7. Extra Sums of Squares

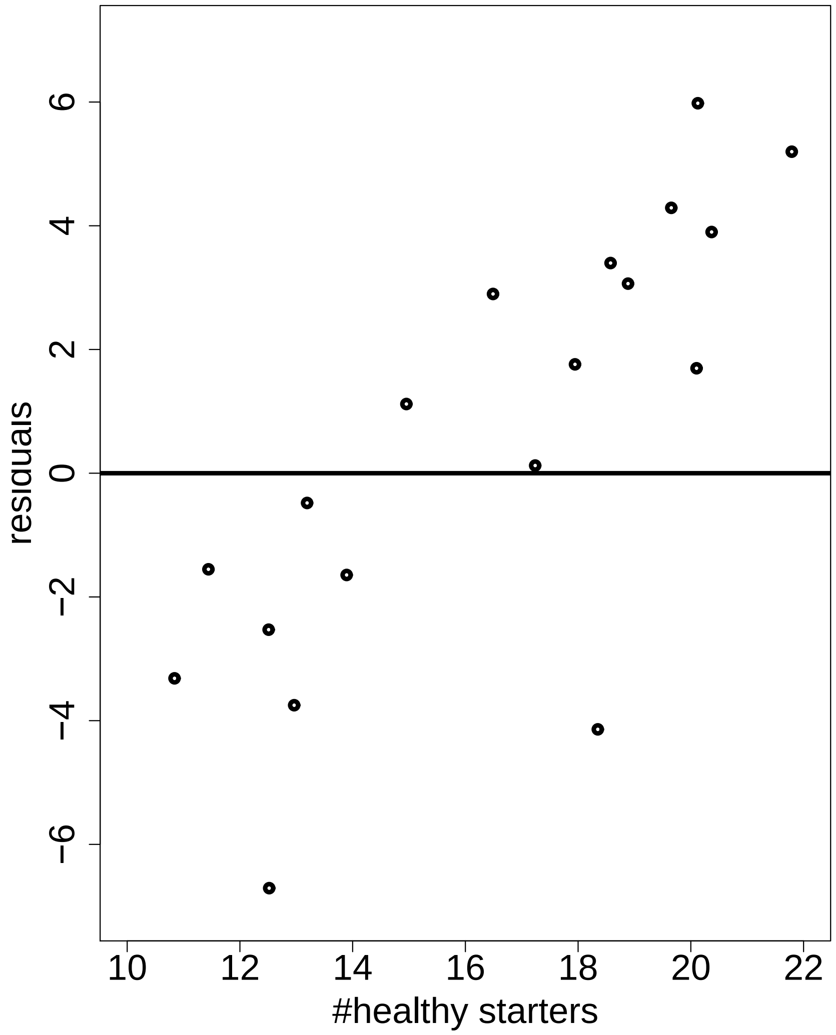
Football Example:

Y_i = #points scored by UF football team in game i
X_{i1} = #games won by opponent in their last 10 games
X_{i2} = #healthy starters for UF (out of 22) in game i

Suppose we fit the SLR

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$$

and plot the residuals e_i against X_{i2}:



Q: What do we conclude from this ?

A: The residuals appear to be linearly related to X_{i2}, thus, X_{i2} should be put into the model.

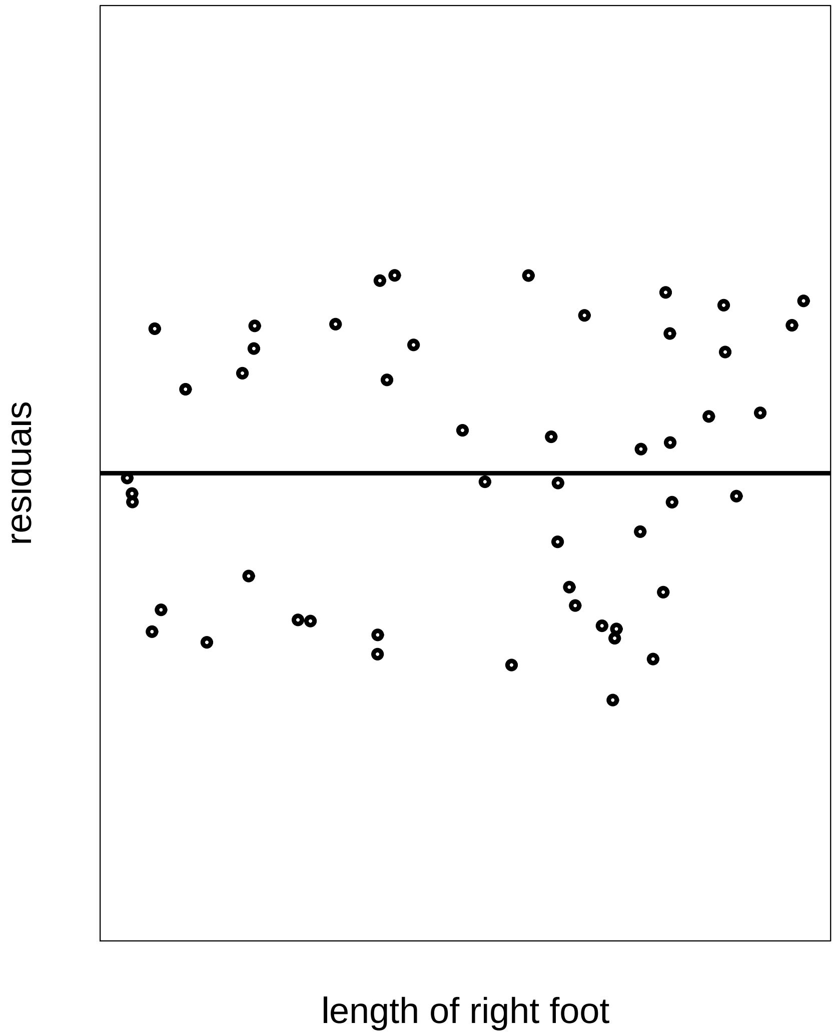
Another Example:

Y_i = height of a person
X_{i1} = length of left foot
X_{i2} = length of right foot

Suppose we fit the SLR

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$$

and plot the residuals e_i against X_{i2}:



Q: Why no pattern?

A: X_{i2} is providing the same information about Y that X_{i1} does. Thus, even though X_{i2} is a good predictor of height, it is unnecessary if X_{i1} is already in the model.

Extra sums of squares provide a means of formally testing whether one set of predictors is necessary **given** that another set is already in the model.

Recall that

$$\begin{aligned} \text{SSTO} &= \text{SSR} + \text{SSE} \\ \sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \\ R^2 &= \frac{\text{SSR}}{\text{SSTO}} \end{aligned}$$

Important Fact: R^2 will never decrease when a predictor is added to a regression model.

Consider the two different models:

$$\begin{aligned} \mathsf{E}(Y_i) &= \beta_0 + \beta_1 X_{i1} \\ \mathsf{E}(Y_i) &= \beta_0^* + \beta_1^* X_{i1} + \beta_2^* X_{i2} \end{aligned}$$

Q: Is SSTO the same for both models?

A: Yes! Thus, SSR will never decrease when a predictor is added to a model.

Since SSE and SSR are different depending upon which predictors are in the model, we use the following notation:

SSR(X_1): SSR for a model with only X_1

SSR(X_1, X_2): SSR for a model with X_1 and X_2

SSE(X_1) and SSE(X_1, X_2) have analogous def's

Note

$$\begin{aligned} \text{SSTO} &= \text{SSR}(X_1) + \text{SSE}(X_1) \\ \text{SSTO} &= \text{SSR}(X_1, X_2) + \text{SSE}(X_1, X_2) \end{aligned}$$

We also know SSR$(X_1, X_2) \geq$ SSR(X_1).

Thus SSE$(X_1, X_2) \leq$ SSE(X_1).

Conclusion: SSE never increases when a predictor is added to a model.

Reconsider the Example:
$Y_i \;=\;$ height of a person
$X_{i1} =$ length of left foot; $X_{i2} =$ length of right foot

Q: What do you think about the quantity

$$\text{SSR}(X_1, X_2) - \text{SSR}(X_1)$$

A: Probably small because if we know the length of the left foot, knowing the length of the right won't help.

Notation: Extra Sum of Squares

$$\text{SSR}(X_2|X_1) = \text{SSR}(X_1, X_2) - \text{SSR}(X_1)$$

$\text{SSR}(X_2|X_1)$ tells us how much we gain by adding X_2 to the model **given** that X_1 is already in the model.

We define $\text{SSR}(X_1|X_2) = \text{SSR}(X_1, X_2) - \text{SSR}(X_2)$

We can do this with as many predictors as we like, e.g.

$$\begin{aligned}
\text{SSR}(X_3, X_5|X_1, X_2, X_4) &= \text{SSR}(X_1, X_2, X_3, X_4, X_5) - \text{SSR}(X_1, X_2, X_4) \\
&= \text{SSR(all predictors)} - \text{SSR(given predictors)}
\end{aligned}$$

Suppose our model is:

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$

Consider tests involving β_1, β_2, and β_3.

One Beta: $H_0 : \beta_k = 0, \quad k = 1, 2,$ **or** 3
H_A : not H_0

In words, this test says "Do we need X_k given that the other two predictors are in the model?"

Can do this with a t-test:

$$t^* = b_k / \sqrt{\mathsf{MSE} \cdot [(\mathbf{X}'\mathbf{X})^{-1}]_{k+1,k+1}}$$

Two Betas: (some of the Betas)

$H_0 : \beta_1 = \beta_2 = 0$	$H_0 : \beta_1 = \beta_3 = 0$	$H_0 : \beta_2 = \beta_3 = 0$
H_A : not H_0	H_A : not H_0	H_A : not H_0

For example, the first of these asks "Do we need X_1 and X_2 given that X_3 is in the model?"

All Betas: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
H_A : not H_0

This is just the overall F-Test

We can do all of these tests using extra sum of squares.

Here is the ANOVA table corresponding to the model

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$

ANOVA Table:

Source of variation	SS	df
Regression	SSR(X_1, X_2, X_3)	$p - 1 = 3$
Error	SSE(X_1, X_2, X_3)	$n - p = n - 4$
Total	SSTO	$n - 1$

Partition SSR(X_1, X_2, X_3) into 3 one df extra sums of squares. One way to do it is:

$$\text{SSR}(X_1, X_2, X_3) = \text{SSR}(X_1) + \text{SSR}(X_2|X_1) + \text{SSR}(X_3|X_1, X_2)$$

Modified ANOVA Table:

Source of variation	SS	df
Regression	SSR(X_1, X_2, X_3)	3
	SSR(X_1)	1
	SSR$(X_2\|X_1)$	1
	SSR$(X_3\|X_1, X_2)$	1
Error	SSE(X_1, X_2, X_3)	$n-4$
Total	SSTO	$n-1$

Note: there are 6 equivalent ways of partitioning SSR(X_1, X_2, X_3).

Three Tests: ($p = 4$ in this example)

- **One Beta:** $H_0 : \beta_2 = 0 \quad$ vs. $\quad H_A :$ not H_0

Test statistic: $F^* = \dfrac{\text{SSR}(X_2|X_1, X_3)/1}{\text{SSE}(X_1, X_2, X_3)/(n-p)}$

Rejection rule: Reject H_0 if $F^* > F(1-\alpha; 1, n-p)$

- **Some Betas:** $H_0 : \beta_2 = \beta_3 = 0 \quad$ vs. $\quad H_A :$ not H_0

Test statistic: $F^* = \dfrac{\text{SSR}(X_2, X_3|X_1)/2}{\text{SSE}(X_1, X_2, X_3)/(n-p)}$

Rejection rule: Reject H_0 if $F^* > F(1-\alpha; 2, n-p)$

- **All Betas:** $H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad$ vs. $\quad H_A :$ not H_0

Test statistic: $F^* = \dfrac{\text{SSR}(X_1, X_2, X_3)/3}{\text{SSE}(X_1, X_2, X_3)/(n-p)}$

Rejection rule: Reject H_0 if $F^* > F(1-\alpha; p-1, n-p)$

Let's return to the model

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$

and think about testing

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad H_A : \text{not } H_0$$

Test statistic: $F^* = \dfrac{\mathsf{SSR}(X_2, X_3|X_1)/2}{\mathsf{MSE}(X_1, X_2, X_3)}$

How do we get $\mathsf{SSR}(X_2, X_3|X_1)$ if we have $\mathsf{SSR}(X_1)$, $\mathsf{SSR}(X_2|X_1)$, and $\mathsf{SSR}(X_3|X_1, X_2)$?

$$\mathsf{SSR}(X_2, X_3|X_1) = \mathsf{SSR}(X_2|X_1) + \mathsf{SSR}(X_3|X_1, X_2)$$

What if we would have $\mathsf{SSR}(X_2)$, $\mathsf{SSR}(X_1|X_2)$, and $\mathsf{SSR}(X_3|X_1, X_2)$? **Stuck!**

lm(Y ~ X1+X2+X3)	lm(Y ~ X2+X1+X3)
SSR(X_1)	SSR(X_2)
SSR$(X_2\|X_1)$	SSR$(X_1\|X_2)$
SSR$(X_3\|X_1, X_2)$	SSR$(X_3\|X_1, X_2)$

Example: Patient Satisfaction

$Y_i \quad =$ patient satisfaction ($n = 23$)
$X_{i1} =$ patient's age in years
$X_{i2} =$ severity of illness (index)
$X_{i3} =$ anxiety level (index)

Model 1: Consider the model with all 3 pairwise interactions included ($p = 7$)

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1} X_{i3} + \beta_6 X_{i2} X_{i3}$$

and think about testing the 3 interaction terms:

$$H_0 : \beta_4 = \beta_5 = \beta_6 = 0 \quad \text{vs.} \quad H_A : \text{not } H_0$$

Denote the interaction $X_j X_k$ by I_{jk}. Then

Test statistic: $F^* = \dfrac{\text{SSR}(I_{12}, I_{13}, I_{23} | X_1, X_2, X_3)/3}{\text{MSE}(X_1, X_2, X_3, I_{12}, I_{13}, I_{23})}$

Rejection rule: Reject H_0 if $F^* > F(1-\alpha; 3, n-p)$

How do we get this extra sum of squares?

Q: How many partitions of SSR$(X_1, X_2, X_3, I_{12}, I_{13}, I_{23})$ into 6 one df extra sums of squares are there?
A: $6 \times 5 \times 4 \times 3 \times 2 = 6! = 720$

Q: Which ones will allow us to compute F^*?
A: The ones with I_{12}, I_{13}, and I_{23} last.

$$
\begin{aligned}
\text{SSR}(\cdot) &= \text{SSR}(X_1) + \text{SSR}(X_2|X_1) + \text{SSR}(X_3|X_1, X_2) \\
&+ \text{SSR}(I_{12}|X_1, X_2, X_3) \\
&+ \text{SSR}(I_{13}|X_1, X_2, X_3, I_{12}) \\
&+ \text{SSR}(I_{23}|X_1, X_2, X_3, I_{12}, I_{13})
\end{aligned}
$$

Add the last 3 (the interaction terms) to get SSR$(I_{12}, I_{13}, I_{23}|X_1, X_2, X_3)$

```
> summary(mod1 <- lm(sat ~ age + sev + anx + age:sev + age:anx + sev:anx))
 Coefficients:
              Estimate Std. Error t value Pr(>|t|)
 (Intercept) 241.57104  169.91520   1.422    0.174
 age           0.28112    4.65467   0.060    0.953
 sev          -6.32700    5.40579  -1.170    0.259
 anx          24.02586  101.65309   0.236    0.816
 age:sev       0.06969    0.10910   0.639    0.532
 age:anx      -2.20711    1.74936  -1.262    0.225
 sev:anx       1.16347    1.98054   0.587    0.565
```

```
> anova(mod1)
 Analysis of Variance Table
 Response: sat
           Df  Sum Sq Mean Sq F value Pr(>F)
 age        1 3678.44 3678.44 32.20  3.45e-05 ***
 sev        1  402.78  402.78  3.53  0.079    .
 anx        1   52.41   52.41  0.46  0.508
 sev:age    1    0.02    0.02  0.00  0.989
 sev:anx    1    1.81    1.81  0.02  0.901
 age:anx    1  181.85  181.85  1.59  0.225
 Residuals 16 1827.90  114.24
```

$F^* = \frac{(0.02+1.81+181.85)/3}{114.24} = 0.54$ is compared to $F(0.95; 3, 16)$

```
> qf(0.95, 3, 16)
 [1] 3.238872
```

Because $F^* < F(0.95; 3, 16) = 3.24$ we fail to reject H_0 (Interactions are not needed).

Model 2: Let's get rid of the interactions and consider

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$$

Do we need X_2 (severity of illness) and X_3 (anxiety level) if X_1 (age) is already in the model?

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad H_A : \text{not } H_0$$

Test statistic: $F^* = \dfrac{\mathsf{SSR}(X_2, X_3|X_1)/2}{\mathsf{MSE}(X_1, X_2, X_3)}$

Rejection rule: Reject H_0 if $F^* > F(1-\alpha; 2, n-p)$

How do we get this extra sum of squares?

$$\mathsf{SSR}(X_2, X_3|X_1) = \mathsf{SSR}(X_2|X_1) + \mathsf{SSR}(X_3|X_1, X_2)$$

```
> summary(mod2 <- lm(sat ~ age + sev + anx))
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) 162.8759    25.7757   6.319 4.59e-06 ***
 age          -1.2103     0.3015  -4.015  0.00074 ***
 sev          -0.6659     0.8210  -0.811  0.42736
 anx          -8.6130    12.2413  -0.704  0.49021

> anova(mod2)
 Analysis of Variance Table
 Response: sat
           Df Sum Sq Mean Sq F value Pr(>F)
 age        1 3678.4  3678.4 34.74  1.e-05 ***
 sev        1  402.8   402.8  3.80  0.0660 .
 anx        1   52.4    52.4  0.49  0.4902
 Residuals 19 2011.6   105.9
```

$F^* = \frac{(402.8+52.4)/2}{105.9} = 2.15$ is compared to

```
> qf(0.95, 2, 19)
[1] 3.521893
```

Because $F^* < F(0.95; 2, 19) = 3.52$ we again fail to reject H_0 (X_2 and X_3 are not needed).

Model 3: Let's get rid of X_2 (severity of illness) and X_3 (anxiety level) and consider the SLR with X_1 (age)

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1}$$

```
> summary(mod3 <- lm(sat ~ age))
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) 121.8318    11.0422  11.033 3.37e-10 ***
 age          -1.5270     0.2729  -5.596 1.49e-05 ***

> anova(mod3)
 Analysis of Variance Table
 Response: sat
           Df Sum Sq Mean Sq F value Pr(>F)
 age        1 3678.4  3678.4 31.315 1.49e-05 ***
 Residuals 21 2466.8   117.5
```

Let's construct 95% CI's for β_1 and for $\mathrm{E}(Y_h) = \mathbf{X}_h'\boldsymbol{\beta}$, where $\mathbf{X}_h' = (1\ 40\ 50\ 2)$, based on these 3 models.

```
> new <- data.frame(age=40, sev=50, anx=2)
```

Model 3: $(p = 2)$ $\qquad b_1 \pm t(0.975; 21)\sqrt{\mathsf{MSE}/S_{XX}} = (-2.09, -0.96)$

```
> predict(mod3,new,interval="confidence",level=0.95)
          fit     lwr      upr
 [1,] 60.75029 56.0453 65.45528
```

Model 2: $(p = 4)$ $\qquad b_1 \pm t(0.975; 19)\sqrt{\mathsf{MSE}[(\mathbf{X}'\mathbf{X})^{-1}]_{22}} = (-1.84, -0.58)$

```
> predict(mod2,new,interval="confidence",level=0.95)
          fit      lwr      upr
 [1,] 63.94183 55.85138 72.03228
```

Model 1: $(p = 7)$ $\qquad b_1 \pm t(0.975; 16)\sqrt{\mathsf{MSE}[(\mathbf{X}'\mathbf{X})^{-1}]_{22}} = (-9.59, 10.15)$

```
> predict(mod1,new,interval="confidence",level=0.95)
          fit     lwr      upr
 [1,] 63.67873 54.9398 72.41767
```

Correlation of Predictors Multicollinearity

Recall the SLR situation: data (X_i, Y_i), $i = 1, \ldots, n$

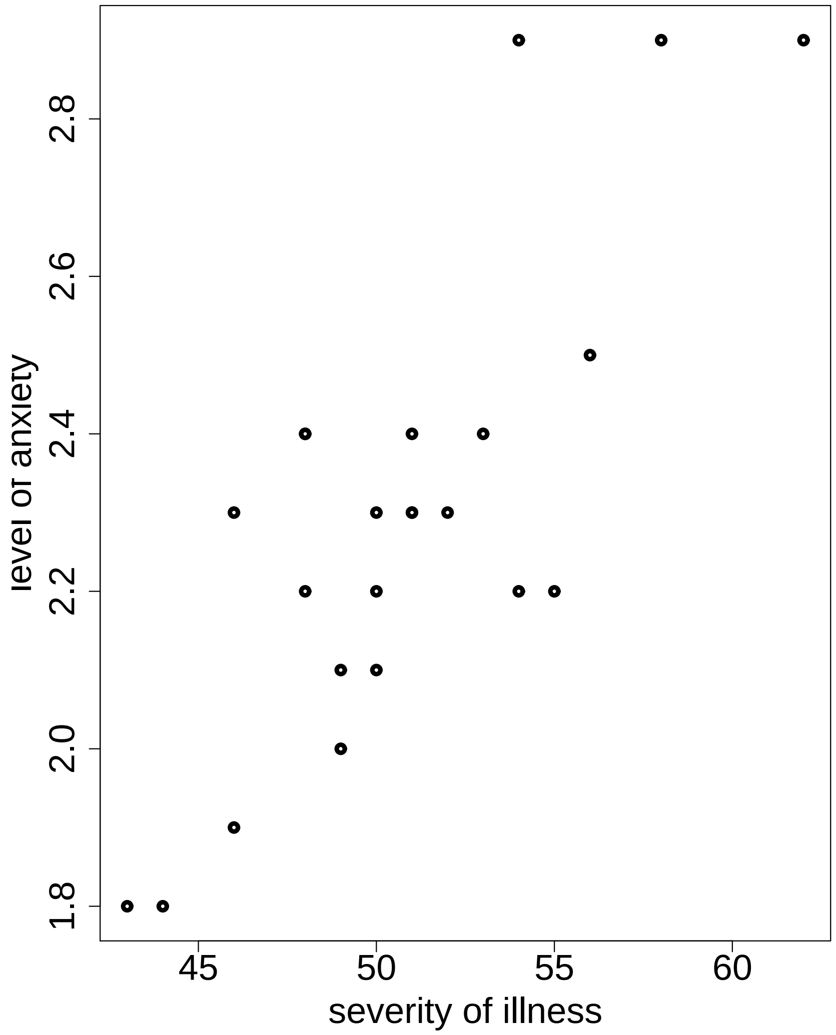
$$r^2 = \text{SSR}/\text{SSTO}$$

describes the amount of total variability in the Y_i's explained by the linear relationship between X and Y.

Because of SSR $= b_1^2 S_{XX}$, where $b_1 = S_{XY}/S_{XX}$, and with $S_{YY} =$ SSTO, the sample coefficient of correlation between X and Y is

$$r = \text{sign}(b_1)\sqrt{r^2} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

and gives us information about the strength of the linear relationship between X and Y, as well as the sign of the slope $(-1 \leq r \leq 1)$.



Patient Satisfaction:

Correlation between

$X_{i2} =$ severity of illness
$X_{i3} =$ anxiety level

$r_{23} = 0.7945$ (see below)

For a multiple regression data set $(X_{i1}, \ldots, X_{i,p-1}, Y_i)$

r_{jY} is the sample correlation coefficient between X_j and Y,

r_{jk} is the sample correlation coefficient between X_j and X_k.

- If $r_{jk} = 0$ then X_j and X_k are **uncorrelated**.

When most of the r_{jk}'s are close to 1 or -1, we say we have **multicollinearity** among the predictors.

```
> cor(patsat)
        sat     age     sev     anx
sat  1.0000 -0.7737 -0.5874 -0.6023
age -0.7737  1.0000  0.4666  0.4977
sev -0.5874  0.4666  1.0000  0.7945
anx -0.6023  0.4977  0.7945  1.0000
```

Uncorrelated vs. correlated predictors

Consider the 3 models:

(1) $\text{E}(Y_i) = \beta_0 + \beta_1 X_{i1}$

(2) $\text{E}(Y_i) = \beta_0 + \qquad\qquad \beta_2 X_{i2}$

(3) $\text{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$

and the 2 cases:

- X_1 and X_2 are uncorrelated ($r_{12} \approx 0$), then
  b_1 will be the same for models (1) and (3)
  b_2 will be the same for models (2) and (3)
  $\text{SSR}(X_1|X_2) = \text{SSR}(X_1)$
  $\text{SSR}(X_2|X_1) = \text{SSR}(X_2)$

- X_1 and X_2 are correlated ($|r_{12}| \approx 1$), then
  b_1 will be different for models (1) and (3)
  b_2 will be different for models (2) and (3)
  SSR$(X_1|X_2) <$ SSR(X_1)
  SSR$(X_2|X_1) <$ SSR(X_2)

When $r_{12} \approx 0$, X_1 and X_2 contain no redundant information about Y.

Thus, X_1 explains the same amount of the SSTO when X_2 is in the model as it does when X_2 is not.

Overview of the Effect of Multicollinearity

The standard errors of the parameter estimates are inflated. Thus, CI's for the regression parameters may be to large to be useful.

Inferences about $\mathsf{E}(Y_h) = \mathbf{X}_h'\boldsymbol{\beta}$, the mean of a response at $\mathbf{X}_h'$, and $Y_{h(new)}$, a new random variable observed at $\mathbf{X}_h$, are unaffected for the most part.

The idea of increasing X_1, when X_2 is fixed, may not be reasonable.

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

Interpretation: β_1 represents "the change in the mean of Y corresponding to a unit increase in X_1 holding X_2 fixed".

Polynomial Regression

Suppose we have SLR type data (X_i, Y_i), $i = 1, \ldots, n$. If $Y_i = f(X_i) + \epsilon_i$, where $f(\cdot)$ is unknown, it may be reasonable to approximate $f(\cdot)$ using a polynomial

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \cdots$$

Usually, you wouldn't go beyond the 3rd power.

Standard Procedure:
- Start with a higher order model and try to simplify.
- If X^k is retained, so are the lower order terms X^{k-1}, $X^{k-2}, \ldots, X$.

Warning:
- The model $\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_i + \cdots + \beta_{n-1} X_i^{n-1}$ always fits perfectly ($p = n$).
- Polynomials in X are highly correlated.

Polynomial Regression Example: Fish Data

$Y_i = \log(\text{species richness} + 1)$ observed at lake i, $i = 1, \ldots, 80$, in NY's Adirondack State Park.

We consider the 3rd order model:

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 pH_i + \beta_2 pH_i^2 + \beta_3 pH_i^3$$

```
> lnsr <- log(rch+1)
> ph2 <- ph*ph; ph3 <- ph2*ph
> summary(m3 <- lm(lnsr ~ ph + ph2 + ph3))
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) -16.82986    7.44163  -2.262   0.0266 *
 ph            7.07937    3.60045   1.966   0.0529 .
 ph2          -0.87458    0.56759  -1.541   0.1275
 ph3           0.03505    0.02930   1.196   0.2354
 ---
```

```
 Residual standard error: 0.4577 on 76 df
 Multiple R-Squared: 0.447, Adjusted R-squared: 0.425
 F-statistic: 20.45 on 3 and 76 df, p-value: 8.24e-10

> anova(m3)
Analysis of Variance Table
Response: lnsr
          Df  Sum Sq Mean Sq F value    Pr(>F)
ph         1  7.9340  7.9340 37.8708 3.280e-08 ***
ph2        1  4.6180  4.6180 22.0428 1.158e-05 ***
ph3        1  0.2998  0.2998  1.4308    0.2354
Residuals 76 15.9221  0.2095
```

Looks like pH^3 is not needed.

Let's see if we can get away with a SLR:

$$H_0 : \beta_2 = \beta_3 = 0 \quad \text{vs.} \quad H_A : \text{not } H_0$$

Test statistic:

$$\begin{aligned} F^* &= \frac{\mathsf{SSR}(pH^2, pH^3|pH)/2}{\mathsf{MSE}(pH, pH^2, pH^3)} \\ &= \frac{(4.6180 + 0.2998)/2}{0.2095} = 11.74 \end{aligned}$$

Rejection rule: Reject H_0 if $F^* > F(0.95; 2, 76) = 3.1$

Thus, a higher order term is necessary.

Let's test

$$H_0 : \beta_3 = 0 \quad \text{vs.} \quad H_A : \beta_3 \neq 0$$

Test statistic: $F^* = \dfrac{\mathsf{SSR}(pH^3|pH, pH^2)/1}{\mathsf{MSE}(pH, pH^2, pH^3)} = 1.43$

Rejection rule: Reject H_0 if $F^* > F(0.95; 1, 76) = 4.0$

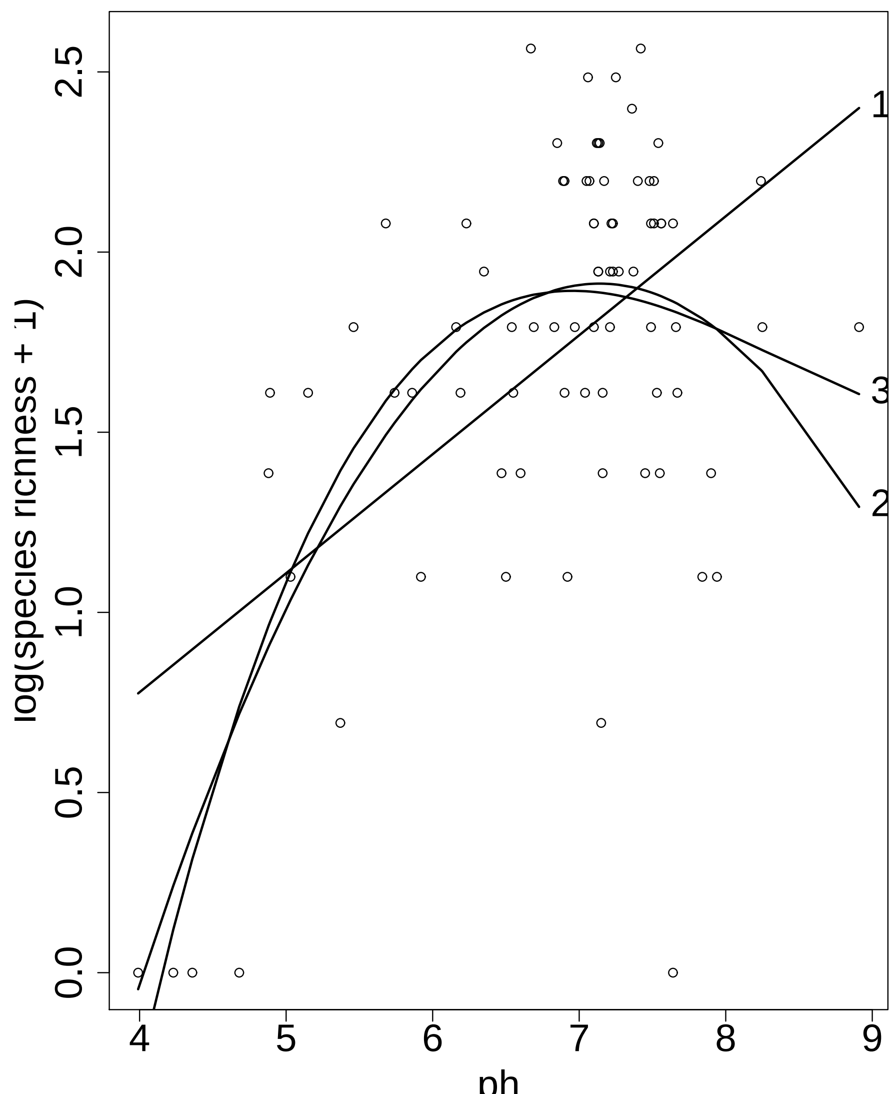
Conclusion: Can't throw away pH and pH^2 so the model we use is

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 pH_i + \beta_2 pH_i^2$$

```
> summary(m2 <- lm(lnsr ~ ph + ph2))
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) -8.1535    1.6675  -4.890 5.40e-06 ***
 ph           2.8201    0.5345   5.276 1.18e-06 ***
 ph2         -0.1975    0.0422  -4.682 1.20e-05 ***
 ---
```

```
Residual standard error: 0.459 on 77 df
Multiple R-Squared: 0.436, Adjusted R-squared: 0.422
F-statistic: 29.79 on 2 and 77 df, p-value: 2.6e-10

> anova(m2)
Analysis of Variance Table
Response: lnsr
          Df  Sum Sq Mean Sq F value    Pr(>F)
ph         1  7.9340  7.9340   37.66 3.396e-08 ***
ph2        1  4.6180  4.6180   21.92 1.198e-05 ***
Residuals 77 16.2218  0.2107
---
```



log(species richness + 1)
0.0
0.5
1.0
1.5
2.0
2.5
4
5
6
7
8
9
ph
1
2
3

Q: What's the big deal? All we did was get rid of the third order term, pH_i^3.

A: Suppose we are interested in a 95% CI for β_1:

Model	b_1	s.e.	CI(β_1)
3rd order	7.08	3.60	$(-0.12, 14.28)$
2nd order	2.82	0.53	$(+1.75, \ \ 3.89)$

We can do all of this stuff with more than 1 predictor. Suppose we have (X_{i1}, X_{i2}, Y_i), $i = 1, \ldots, n$.

2nd order model:

$$\mathsf{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2}$$

We could test $H_0 : \beta_3 = \beta_4 = \beta_5 = 0$. That is: "Is a 1st order model sufficient?"

Test statistic:

$$F^* = \frac{\mathsf{SSR}(X_1^2, X_2^2, X_1 X_2 | X_1, X_2)/3}{\mathsf{MSE}(X_1, X_2, X_1^2, X_2^2, X_1 X_2)}$$

Rejection rule: Reject H_0 if $F^* > F(0.95; 3, n-6)$.