STAT 571A — Advanced Statistical Regression Analysis

Chapter 7 NOTES

Multiple Regression – II

© 2015 University of Arizona Statistics GIDP. All rights reserved, except where previous rights exist. No part of this material may be reproduced, stored in a retrieval system, or transmitted in any form or by any means — electronic, online, mechanical, photoreproduction, recording, or scanning — without the prior written consent of the course instructor.

Extra Sums of Squares in MLR

- Recall: the (p–1)-variable **Multiple Linear Regression (MLR)** model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

  where $\varepsilon_i \sim$ i.i.d.$N(0, \sigma^2)$; $i = 1, \ldots, n$.
- The MLR ANOVA decomposes SSTO into SSR + SSE.
- Consider: each X_k contributes its own particular sum of squares to the SSR term. Let's explore this in more detail...

Extra Sums of Squares (cont'd)

For instance in the SLR ANOVA, if we drill down into the calculations we find that X_1 contributes

$$SSR(X_1) = \sum(b_0 + b_1X_{i1} - \bar{Y})^2$$

to SSR, with 1 d.f. We could write the SLR ANOVA decomposition as

$$SSTO = SSR(X_1) + SSE(X_1)$$

Extra Sums of Squares (cont'd)

Recall that SSTO = $\sum(Y_i - \bar{Y})^2$ doesn't change with the X_k's. Say we add X_2 to the model. Then SSR is now SSR(X_1,X_2).

But SSTO = SSR(X_1,X_2) + SSE(X_1,X_2) is fixed so

SSR(X_1,X_2) $\uparrow$ must force SSE(X_1,X_2) $\downarrow$

Extra Sums of Squares (cont'd)

Recall that SSTO = $\sum(Y_i - \bar{Y})^2$ doesn't change with the X_k's. Say we add X_2 to the model. Then SSR is now $SSR(X_1,X_2)$.

But SSTO = $SSR(X_1,X_2) + SSE(X_1,X_2)$ is fixed so

$$SSR(X_1,X_2) \uparrow \text{ must force } SSE(X_1,X_2) \downarrow$$

$\Rightarrow$ SS info. has been transferred <u>from</u> SSE <u>to</u> SSR.

Extra Sums of Squares (cont'd)

The additional SS moved from SSE to SSR is called an Extra Sum of Squares:

- **For X_1 already in the model, adding X_2 generates an extra SS of $SSR(X_2|X_1)$, so write**

  $$SSR(X_1,X_2) = SSR(X_1) + SSR(X_2|X_1)$$

  in $SSTO = SSR(X_1,X_2) + SSE(X_1,X_2)$.

- **Order is (usually) important, so $SSR(X_2|X_1) \neq SSR(X_1|X_2)$. But we do see that**

  $$SSR(X_1|X_2) = SSE(X_2) - SSE(X_1,X_2)$$
  $$= SSR(X_1,X_2) - SSR(X_2).$$

Extra Sums of Squares (cont'd)

Moving to p=4 with 3 X_k's gives, e.g.,

$$SSR(X_3|X_1,X_2) = SSE(X_1,X_2) - SSE(X_1,X_2,X_3)$$
$$= SSR(X_1,X_2,X_3) - SSR(X_1,X_2)$$

And, for that matter,

$$SSR(X_2,X_3|X_1) = SSE(X_1) - SSE(X_1,X_2,X_3)$$
$$= SSR(X_1,X_2,X_3) - SSR(X_1)$$

(You get the idea...)

Extra Sums of Squares (cont'd)

Given a full model (FM) with p–1 X_k's, the extra SS terms in effect decompose the full SSR:

$$SSR(F) = SSR(X_1) + SSR(X_2|X_1) + \cdots + SSR(X_{p-1}|X_1,\ldots,X_{p-2})$$

Indeed, we can also write

$$SSR(X_2,\ldots,X_{p-1}|X_1) = SSR(X_2|X_1) + \cdots + SSR(X_{p-1}|X_1,\ldots,X_{p-2})$$

etc., etc.,...

ANOVA Table

The ANOVA table becomes (Table 7.3):

Source	d.f.	SS	MS
Regr.	$p-1$	$SSR(F)$	$MSR(F)$
X_1	1	$SSR(X_1)$	$MSR(X_1)$
$X_2 \| X_1$	1	$SSR(X_2 \| X_1)$	$MSR(X_2 \| X_1)$
⋮	⋮	⋮	⋮
$X_{p-1} \| X_1 \ldots X_{p-2}$	1	$SSR(X_{p-1} \| X_1 \ldots X_{p-2})$	$MSR(X_{p-1} \| X_1 \ldots X_{p-2})$
Error	$n-p$	SSE	MSE
Total	$n-1$	$SSTO$	

Each indiv. term is called a **sequential SS**

The last sequential SS is called the **partial SS**

ANOVA Table (cont'd)

In the ANOVA table:

- Recall that order is important! So, re-arranging the order of where each X_k is fit will (usually) change the sequential SS terms.
- Notice: the indiv. <u>sequential</u> MSR terms each have 1 d.f. So, e.g.,

$$MSR(X_2|X_1) = SSR(X_2|X_1)/1.$$

- We <u>can</u> pool additive terms. For instance

$$MSR(X_2,X_3|X_1) = SSR(X_2,X_3|X_1)/2$$
$$= \tfrac{1}{2}\{SSR(X_2|X_1) + SSR(X_3|X_1, X_2)\}$$

§7.2: Hypothesis Testing

To <u>test</u> H_o:$\beta_k = 0$ vs. H_a:$\beta_k \neq 0$ we already know that $t_k^* = b_k/s\{b_k\} \sim t(n-p)$ (under H_o) provides a test statistic.

It can be shown that

$$(t_k^*)^2 = \frac{SSR(X_k|X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_{p-1})/(1)}{MSE} = F_k^*$$

has $F_k^* \sim F(1,n-p)$ (under H_o).

(Notice that the SSR in the F_k^* numerator is the *k*th partial SS.)

t^2-to-F equivalence

- So, we can reject $H_o:\beta_k = 0$ vs. $H_a:\beta_k \neq 0$ whenever $F_k^* > F(1-\alpha;1,n-p)$.
- But since $(t_k^*)^2 = F_k^*$ and we know $t(n-p)^2 = F(1,n-p)$, we see the *k*th partial F-test and the *k*th t-test are equivalent!
- Which to use? Whichever is handy (i.e., fastest on the computer...).
- But, for one-sided tests of, say, $H_o:\beta_k=0$ vs. $H_a:\beta_k>0$, can only use t_k^*.

Multi-d.f. F-tests

Now, recall that we can build multiple β_k's into the extra SSR terms. From this, we can test multi-d.f. hypotheses.

For instance, to test
$H_o: \beta_k = \beta_j = 0$ vs. H_a: any difference, use

$$F_{kj}^* = \frac{SSR(X_k, X_j | X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_{j-1}, X_{j+1}, \ldots, X_{p-1})/(2)}{MSE}$$

Under H_o, $F_{kj}^* \sim F(2, n-p)$ so reject H_o when $F_{kj}^* > F(1-\alpha; 2, n-p)$.

Multi-d.f. F-tests (cont'd)

- **Notice what this is doing: the SSR in the numerator of the F_{kj}-statistic, $SSR(X_k, X_j | X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_{j-1}, X_{j+1}, \ldots, X_{p-1})$, fits X_k and X_j last and is the 2 d.f. partial SSR.**
- **Then, it builds the MSR and divides by the MSE to create a 2 d.f. partial F-test.**
- **(There is *no equivalent* 2 d.f. t-test here.)**

Example: Body Fat Data (CH07TA01)

- Example: from table 7.1, let
  Y = % body fat in adult women
  X_1 = tricep thickness
  X_2 = thigh circumf.
  X_3 = midarm circumf.
- Sample size is n = 20.
- Test if these predictor variables affect E{Y}.

Body Fat Data (CH07TA01) (cont'd)

In the Body Fat Data example, p = 4 and we produce an overall ANOVA as seen earlier:

```
> CH07TA01.lm = lm( Y ~ X1 + X2 + X3 )
> anova( lm(Y ~ 1), CH07TA01.lm )
```

```
Analysis of Variance Table
Model 1: Y ~ 1
Model 2: Y ~ X1 + X2 + X3
 Res.Df     RSS Df   [SSR]      F    Pr(>F)
1     19 495.39
2     16  98.40  3  396.98 21.516 7.343e-06
```

$F^* = 21.516$ tests $H_o{:}\beta_1=\beta_2=\beta_3=0$ ($P = 7.3\times10^{-6}$)

Body Fat Data (CH07TA01) (cont'd)

The sequential SSR terms (with the ANOVA decomposition) are found using:

```
> anova( CH07TA01.lm )

Analysis of Variance Table
Response: Y
        Df  Sum Sq Mean Sq F value    Pr(>F)
X1       1  352.27  352.27 57.2768 1.131e-06
X2       1   33.17   33.17  5.3931   0.03373
X3       1   11.55   11.55  1.8773   0.18956
Resid. 16   98.40    6.15
```

Partial F_3^* = 1.8773 tests H_o:β_3=0 (P = .1896)

Body Fat Data (CH07TA01) (cont'd)

A 2 df partial F-test – using $SSR(X_2,X_3|X_1)$ – is also easy to produce:

```
> anova( lm(Y ~ X1), CH07TA01.lm )
Analysis of Variance Table
Model 1: Y ~ X1
Model 2: Y ~ X1 + X2 + X3
 Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     18 143.120
2     16  98.405  2    44.715 3.6352 0.04995
```

$F^* = 3.6352$ tests $H_o{:}\beta_2=\beta_3=0$ ($P = 0.04995$)

Sequential Sums of Squares

- After fitting X_1, we can sequentially fit X_2, and then X_3, etc. Each term produces a sequential SSR in the order that they are fit (so order is important): $SSR(X_2|X_1)$, and $SSR(X_3|X_1,X_2)$.
- The sequential SSR terms add up to the full SSR available in the ANOVA:

$$SSR(F) = SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1,X_2)$$
$$= SSTO - SSE(F)$$

- Sequential SSR's allow for "sequential" testing of the X_k's <u>in the order they enter the model</u>, using the ANOVA decomposition.

Sequential Sums of Squares (cont'd)

Recall the (sequenced) ANOVA table:

Source	d.f.	SS	MS
Regr.	p–1	SSR(F)	MSR(F)
X_1	1	$SSR(X_1)$	$MSR(X_1)$
$X_2\|X_1$	1	$SSR(X_2\|X_1)$	$MSR(X_2\|X_1)$
⋮	⋮	⋮	⋮
$X_{p-2}\|X_1 \ldots X_{p-3}$	1	$SSR(X_{p-2}\|X_1 \ldots X_{p-3})$	$MSR(X_{p-2}\|X_1 \ldots X_{p-3})$
$X_{p-1}\|X_1 \ldots X_{p-2}$	1	$SSR(X_{p-1}\|X_1 \ldots X_{p-2})$	$MSR(X_{p-1}\|X_1 \ldots X_{p-2})$
Error	n–p	SSE	MSE
Total	n–1	SSTO	

Sequential Sums of Squares (cont'd)

- **In general, for sequential testing via the SSR's:**
- **Step 1. Start <u>at</u> <u>the</u> <u>bottom</u> with the partial test of $H_o{:}\beta_{p-1} = 0$ vs. $H_a{:}\beta_{p-1} \neq 0$ (all tests are two-sided). Find $F_{p-1}^* = MSR(X_{p-1}|X_1 \ldots X_{p-2})/MSE$.**
- **Step 2a. <u>If</u> $F_{p-1}^* > F(1-\alpha;1,n-p)$ then reject H_o and STOP. (Cannot proceed further 'up'.)**

Sequential Sums of Squares (cont'd)

- **Step 2b**. But if $F_{p-1}^* \leq F(1-\alpha;1,n-p)$ then fail to reject H_o and conclude $\beta_{p-1} = 0$.
- **Step 3**. Now, if $\beta_{p-1} = 0$, view $SSR(X_{p-1}|X_1 \ldots X_{p-2})$ as inconsequential and proceed 'up' to test $H_o:\beta_{p-2} = 0$ via

$$F_{p-2}^* = MSR(X_{p-2}|X_1 \ldots X_{p-3})/MSE.$$

[Technically, we really should resorb $SSR(X_{p-1}|X_1 \ldots X_{p-2})$ back into SSE, so this is a bit of a short-cut approximation.]

Sequential Sums of Squares (cont'd)

- **Step 4a**. <u>If</u> $F_{p-2}^* > F(1-\alpha;1,n-p)$ then reject $H_o:\beta_{p-2} = 0$ and STOP. (Cannot proceed further 'up'.)
- **Step 4b**. But if $F_{p-2}^* \leq F(1-\alpha;1,n-p)$ then fail to reject H_o and conclude $\beta_{p-2} = 0$.
- **Step 5**. Now, if $\beta_{p-2} = 0$, view $SSR(X_{p-2}|X_1 \ldots X_{p-3})$ as inconsequential and proceed 'up' to test $H_o:\beta_{p-3} = 0$ via

$$F_{p-3}^* = MSR(X_{p-3}|X_1 \ldots X_{p-4})/MSE.$$

Sequential Sums of Squares (cont'd)

- **Step 5+**. Continue 'up the ladder' in this fashion until the first rejection occurs – stop there.
- Note that there is an issue of *multiplicity* here (same data are used to perform all the sequential tests). If felt to be an issue, can apply a Bonferroni correction, but...that's awfully conservative!
- If the SSR terms are orthogonal (see below), then the Kimball Inequality may be applicable.

§7.3: Summary of β_k Testing

(A) To test H_o:$\beta_1 = \cdots = \beta_{p-1} = 0$ use "full" F-test via $F^* = MSR(F)/MSE \sim F(p-1,n-p)$.

(B) To test a <u>single</u> H_o:$\beta_k = 0$ use "partial" F-test via

$F_k^* = MSR(X_k|X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_{p-1})/MSE$

with $F_k^* \sim F(1,n-p)$

$\Leftrightarrow$ Equiv. to $t^* = b_k/s\{b_k\} \sim t(n-p)$.

β_k Testing (cont'd)

(C) To test a <u>subset</u> of β_k's, say (after reordering) $H_o: \beta_q = \beta_{q+1} = \cdots = \beta_{p-1} = 0$ use the p–q d.f. partial F-test via

$F_{p-q}^* = MSR(X_q, \ldots, X_{p-1} | X_1, \ldots, X_{q-1})/MSE$

with $F_{p-q}^* \sim F(p-q, n-p)$.

(D) To test something funkier, e.g., $H_o: \beta_3 = \beta_4$, need to build a reduced model (RM) under H_o and apply the FM-RM discrepancy approach from (2.70).

§7.4: Partial R^2

- The quantity $R^2 = SSR/(SSR + SSE)$ can be manipulated in similar "sequential" or "partial" fashion, since it derives from SSR and SSE.
- For instance, suppose p–1 X_k's make up FM. Consider the 4 predictors X_k, X_ℓ, X_m, X_q. The <u>partial</u> R^2 for X_k, <u>given</u> X_ℓ, X_m, X_q is

$$R_{Yk|\ell,m,q}^{2} = SSR(X_k|X_\ell,X_m,X_q)/SSE(X_\ell,X_m,X_q)$$

Partial R^2 (cont'd)

- $R_{Yk|\ell,m,q}^2$ is called a **Coefficient of Partial Determination**.
- Interpretation: % variation in Y explained by X_k given that X_ℓ, X_m, X_q have already been fit in the MLR model.

Body Fat Data (CH07TA01) (cont'd)

The **partial R^2 values** are available from the **anova()** components: e.g., $R^2_{Y3|12}$ is

```
> CH07TA01.aov = anova(lm(Y~X1+X2+X3))
> CH07TA01.aov[3,2]/anova(lm(Y~X1+X2))[3,2]
[1] 0.1050097
```

while $R^2_{Y2|1}$ is

```
> CH07TA01.aov[2,2]/anova(lm(Y~X1))[2,2]
[1] 0.2317564
```

etc.

§7.6: Multicollinearity

- **The MLR calculations run into trouble when two different X_k's represent the same information.**
- **For instance, if $X_3 = 2X_2$, there is <u>no new info</u>. in X_3. [Technically, rank($\mathbf{X'X}$) < p.] So, the ANOVA breaks down – cf. Table 7.8. Most programs spot this and just drop X_3.**
- **But, this is pretty obvious...**

Multicollinearity (cont'd)

- What if 2 (or more!) X_k's represent almost the same info.? We can still fit them in the MLR model, but they aren't really helping that much.
- Usual consequence: the sequential SSR's get all mucked up. E.g., suppose X_1 and X_2 are highly correlated & represent very similar info. We might find $SSR(X_1) = 352.27$, but $SSR(X_1|X_2) = 3.47$. Weird? No: X_1 fits fine until it's swamped out by X_2.

Multicollinearity (cont'd)

- So, when $SSR(X_1) >> SSR(X_1|X_2)$, it's possible that the conclusions of the sequential F-tests could rely *solely on the order* under which the X's are fit.

  $\Rightarrow$ Seems capricious!

- We say then that X_1 & X_2 are **Multicollinear** (a bad thing).

Effects of Multicollinearity

Multicollinearity can:

- substantially affect the partial F-tests and how ordering of the X_k's impacts the inferences;
- destabilize point estimates of b_k [since $(\mathbf{X}'\mathbf{X})^{-1}$ is "ill-conditioned"];
- destabilize (usually inflate!) $s\{b_k\}$, s{pred}, etc.;
- botch up the partial R^2 values.

Example: Body Fat Data (CH07TA01)

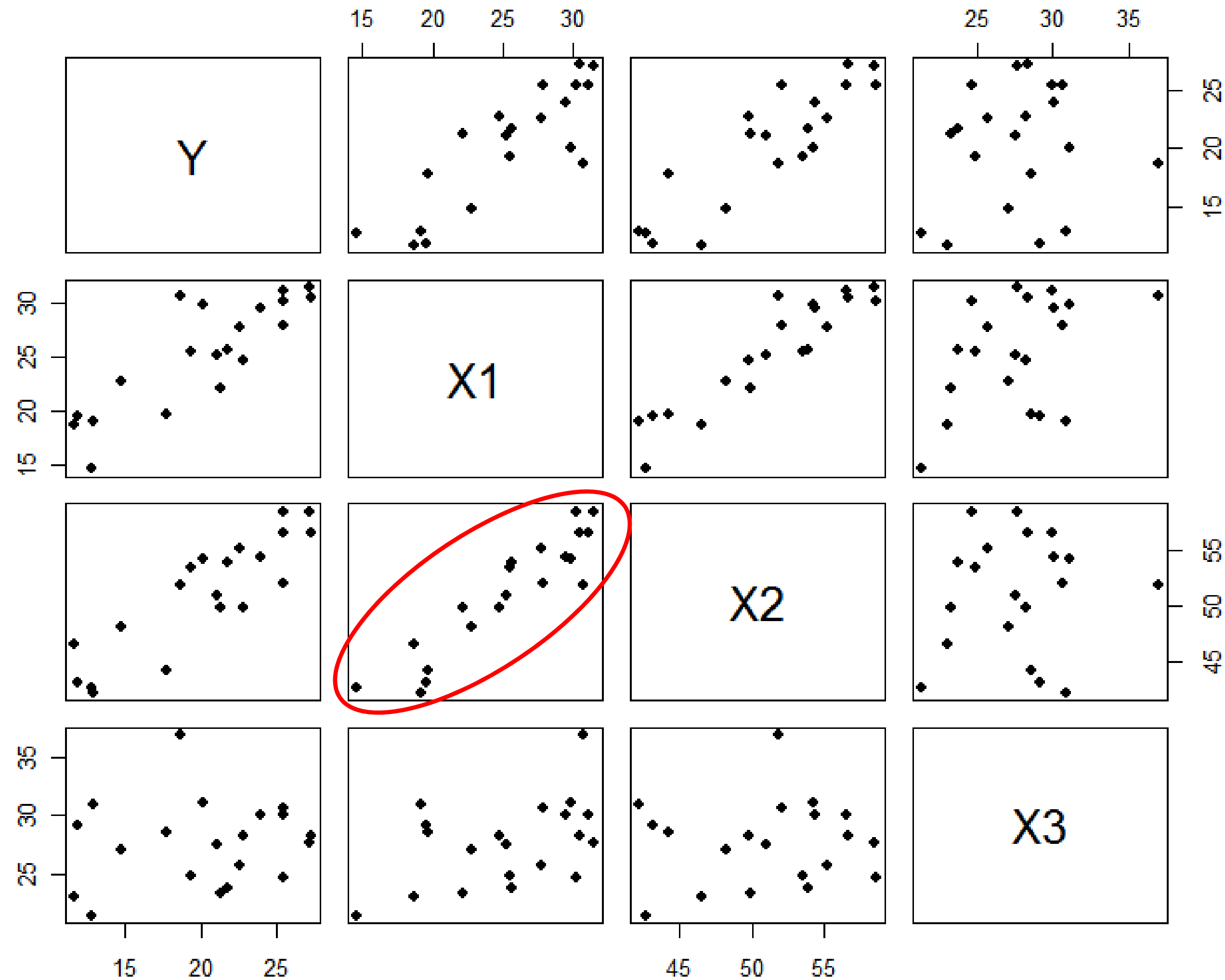
Find the correlations between Y and the X_k predictor variables via the cor() command:

```
> cor( CH07TA01.df )
           Y        X1        X2        X3
Y   1.000000  0.843265  0.8780896  0.1424440
X1  0.843265  1.000000  0.9238425  0.4577772
X2  0.878090  0.923843  1.0000000  0.0846675
X3  0.142444  0.457777  0.0846675  1.0000000
```

Large correlation between X_1 and X_2
$\Rightarrow$ possible multicollinearity!

Scatterplot Matrix via `pairs(CH07TA01.df)`

High **linear relationship** **between X_1 and X_2**



Body Fat Data (CH07TA01) (cont'd)

Multicollinearity between X_1 and X_2 disturbs inferences from the ANOVA. Compare the

```
lm( Y ~ X1 + X2 + X3 )
```

ordering with

```
lm( Y ~ X3 + X2 + X1 )
```

in terms of the sequential SSR's

(see next slides →)

Body Fat Data (CH07TA01) (cont'd)

```
> anova( lm(Y ~ X1 + X2 + X3) )
Analysis of Variance Table
Response: Y
       Df  Sum Sq  Mean Sq F value    Pr(>F)
X1      1  352.27   352.27 57.2768 1.131e-06
X2      1   33.17    33.17  5.3931   0.03373
X3      1   11.55    11.55  1.8773   0.18956
Resid. 16   98.40     6.15
```

Sequencing "up the ladder":

X_3 appears insignif., then X_2 (weakly) significant (so stop there)

Body Fat Data (CH07TA01) (cont'd)

```
> anova( lm(Y ~ X3 + X2 + X1) )
Analysis of Variance Table
Response: Y
        Df  Sum Sq  Mean Sq  F value    Pr(>F)
X3       1    10.1     10.1    1.634     0.219
X2       1   374.2    374.2   60.847  7.68e-07
X1       1    12.7     12.7    2.066     0.170
Resid. 16   98.40     6.15
```

Sequencing "up the ladder":
X_1 appears insignif., then X_2 (strongly) significant (so stop there) ⇒ multicollin. is quite confusing!

Multicollinearity Control

- **Are there remedies for multicollinearity? Not really. (Too bad!)**
- **<u>If</u> selection of the X_k's can be controlled, we can try to minimize multicollinearity amongst them.**
- **Easiest way: drive $corr(X_k, X_m) \to 0$ so that X_k and X_m are orthogonal.**

Multicollinearity Control (cont'd)

- In fact, <u>when</u> $corr(X_k,X_m) = 0$, $SSR(X_k|X_m) = SSR(X_k)$ and $SSR(X_m|X_k) = SSR(X_m)$.
  $\Rightarrow$ "the sequentials equal the partials"
- If so, no multicollinearity exists between them! (A good thing.)
- Otherwise, can try manipulating the ANOVA sequencing order of the X_k's to isolate any strange inferences/collinear effects.

Example: Work Crew Data

In the Work Crew Data example (CH07TA06), the $p-1 = 2$ X_k-variables are uncorrelated:

```
> X1 = c( rep(4,4),rep(6,4) )
> X2 = rep( c(2,2,3,3), 2 )
> Y = c(42, 39, 48, 51, 49, 53, 61, 60)
> cor( cbind(Y,X1,X2) )
```

	Y	X1	X2
Y	1.0000000	0.7419309	0.6384057
X1	0.7419309	1.0000000	0.0000000
X2	0.6384057	0.0000000	1.0000000

Work Crew Data (CH07TA06) (cont'd)

X_1 and X_2 exhibit <u>no collinearity</u>, so the resulting sequential SSRs are orthogonal and unaffected by entry order in the ANOVA. Start with X_1-then-X_2:

```
> anova( lm(Y ~ X1 + X2) )
Analysis of Variance Table
Response: Y
      Df  Sum Sq Mean Sq F value   Pr(>F)
X1     1 231.125 231.125  65.567 0.000466
X2     1 171.125 171.125  48.546 0.000937
Resid  5  17.625   3.525
```

Work Crew Data (CH07TA06) (cont'd)

Now fit X_2-then-X_1:

```
> anova( lm(Y ~ X2 + X1) )
Analysis of Variance Table
Response: Y
      Df  Sum Sq Mean Sq F value   Pr(>F)
X2     1 171.125 171.125  48.546 0.000937
X1     1 231.125 231.125  65.567 0.000466
Resid  5  17.625   3.525
```

Notice that $SSR(X_2) = SSR(X_2|X_1)$ and $SSR(X_1|X_2) = SSR(X_1)$ (see previous slide).

Work Crew Data (CH07TA06) (cont'd)

Now just fit X_2:

```
> anova( lm(Y ~ X2) )
Analysis of Variance Table
Response: Y
      Df  Sum Sq Mean Sq F value   Pr(>F)
X2     1 171.12  171.125  4.1276  0.08846
Resid  6 248.75   41.458
```

$SSR(X_2)$ is unchanged but $SSR(X_1|X_2)$ has been absorbed into SSE ($\Rightarrow$ MSE rises sharply, so β_2 no longer significant at $\alpha = .05$!); cf. Table 7.7.