Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ where $\boldsymbol{X}_1$ is $n \times (k+1-r)$, $\boldsymbol{X}_2$ is $n \times r$ and $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ to conform, so $\boldsymbol{\beta}_1$ is $(k+1-r) \times 1$ and $\boldsymbol{\beta}_2$ is $r \times 1$.

The following notes:

- justify the partial F test for testing $H_0 : \boldsymbol{\beta}_2 = 0$ when the regression model already includes X_1.
- prove the equivalence of the t test for a single variable, and the associated F test
- motivate the following sequential 3 step testing procedure:
  1. Regress $\boldsymbol{y}$ on $\boldsymbol{X}_1$ to get residuals $\boldsymbol{e}_1$
  2. Regress $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$ (each column) to get residuals $\boldsymbol{e}_2$
  3. Regress $\boldsymbol{e}_1$ on $\boldsymbol{e}_2$ to get $\hat{\boldsymbol{\beta}}_2$.

- motivate the partitioning of the regression sum of squares, and construction of the following ANOVA table

Source	SS	D.F.	MS
$\boldsymbol{X}_1$	$SSR(\beta_1)$	$k-r$	$SSR(\beta_1)/(k-r)$
$\boldsymbol{X}_2 \vert \boldsymbol{X}_1$	$SSR(\beta_2 \vert \beta_1)$	r	$SSR(\beta_2 \vert \beta_1)/r$
Error	SSE	$n-1-k$	MSE
Total	SST	$n-1$	

The F statistic for the test that $H_0 : \beta_2 = 0$, when the variables in $\boldsymbol{X}_2$ are entered into the regression after the variables in $\boldsymbol{X}_1$ is given by

$$F = \frac{MSR(\beta_2|\beta_1)}{MSE} = \frac{(SSE(\beta_1) - SSE(\beta_1, \beta_2))/r}{MSE} \sim F_{r,n-1-k}$$

- In the case that X_2 consists of a single regressor, the plot of e_1 vs e_2 is called an **added variable plot**. It is useful to diagnose the functional form of the relationship between X_2 and y given that the variable in X_1 are already included in the regression.
- If a linear regression line is fit to the added variable plot, it can be shown that:the slope of the line is the coefficient of X_2 in the multivariate regresson containing both X_1 and X_2.

Partial F test - Review - you are not responsible for remembering the algebraic derivations

- Suppose we have the model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$$

  and want to add the r predictors $\boldsymbol{X}_2$.
- For example, we may wish to test the hypotheses
  $H_0 : \boldsymbol{\beta}_2 = 0$
  $H_A : \boldsymbol{\beta}_{2j} \neq 0$ for some j
- Then we want to compare the fit of the reduced model under H_0 to that of the full model under H_1.
- In total there are k predictors, so $\boldsymbol{X}_1$ consists of the column of 1's and $k - r$ columns of predictors.
- Write $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ where $\boldsymbol{X}_1$ is $n \times (k + 1 - r)$, $\boldsymbol{X}_2$ is $n \times r$ and $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ to conform, so $\boldsymbol{\beta}_1$ is $(k + 1 - r) \times 1$ and $\boldsymbol{\beta}_2$ is $r \times 1$.

- Then the model containing $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ can be written

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}.$$

Case 1: Predictors orthogonal

- If the new predictors $\boldsymbol{X}_2$ are orthogonal to the old ones $\boldsymbol{X}_1^T \boldsymbol{X}_2 = 0$ and

$$\boldsymbol{X}^T \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1^T \boldsymbol{X}_1 & 0 \\ 0 & \boldsymbol{X}_2^T \boldsymbol{X}_2 \end{pmatrix}$$

  which has inverse

$$(\boldsymbol{X}^T \boldsymbol{X})^{-1} = \begin{pmatrix} (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} & 0 \\ 0 & (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \end{pmatrix}.$$

- The least squares estimates are

$$\begin{aligned} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} &= \begin{pmatrix} (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} & 0 \\ 0 & (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_1^T \boldsymbol{y} \\ \boldsymbol{X}_2^T \boldsymbol{y} \end{pmatrix} \\ &= \begin{pmatrix} (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{y} \\ (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^T \boldsymbol{y} \end{pmatrix}. \end{aligned}$$

- The estimates of $\boldsymbol{\beta}_1$ are unchanged and $\boldsymbol{\beta}_2$ is estimated separately from the new columns.
- The regression sum of squares is

$$\begin{aligned} SSR(\boldsymbol{\beta}) &= \hat{\boldsymbol{\beta}}^T \boldsymbol{X}^T \boldsymbol{y} - n\bar{y}^2 \\ &= \hat{\boldsymbol{\beta}}_1^T \boldsymbol{X}_1^T \boldsymbol{y} - n\bar{y}^2 + \hat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_2^T \boldsymbol{y} \\ &= SSR(\boldsymbol{\beta}_1) + SSR(\boldsymbol{\beta}_2) \end{aligned}$$

  and factors into two parts depending on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ separately.
- The extra regression sum of squares for $\boldsymbol{X}_2$ given that $\boldsymbol{X}_1$ is already in the model can be written

$$\begin{aligned} SSR(\boldsymbol{\beta}_2) &= \hat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_2^T \boldsymbol{y} \\ &= \boldsymbol{y}^T \boldsymbol{X}_2 (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^T \boldsymbol{y} \\ &= \boldsymbol{y}^T \boldsymbol{H}_2 \boldsymbol{y} \end{aligned}$$

  where $\boldsymbol{H}_2 = \boldsymbol{X}_2 (\boldsymbol{X}_2^T \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^T$ is the projection onto the subspace spanned by the columns of $\boldsymbol{X}_2$ (which is orthogonal to $\boldsymbol{X}_1$).

- Under the null hypothesis that $\beta_2 = 0$

$$\frac{SSR(\beta_2)}{\sigma^2} \sim \chi^2_r$$

and

$$F = \frac{MSR(\beta_2)}{MS_{Res}} \sim F_{r,n-1-k}$$

and large F gives evidence against H_0.

Case 2: Predictors not orthogonal

- When the new predictors are not orthogonal to the old ones, $\boldsymbol{X}_1^T \boldsymbol{X}_2 \neq 0$, the situation is more complicated.
- The model can be written as before, and then manipulated to create new predictors which are orthogonal

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{\epsilon} \\ &= \boldsymbol{X}_1\beta_1 + (\boldsymbol{H}_1 + \boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2\beta_2 + \boldsymbol{\epsilon} \\ &= \boldsymbol{X}_1\boldsymbol{\theta} + (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2\beta_2 + \boldsymbol{\epsilon}, \end{aligned}$$

where

$$\boldsymbol{H}_1 = \boldsymbol{X}_1(\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T$$

is the projection on the subspace spanned by the predictors $\boldsymbol{X}_1$, and

$$\boldsymbol{\theta} = \beta_1 + (\boldsymbol{X}_1^T\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2\beta_2 \qquad (1)$$

is a new parameter created from β_1 and β_2.

- The matrices $\boldsymbol{X}_1$ and $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2$ are orthogonal, so estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}_2$ can be obtained separately, as above:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{y} \tag{2}$$

and

$$\hat{\boldsymbol{\beta}}_2 = [\boldsymbol{X}_2^T (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2]^{-1} \boldsymbol{X}_2^T (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y}. \tag{3}$$

- Rearranging (1) gives

$$\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\theta}} - (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2$$

or

$$\hat{\boldsymbol{\beta}}_1 = [\boldsymbol{X}_1^T \boldsymbol{X}_1]^{-1} \boldsymbol{X}_1^T (\boldsymbol{y} - \boldsymbol{X}_2 \hat{\boldsymbol{\beta}}_2). \tag{4}$$

- From (3) we see that $\hat{\boldsymbol{\beta}}_2$ is the result of regressing one set of residuals, $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y}$ on another $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2$.

- The latter is a matrix of residuals obtained by regressing each column of $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$.
- It contains the information from $\boldsymbol{X}_2$ not already explained by $\boldsymbol{X}_1$.

Example:

Suppose that $\boldsymbol{\beta}_1 = (\beta_0, \beta_1)^T$ and $\boldsymbol{\beta}_2$ has just one element β_2. That is, the first regression is $y = \beta_0 + \beta_1 x_1 + e$, and we are looking at the effect of adding a second variable x_2 to the model. The matrix $\boldsymbol{X}_1$ consists of a column of 1's, and a column containing data on x_1. $\boldsymbol{X}_2$ has just one column, the data on x_2. In this case:

1. $\hat{\boldsymbol{\theta}}$ are the least squares estimates for the model $y = \theta_0 + \theta_1 x_1 + e$. Call them $\hat{\theta}_0$ and $\hat{\theta}_1$.
1. $(\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{X}_2$ are the regression coefficients from the model $x_2 = \gamma_0 + \gamma_1 x_1 + e$. Call them $\hat{\gamma}_0$ and $\hat{\gamma}_1$.
1. $\hat{\beta}_2$ is the least squares estimator from the regression of $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y}$ on $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2$, which does NOT include an intercept. Call the estimate $\hat{\alpha}$.
   - Then the line before (4) says that $\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_0, \hat{\beta}_1)^T = \hat{\theta} - \hat{\alpha}\hat{\boldsymbol{\gamma}} = (\hat{\theta}_0 - \hat{\alpha}\hat{\gamma}_0, \hat{\theta}_1 - \hat{\alpha}\hat{\gamma}_1)$
   - $\hat{\beta}_2$ was given by (3).
   - Together, these are $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T$.

example, continued

```
data=read.csv(
"http://chase.mathstat.dal.ca/~bsmith/stat3340/Data/NFLdata.
attach(data)
lm1=lm(y~x1)
e1=resid(lm1)
coef(lm1)  #theta hat in notes
>  (Intercept)            x1
-4.330015011  0.005352206

lm2=lm(x2~x1)
e2=resid(lm2)
coef(lm2)
> > coef(lm2)
  (Intercept)            x1
2227.55824645   -0.04755155
```

example, continued

```
lm3=lm(e1~e2-1)
b2=coef(lm3) #alpha hat in notes, which is betahat_2 in full
b0=coef(lm1)[1]-coef(lm3)*coef(lm2)[1] #gives bethat_0
b1=coef(lm1)[2]-coef(lm3)*coef(lm2)[2] #gives bethat_1
c(b0,b1,b2)
  (Intercept)             x1             e2
-12.176470327   0.005519703   0.003522447
>

> > #check by fitting full model
lm(y~x1+x2,data=data)

> Call: lm(formula = y ~ x1 + x2, data = data)

Coefficients:
(Intercept)             x1             x2
 -12.176470      0.005520      0.003522
```

- This verifies the 3 step procedure for this particular example.
- See the "trees" example for another case when adding a single variable to a simple linear regression.
- The added variable plot is a plot of the first set of residuals against the second. It is used to suggest the functional form of $x2$ to add.

Know how to carry out the 3 step procedure - see trees example

- Regression on both sets of variables can be thought of as a sequential three step procedure
  1. Regress $\boldsymbol{y}$ on $\boldsymbol{X}_1$ to get residuals $\boldsymbol{e}_1 = (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y}$ and estimates $\hat{\boldsymbol{\theta}}$.
  2. Regress $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$ (each column) to get residuals $\boldsymbol{e}_2 = (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2$.
  3. Regress $\boldsymbol{e}_1$ on $\boldsymbol{e}_2$ to get $\hat{\boldsymbol{\beta}}_2$ as in (3) above and solve for $\hat{\boldsymbol{\beta}}_1$ as in (4) above.

Partitioning the regression sum of squares

- Step 1 gives $SSR(\beta_1)$, the amount explained by $\boldsymbol{X}_1$ in the regression of y on $\boldsymbol{X}_1$, and

$$SST = SSR(\beta_1) + SSE(\beta_1)$$

.

In the ANOVA table for the regression of y on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, $SSR(\beta_1)$ is the regression SS entry for the terms $\boldsymbol{X}_1$.

- Step 3 gives $SSR(\beta_2|\beta_1)$, the regression SS for $\boldsymbol{X}_2$ given that $\boldsymbol{X}_1$ is already accounted for. This is the regression SS entry in the ANOVA table for the regression of y on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

- When additional variables $\boldsymbol{X}_2$ are added to a reduced model containing $\boldsymbol{X}_1$, the error SS in the reduced model is partitioned into a reduced error SS for the full model, plus the regression SS for the variables added.
- Equivalently, the regression SS for the terms added is the difference in error SS of the "reduced" and "full" models.

$$SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = SSE(\boldsymbol{\beta}_1) - SSE(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$$

example, continued

```
>lm1=lm(y~x1)
> anova(lm1)
Analysis of Variance Table

          Df Sum Sq Mean Sq F value   Pr(>F)
x1         1 115.07  115.07  14.119 0.000877 ***
Residuals 26 211.90    8.15

> lmbig=lm(y~x1+x2)
> anova(lmbig)
Analysis of Variance Table

          Df  Sum Sq Mean Sq F value    Pr(>F)
x1         1 115.068 115.068  22.378 7.492e-05 ***
x2         1  83.343  83.343  16.208 0.0004635 ***
Residuals 25 128.553   5.142
```

- $SSE(\beta_1) = SSE_{reduced} = 211.90$
- $SSE(\beta_1, \beta_2) = SSE_{full} = 128.553$
- $SSR(\beta_2|\beta_1) = 211.90 - 128.553 \approx 83.343$

- Equivalently, the total regression sum of squares (when both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are included), is

$$SSR(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = SSR(\boldsymbol{\beta}_1) + SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$$

where the extra sum of squares for regression explained by $\boldsymbol{X}_2$ given that $\boldsymbol{X}_1$ is in the model is

$$SSR(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}_2^T \boldsymbol{X}_2^T (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{H}_{2.1} \boldsymbol{y}$$

with

$$\boldsymbol{H}_{2.1} = (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2[\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2]^{-1}\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)$$

- $\boldsymbol{H}_{2.1}$ is the projection onto the component of the subspace spanned by $\boldsymbol{X}_2$ which is orthogonal to the subspace spanned by the columns of $\boldsymbol{X}_1$.
- **The analysis above shows that the partition of the sum of squares into two parts depends on the order in which the variables are added into the model when the columns of $\boldsymbol{X}_1$ are not orthogonal to the columns of $\boldsymbol{X}_2$**
- The numerator of the F test of the null hypothesis $H_0 : \beta_2 = 0$ is

$$(SSE(\beta_1) - SSE(\beta_1, \beta_2))/r = SSR(\beta_2|\beta_1)/r$$

  where the regression sum of squares is for those terms which were added last.
- If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are orthogonal, $(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2 = \boldsymbol{X}_2$ so that $SSR(\beta_2|\beta_1) = SSR(\beta_2)$, and the order of inclusion doesn't matter.

ANOVA table showing the decomposition of SSR - know how to construct/use the table

Source	SS	D.F.	MS
$\boldsymbol{X}_1$	$SSR(\boldsymbol{\beta}_1)$	$k-r$	$SSR(\boldsymbol{\beta}_1)/(k-r)$
$\boldsymbol{X}_2 \vert \boldsymbol{X}_1$	$SSR(\boldsymbol{\beta}_2 \vert \boldsymbol{\beta}_1)$	r	$SSR(\boldsymbol{\beta}_2 \vert \boldsymbol{\beta}_1)/r$
Error	SSE	$n-1-k$	MSE
Total	SST	$n-1$	

- **Takeaway message:** When new variables X_2 are added to a regression of y on X_1, the error sum of squares for the original regression is partitioned into a new, smaller error sum of squares, and the (sequential) regression sum of squares for X_2.

Justification for the partial F test

This is similar to what we saw previously. It's justification is once again Cochran's theorem.

- Under the null hypothesis $H_0 : \beta_2 = 0$

$$\frac{SSR(\beta_2|(\beta_1)}{\sigma^2} \sim \chi^2_r$$

and

$$F = \frac{MSR(\beta_2|\beta_1)}{MSE} = \frac{(SSE(\beta_1) - SSE(\beta_1, \beta_2))/r}{MSE} \sim F_{r,n-1-k}$$

and large F gives evidence against H_0.

Explicit formula for the covariance matrix - no need to remember these

- From (3) and (4), the fitted values are

$$\hat{\boldsymbol{y}} = (\boldsymbol{H}_1 + \boldsymbol{H}_{2.1})\boldsymbol{y},$$

so the hat matrix is

$$\boldsymbol{H} = \boldsymbol{H}_1 + \boldsymbol{H}_{2.1}.$$

- The covariance matrix of $\hat{\boldsymbol{\beta}}_2$ can be calculated from equation (3)

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}_2) &= Var([\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2]^{-1}\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y}) \\ &= [\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2]^{-1}\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)\sigma^2\boldsymbol{I} \\ &\quad (\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2[\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2]^{-1} \\ &= \sigma^2[\boldsymbol{X}_2^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{X}_2]^{-1} \end{aligned}$$

.

- We have used the trick of orthogonalizing the two sets of predictors, and this has allowed us to avoid inverting the $\boldsymbol{X}^T\boldsymbol{X}$ matrix, which is difficult when it is not block diagonal.
- In this way we have also been able to obtain $Var(\hat{\beta}_2)$, which is the bottom right corner of $\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.
- By symmetry it follows that

$$Var(\hat{\beta}_1) = \sigma^2[\boldsymbol{X}_1^T(\boldsymbol{I} - \boldsymbol{H}_2)\boldsymbol{X}_1]^{-1}$$

Equivalence of the partial F test and the t test - know that they are equivalent

- The partial F test for $H_0 : \beta_k = 0$ is equivalent to the t test.
- To see this, note that

$$\begin{aligned} SSR(\beta_k|\boldsymbol{\beta}_1) &= \hat{\beta}_k x_k^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{y} \\ &= \hat{\beta}_k^2/(x_k^T(\boldsymbol{I} - \boldsymbol{H}_1)x_k)^{-1} \end{aligned}$$

so

$$F = \frac{\hat{\beta}_k^2}{s^2 x_k^T(\boldsymbol{I} - \boldsymbol{H}_1)x_k}$$

on 1 and $n - 1 - k$ degrees of freedom.

- We saw that

$$C_{k,k} = (x_k^T(\boldsymbol{I} - \boldsymbol{H}_1)x_k)^{-1}$$

so

$$F = \frac{\hat{\beta}_k^2}{(s\sqrt{C_{k,k}})^2} = \left(\frac{\beta_k}{s\sqrt{C_{k,k}}}\right)^2 = t^2$$

- From this we see that the usual t ratio for testing $H_0 : \beta_j = 0$, when squared, gives the F statistic.
- We also see that the t ratio assumes that all the other variables are included in the model first.
- In other words, when we look at the t statistic we must consider that all other variables have been included in the model.

Added variable plot

- When $\boldsymbol{X}_2$ has only 1 column, the residuals at step 1 can be plotted against the residuals at step 2, showing exactly how the coefficient of the new variable, X_k, is obtained in the full model.
- This is called the **added variable** or **partial leverage** plot.
- The slope of the least squares line for this plot equals the coefficient of X_k.
- The correlation between $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ is called the **partial correlation** between y and X_k given $\boldsymbol{X}_1$.
- see the example using the trees dataset for an added variable plot. In that example, the plot looks linear, which suggests adding a linear function of x_2 to the model.
- On assignment 5, you're given the added variable plot, and asked to suggest which function of "age" - linear or quadratic - is most appropriate.