ANOVA Table with Degrees of Freedom

each line is 1 df and we have a total of r 1-df lines

Sum of Squares	Degrees of Freedom	DF
$\boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$	$\text{rank}(\boldsymbol{X}_2) - \text{rank}(\boldsymbol{X}_1)$	$r_2 - 1$
$\boldsymbol{y}^\top(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y}$	$\text{rank}(\boldsymbol{X}_3) - \text{rank}(\boldsymbol{X}_2)$	$r_3 - r_2$
$\vdots$	$\vdots$	$\vdots$
$\boldsymbol{y}^\top(\boldsymbol{P}_m - \boldsymbol{P}_{m-1})\boldsymbol{y}$	$\text{rank}(\boldsymbol{X}_m) - \text{rank}(\boldsymbol{X}_{m-1})$	$r - r_{m-1}$
$\boldsymbol{y}^\top(\boldsymbol{P}_{m+1} - \boldsymbol{P}_m)\boldsymbol{y}$	$\text{rank}(\boldsymbol{X}_{m+1}) - \text{rank}(\boldsymbol{X}_m)$	$n - r$
$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$	$\text{rank}(\boldsymbol{X}_{m+1}) - \text{rank}(\boldsymbol{X}_1)$	$n - 1$

Mean Squares

For $j = 1, \ldots, m-1$, define

$$MS(j+1 \mid j) = \frac{SS(j+1 \mid j)}{r_{j+1} - r_j} = \frac{\boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}}{r_{j+1} - r_j}.$$

These sums of squares divided by their degrees of freedom are known as *mean squares*.

ANOVA Table with Mean Squares

$MS(j+1|j) / MSE = F$

Sum of Squares	Degrees of Freedom	Mean Square
$SS(2 \mid 1)$	$r_2 - 1$	$MS(2\|1)$
$SS(3 \mid 2)$	$r_3 - r_2$	$MS(3\|2)$
$\vdots$	$\vdots$	$\vdots$
$SS(m \mid m-1)$	$r - r_{m-1}$	$MS(m\|m-1)$
SSE	$n - r$	$MSE - \hat{\sigma}^2$
$SSTo$	$n - 1$	

Independence of ANOVA Sums of Squares

Because

$$(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\left(\sigma^2 \boldsymbol{I}\right)(\boldsymbol{P}_{\ell+1} - \boldsymbol{P}_\ell) = \boldsymbol{0}$$

for all $j \neq \ell$, any two ANOVA sums of squares (not including $SSTo$) are independent.

It is also true that the ANOVA sums of squares (not including $SSTo$) are mutually independent by Cochran's Theorem, but that stronger result is not usually needed.

Stat 6110

For $j = 1, \ldots, m-1$ we have

$$F_j = \frac{MS(j+1 \mid j)}{MSE} = \frac{\boldsymbol{y}^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}/(r_{j+1} - r_j)}{\boldsymbol{y}^\top (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}/(n-r)}$$

den df

$$\sim F_{\boxed{r_{j+1}-r_j}, n-r}\left(\underbrace{\frac{\boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2}}_{\text{ncp}}\right).$$

n df

ANOVA Table with F-Statistics

Sum of Squares	Degrees of Freedom	Mean Square	F Stat
$SS(2 \mid 1)$	$r_2 - 1$	$MS(2\|1)$	F_1
$SS(3 \mid 2)$	$r_3 - r_2$	$MS(3\|2)$	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$SS(m \mid m-1)$	$r - r_{m-1}$	$MS(m\|m-1)$	F_{m-1}
SSE	$n - r$	MSE	
$SSTo$	$n - 1$		

Relationship with Reduced vs. Full Model F-Statistic

The ANOVA F_j statistic:

$$F_j = \frac{\boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}/(r_{j+1} - r_j)}{\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}/(n-r)} = \frac{MS(j+1 \mid j)}{MSE}$$

The reduced vs. full model F statistic:

$$F = \frac{\boldsymbol{y}^\top(\boldsymbol{P}_{\boldsymbol{X}} - \boldsymbol{P}_{\boldsymbol{X}_0})\boldsymbol{y}/(r - r_0)}{\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}/(n-r)}$$

What do ANOVA F-statistics test?

In general, an F-statistic is used to test

$\delta = n p c = 0$

H_0 : "The non-centrality parameter of the F-statistic is zero."

vs.

H_A : "The non-centrality parameter of the F-statistic is not zero."

$\delta = n p c \neq 0$

What do ANOVA F-statistics test?

The ANOVA F-statistic

$$F_j = \frac{\boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}/(r_{j+1} - r_j)}{\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}/(n - r)} = \frac{MS(j+1 \mid j)}{MSE}$$

has non-centrality parameter

ncp will be zero as long as numerator = 0

$$\text{ncp} = \frac{\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2}.$$

Thus, F_j can be used to test

$$\text{H}_{0j} : \frac{\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2} = 0 \text{ versus}$$

$$\text{H}_{Aj} : \frac{\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2} \neq 0.$$

$= a$

$\beta^T X^T (P_{j+1} - P_j)^T \boxed{(P_{j+1} - P_j) X\beta}$ due to symmetry & idempotency which is of form $a^T a$

The following are equivalent ways to write the null and alternative hypotheses tested by F_j.

H_{0j} : ncp $= 0$	H_{Aj} : ncp $\neq 0$
follows from previous slide (not considering $2\sigma^2$ in denom)	
$\boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} = 0$	$\boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} \neq 0$
$\boxed{(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta}} = \boldsymbol{0}$ a	$(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} \neq \boldsymbol{0}$ which can only $= 0$ if $a = 0$
$\boldsymbol{P}_j E(\boldsymbol{y}) = \boldsymbol{P}_{j+1} E(\boldsymbol{y})$ $X\beta = E(y)$	$\boldsymbol{P}_j E(\boldsymbol{y}) \neq \boldsymbol{P}_{j+1} E(\boldsymbol{y})$
$\boldsymbol{P}_{j+1} E(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}_j)$	$\boldsymbol{P}_{j+1} E(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}_{j+1}) \setminus \mathcal{C}(\boldsymbol{X}_j)$ for $E(y)$

$E(y)$ is found in the smaller of the two column spaces, thus it does matter in which column space we look

What do ANOVA F-statistics test?

Recall $C\beta$ is estimable $\Longleftrightarrow$ $C = AX$

$$H_{0j} : (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0} \quad \text{vs.} \quad H_{Aj} : (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\underline{\boldsymbol{X}}\boldsymbol{\beta} \neq \boldsymbol{0}$$

is of the form

$$H_{0j} : \boldsymbol{C}_j^*\boldsymbol{\beta} = \boldsymbol{0} \quad \text{vs.} \quad H_{Aj} : \boldsymbol{C}_j^*\boldsymbol{\beta} \neq \boldsymbol{0},$$

$$\text{where } \boldsymbol{C}_j^* = \underbrace{\boxed{(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)}}_{=A}\underline{\boldsymbol{X}}.$$

As written, H_{0j} is not a testable hypothesis because $\boldsymbol{C}_j^*$ has n rows but rank $r_{j+1} - r_j < n$ (homework problem).

We can rewrite H_{0j} as a testable hypothesis by replacing $\boldsymbol{C}_j^*$ with any matrix $\boldsymbol{C}_j$ whose $q = r_{j+1} - r_j$ rows form a basis for the row space of $\boldsymbol{C}_j^*$.

Example: Multiple Regression

intercept only model

adding a covariate assuming linear relationship with y
$\Rightarrow$ SLR

multiple linear regression

$$
\begin{aligned}
\boldsymbol{X}_1 &= \mathbf{1} \\
\boldsymbol{X}_2 &= [\mathbf{1},\ \boldsymbol{x}_1] \\
\boldsymbol{X}_3 &= [\mathbf{1},\ \boldsymbol{x}_1,\ \boldsymbol{x}_2] \\
&\ \vdots \\
\boldsymbol{X}_m &= [\mathbf{1},\ \boldsymbol{x}_1,\ \ldots,\ \boldsymbol{x}_{m-1}]
\end{aligned}
$$

$\boxed{SS(j+1 \mid j)}$ is the decrease in SSE that results when the explanatory variable x_j is added to a model containing an intercept and explanatory variables $x_1, \ldots, x_{j-1}$.

Example: Polynomial Regression

quadratic model

$$\begin{aligned} \boldsymbol{X}_1 &= \mathbf{1} \\ \boldsymbol{X}_2 &= [\mathbf{1},\ \boldsymbol{x}] \\ \boldsymbol{X}_3 &= [\mathbf{1},\ \boldsymbol{x},\ \boldsymbol{x}^2] \\ &\vdots \\ \boldsymbol{X}_m &= [\mathbf{1},\ \boldsymbol{x},\ \boldsymbol{x}^2,\ \ldots,\ \boldsymbol{x}^{m-1}] \end{aligned}$$

$SS(j+1 \mid j)$ is the decrease in SSE that results when the explanatory variable x^j is added to a model containing an intercept and explanatory variables $x, x^2, \ldots, x^{j-1}$.

An Example in R

```
> #An example from "Design of Experiments: Statistical
> #Principles of Research Design and Analysis"
> #2nd Edition by Robert O. Kuehl
>
> d=read.delim("https://.../S510/PlantDensity.txt")
```

The Data

X

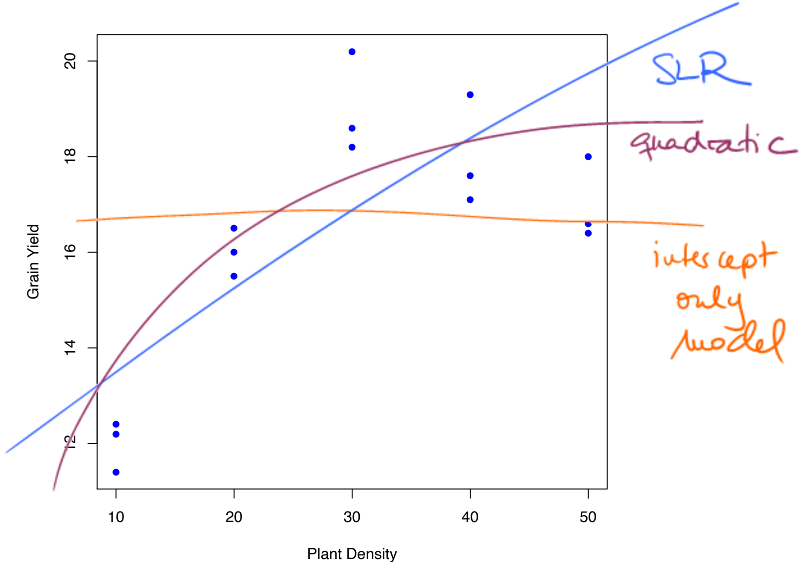
```
> d
   PlantDensity GrainYield
1            10       12.2
2            10       11.4
3            10       12.4
4            20       16.0
5            20       15.5
6            20       16.5
7            30       18.6
8            30       20.2
9            30       18.2
10           40       17.6
11           40       19.3
12           40       17.1
13           50       18.0
14           50       16.4
15           50       16.6
```

5 different levels of treatment

3 replicates per level

Renaming the Variables and Plotting the Data

```
> names(d)=c("x","y")
> head(d)
   x    y
1 10 12.2
2 10 11.4
3 10 12.4
4 20 16.0
5 20 15.5
6 20 16.5
>
> plot(d[,1],d[,2],col=4,pch=16,xlab="Plant Density",
+      ylab="Grain Yield")
```


Grain Yield
20
18
16
14
12
10
20
30
40
50
Plant Density
SLR
quadratic
intercept only model

Matrices with Nested Column Spaces

$$\boldsymbol{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \boldsymbol{X}_2 = \begin{bmatrix} 1 & 10 \\ 1 & 10 \\ 1 & 10 \\ 1 & 20 \\ 1 & 20 \\ 1 & 20 \\ 1 & 30 \\ 1 & 30 \\ 1 & 30 \\ 1 & 40 \\ 1 & 40 \\ 1 & 40 \\ 1 & 50 \\ 1 & 50 \\ 1 & 50 \end{bmatrix}, \ \boldsymbol{X}_3 = \begin{bmatrix} 1 & 10 & 100 \\ 1 & 10 & 100 \\ 1 & 10 & 100 \\ 1 & 20 & 400 \\ 1 & 20 & 400 \\ 1 & 20 & 400 \\ 1 & 30 & 900 \\ 1 & 30 & 900 \\ 1 & 30 & 900 \\ 1 & 40 & 1600 \\ 1 & 40 & 1600 \\ 1 & 40 & 1600 \\ 1 & 50 & 2500 \\ 1 & 50 & 2500 \\ 1 & 50 & 2500 \end{bmatrix},$$

Matrices with Nested Column Spaces

$$\boldsymbol{X}_4 = \begin{bmatrix} 1 & 10 & 100 & 1000 \\ 1 & 10 & 100 & 1000 \\ 1 & 10 & 100 & 1000 \\ 1 & 20 & 400 & 8000 \\ 1 & 20 & 400 & 8000 \\ 1 & 20 & 400 & 8000 \\ 1 & 30 & 900 & 27000 \\ 1 & 30 & 900 & 27000 \\ 1 & 30 & 900 & 27000 \\ 1 & 40 & 1600 & 64000 \\ 1 & 40 & 1600 & 64000 \\ 1 & 40 & 1600 & 64000 \\ 1 & 50 & 2500 & 125000 \\ 1 & 50 & 2500 & 125000 \\ 1 & 50 & 2500 & 125000 \end{bmatrix}, \quad \boldsymbol{X}_5 = \begin{bmatrix} 1 & 10 & 100 & 1000 & 10000 \\ 1 & 10 & 100 & 1000 & 10000 \\ 1 & 10 & 100 & 1000 & 10000 \\ 1 & 20 & 400 & 8000 & 160000 \\ 1 & 20 & 400 & 8000 & 160000 \\ 1 & 20 & 400 & 8000 & 160000 \\ 1 & 30 & 900 & 27000 & 810000 \\ 1 & 30 & 900 & 27000 & 810000 \\ 1 & 30 & 900 & 27000 & 810000 \\ 1 & 40 & 1600 & 64000 & 2560000 \\ 1 & 40 & 1600 & 64000 & 2560000 \\ 1 & 40 & 1600 & 64000 & 2560000 \\ 1 & 50 & 2500 & 125000 & 6250000 \\ 1 & 50 & 2500 & 125000 & 6250000 \\ 1 & 50 & 2500 & 125000 & 6250000 \end{bmatrix}$$

Centering and Standardizing for Numerical Stability

It is typically best for $\boxed{\text{numerical stability}}$ to center and scale a quantitative explanatory variable prior to computing higher order terms.

In the plant density example, we could replace x by $(x - 30)/10$ and work with the matrices on the next two slides.

Because these matrices have the same column spaces as the original matrices, the ANOVA table entries are mathematically identical for either set of matrices.

Matrices with Centered and Scaled x

$$
\boldsymbol{X}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad
\boldsymbol{X}_2 = \begin{bmatrix} 1 & -2 \\ 1 & -2 \\ 1 & -2 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}, \quad
\boldsymbol{X}_3 = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -2 & 4 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 2 & 4 \end{bmatrix},
$$

Matrices with Centered and Scaled x

a different mean for each treatments

$$
\boldsymbol{X}_4 = \begin{bmatrix}
1 & -2 & 4 & -8 \\
1 & -2 & 4 & -8 \\
1 & -2 & 4 & -8 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & 2 & 4 & 8 \\
1 & 2 & 4 & 8 \\
1 & 2 & 4 & 8
\end{bmatrix}, \quad
\boldsymbol{X}_5 = \begin{bmatrix}
1 & -2 & 4 & -8 & 16 \\
1 & -2 & 4 & -8 & 16 \\
1 & -2 & 4 & -8 & 16 \\
1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & 1 \\
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 2 & 4 & 8 & 16 \\
1 & 2 & 4 & 8 & 16 \\
1 & 2 & 4 & 8 & 16
\end{bmatrix}
$$

$\Updownarrow$

5 different means

Regardless of whether we center and scale x, the column space of $\boldsymbol{X}_5$ is the same as the column space of the cell means model matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

ANOVA Table for the Plant Density Data

by how much will the SSE ↓ if we add x to the model that contains only intercept

Source	Sum of Squares	DF
$x\|1$ intercept only	$\boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$	$2-1=1$
$x^2\|1,x$	$\boldsymbol{y}^\top(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y}$	$3-2=1$
$x^3\|1,x,x^2$	$\boldsymbol{y}^\top(\boldsymbol{P}_4 - \boldsymbol{P}_3)\boldsymbol{y}$	$4-3=1$
$x^4\|1,x,x^2,x^3$	$\boldsymbol{y}^\top(\boldsymbol{P}_5 - \boldsymbol{P}_4)\boldsymbol{y}$	$5-4=1$
Error	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_5)\boldsymbol{y}$	$15-5=10$
C. Total	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$	$15-1=14$

Creating the Matrices in R

```
> y=d$y
> x=(d$x-mean(d$x))/10
> x
 [1] -2 -2 -2 -1 -1 -1  0  0  0  1  1  1  2  2  2
>
> n=nrow(d)
>
> x1=matrix(1,nrow=n,ncol=1)
> x2=cbind(x1,x)
> x3=cbind(x2,x^2)
> x4=cbind(x3,x^3)
> x5=matrix(model.matrix(~0+factor(x)),nrow=n)
> I=diag(rep(1,n))
```

Creating the Projection Matrices in R

```
> library(MASS)
> proj=function(x){
+   x%*%ginv(t(x)%*%x)%*%t(x)
+ }
>
> p1=proj(x1)
> p2=proj(x2)
> p3=proj(x3)
> p4=proj(x4)
> p5=proj(x5)
```

P_1

$\vdots$

P_5

Computing the Sums of Squares in R

```
> t(y)%*%(p2-p1)%*%y
      [,1]
[1,] 43.2
> t(y)%*%(p3-p2)%*%y
      [,1]
[1,]   42
> t(y)%*%(p4-p3)%*%y
      [,1]
[1,]  0.3
> t(y)%*%(p5-p4)%*%y
      [,1]
[1,]  2.1
> t(y)%*%(I-p5)%*%y
      [,1]
[1,] 7.48
> t(y)%*%(I-p1)%*%y
       [,1]
[1,] 95.08
```

$= y^T(P_2 - P_1)y$

$= y^T(P_3 - P_2)y$

sizable reductions in SSE

reductions are neglig– compared to SLR & quadratic regression model

// end lecture 14