{MSR}{MSE}$	$P=\mathrm{P}(F>F_{k,\,n-k-1;\alpha})$
Residual (error)	SSE	n-k-1	$MSE=SSE/(n-k-1)$		
Corrected Total	$SST=\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2$	n-1			

F tests in multiple regression

Test of significance of overall regression. With similar methods as in the simple linear regression case, we can carry out an overall (omnibus) F test. This is based on the statistic

$$F = \frac{MSR}{MSE} = \frac{\sum \left(\hat{Y}_i - \overline{Y}\right)^2 / k}{\sum \left(Y_i - \hat{Y}_i\right)^2 / (n-k-1)} = \frac{R^2/k}{\left(1-R^2\right)/(n-k-1)}$$

This statistic is compared against the tail of the F distribution with k and n-k-1 degrees of freedom. The regression sum of squares (SSR) receives contributions from all the predictors. However, not all contributions are equally important. Another problem involves the fact that the predictors themselves may be correlated to one another. Thus, including one predictor in the model provides some information about the other predictor as well. Then, when the second predictor is included, its individual contribution (in the presence of the first predictor) may not be as significant as it would have been if the second were the only predictor in the model. We formalize these ideas below.

<u>*Partial F* tests.</u> The partial contributions by each individual predictor to the regression (model) sum of squares can be explored by partial *F* tests. As we see in the table above, the predictors can be included in the model sequentially. Thus, X_1 is entered first, then X_2, and so on up to X_k. These partial *F* tests are called *variables-added-in-order* or *Type I F* tests. Note that the order of addition of variable in the model is critically important when computing these partial *F* tests. The model sum of squares can be broken up into the following parts:

1. $SS(\beta_1)$ is the sum of squares (variability in Y) explained by only using X_1 to predict Y.
2. $SS(\beta_2|\beta_1)$ is the *additional* variability in Y explained by adding X_2 into the model *after* X_1.
3. $SS(\beta_k|\beta_1,\beta_2,\ldots,\beta_{k-1})$ is the additional variability explained by X_k after X_1, $X_2,\ldots,X_{k-1}$ are already in the model.

We cannot decompose the model sum of squares into k separate sums of squares (i.e., *unconditional* sums of squares) because the predictors are not independent from one another (we can redefine the predictors and obtain an “orthogonal” decomposition but this is beyond the scope of this lecture).

<u>Type I *F* tests (continued):</u>

1. This test addresses the question of whether X_1 alone can significantly predict Y. It can also be obtained by a simple regression with X_1 as the only predictor.
2. The sum of squares addresses the question of whether adding X_2 significantly contributes to the prediction of Y after accounting for the contribution of X_1. To test we use a *partial F* test:

$$F = \frac{\text{Regression } SS(\beta_1, \beta_2) - \text{Regression } SS(\beta_1)}{\text{Residual } SS(\beta_1, \beta_2)/(n-k-1)} = \frac{\text{Residual } SS(\beta_1) - \text{Residual } SS(\beta_1, \beta_2)}{\text{Residual } SS(\beta_1, \beta_2)/(n-k-1)}$$

The Regression $SS(\beta_1,\beta_2)$ and Residual $SS(\beta_1,\beta_2)$ are derived from a model with both X_1 and X_2, while the Regression $SS(\beta_1)$ and Residual $SS(\beta_1)$ come from the simple linear regression model.

3. In general, to answer whether a contribution of a single variable or a number of variables contributes significantly in the prediction of Y after controlling for a number of other predictors is given by the (multiple) partial F test,

$$F\left(X_1^*, X_2^*, \ldots, X_k^* \mid X_1, X_2, \ldots, X_p\right) = \frac{\left[\text{Regression } SS\left(\beta_1^*, \beta_2^*, \ldots, \beta_k^*, \beta_1, \beta_2, \ldots, \beta_p\right) - \text{Regression } SS\left(\beta_1, \beta_2, \ldots, \beta_p\right)\right]/k}{\text{Regression } SS\left(\beta_1^*, \beta_2^*, \ldots, \beta_k^*, \beta_1, \beta_2, \ldots, \beta_p\right) \big/ (n-p-k-1)}$$

The *t* test as an alternative to a partial *F* test

Another way to test whether the addition of a new variable X^*, after p variables X_1, X_2, … , X_p already in the model, significantly predicts Y, is to use a t test (recall that a t test is equivalent to an F test with 1 degree of freedom in the numerator). This test is defined as follows:

1. H_o: $\beta^*=0$ (i.e., addition of X^* to the model does not add significantly to the prediction of Y)

2. $$H_a: \begin{cases} \beta^* \neq 0 & \text{Two-sided test} \\ \beta^* > 0 & \multirow{2}{*}{One-sided tests} \\ \beta^* < 0 & \end{cases}$$

3. Specify the significance level (1-α)%

4. The test statistic is $T=\dfrac{\hat{\beta}^*}{S_{\beta^*}} \sim t_{n-p-2}$

5. Decision rule: Reject H_o: $\beta^*=0$ if $$\begin{cases} T>t_{n-p-2,1-\alpha/2} \text{ or if } T<-t_{n-p-2,1-\alpha/2} & \text{(two-sided test: } H_a\text{:}\beta^*\neq 0) \\ T>t_{n-p-2,1-\alpha} & \text{(upper one-sided test: } H_a\text{:}\beta^*>0) \\ T<-t_{n-p-2,1-\alpha} & \text{(lower one-sided test: } H_a\text{:}\beta^*<0) \end{cases}$$

Notice that T^2=partial F($X^*|X_1$, X_2, … , X_p).

Variables-added-last or Type III *F* tests

A final type of partial *F* tests that we will review is the "variables-added-last" or "Type III" *F* tests. These are tests based on the sums of squares of each variable *conditional* (or accounting for) *all other variables in the model*. In other words, if we have *k* variables in the model, the Type III *F* tests are given as follows:

$$\begin{aligned} X_1 &: SS(X_1 | X_2 X_3, \cdots, X_k) \\ X_2 &: SS(X_2 | X_1 X_3, \cdots, X_k) \\ &\vdots \\ X_k &: SS(X_k | X_1 X_2, \cdots, X_{k-1}) \end{aligned}$$

These sums of squares can be computed in models where the variable in question is added *last*, that is, after all the others are already present in the model. The primary advantage of these sums of squares is that order of entry into the model is no longer important.

Criteria of inclusion of additional variables in the model

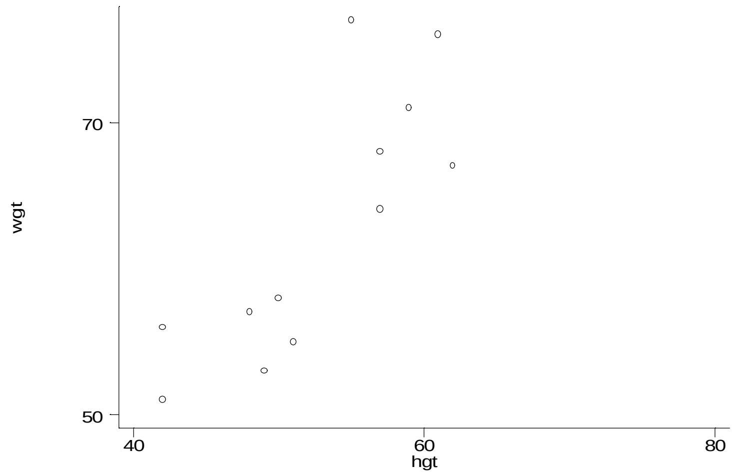
1. *Variables added in order*:
   i. The order of addition is specified
   ii. The significance of the (straight-line) model involving only the first variable is assessed
   iii. The significance of adding the second variable to the model involving only the first variable is assessed
   iv. The significance of adding the third variable to the model containing the first and second variables is assessed; and so on.

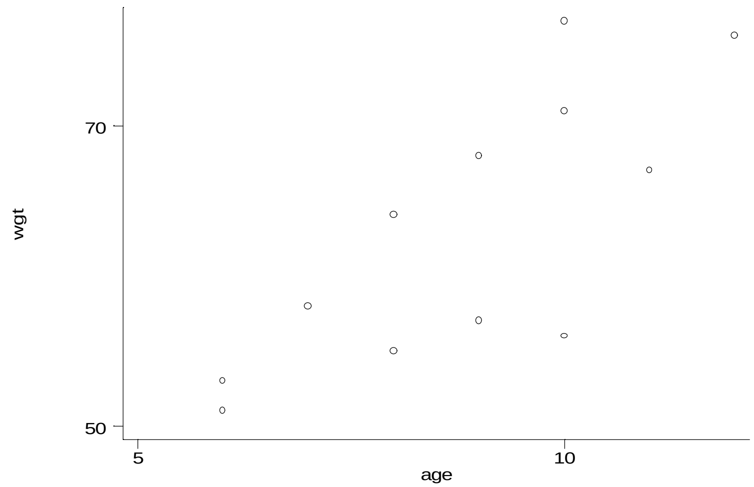
2. *Variables added last*:
   i. An initial model containing several (more than one) variables is specified.
   ii. The significance of each variable in the model is assessed separately, as if it were the last variable added to the model (thus, k variables-added-last tests are carried out, as many as the variables under review).

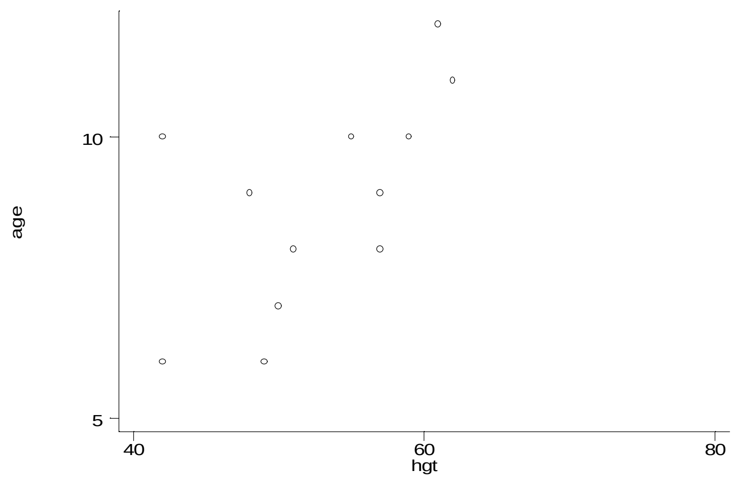
Example: The weight (`wgt`), height (`hgt`) and age (`age`) data (Table 8-1, page 112).

```
. list

            wgt         hgt         age        age2
  1.         64          57           8          64
  2.         71          59          10         100
  3.         53          49           6          36
  4.         67          62          11         121
  5.         55          51           8          64
  6.         58          50           7          49
  7.         77          55          10         100
  8.         57          48           9          81
  9.         56          42          10         100
 10.         51          42           6          36
 11.         76          61          12         144
 12.         68          57           9          81
```








```
. pwcorr hgt age, sig

             |      hgt      age
-------------+------------------
         hgt |   1.0000
             |
             |
         age |   0.6138   1.0000
             |   0.0337
             |
```

Model 1: WGT= $\beta_o+\beta_1$HGT+ε

```
. anova wgt hgt, continuous(hgt) regress

  Source |       SS       df       MS                  Number of obs =      12
---------+------------------------------               F(  1,    10) =   19.67
   Model |  588.922523     1  588.922523               Prob > F      =  0.0013
Residual |  299.327477    10  29.9327477               R-squared     =  0.6630
---------+------------------------------               Adj R-squared =  0.6293
   Total |      888.25    11       80.75               Root MSE      =  5.4711

------------------------------------------------------------------------------
     wgt        Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
------------------------------------------------------------------------------
_cons        6.189849   12.84875      0.482   0.640      -22.43894    34.81864
hgt           1.07223    .241731      4.436   0.001       .5336202    1.610841
------------------------------------------------------------------------------
. anova, sequential

                           Number of obs =      12     R-squared     =  0.6630
                           Root MSE      = 5.47108     Adj R-squared =  0.6293

                  Source |    Seq. SS     df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  588.922523     1  588.922523      19.67     0.0013
                         |
                     hgt |  588.922523     1  588.922523      19.67     0.0013
                         |
                Residual |  299.327477    10  29.9327477
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

Model 2: WGT= $\beta_o+\beta_2$AGE+ε

```
. anova wgt age, continuous(age) regress

  Source |       SS       df       MS                  Number of obs =      12
---------+-----------------------------                F(  1,    10) =   14.55
   Model |  526.392857     1  526.392857               Prob > F      =  0.0034
Residual |  361.857143    10  36.1857143               R-squared     =  0.5926
---------+-----------------------------                Adj R-squared =  0.5519
   Total |      888.25    11       80.75               Root MSE      =  6.0155

------------------------------------------------------------------------------
     wgt        Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
------------------------------------------------------------------------------
_cons        30.57143   8.613705      3.549   0.005        11.3789    49.76396
age          3.642857   .9551151      3.814   0.003       1.514728    5.770986
------------------------------------------------------------------------------
. anova, sequential

                           Number of obs =      12     R-squared     =  0.5926
                           Root MSE      = 6.01546     Adj R-squared =  0.5519

                  Source |    Seq. SS     df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  526.392857     1  526.392857      14.55     0.0034
                         |
                     age |  526.392857     1  526.392857      14.55     0.0034
                         |
                Residual |  361.857143    10  36.1857143
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

Model 3: WGT= $\beta_o+\beta_3(AGE)^2+\varepsilon$

```
. anova wgt age2, continuous(age2) regress

  Source |       SS       df       MS                  Number of obs =      12
---------+------------------------------               F(  1,    10) =   14.25
   Model |  521.932047     1  521.932047               Prob > F      =  0.0036
Residual |  366.317953    10  36.6317953               R-squared     =  0.5876
---------+------------------------------               Adj R-squared =  0.5464
   Total |      888.25    11       80.75               Root MSE      =  6.0524

------------------------------------------------------------------------------
     wgt        Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
------------------------------------------------------------------------------
_cons        45.99764    4.76964      9.644   0.000       35.37022    56.62506
age2         .2059716   .0545669      3.775   0.004       .0843889    .3275543
------------------------------------------------------------------------------
. anova, sequential

                           Number of obs =      12     R-squared     =  0.5876
                           Root MSE      = 6.05242     Adj R-squared =  0.5464

                  Source |    Seq. SS     df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  521.932047     1  521.932047      14.25     0.0036
                         |
                    age2 |  521.932047     1  521.932047      14.25     0.0036
                         |
                Residual |  366.317953    10  36.6317953
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

Model 4:WGT= $\beta_o + \beta_1 HGT + \beta_2 AGE + \varepsilon$

```
. anova wgt hgt age, continuous(hgt age) regress

  Source |       SS       df       MS                  Number of obs =      12
---------+------------------------------               F(  2,     9) =   15.95
   Model |  692.822607     2  346.411303               Prob > F      =  0.0011
Residual |  195.427393     9  21.7141548               R-squared     =  0.7800
---------+------------------------------               Adj R-squared =  0.7311
   Total |      888.25    11       80.75               Root MSE      =  4.6598
------------------------------------------------------------------------------
     wgt        Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
------------------------------------------------------------------------------
_cons        6.553048   10.94483      0.599   0.564      -18.20587    31.31197
hgt           .722038   .2608051      2.768   0.022       .1320559     1.31202
age          2.050126   .9372256      2.187   0.056      -.0700253    4.170278
------------------------------------------------------------------------------
. anova, sequential
                           Number of obs =      12     R-squared     =  0.7800
                           Root MSE      = 4.65984     Adj R-squared =  0.7311

                  Source |    Seq. SS     df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  692.822607     2  346.411303      15.95     0.0011
                         |
                     hgt |  588.922523     1  588.922523      27.12     0.0006
                     age |  103.900083     1  103.900083       4.78     0.0565
                Residual |  195.427393     9  21.7141548
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

Model 5: WGT= $\beta_o + \beta_1 HGT + \beta_3 (AGE)^2 + \varepsilon$

```
. anova wgt hgt age2, continuous(hgt age2) regress

  Source |       SS       df       MS              Number of obs =      12
---------+------------------------------           F(  2,     9) =   15.63
   Model |  689.649951     2  344.824976           Prob > F      =  0.0012
Residual |  198.600049     9  22.0666721           R-squared     =  0.7764
---------+------------------------------           Adj R-squared =  0.7267
   Total |      888.25    11       80.75           Root MSE      =  4.6975

------------------------------------------------------------------------------
     wgt        Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
------------------------------------------------------------------------------
_cons        15.11754    11.7969      1.281   0.232       -11.5689    41.80398
hgt          .7259765   .2633306      2.757   0.022       .1302814    1.321672
age2         .1148016   .0537332      2.137   0.061      -.0067513    .2363546
------------------------------------------------------------------------------
. anova, sequential
                           Number of obs =      12     R-squared     =  0.7764
                           Root MSE      = 4.69752     Adj R-squared =  0.7267

                  Source |    Seq. SS     df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  689.649951     2  344.824976      15.63     0.0012
                     hgt |  588.922523     1  588.922523      26.69     0.0006
                    age2 |  100.727428     1  100.727428       4.56     0.0614
                Residual |  198.600049     9  22.0666721
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

Model 6: $WGT=\beta_o+\beta_1 HGT+\beta_2 AGE+\beta_3(AGE)^2+\varepsilon$

```
. anova wgt hgt age age2, continuous(hgt age age2) regress
     Source |       SS       df       MS                  Number of obs =      12
  ---------+------------------------------               F(  3,     8) =    9.47
     Model |  693.060463     3  231.020154               Prob > F      =  0.0052
  Residual |  195.189537     8  24.3986921               R-squared     =  0.7803
  ---------+------------------------------               Adj R-squared =  0.6978
     Total |      888.25    11       80.75               Root MSE      =  4.9395

  ------------------------------------------------------------------------------
       wgt        Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
  ------------------------------------------------------------------------------
  _cons        3.438426   33.61082      0.102   0.921      -74.06826    80.94512
  hgt          .7236902   .2769632      2.613   0.031        .085012    1.362368
  age          2.776875   7.427279      0.374   0.718      -14.35046    19.90421
  age2        -.0417067   .4224071     -0.099   0.924      -1.015779    .9323659
  ------------------------------------------------------------------------------
. anova, sequential
                      Number of obs =      12     R-squared     =  0.7803
                     Root MSE      =  4.9395     Adj R-squared =  0.6978

            Source |    Seq. SS     df       MS            F     Prob > F
       -----------+---------------------------------------------------
             Model |  693.060463     3  231.020154       9.47     0.0052
               hgt |  588.922523     1  588.922523      24.14     0.0012
               age |  103.900083     1  103.900083       4.26     0.0730
              age2 |  .237856856     1  .237856856       0.01     0.9238
          Residual |  195.189537     8  24.3986921
       -----------+---------------------------------------------------
             Total |      888.25    11       80.75
```

Analysis results

1. Models 1 and 2 show a significant association between weight and height (overall *F* p-value 0.0013) and between weight and age (overall *F* p-value 0.0034) respectively.

2. Model 3 shows a significant association between weight and $(AGE)^2$ (overall *F* p-value 0.0036) implying a possible curvilinear (quadratic) relationship.

3. Models 4 and 5 investigate the two-predictor cases, with height as the first predictor entered, and AGE and $(AGE)^2$ the second predictors respectively. In both cases the overall *F* test is highly significant implying that the two variables are significant predictors of weight (p-values are 0.0011 and 0.0012 respectively). Note however, that we have not answered whether addition of the second variable contributes substantially to the prediction of weight beyond the first variable.

4. Model 6 shows the result of adding all three predictors. The overall *F* test p-value is 0.0052 indicating that a significant part of the variability in the data is explained by the regression model.

***Type I F* tests**

1. To decide whether adding age to the model after controlling for height (age and height should be correlated), we can use a *Type I* test. The test is computed from models 1 and 4 as follows:

$$F(\text{AGE}|\text{HGT}) = \frac{\text{Regression } SS(\text{AGE,HGT}) - \text{Regression } SS(\text{HGT})}{\text{Residual } SS(\text{AGE,HGT})/n-k-1} = \frac{692.8226 - 588.9225}{195.4274/9} = 4.78 .$$

Since 3.36=$F_{1,9;0.10}$ <4.78< $F_{1,9;0.05}$ =5.12, adding age to the model significantly improves prediction of Y at the 10% α level, but not at the 5% α level. Notice that the t test p value for β_2 (the regression coefficient associated with age, is 0.056, and T^2=(2.187)^2=4.78=F.

2. To answer the same question about (AGE)^2 after controlling both for height and age, we consider models 4 and 6. The partial (Type I) F test is computed as above. $F(\text{AGE}^2|\text{HGT, AGE}) = 0.01$, which is not significant. Thus, even though AGE^2 was significant as a single predictor of weight, it is not significant after controlling for height and age. Thus, a quadratic relationship between weight and age is probably not born out by the data.

Model 7: $WGT=\beta_o + \beta_1 HGT + \beta_3 (AGE)^2 + \beta_2 AGE + \varepsilon$ (AGE is entered last)

```
. anova wgt  hgt age2 age, continuous(hgt age age2) sequential

                           Number of obs =      12     R-squared     =  0.7803
                           Root MSE      =  4.9395     Adj R-squared =  0.6978

                  Source |    Seq. SS     df       MS            F     Prob > F
              -----------+----------------------------------------------------
                   Model |  693.060463     3  231.020154       9.47     0.0052
                         |
                     hgt |  588.922523     1  588.922523      24.14     0.0012
                    age2 |  100.727428     1  100.727428       4.13     0.0766
                     age |  3.41051231     1  3.41051231       0.14     0.7182
                         |
                Residual |  195.189537     8  24.3986921
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

$SS(age \mid hgt, age2)$ → 3.41051231

Model 8: WGT=β_o+ β_2AGE+β_3(AGE)^2+ β_1HGT+ε (HGT is entered last)

```
. anova wgt  age age2 hgt, continuous(hgt age age2) sequential

                           Number of obs =      12     R-squared     =  0.7803
                           Root MSE      =  4.9395     Adj R-squared =  0.6978

                  Source |    Seq. SS     df       MS            F     Prob > F
              -----------+----------------------------------------------------
                   Model |  693.060463     3  231.020154       9.47     0.0052
                         |
                     age |  526.392857     1  526.392857      21.57     0.0017
                    age2 |  .085651307     1  .085651307       0.00     0.9542
                     hgt |  166.581955     1  166.581955       6.83     0.0310
                         |
                Residual |  195.189537     8  24.3986921
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

$SS(hgt|age, age2)$

Model 9: $WGT=\beta_o+\beta_1 HGT+\beta_2 AGE+\beta_3 (AGE)^2+\varepsilon$

```
. anova wgt hgt age age2, continuous(hgt age age2) partial

                           Number of obs =      12     R-squared     =  0.7803
                           Root MSE      =  4.9395     Adj R-squared =  0.6978

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  693.060463     3  231.020154       9.47     0.0052
                         |
                     hgt |  166.581955     1  166.581955       6.83     0.0310
                     age |  3.41051231     1  3.41051231       0.14     0.7182
                    age2 |  .237856856     1  .237856856       0.01     0.9238
                         |
                Residual |  195.189537     8  24.3986921
              -----------+----------------------------------------------------
                   Total |      888.25    11       80.75
```

Type III F tests

We can address the same question as 1 and 2 above with Type III partial *F* tests. These can be derived by running several regressions each time entering the variable in question last. For our example consider models 6, 7 and 8. $(\text{AGE})^2$, age and height were entered last in each model respectively. We did not print the regression ANOVA table for models 7, 8 and 9 since they are identical to that in model 6. The type sums of squares are derived in each case as follows:

$SS\left((\text{AGE})^2|\text{HGT},\text{AGE}\right) = 0.24$. HGT is entered first, then AGE and finally $(\text{AGE})^2$ (model 6).

$SS\left(\text{AGE}|\text{HGT},(\text{AGE})^2\right) = 3.41$. HGT is entered first, then $(\text{AGE})^2$ and finally AGE (model 7).

$SS\left(\text{HGT}|\text{AGE},(\text{AGE})^2\right) = 166.58$. AGE is entered first, then $(\text{AGE})^2$ and finally HGT (model 8)

The Type III *F* tests are derived by dividing the above sums of squares by the full model mean square error: $F\left(\text{HGT}|\text{AGE},(\text{AGE})^2\right) = \frac{166.58}{195.19/8} = 6.83$, $F\left(\text{AGE}|\text{HGT},(\text{AGE})^2\right) = \frac{3.41}{195.19/8} 0.14$ and $F\left((\text{AGE})^2|\text{HGT},\text{AGE}\right) = \left(\frac{0.24}{195.19/8}\right) = 0.01$ (as before). Notice that we can derive the Type III *F* tests immediately by specifying the `partial` option or by not specifying an option at all as `partial` is the default (model 9).