Stat 206: Linear Models

Lecture 11

Nov. 4, 2019

General Linear Tests

$\mathcal{I}$ and $\mathcal{J}$ are two non-overlapping index sets.

- **Full model**: Contain both $X_{\mathcal{I}}$ and $X_{\mathcal{J}}$.
- **Reduced model**: Contain only $X_{\mathcal{I}}$.
- Test whether $X_{\mathcal{J}}$ may be dropped out of the full model:

$$H_0 : \beta_j = 0, \ \text{ for } \textbf{all } j \in \mathcal{J}$$

vs.

$$H_a : \text{some } \beta_j : j \in \mathcal{J} \text{ are nonzero.}$$

Basic idea: Compare under the full model with *SSE* under the reduced model by an F ratio:

- Under H_0 (i.e., the model):

$$F^* \sim_{H_0} \quad .$$

- Reject H_0 at level α if the observed F^* .

Basic idea: Compare SSE under the full model with SSE under the reduced model by an F ratio:

$$F^* = \frac{\frac{SSE(R)-SSE(F)}{df_R-df_F}}{\frac{SSE(F)}{df_F}} = \frac{MSR(X_{\mathcal{J}}|X_{\mathcal{I}})}{MSE(F)}.$$

- Under H_0 (i.e., the reduced model):

$$F^* \sim_{H_0} F_{df_R-df_F,df_F}.$$

- Reject H_0 at level α if the observed $F^* > F(1-\alpha; df_R - df_F, df_F)$.

F-test for Regression Relation

- Full model with $X_1, \cdots, X_{p-1}$:

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

- Reduced model with no X variable:

$$Y_i = \beta_0 + \epsilon_i, \quad i = 1, \cdots, n.$$

  So $SSE(R) =$,and $df_R =$.

- $SSE(R) - SSE(F) =$, and $df_R - df_F =$.
- F ratio

$$F^* = \qquad .$$

F-test for Regression Relation

- Full model with $X_1, \cdots, X_{p-1}$:

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

- Reduced model with no X variable:

$$Y_i = \beta_0 + \epsilon_i, \quad i = 1, \cdots, n.$$

  So $SSE(R) = SSTO$ and $df_R = n - 1$.

- $SSE(R) - SSE(F) = SSTO - SSE(F) = SSR(F)$, and $df_R - df_F = (n-1) - (n-p) = p - 1 = d.f.(SSR(F))$.
- F ratio

$$F^* = \frac{SSR(F)/(p-1)}{SSE(F)/(n-p)} = \frac{MSR(F)}{MSE(F)}.$$

Test whether a Single $\beta_k = 0$

Body fat: Test for the model with all three predictors whether the midarm circumference (X_3) can be dropped.

- Full model: $SSE(F) = 98.40$ with d.f. 16.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Null and alternative hypotheses:

$$H_0: \qquad\qquad vs. \quad H_a: \qquad\qquad .$$

- Reduced model: $SSE(R) =$ with d.f.

- $F^* =$.

- Pvalue= . So we
  X_3 from the full model.

Test whether a Single $\beta_k = 0$

Body fat: Test for the model with all three predictors whether the midarm circumference (X_3) can be dropped.

- Full model: $SSE(F) = 98.40$ with d.f. 16.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Null and alternative hypotheses:

$$H_0 : \beta_3 = 0 \;\; vs. \quad H_a : \beta_3 \neq 0.$$

- Reduced model: $SSE(R) = 109.95$ with d.f. 17.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- $F^* = \frac{11.55/1}{98.40/16} = 1.88.$
- Pvalue=$P(F_{1,16} > 1.88) = 0.189$. So we can drop X_3 from the full model.

Equivalence between F-test and T-test

- Test whether X_k can be dropped from a regression model with $p-1$ X variables:

$$H_0 : \beta_k = 0 \;\; vs. \;\; H_a : \beta_k \neq 0.$$

- T-test:

$$T^* = \frac{\hat{\beta}_k}{s\{\hat{\beta}_k\}} \underset{H_0}{\sim} t_{(n-p)},$$

where $\hat{\beta}_k$ is the LS estimator of β_k and $s\{\hat{\beta}_k\}$ is its standard error under the full model. Reject H_0 when $|T^*| > t(1-\alpha/2; n-p)$.

- $F^* = (T^*)^2$ and $F(1-\alpha; 1, n-p) = (t(1-\alpha/2; n-p))^2$. So for this test, F-test and T-test are equivalent.

Notes: for one one-sided alternatives, we still need the T-tests.

Test whether Several $\beta_k = 0$

Body fat: Test whether both thigh circumference (X_2) and midarm circumference (X_3) can be dropped from the model with all three predictors.

- Full model: $SSE(F) = 98.40$ with d.f. 16.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Null and alternative hypotheses:

$$H_0 : \qquad\qquad vs. \quad H_a :$$

- Reduced model: $SSE(R) =$ with d.f. .

- $F^* =$.
- Pvalue= . The result is at $\alpha = 0.05$.

Test whether Several $\beta_k = 0$

Body fat: Test whether both thigh circumference (X_2) and midarm circumference (X_3) can be dropped from the model with all three predictors.

- Full model: $SSE(F) = 98.40$ with d.f. 16.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Null and alternative hypotheses:

$$H_0 : \beta_2 = \beta_3 = 0 \;\; vs. \quad H_a : \text{not both } \beta_2 \text{ and } \beta_3 \text{ equal zero.}$$

- Reduced model: $SSE(R) = 143.12$ with d.f. 18.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- $F^* = \frac{44.72/2}{98.40/16} = 3.635.$
- Pvalue$= P(F_{2,16} > 3.635) = 0.0499$. The result is barely significant at $\alpha = 0.05$.

Standardization

Different X variables often have different units which could make their values vastly different.

- Regression coefficients are not comparable.
- Elements of $\mathbf{X}'\mathbf{X}$ could differ substantially in order of magnitude, causing numerical instability.
- A regression model can be reparametrized into a standardized regression model through centering and rescaling.
- This process is called **standardization**, a.k.a. **correlation transformation**.

Correlation Transformation

Define transformed variables:

$$X_{ik}^{*} = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik}-\overline{X}_k}{s_{X_k}}\right), \quad k = 1, \cdots, p-1,$$

where

$$\overline{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ik}, \quad s_{X_k} = \sqrt{\frac{\sum_{i=1}^{n}(X_{ik}-\overline{X}_k)^2}{n-1}}, \quad (k = 1, \cdots, p-1).$$

are sample means and sample standard deviations, respectively.

- The sample means of the transformed variables are .
- The sample standard deviations of the transformed variables are .
- So all variables are and are .
- Correlation transformation the pairwise (sample) correlations among the X variables, the (sample) correlations between the X variables and the response variable.

- The sample means of the transformed variables are all zero.
- The sample standard deviations of the transformed variables are all $\frac{1}{\sqrt{n-1}}$.
- So all variables are centered and are on the same scale.
- Correlation transformation does not change the pairwise (sample) correlations among the X variables, nor does it change the (sample) correlations between the X variables and the response variable.

Standardized Regression Model

Rewrite the regression model in terms of standardized variables:

$$Y_i = \beta_0^* + \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \cdots + \beta_{p-1}^* X_{i,p-1}^* + \epsilon_i, \quad i = 1, \cdots n,$$

where

$$\beta_k^* = \qquad (k = 1, \cdots, p-1), \quad \beta_0^* =$$

is a “reparametrization” of the original model.

Standardized Regression Model

Rewrite the regression model in terms of standardized variables:

$$Y_i = \beta_0^* + \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \cdots + \beta_{p-1}^* X_{i,p-1}^* + \epsilon_i, \quad i = 1, \cdots n,$$

where

$$\beta_k^* = \sqrt{n-1} s_{X_k} \beta_k \;\; (k = 1, \cdots, p-1), \quad \beta_0^* = \beta_0 + \sum_{k=1}^{p-1} \beta_k \bar{X}_k$$

is a “reparametrization” of the original model.

Design Matrix of Standardized Model

$$\underset{n\times p}{\mathbf{X}^*} = \begin{bmatrix} 1 & X^*_{11} & \cdots & X^*_{1,p-1} \\ 1 & X^*_{21} & \cdots & X^*_{2,p-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X^*_{n1} & \cdots & X^*_{n,p-1} \end{bmatrix}.$$

$$\underset{p\times p}{\mathbf{X}^{*\prime}\mathbf{X}^*} = \begin{bmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & 1 & r_{12} & \cdots & r_{1,p-1} \\ 0 & r_{21} & 1 & \cdots & r_{2,p-1} \\ 0 & \vdots & \cdots & \vdots & \\ 0 & r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} n & \mathbf{0}^T \\ \mathbf{0} & \underset{(p-1)\times(p-1)}{\mathbf{r}_{XX}} \end{bmatrix},$$

where $\mathbf{r}_{XX}$ is the sample correlation matrix of the X variables.

Correlation Matrix

- Its (k, l)-element r_{kl} is the sample correlation coefficient between X_k, X_l:

- All its elements are numbers
  .
- Its diagonal elements are , since the correlation of a variable with itself is .
- Correlation matrix is a matrix:

Correlation Matrix

- Its (k, l)-element r_{kl} is the sample correlation coefficient between X_k, X_l:

$$r_{kl} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_{ik} - \overline{X}_k)(X_{il} - \overline{X}_l)}{s_{X_k} s_{X_l}}, \quad 1 \leq k, l \leq p - 1.$$

- All its elements are unit-less numbers in between -1 and 1.
- Its diagonal elements are all one, since the correlation of a variable with itself is one, i.e., $r_{kk} \equiv 1$ for $k = 1, \cdots, p - 1$.
- Correlation matrix is a symmetric matrix: $r_{kl} = r_{lk}$.

$\mathbf{X}'\mathbf{Y}$ Matrix of Standardized Model

$$\underset{p\times 1}{\mathbf{X}^{*\prime}\mathbf{Y}} = \begin{bmatrix} n\bar{Y} \\ \sqrt{n-1}s_Y r_{Y1} \\ \sqrt{n-1}s_Y r_{Y2} \\ \vdots \\ \sqrt{n-1}s_Y r_{Y,p-1} \end{bmatrix} = \sqrt{n-1}s_Y \begin{bmatrix} \frac{n}{\sqrt{n-1}s_Y}\bar{Y} \\ \underset{(p-1)\times 1}{\mathbf{r}_{XY}} \end{bmatrix}$$

where r_{Yk} is the sample correlation coefficient between Y and X_k:

$$r_{Yk} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_{ik}-\overline{X}_k)(Y_i-\overline{Y})}{s_{X_k}s_Y}, \quad k = 1, \cdots p-1.$$

LS Fit of Standardized Model

$$\underset{p\times 1}{\hat{\boldsymbol{\beta}}^*} = \begin{bmatrix} \hat{\beta}_0^* \\ \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \hat{\beta}_{p-1}^* \end{bmatrix} = \begin{bmatrix} \bar{Y} \\ \sqrt{n-1}\, s_Y \underset{(p-1)\times 1}{\mathbf{r}_{XX}^{-1}\mathbf{r}_{XY}}. \end{bmatrix}$$

- These are called *fitted standardized regression coefficients*.
- Relationships with the LS estimators of the original model:

$$\begin{aligned} \hat{\beta}_k &= \frac{1}{\sqrt{n-1}\, s_{X_k}} \hat{\beta}_k^*, \quad k = 1, \cdots, p-1 \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \cdots - \hat{\beta}_{p-1} \bar{X}_{p-1}. \end{aligned}$$

Do fitted values, residuals and sums of squares change due to standardization of the X variables?

Body Fat

Sample means and sample standard deviations ($n = 20$):

$$\overline{Y} = 20.20, \quad \overline{X}_1 = 25.30, \quad \overline{X}_2 = 51.17, \quad \overline{X}_3 = 27.62;$$

$$s_Y = 5.11, \quad s_{X_1} = 5.02, \quad s_{X_2} = 5.23, \quad s_{X_3} = 3.65.$$

Correlation matrices:

$$\mathbf{r}_{XX} = \begin{bmatrix} 1.00 & 0.92 & 0.46 \\ 0.92 & 1.00 & 0.08 \\ 0.46 & 0.08 & 1.00 \end{bmatrix}, \quad \mathbf{r}_{XY} = \begin{bmatrix} 0.84 \\ 0.88 \\ 0.14 \end{bmatrix}.$$

Least-squares estimators of the standardized model:

$$\hat{\beta}_0^* = \overline{Y} = 20.20, \quad \begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \hat{\beta}_3^* \end{bmatrix} = \sqrt{n-1}\, s_Y \mathbf{r}_{XX}^{-1} \mathbf{r}_{XY} = 27.5 \times \begin{bmatrix} 4.26 \\ -2.93 \\ -1.56 \end{bmatrix}.$$

Least-squares estimators of the original model:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 4.33 \\ -2.86 \\ -2.18 \end{bmatrix} = \begin{bmatrix} \frac{5.11}{5.02} \times 4.26 \\ \frac{5.11}{5.23} \times (-2.93) \\ \frac{5.11}{3.65} \times (-1.56) \end{bmatrix}.$$

Multicollinearity

Multicollinearity refers to the situation when the X variables are among themselves.

- This term is often reserved for the situation when the inter-correlation/collinearity among the X variables is .
- X variables being nearly collinear means

Multicollinearity

Multicollinearity refers to the situation when the X variables are **intercorrelated** among themselves.

- This term is often reserved for the situation when the inter-correlation/collinearity among the X variables is **very high**.
- X variables being nearly collinear/highly intercorrelated means that there exist constants $c_0, c_1, \cdots, c_{p-1}$ not all zero such that

$$c_0 + c_1 X_{i1} + \cdots + c_{p-1} X_{i,p-1} \approx 0, \quad i = 1, \cdots, n.$$

i.e., there exists a nonzero vector $\mathbf{c}$ such that $\underset{n\times p}{\mathbf{X}}\underset{p\times 1}{\mathbf{c}} \approx \mathbf{0}_n$.

- To understand the effects of multicollinearity, we consider two extreme situations:
  1. (i) When the X variables are not correlated with each other at all
  2. (ii) When they are perfectly intercorrelated.
- In practice, it is usually somewhere in between (i) and (ii).

Uncorrelated X Variables

- $\mathbf{r}_{XX} =$
- Fitted standardized regression coefficients:

$$\hat{\beta}_k^* = \qquad\qquad , \quad k = 1, \cdots, p-1$$

are the between the response variable Y and individual X variables.

- Variance-covariance matrix:

$$\sigma^2\{\begin{bmatrix} \hat{\beta}_0^* \\ \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \hat{\beta}_{p-1}^* \end{bmatrix}\} =$$

- So the LS estimators of the standardized model are . *How about the LS estimators of the original model?*

Uncorrelated X Variables

- $\mathbf{r}_{XX} = \mathbf{I}_{p-1}$
- Fitted standardized regression coefficients:

$$\hat{\beta}_k^* = \sqrt{n-1} s_Y \times r_{YX_k}, \quad k = 1, \cdots, p-1$$

are the sample correlation coefficients (up to a scaling factor) between the response variable Y and the respective X variables.

- Variance-covariance matrix:

$$\sigma^2 \{ \begin{bmatrix} \hat{\beta}_0^* \\ \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \hat{\beta}_{p-1}^* \end{bmatrix} \} = \sigma^2 (X^{*,T} X^*)^{-1} = \sigma^2 \begin{bmatrix} \frac{1}{n} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I}_{p-1} \end{bmatrix}.$$

- So the LS estimators of the standardized model are uncorrelated. *How about the LS estimators of the original model?*

When the X variables are uncorrelated, the effect of an X variable other X variables in the model.

- The LS fitted regression coefficient of an X variable is by which other (uncorrelated) X variables are in the model.
- The LS fitted regression coefficients of the X variables are with each other.
- The contribution of an X variable in reducing the error sum of squares is the other (uncorrelated) X variables in the model, i.e.

When the X variables are uncorrelated, the effect of an X variable does **not** depend on other X variables in the model.

- The LS fitted regression coefficient of an X variable is **not** affected by which other (uncorrelated) X variables are in the model.
- The LS fitted regression coefficients of the X variables are uncorrelated with each other.
- The contribution of an X variable in reducing the error sum of squares is the **same** with or without other (uncorrelated) X variables in the model, i.e.

$$SSR(X_j|X_{\mathcal{I}}) = SSR(X_j).$$

Notes: This is a strong advocate for controlled experiments, since there it may be possible to use an ***orthogonal design*** *where the levels of the X variables are chosen such that their sample correlations are (nearly) zero.*

Crew Productivity

A study on the effect of work crew size (X_1) and level of bonus pay (X_2) on productivity (Y).

case	X1 crew-size	X2 bonus-pay	Y productivity
1	4	2	42
2	4	2	39
3	4	3	48
4	4	3	51
5	6	2	49
6	6	2	53
7	6	3	61
8	6	3	60

Pairwise correlation matrix.

	X1	X2	Y
X1	1.00	0.00	0.74
X2	0.00	1.00	0.64
Y	0.74	0.64	1.00

X_1 and X_2 are uncorrelated.

Crew Productivity: Model 1

```
Call:
lm(formula = Y ~ X1, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-6.750 -3.750  0.125  4.500  6.000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   23.500     10.111   2.324   0.0591 .
X1             5.375      1.983   2.711   0.0351 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 5.609 on 6 degrees of freedom
Multiple R-squared: 0.5505,     Adjusted R-squared: 0.4755
F-statistic: 7.347 on 1 and 6 DF,  p-value: 0.03508

> anova(fit1)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
X1         1 231.12 231.125   7.347 0.03508 *
Residuals  6 188.75  31.458
```

Crew Productivity: Model 2

```
Call:
lm(formula = Y ~ X2, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-7.000 -4.688 -0.250  5.250  7.250

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   27.250     11.608   2.348   0.0572 .
X2             9.250      4.553   2.032   0.0885 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 6.439 on 6 degrees of freedom
Multiple R-squared: 0.4076,     Adjusted R-squared: 0.3088
F-statistic: 4.128 on 1 and 6 DF,  p-value: 0.08846

> anova(fit2)
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 .
Residuals  6 248.75  41.458
```

Crew Productivity: Model 3

```
Call:
lm(formula = Y ~ X1 + X2, data = data)

Residuals:
     1      2      3      4      5      6      7      8
 1.625 -1.375 -1.625  1.375 -2.125  1.875  0.625 -0.375

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.3750     4.7405   0.079 0.940016
X1            5.3750     0.6638   8.097 0.000466 ***
X2            9.2500     1.3276   6.968 0.000937 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.877 on 5 degrees of freedom
Multiple R-squared: 0.958,      Adjusted R-squared: 0.9412
F-statistic: 57.06 on 2 and 5 DF,  p-value: 0.000361

> anova(fit3)
Analysis of Variance Table

Response: Y
          Df  Sum Sq Mean Sq F value    Pr(>F)
X1         1 231.125 231.125  65.567 0.0004657 ***
X2         1 171.125 171.125  48.546 0.0009366 ***
Residuals  5  17.625   3.525
```

Perfectly Correlated X variables

A set of X variables is said to be *collinear* if one or several of them may be expressed as a linear combination of the other X variables (including $\mathbf{1}_n$).

- The design matrix $\underset{n\times p}{\mathbf{X}}$ is . So the matrix $\underset{p\times p}{\mathbf{X}'\mathbf{X}}$ is .
- LS estimators are because the least-squares equation

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

  has solutions.
- This means that there exist vectors $\mathbf{b}$ that minimize the least squares criterion:

$$Q(\mathbf{b}) \quad = \quad \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - \cdots - b_{p-1} X_{i,p-1})^2 \,.$$

Perfectly Correlated X variables

A set of X variables is said to be *collinear* if one or several of them may be expressed as a linear combination of the other X variables (including $\mathbf{1}_n$).

- The design matrix $\underset{n\times p}{\mathbf{X}}$ is not of full column rank: $rank(\mathbf{X}) < p$. So the matrix $\underset{p\times p}{\mathbf{X}'\mathbf{X}}$ is not invertible.
- LS estimators are not well-defined because the least-squares equation
$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$
has many solutions.
- This means that there exist many vectors $\mathbf{b}$ that minimize the least squares criterion:

$$Q(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - \cdots - b_{p-1} X_{i,p-1})^2 .$$

- If X variables are perfectly correlated, then there exists a nonzero vector $\underset{p\times 1}{\mathbf{c}}$ such that

$$\underset{n\times p}{\mathbf{X}}\underset{p\times 1}{\mathbf{c}} = \mathbf{0}_n.$$

- If $\mathbf{b}$ is a solution to the least-squares equation, i.e.,

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y},$$

  then $\mathbf{b} + k\mathbf{c}$ is also a solution where $k \in \mathbb{R}$ is an arbitrary scalar since

$$\begin{aligned} \mathbf{X}'\mathbf{X}\left(\mathbf{b} + k\mathbf{c}\right) &= \mathbf{X}'\mathbf{X}\mathbf{b} + k\mathbf{X}'\mathbf{X}\mathbf{c} \\ &= \mathbf{X}'\mathbf{Y} + k\mathbf{X}'\mathbf{0}_n = \mathbf{X}'\mathbf{Y}. \end{aligned}$$

- Similarly, if $\mathbf{b}$ minimizes the least-squares criterion function $Q(\cdot)$, then $\mathbf{b} + k\mathbf{c}$ also minimizes $Q(\cdot)$ since

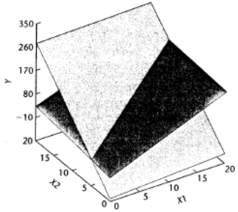
$$\begin{aligned} Q(\mathbf{b}) &= \left(\mathbf{Y} - \mathbf{X}\mathbf{b}\right)'\left(\mathbf{Y} - \mathbf{X}\mathbf{b}\right) \\ &= \left(\mathbf{Y} - \mathbf{X}(\mathbf{b} + k\mathbf{c})\right)'\left(\mathbf{Y} - \mathbf{X}(\mathbf{b} + k\mathbf{c})\right) = Q(\mathbf{b} + k\mathbf{c}). \end{aligned}$$

Example

case	X1	X2	Y
1	2	6	24
2	8	9	82
3	6	8	66
4	10	10	98

- X variables (including the column of 1) are perfectly correlated since $X_2 = 5 + 0.5X_1$.
- There are infinitely many response functions that fit this data equally "best" (with $SSE = 17.14$).

FIGURE 7.2 Two Response Planes That Intersect when $X_2 = 5 + .5X_1$.



- The two response surfaces in the figure are completely different, but they have the same y values on $X_2 = 5 + 0.5X_1$: $y = 7.14 + 9.29X_1$.
- Actually, any response surface that passes the intersecting line will fit the data equally well as these two, e.g.,

$$\widehat{Y} = 7.14 + 9.29X_1, \quad \widehat{Y} = -85.71 + 18.57X_2.$$

Can you think about some others?

```
Call:
lm(formula = Y ~ X1, data = data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.1429     3.5341   2.021  0.18066
X1            9.2857     0.4949  18.764  0.00283 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.928 on 2 degrees of freedom
Multiple R-squared: 0.9944,     Adjusted R-squared: 0.9915
F-statistic: 352.1 on 1 and 2 DF,  p-value: 0.002828

Call:
lm(formula = Y ~ X2, data = data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -85.7143     8.2956  -10.33  0.00924 **
X2           18.5714     0.9897   18.76  0.00283 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.928 on 2 degrees of freedom
Multiple R-squared: 0.9944,     Adjusted R-squared: 0.9915
F-statistic: 352.1 on 1 and 2 DF,  p-value: 0.002828
```

```
Call:
lm(formula = Y ~ X1 + X2, data = data)

Residuals:
      1       2       3       4
-1.7143  0.5714  3.1429 -2.0000

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.1429     3.5341   2.021  0.18066
X1            9.2857     0.4949  18.764  0.00283 **
X2                NA         NA      NA       NA
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.928 on 2 degrees of freedom
Multiple R-squared: 0.9944,     Adjusted R-squared: 0.9915
F-statistic: 352.1 on 1 and 2 DF,  p-value: 0.002828
```

Here, R discards X_2 and fits a model only using X_1.

When X variables are perfectly correlated, we may still get a fit of the data.

- The least-squares fitted values $\widehat{\mathbf{Y}}$ is and is the

  of the response vector $\mathbf{Y}$ to the linear subspace of $\mathbb{R}^n$ generated by the columns of the design matrix $\mathbf{X}$ (the column space).
- However, the regression coefficients are

  .

When X variables are perfectly correlated, we may still get a good fit of the data.

- The least-squares fitted values $\widehat{\mathbf{Y}}$ is uniquely defined and is the orthogonal projection of the response vector $\mathbf{Y}$ to the linear subspace of $\mathbb{R}^n$ generated by the columns of the design matrix $\mathbf{X}$ (the column space).
- However, the regression coefficients are not meaningful anymore without additional constraints.