STAT22200 Spring 2014 Lecture 1B

Yibi Huang

April 3, 2014

- Linear Regression Models in Matrix Representation
- Random Vectors
- t-Tests on Individual Regression Coefficients
- F-Tests on Multiple Regression Coefficients
- Regression Model Without Intercept

Linear Regression Models in Matrix Representation

The simple linear regression model $Y_i = \alpha + \beta X_i + \varepsilon_i$, $i = 1, \cdots, n$ is a shorthand for n linear relationships

$$\begin{aligned} Y_1 &= \alpha + \beta X_1 + \varepsilon_1 \\ Y_2 &= \alpha + \beta X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \alpha + \beta X_n + \varepsilon_n \end{aligned}$$

which can be written as

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

because, by matrix multiplication,

$$\begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \times \alpha + X_1 \times \beta \\ 1 \times \alpha + X_2 \times \beta \\ \vdots \\ 1 \times \alpha + X_n \times \beta \end{pmatrix} = \begin{pmatrix} \alpha + X_1\beta \\ \alpha + X_2\beta \\ \vdots \\ \alpha + X_n\beta \end{pmatrix}$$

For a multiple linear regression model

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj} + \varepsilon_j,$$

the matrix representation is

$$\overbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}^{Y} = \overbrace{\begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}}^{X} \overbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}^{\beta} + \overbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}^{\varepsilon}$$

dimensions: $[n \times 1]$ $\quad$ $[n \times (p+1)]$ $\quad$ $[(p+1) \times 1]$ $\quad$ $[n \times 1]$

This is often written as

$$Y = X\beta + \varepsilon$$

for short and X is often called the *model matrix* (or the *design matrix*).

How to View the Design Matrix In R?

In R, one can use the `model.matrix()` command to view the design matrix.

```
> lm1 = lm(Price ~ FLR+RMS+BDR+GAR+LOT+ST+CON+LOC, data = housing)
> model.matrix(lm1)
   (Intercept)  FLR RMS BDR GAR LOT ST CON LOC
1            1  967   5   2 0.0  39  0   1   0
2            1  815   5   2 2.0  33  0   1   0
3            1  900   5   3 1.0  35  1   1   0
4            1 1007   6   3 2.0  24  1   0   0
5            1 1100   7   3 1.5  50  0   1   0
6            1  897   7   4 1.0  25  0   0   0
7            1 1400   8   5 1.0  30  0   0   0
8            1 2261   6   3 2.0  29  0   0   0
9            1 1290   8   4 1.5  33  1   1   0
10           1 2104   9   4 1.0  40  0   1   0
11           1 2240  12   8 2.0  50  1   0   0
12           1  641   5   2 0.0  25  0   0   1
...
... (omitted)
...
26           1 1023   7   3 1.0  30  1   1   0
attr(,"assign")
[1] 0 1 2 3 4 5 6 7 8
```

The explanatory variables are expressed in matrix notation as column vectors

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{\mathbf{Y}} = \beta_0 \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{\mathbf{1}} + \beta_1 \underbrace{\begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{n1} \end{pmatrix}}_{\mathbf{X}_1} + \beta_2 \underbrace{\begin{pmatrix} X_{12} \\ X_{22} \\ \vdots \\ X_{n2} \end{pmatrix}}_{\mathbf{X}_2} + \cdots + \beta_p \underbrace{\begin{pmatrix} X_{1p} \\ X_{2p} \\ \vdots \\ X_{np} \end{pmatrix}}_{\mathbf{X}_p} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}_{\varepsilon}$$

The linear regression model can also be written as

$$\mathbf{Y} = \beta_0 \mathbf{1} + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \cdots + \beta_p \mathbf{X}_p + \varepsilon.$$

Note that the $(p+1)$ vectors above simply the column vectors of the design matrix X

$$X = (\mathbf{1}\ X_1\ X_2\ \cdots\ X_p).$$

Least Squares Method In Matrix Notation

- The sum of squares $\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_p X_{ip})^2$ can be written as

$$SSE = (Y - X\widehat{\beta})^T (Y - X\widehat{\beta})$$

- The normal equations can be written as:

$$X^T X \widehat{\beta} = X^T Y$$

- Least squares estimate for β:

$$\widehat{\beta} = (X^T X)^{-1} X^T Y$$

- Predicted Value for $\widehat{Y}$:

$$\widehat{Y} = X\widehat{\beta} = X(X^T X)^{-1} X^T Y$$

- Residuals:

$$e = Y - \widehat{Y} = (I - X(X^T X)^{-1} X^T) Y$$

Random Vectors

- We say $\mathbf{U}$ is a *random vector* if $\mathbf{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \end{pmatrix}$ is a vector that each of its element U_i is a random variable.
- The *expected value of a random vector* $\mathbf{U}$, denoted as $\mathbb{E}[\mathbf{U}]$, is defined as the vector $\mathbb{E}[\mathbf{U}] = \begin{pmatrix} \mathbb{E}[U_1] \\ \mathbb{E}[U_2] \\ \vdots \\ \mathbb{E}[U_k] \end{pmatrix}$
- Linear property of expected value: $\mathbb{E}[\mathbf{AU} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{U}] + \mathbf{b}$, in which $\mathbf{U}_{n\times 1}$ is a random vector, $\mathbf{A}_{m\times n}$ and $\mathbf{b}_{n\times 1}$ are constant matrices.

  Recall the expected value of random variables has a similar linear property $\mathbb{E}[aU + b] = a\mathbb{E}[U] + b$ in which U is a random variable and a, b are constants.

Least Squared Estimate for β is Unbiased

Observed that the least squares estimate for β is a random vector of length $p+1$

$$\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_p \end{pmatrix} = (X^T X)^{-1} X^T Y$$

Since X is a constant matrix (and so does the matrix $(X^T X)^{-1} X^T$), by the linear property of expected values, we can see that

$$\begin{aligned} \mathbb{E}[\widehat{\beta}] &= (X^T X)^{-1} X^T \mathbb{E}[Y] \\ &= (X^T X)^{-1} X^T \mathbb{E}(X\beta + \varepsilon) \\ &= (X^T X)^{-1} X^T (X\beta + \underbrace{\mathbb{E}[\varepsilon]}_{=0}) \\ &= (X^T X)^{-1} (X^T X) \beta = \beta. \end{aligned}$$

So $\widehat{\beta}$ is an unbiased estimate for β.

Variance-Covariance Matrix of a Random Vector

For any random vector $\mathbf{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \end{pmatrix}$, the **variance-covariance matrix** is defined to be a $k \times k$ matrix, with the (i, j)th entry being $\mathrm{Cov}(U_i, U_j)$.

$$\mathrm{Var}(\mathbf{U}) = \begin{pmatrix} \mathrm{Var}(U_1) & \mathrm{Cov}(U_1, U_2) & \mathrm{Cov}(U_1, U_3) & \cdots & \mathrm{Cov}(U_1, U_k) \\ \mathrm{Cov}(U_2, U_1) & \mathrm{Var}(U_2) & \mathrm{Cov}(U_2, U_3) & \cdots & \mathrm{Cov}(U_2, U_k) \\ \mathrm{Cov}(U_3, U_1) & \mathrm{Cov}(U_3, U_2) & \mathrm{Var}(U_3) & \cdots & \mathrm{Cov}(U_3, U_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(U_k, U_1) & \mathrm{Cov}(U_k, U_2) & \mathrm{Cov}(U_k, U_3) & \cdots & \mathrm{Var}(U_k) \end{pmatrix}.$$

Observe

- the diagonal elements are $\mathrm{Cov}(U_i, U_i) = \mathrm{Var}(U_i) \geq 0$.
- the variance-covariance matrix is symmetric since $\mathrm{Cov}(U_i, U_j) = \mathrm{Cov}(U_j, U_i)$.

Variance-Covariance Matrix for $\widehat{\beta}$

It can be shown that the variance-covariance matrix for $\widehat{\beta}$ is

$$\mathrm{Var}(\widehat{\beta}) = \sigma^2 (X^T X)^{-1}$$

So $\mathrm{Var}(\widehat{\beta}_j) = \sigma^2 \times$ the j^{th} diagonal element of the matrix $(X^T X)^{-1}$ and hence

$$\begin{aligned} SD(\widehat{\beta}_j) &= \sqrt{\mathrm{Var}(\widehat{\beta}_j)} \\ &= \sqrt{\sigma^2 \times \text{ the } j^{th} \text{ diagonal element of the matrix } (X^T X)^{-1}} \end{aligned}$$

in which the unknown σ^2 is estimated by MSE.
So the **standard error** for $\widehat{\beta}_j$ is

$$\begin{aligned} SE(\widehat{\beta}_j) &= \widehat{SD(\widehat{\beta}_j)} \\ &= \sqrt{MSE \times \text{the } j^{th} \text{ diagonal element of the matrix } (X^T X)^{-1}} \end{aligned}$$

Don't worry about the formulas.

R and other statistical software can crunch all the numbers for us.

One Sample t-Test (Review)

Given a sample of size n, $Y_1, Y_2, \ldots, Y_n$, from some (normal) population with unknown mean μ and unknown variance σ^2. Want to test

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu \neq \mu_0$$

The t-statistic is

$$t = \frac{\overline{Y} - \mu_0}{\mathrm{SE}(\overline{Y})} = \frac{\overline{Y} - \mu_0}{s/\sqrt{n}} \quad \text{where } s = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}}$$

If the population is normal, the t-statistic has a t-distribution with $n - 1$ degrees of freedom

t-Tests on Individual Regression Coefficients

For a MLR model,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i$$

To test the hypotheses,

$$H_0 : \beta_j = 0 \quad \text{v.s.} \quad H_a : \beta_j \neq 0$$

the t-statistic is

$$t = \frac{\widehat{\beta}_j}{\text{SE}(\widehat{\beta}_j)}$$

in which the $\text{SE}(\widehat{\beta}_j)$ is defined on page 10 in this slide.
This t-statistic also has a t-distribution with $n - p - 1$ degrees of freedom (where $p + 1$ is the number of parameter β's)

```
> lm1 = lm(Price ~ FLR+RMS+BDR+GAR+LOT+ST+CON+LOC, data=housing)
> summary(lm1)
Call:
lm(formula = Price ~ FLR + RMS + BDR + GAR + LOT + ST + CON +
    LOC, data = housing)

Residuals:
   Min     1Q Median     3Q    Max
-6.020 -2.129 -0.213  2.147  6.492

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.479799   4.443985   2.808 0.012094 *
FLR          0.017038   0.002751   6.195  9.8e-06 ***
RMS          2.382638   1.418290   1.680 0.111251
BDR         -4.543550   1.781145  -2.551 0.020671 *
GAR          5.078729   1.209692   4.198 0.000604 ***
LOT          0.382411   0.106832   3.580 0.002309 **
ST           9.827572   1.929232   5.094  9.0e-05 ***
CON          4.865071   1.890718   2.573 0.019746 *
LOC          6.957007   2.044084   3.403 0.003382 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.021 on 17 degrees of freedom
Multiple R-squared: 0.9305,     Adjusted R-squared: 0.8979
F-statistic: 28.47 on 8 and 17 DF,  p-value: 2.25e-08
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.479799   4.443985   2.808 0.012094 *
FLR          0.017038   0.002751   6.195  9.8e-06 ***
RMS          2.382638   1.418290   1.680 0.111251
......(some rows are omitted)
LOC          6.957007   2.044084   3.403 0.003382 **
---
```

- first column = variable name
- the column `Estimate` gives the LS estimate $\widehat{\beta}_j$'s
- the column `Std.Error` gives SE($\widehat{\beta}_j$), the standard error of $\widehat{\beta}_j$
- the column `t value` gives t-value $= \frac{\widehat{\beta}_j}{SE(\widehat{\beta}_j)}$
- column `Pr(>|t|)` gives the P-value for testing H_0: $\beta_j = 0$ v.s. H_a: $\beta_j \neq 0$.

E.g., from the row for the variable RMS, we know $\widehat{\beta}_{RMS} = 2.383$, with SE($\widehat{\beta}_{RMS}$) $= 1.418$. The t-value 1.680 is the ratio of the estimate and the SE 2.383/1.418. The P-value 0.111 says that the term RMS is not significant at 5% level to predict housing prices.

Types of Tests for MLR

The t-test can only tests for a single coefficient. One may also want to test for multiple coefficients.
Say for the following MLR model with 6 covariates

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_6 X_6 + \varepsilon,$$

one may want to test

1. all the regression coefficients are equal to zero
   H_0: $\beta_1 = \beta_2 = \cdots = \beta_6 = 0$ v.s. H_a: not all $\beta_1 \ldots, \beta_6 = 0$
2. some subset of the regression coefficients are equal to zero
   e.g. H_0: $\beta_2 = \beta_3 = \beta_5 = 0$ v.s.
   H_a: at least one of $\beta_2, \beta_3, \beta_5 \neq 0$
3. some of the regression coefficients are equal to each other
   e.g. H_0: $\beta_2 = \beta_3 = \beta_5$ v.s.
   H_a: β_2, β_3 and β_5 are not all equal
4. the regression coefficients satisfy certain specified constraints
   e.g. $H_0 : \beta_2 = \beta_3 + \beta_5$ v.s. H_a: $\beta_2 \neq \beta_3 + \beta_5$

Tests on Multiple Coefficients are Model Comparisons

Each of the four tests in the previous slide can be viewed as **a comparison of 2 models**, a **full model** and a **reduced model**.

- Testing $\beta_1 = \beta_2 = \ldots = \beta_6 = 0$

$$\begin{aligned} \text{Full}: & Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_6 X_6 + \varepsilon \\ \text{Reduced}: & Y = \beta_0 + \varepsilon \text{ (All covariates are redundant)} \end{aligned}$$

- Testing $\beta_2 = \beta_3 = \beta_5 = 0$

$$\begin{aligned} \text{Full}: & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \\ \text{Reduced}: & Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \beta_6 X_6 + \varepsilon \text{ } (X_2, X_3, X_5 \text{ are redundant}) \end{aligned}$$

- Testing $\beta_2 = \beta_3 = \beta_5$

$$\begin{aligned} \text{Full}: & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \\ \text{Reduced}: & Y = \beta_0 + \beta_1 X_1 + \beta_2 (X_2 + X_3 + X_5) + \beta_4 X_4 + \beta_6 X_6 + \varepsilon \end{aligned}$$

Tests on Multiple Coefficients are Model Comparisons (2)

- Testing $\beta_2 = \beta_3 + \beta_5$

$$\begin{aligned}
\text{Full}: \ Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \\
&\quad + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \\
\text{Reduced}: \ Y &= \beta_0 + \beta_1 X_1 + (\beta_3 + \beta_5) X_2 + \beta_3 X_3 + \\
&\quad + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon \\
&= \beta_0 + \beta_1 X_1 + \beta_3 (X_2 + X_3) + \beta_4 X_4 + \\
&\quad + \beta_5 (X_2 + X_5) + \beta_6 X_6 + \varepsilon
\end{aligned}$$

Observed the reduced model is always *nested* in the full model.

General Framework for Testing Nested Models

$$H_0\text{: the reduced model is true} \quad \text{v.s.}$$
$$H_a\text{ : the full model is true}$$

- As the reduced model is nested in the full model, it is always true that
$$SSE_{reduced} \geq SSE_{full}$$
- Trade-off: Simplicity v.s. Precision
  - Full model fits the data better (with smaller SSE) but is more complicate
  - Reduced model doesn't fit as well but is simpler.
- How to choose between the full and the reduced models?
  - If $\text{SSE}_{reduced} \approx \text{SSE}_{full}$, one can sacrifice a bit precision in exchange for simplicity
  - If $\text{SSE}_{reduced} \gg \text{SSE}_{full}$, it costs a great reduction in precision in exchange for simplicity. We will have to choose the full model.

The F-Statistic

$$F = \frac{(SSE_{reduced} - SSE_{full})/(df_{reduced} - df_{full})}{SSE_{full}/df_{full}}$$

- $SSE_{reduced} - SSE_{full}$ is the reduction in SSE from by replacing the reduced model with the full model.
- $df_{full}/df_{reduced}$ is the dfE for the full/reduced model.
- The denominator is MSE for full model
- $F \geq 0$ since $SSE_{reduced} \geq SSE_{full} \geq 0$
- The smaller the F-statistic, the more we favor the reduced model
- Under H_0, the F-statistic has an F-distribution with $df_{reduced} - df_{full}$ and df_{full} degrees of freedom.

Testing All Coefficients Equal Zero

Testing the hypotheses

$$H_0\colon\ \beta_1 = \cdots = \beta_p = 0 \text{ v.s. } H_a\colon \text{ not all } \beta_1 \ldots, \beta_p = 0$$

is a test to evaluate the **overall significance** of a model.

$$\text{Full}: Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon$$
$$\text{Reduced}: Y = \beta_0 + \varepsilon \quad \text{(all covariates are unnecessary)}$$

- The LS estimate for β_0 in the reduced model is $\widehat{\beta}_0 = \overline{Y}$, so

$$SSE_{reduced} = \sum_{i=1}^{n}(Y_i - \widehat{\beta}_0)^2 = \sum_{i}(Y_i - \overline{Y})^2 = SST_{full}$$

- $df_{reduced} = dfE_{reduced} = n - 1$,
  because the reduced model has only one coefficient β_0
- $df_{full} = dfE_{full} = n - p - 1$.

Testing All Coefficients Equal Zero

So

$$\begin{aligned} F &= \frac{(SSE_{reduced} - SSE_{full})/(df_{reduced} - df_{full})}{SSE_{full}/df_{full}} \\ &= \frac{(SST_{full} - SSE_{full})/[n-1-(n-p-1)]}{SSE_{full}/(n-p-1)} \\ &= \frac{SSR_{full}/p}{SSE_{full}/(n-p-1)}. \end{aligned}$$

Moreover, $F \sim F_{p,n-p-1}$ under H_0: $\beta_1 = \beta_2 = \cdots = \beta_p = 0$.

In R, the F statistic and p-value are displayed in the last line of the output of the `summary()` command.

```
> lm1 = lm(Price ~ FLR+RMS+BDR+GAR+LOT+ST+CON+LOC, data=housing)
> summary(lm1)
... (output omitted)

Residual standard error: 4.021 on 17 degrees of freedom
Multiple R-squared: 0.9305,     Adjusted R-squared: 0.8979
F-statistic: 28.47 on 8 and 17 DF,  p-value: 2.25e-08
```

ANOVA and the F-Test

The test of all coefficients equal zero is often summarized in an ANOVA table.

Source	df	Sum of Squares	Mean Squares	F
Regression	$dfR = p$	SSR	$MSR = \dfrac{SSR}{dfR}$	$F = \dfrac{MSR}{MSE}$
Error	$dfE = n - p - 1$	SSE	$SSE = \dfrac{SSE}{dfE}$	
Total	$dfT = n - 1$	SST		

ANOVA is the shorthand for **analysis of variance**.
It decomposes the total variation in the response (SST) into separate pieces that correspond to different sources of variation, like $SST = SSR + SSE$ in the regression setting.
Throughout STAT222, we will introduce several other ANOVA tables.

Testing Some Coefficients Equal to Zero

Say in the housing price example, want to test whether RMS and CON as a group are important, i.e., H_0: $\beta_{RMS} = \beta_{CON} = 0$.

```
> lm1 = lm(Price ~ FLR+RMS+BDR+GAR+LOT+ST+CON+LOC, data=housing)
> lm3 = lm(Price ~ FLR+BDR+GAR+LOT+ST+LOC, data=housing)
> anova(lm3,lm1)
Analysis of Variance Table

Model 1: Price ~ FLR + BDR + GAR + LOT + ST + LOC
Model 2: Price ~ FLR + RMS + BDR + GAR + LOT + ST + CON + LOC
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     19 472.03
2     17 274.84  2    197.19 6.0985 0.01008 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Testing Equality of Coefficients (1)

$$\text{Full model}: \; Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

Example 1: want to test H_0: $\beta_1 = \beta_4$, then the reduced model is

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_1 X_4 + \varepsilon \\ &= \beta_0 + \beta_1 (X_1 + X_4) + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \end{aligned}$$

1. Make a new variable $W = X_1 + X_4$
2. Fit the reduced model by regressing Y on W, X_2, and X_3
3. Find $\text{SSE}_{reduced}$ and $\text{df}_{reduced} - \text{df}_{full} =$ __1__
4. Can be done in R as follows:

```
> lmfull = lm(Y ~ X1 + X2 + X3 + X4)
> lmreduced = lm(Y ~ I(X1 + X4) + X2 + X3)
> anova(lmreduced,lmfull)
```

The line `lmreduced = lm(Y ~ I(X1 + X4) + X2 + X3)` is equivalent to

```
> W = X1 + X4
> lmreduced = lm(Y ~ W + X2 + X3)
```

Testing Equality of Coefficients (2)

Example 2: want to test H_0: $\beta_1 = \beta_2 = \beta_3$, then the reduced model is

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \beta_1 X_3 + \beta_4 X_4 + \varepsilon \\ &= \beta_0 + \beta_1 (X_1 + X_2 + X_3) + \beta_4 X_4 + \varepsilon \end{aligned}$$

1. Make a new variable $W = X_1 + X_2 + X_3$
2. Fit the reduced model by regressing Y on W and X_4
3. Find $\text{SSE}_{reduced}$ and $\text{df}_{reduced} - \text{df}_{full} =$ __2__
4. Can be done in R as follows

```
> lmfull = lm(Y ~ X1 + X2 + X3 + X4)
> lmreduced = lm(Y ~ I(X1 + X2 + X3) + X4)
> anova(lmreduced, lmfull)
```

Testing Equality of Coefficients (3)

Example 3: want to test H_0: $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$, then the reduced model is

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_1 X_2 + \beta_3 X_3 + \beta_3 X_4 + \varepsilon \\ &= \beta_0 + \beta_1 (X_1 + X_2) + \beta_3 (X_3 + X_4) + \varepsilon \end{aligned}$$

1. Make new variables $W_1 = X_1 + X_2$, $W_2 = X_3 + X_4$
2. Fit the reduced model by regressing Y on W_1 and W_2
3. Find $\text{SSE}_{reduced}$ and $\text{df}_{reduced} - \text{df}_{full} = \underline{\ 2\ }$
4. Can be done in R as follows

```
> lmfull = lm(Y ~ X1 + X2 + X3 + X4)
> lmreduced = lm(Y ~ I(X1 + X2) + I(X3 + X4))
> anova(lmreduced, lmfull)
```

Testing Coefficients under Constraints (1)

Again say the full model is

$$\text{Full model}: \; Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

Example 4: If H_0: $\beta_2 = \beta_3 + \beta_4$, then the reduced model is

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + (\beta_3 + \beta_4) X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \\ &= \beta_0 + \beta_1 X_1 + \beta_3 (X_2 + X_3) + \beta_4 (X_2 + X_4) + \varepsilon \end{aligned}$$

1. Make new variables $W_1 = X_2 + X_3$, $W_2 = X_2 + X_4$
2. Fit the reduced model by regressing Y on X_1, W_1 and W_2
3. Find $SSE_{reduced}$ and $df_{reduced} - df_{full} = \underline{\;1\;}$
4. Can be done in R as follows

```
> lmfull = lm(Y ~ X1 + X2 + X3 + X4)
> lmreduced = lm(Y ~ X1 + I(X2 + X3) + I(X2 + X4))
> anova(lmreduced, lmfull)
```

Testing Coefficients under Constraints (2)

Example 5: If we think $\beta_2 = 2\beta_1$, then the reduced model is

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \\ &= \beta_0 + \beta_1 X_1 + 2\beta_1 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \\ &= \beta_0 + \beta_1 (X_1 + 2X_2) + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \end{aligned}$$

1. Make a new variable $W = X_1 + 2X_2$
2. Fit the reduced model by regressing Y on W, X_3 and X_4
3. Find $\text{SSE}_{reduced}$ and $\text{df}_{reduced} - \text{df}_{full} =$ <u>1</u>
4. Can be done in R as follows

```
> lmfull = lm(Y ~ X1 + X2 + X3 + X4)
> lmreduced = lm(Y ~ I(X1 + 2*X2) + X3 + X4)
> anova(lmreduced, lmfull)
```

Relationship Between t-tests and F-tests (Optional)

The F-test can also test for a single coefficient, and the result is equivalent to the t-test. E.g., if one wants to test a single coefficient $\beta_3 = 0$ in the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

one way is to do a t-test using the command `summary(lm(Y ~ X1 + X2 + X3 + X4))` and read the t-statistic and P-value for X_3. Alternatively, one can also be viewed as a model comparison between

$$\text{Full model}: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$
$$\text{Reduced model}: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \varepsilon$$

```
> anova(lm(Y ~ X1 + X2 + X3 + X4), lm(Y ~ X1 + X2 + X4))
```

One can show that the F-statistic $= (t\text{-statistic})^2$ and the P-values are the same, and thus the two tests are equivalent.

The proof involves complicate matrix algebra and is hence omitted.

Consider again the model

$$\begin{aligned}\text{Price} = \beta_0 + \beta_1\text{FLR} + \beta_2\text{RMS} + \beta_3\text{BDR} + \beta_4\text{GAR} \\ + \beta_5\text{LOT} + \beta_6\text{ST} + \beta_7\text{CON} + \beta_8\text{LOC} + \varepsilon\end{aligned}$$

for the housing price data and want to test $\beta_{BDR} = 0$.

From the output on Slide 13, we see the t-statistic for test $\beta_{BDR} = 0$ is -2.551 with P-value 0.020671.

If using an F-test,

```
> lmfull = lm(Price ~ FLR+RMS+BDR+GAR+LOT+ST+CON+LOC, data=housing)
> lmreduced = lm(Price ~ FLR+RMS+GAR+LOT+ST+CON+LOC, data=housing)
> anova(lmreduced,lmfull)
Analysis of Variance Table

  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     18 380.04
2     17 274.84  1     105.2 6.5072 0.02067 *
```

we see $t^2 = (-2.551)^2 = 6.507601 \approx 6.5072 = F$ (the subtle difference is due to rounding), and the P-value is also 0.02067.

Regression Model Without Intercept

It is possible to fit an MLR model with no intercept β_0

$$y_i = \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i$$

- can still be written in matrix notation as $Y = X\beta + \varepsilon$ in which

$$\overbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}^{Y} = \overbrace{\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}}^{X} \overbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}^{\beta} + \overbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}^{\varepsilon}$$

dimensions: $[n \times 1]$ $\quad [n \times p]$ $\quad [p \times 1]$ $\quad [n \times 1]$

- the least square estimate for β is also of the same form

$$\widehat{\beta} = (X^T X)^{-1} X^T Y$$

(but note the design matrix X doesn't have the **1** column.)

Sum of Squares Without Intercept

Without intercept, the identity $SST = SSR + SSE$ is still valid but with a different definition

- SST becomes $\sum_{i=1}^{n} y_i^2$
- SSR becomes $\sum_{i=1}^{n} \widehat{y}_i^2 = \sum_{i=1}^{n}(\widehat{\beta}_1 + \ldots + \widehat{\beta}_p x_{ip})^2$
- SSE becomes $\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \widehat{\beta}_1 - \ldots - \widehat{\beta}_p x_{ip})^2$

Degrees of freedoms also change.

$$dfT = n, \quad dfR = p, \quad dfE = n - p$$

The mean squares (MS) remain to be the sum of squares (SS) divided by the degrees of freedom (df), but the SS and df are defined differently as above.

In particular, $MSE = \dfrac{SSE}{n-p}$, not $\dfrac{SSE}{n-p-1}$

- Without intercept, it is no longer true that $\sum_i e_i = 0$ but $\sum_i e_i x_{ik} = 0$ still holds for $k = 1, \ldots, p$
- R command to fit without intercept:

```
lm(y ~ -1 + X1 + X2 + X3)
```

- For the F-test of multiple coefficient, if the full model has no intercept, the reduced model has no intercept either.
  In particular, for the test the **overall significance** of a model H_0: $\beta_1 = \cdots = \beta_p = 0$, the full model and reduced model are

$$\begin{aligned} \text{Full} : & y = \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon \\ \text{Reduced} : & y = \varepsilon \qquad (\text{not } Y = \beta_0 + \varepsilon) \end{aligned}$$

Upcoming in Lecture 3

- Dummy Variables