Fitting the Linear regression Model by Using Matrices: learn it by examples

(WMS) **Example 11.1** Fit the simple linear regression model to the $n = 5$ data points:

x	y
-2	0
-1	0
0	1
1	1
2	3

$$y_i = \beta_0 + x_i \beta_1 + \varepsilon_i \quad i = 1, \cdots, n$$

(a). Write the linear model in a matrix form: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$; i.e., what are $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ in this context? What assumption did we assume for $\boldsymbol{\epsilon}$?

$$\underset{\sim}{Y} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix} \qquad \underset{\sim}{X} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

we assume ε_i's $\overset{iid}{\sim} N(0, \sigma^2)$

unknown : $\underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ and σ^2

(b). Find $\mathbf{X}'\mathbf{X}$, $(\mathbf{X}'\mathbf{X})^{-1}$, $\mathbf{X}'\mathbf{Y}$, $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Are they symmetric? If a matrix $\mathbf{A}$ is symmetric, then $\mathbf{A}' = \mathbf{A}$. What is $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$? Further show that $(\mathbf{X}'\mathbf{Y})' = \mathbf{Y}'\mathbf{X}$. What is $(\mathbf{I}-\mathbf{M})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$? What is $\mathrm{tr}(\mathbf{I}-\mathbf{M})$?

- $$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}$$ $\Leftarrow$ diagnol matrix

- $$(X'X)^{-1} = \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{pmatrix}$$

if a matrix is not diagnol, its inverse is not easy to calculate.

Generally: $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is any 2 by 2 matrix

$$A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- $$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

- $$M = X(X'X)^{-1}X' = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \\ & \frac{1}{10} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 0.6 & 0.4 & 0.2 & 0 & -0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0 & 0.1 & 0.2 & 0.3 & 0.4 \\ -0.2 & 0 & 0.2 & 0.4 & 0.6 \end{pmatrix}$$

We say a matrix A is symmetric, if and only if $A = A'$

- So $(X'X)$, $(X'X)^{-1}$, M are all symmetric.

In fact, if A is symmetric and A^{-1} exists then A^{-1} is symmetric

- $(X'X)(X'X)^{-1} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} = I_2$

since $X'X$ is a 2 by 2 matrix

- $(X'Y)' = (5, 7)$

$$Y'X = (0,0,1,1,3)\begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = (5, 7)$$

So $(X'Y)' = Y'X$

Generally: $(AB)' = B'A'$, $(A')' = A$

- $(I-M)X(X'X)^{-1}$

$$= IX(X'X)^{-1} - MX(X'X)^{-1}$$

$$= X(X'X)^{-1} - X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}$$

$$= X(X'X)^{-1} - X(X'X)^{-1} = 0$$

- $tr(I-M)$

here I is a 5 by 5 Identity matrix

because M is 5×5

Generally:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \vdots & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

$$tr(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

$$tr(A-B) = tr(A) - tr(B)$$

$tr(AB) = tr(BA)$ provided AB, BA exist

hence $tr(I_5 - M) = tr(I_5) - tr(M)$

$$= 5 - tr\left(X(X'X)^{-1}X'\right) = 5 - tr\left((X'X)^{-1}X'X\right) = 5 - tr(I_2)$$

$$= 5 - 2 = 3$$

(c). Find the least-square estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. What is the fitted model?

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$= \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{pmatrix} \begin{pmatrix} 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.7 \end{pmatrix}$$

fitted model: $\hat{Y} = X\hat{\beta}$

or. $\hat{Y}_i = \hat{\beta}_0 + X_i \hat{\beta}_1$

$= 1 + 0.7 X_i$.

(d). The "hat" matrix $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Verify the following properties:

- $\mathbf{M}$ is symmetric; i.e., $\mathbf{M}' = \mathbf{M}$.
- $\mathbf{M}$ is idempotent; i.e., $\mathbf{M}^2 = \mathbf{M}$.
- $\mathbf{MX} = \mathbf{X}$.
- $(\mathbf{I} - \mathbf{M})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$.

$$M' = \left(\underbrace{X(X'X)^{-1}}_{A}\underbrace{X'}_{B}\right)'$$

$(AB)' = B'A'$

$$= (X')'\left(X(X'X)^{-1}\right)'$$

$(A')' = A$

$$= X\left[(X'X)^{-1}\right]'X'$$

because $(X'X)^{-1}$ is symmetric

$$= X(X'X)^{-1}X' = M$$

$$M^2 = X(X'X)^{-1}\underbrace{X'X}(X'X)^{-1}X'$$

$$= X(X'X)^{-1}\cancel{(X'X)}\cancel{(X'X)^{-1}}X'$$

$$= X(X'X)^{-1}X' = M$$

- $$MX = X\underbrace{(X'X)^{-1}X'X}_{I} = X$$

- $$(I - M)X = IX - MX = X - X = 0$$

(e). Instead of using a simple linear model, we fit the data using the model:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

Redo (a). Write the linear model in a matrix form: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$; i.e., what are $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\epsilon}$ in this context? and
(c). Find the least-square estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. What is the fitted model?

$$\underset{\sim}{Y} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix} \quad \underset{\sim}{X} = \begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_5 \end{pmatrix}$$

$$\hat{\beta} = (X'X)^{-1} X' Y$$

$$\text{TI-84} = \begin{pmatrix} 0.5714 \\ 0.7 \\ 0.2143 \end{pmatrix}$$

A summary page of last lecture: what we have observed from Example 11.1:

- For two matrices $\mathbf{A}$ and $\mathbf{B}$, generally $\mathbf{AB} \neq \mathbf{BA}$. But
$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \quad \text{and} \quad \mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}).$$
Further
$$\mathrm{tr}(\mathbf{A} - \mathbf{B}) = \mathrm{tr}(\mathbf{A}) - \mathrm{tr}(\mathbf{B}).$$
- $\mathbf{X}'\mathbf{X}$, $(\mathbf{X}'\mathbf{X})^{-1}$, $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ are symmetric matrices.
- If $\mathbf{I}_n$ is the n-dimensional identity matrix, $\mathbf{X}$ is of dimension $n \times p$, then $\mathrm{tr}(\mathbf{I}_n - \mathbf{M}) = \underset{\mathrm{tr}(I_n)}{\mathrm{tr}(\mathbf{I})} - \mathrm{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}(\mathbf{I}_p) = n - p$.
- $\mathbf{M}$ is symmetric; i.e., $\mathbf{M}' = \mathbf{M}$.
- $\mathbf{M}$ is idempotent; i.e., $\mathbf{M}^2 = \mathbf{M}$.
- $\mathbf{MX} = \mathbf{X}$.
- $(\mathbf{I} - \mathbf{M})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$.

A quick quiz: If $Z \sim N(0,1)$, $W \sim \chi^2(\nu)$, Z and W are independent, what is the distribution of
$$T = \frac{Z}{\sqrt{W/\nu}}$$

Now we know how to estimate $\boldsymbol{\beta}$. The estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Is it a good estimator? Unbiased? What are its sampling distribution? How can we make inference (confidence intervals, hypothesis testing)? To answer these questions, we need some knowledge to help us handle random vectors, more specifically, when these random vectors are from a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$. (Check this for Bivariate normal: https://www.geogebra.org/m/pO4JcWPz).

Generally: Let $\underset{\sim}{\varepsilon}$ be an n dimensional random vector

$$\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad n \times 1 \text{ random vector}$$

- $$E(\underset{\sim}{\varepsilon}) = \begin{pmatrix} E(\varepsilon_1) \\ \vdots \\ E(\varepsilon_n) \end{pmatrix} \quad n \times 1 \text{ vector}$$

- $$V(\underset{\sim}{\varepsilon}) = \begin{pmatrix} V(\varepsilon_1) & Cov(\varepsilon_1, \varepsilon_2) & Cov(\varepsilon_1, \varepsilon_3) & \cdots & Cov(\varepsilon_1, \varepsilon_n) \\ Cov(\varepsilon_2, \varepsilon_1) & V(\varepsilon_2) & Cov(\varepsilon_2, \varepsilon_3) & \cdots & Cov(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & & \ddots & \vdots \\ Cov(\varepsilon_n, \varepsilon_1) & Cov(\varepsilon_n, \varepsilon_2) & \cdots & & V(\varepsilon_n) \end{pmatrix}$$

$\rightarrow$ an $n \times n$ symmetric matrix

diagnol of $V(\underset{\sim}{\varepsilon})$ are variances of ε_i's

In univariate case.

$$E[a+bX] = a + bE[X]$$

$$V(a+bX) = b^2 V(X)$$

$$Cov(a+bX, c+dY) = Cov(X, Y)$$

For random vectors

$$E[\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\varepsilon}] = \underset{\sim}{a} + \underset{\sim}{B}E(\underset{\sim}{\varepsilon})$$

$$V(\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\varepsilon}) = \underset{\sim}{B}\, V(\underset{\sim}{\varepsilon})\, \underset{\sim}{B}'$$

$$Cov(\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\varepsilon}, \underset{\sim}{c} + \underset{\sim}{D}\underset{\sim}{\varepsilon}^*) = \underset{\sim}{B}\, Cov(\underset{\sim}{\varepsilon}, \underset{\sim}{\varepsilon}^*)\, \underset{\sim}{D}'$$

Most commonly-used multivariate distribution.

Multivariate Normal

$$N(\underset{\sim}{\mu}, \underset{\sim}{V})$$

- Totally decided. by $\underset{\sim}{\mu}$, $\underset{\sim}{V}$

$\uparrow$ mean vector, $\uparrow$ covariance matrix

If $\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ where ε_i's are iid $N(0, \sigma^2)$

then $\underset{\sim}{\varepsilon}$ is also jointly normal.

Further $E(\underset{\sim}{\varepsilon}) = \begin{pmatrix} E(\varepsilon_1) \\ \vdots \\ E(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \underset{\sim}{0}$

because $\varepsilon_i, \varepsilon_j$ are independent when $i \neq j$
$Cov(\varepsilon_i, \varepsilon_j) = 0$

$$V(\underset{\sim}{\varepsilon}) = \begin{pmatrix} V(\varepsilon_1) & Cov(\varepsilon_1, \varepsilon_2) & \cdots \\ \vdots & \ddots & \vdots \\ \cdots & \cdots & V(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$$

$$= \sigma^2 I_n$$

So $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$

Practice: Consider $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2)' \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_2)$ where Let

$$\mathbf{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \text{ and } \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

1. What is the distribution of $\mathbf{Y} = (Y_1, Y_2)' = \mathbf{a} + \mathbf{B}\boldsymbol{\epsilon}$? What is the distribution of Y_1? How about the distribution of Y_2?
2. What is the distribution of $\mathbf{Z} = (Z_1, Z_2)' = \mathbf{b} + \mathbf{A}\boldsymbol{\epsilon}$? What is the distribution of Z_1? How about the distribution of Z_2?
3. What is the distribution of $\mathbf{a}'\boldsymbol{\epsilon}$?
4. What is $\mathrm{Cov}(\mathbf{Y}, \mathbf{Z})$? Are $\mathbf{Y}$ and $\mathbf{Z}$ independent?

1. $\underset{\sim}{Y} = \underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\varepsilon}$, because $\underset{\sim}{\varepsilon}$ is normal $N(0, 6^2 I_n)$

linear transformation of $\underset{\sim}{\varepsilon}$ is also normal.

To find the exact distribution, for normal, all you need are mean, variance!

$$E[\underset{\sim}{Y}] = E[\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\varepsilon}] = \underset{\sim}{a} + \underset{\sim}{B} E[\underset{\sim}{\varepsilon}] = \underset{\sim}{a} + \underset{\sim}{B}\,\underset{\sim}{0} = \underset{\sim}{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$V[\underset{\sim}{Y}] = V[\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\varepsilon}] = \underset{\sim}{B}\, V[\underset{\sim}{\varepsilon}]\, \underset{\sim}{B}' = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} 6^2 I_2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}'$$

$$= 6^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 6^2 \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$

So $\underset{\sim}{Y} \sim N\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, 6^2 \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \right)$

2. Similarly $\underset{\sim}{Z} \sim N\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}, 6^2 \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \right)$

3 $\quad a'\underset{\sim}{\varepsilon} \sim N\left(0, \; 2\sigma^2\right)$

since $\quad E[a'\underset{\sim}{\varepsilon}] = a'E[\underset{\sim}{\varepsilon}] = (1, 1)\begin{pmatrix}0\\0\end{pmatrix} = 0$

$$V[a'\underset{\sim}{\varepsilon}] = a' \, V(\underset{\sim}{\varepsilon})(a')' = (1, 1) \, \sigma^2 I_2 \begin{pmatrix}1\\1\end{pmatrix}$$

$$= 2\sigma^2$$

4. $\quad \mathrm{Cov}(\underset{\sim}{Y}, \underset{\sim}{Z}) = \mathrm{Cov}(\underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\varepsilon}, \; \underset{\sim}{b} + \underset{\sim}{A}\underset{\sim}{\varepsilon})$

$$= \underset{\sim}{B} \, \mathrm{Cov}(\underset{\sim}{\varepsilon}, \underset{\sim}{\varepsilon}) \underset{\sim}{A}'$$

$$= \underset{\sim}{B} \, V(\underset{\sim}{\varepsilon}) \underset{\sim}{A}'$$

$$= \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} \sigma^2 I_2 \begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}$$

$$= \sigma^2 \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix} = \sigma^2\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}$$

$$= \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}$$

Note: $\underset{\sim}{Y}, \underset{\sim}{Z}$ are from the same $\underset{\sim}{\varepsilon}$

$\Rightarrow \underset{\sim}{Y}, \underset{\sim}{Z}$ are jointly normal

only because they are jointly normal

then $\mathrm{Cov}(\underset{\sim}{Y}, \underset{\sim}{Z}) = \underset{\sim}{0}$

$\Rightarrow \underset{\sim}{Y}$ and $\underset{\sim}{Z}$ are independent

We now prove that

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \text{Note } Y = X\beta + \varepsilon$$

$$= (X'X)^{-1}X'(X\beta + \varepsilon)$$

$$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon$$

$$= \beta + (X'X)^{-1}X'\varepsilon \qquad \text{where } \varepsilon \sim N(0, \sigma^2 I_n)$$

$\hat{\beta}$ is a linear transformation of ε

So $\hat{\beta}$ is also normal

$$E[\hat{\beta}] = \beta + (X'X)^{-1}X' E[\varepsilon] = \beta$$

$$V(\hat{\beta}) = (X'X)^{-1}X' V(\varepsilon) [\underbrace{(X'X)^{-1}}_{A}\underbrace{X'}_{B}]'$$

$(AB)' = B'A'$

$$= \sigma^2 (X'X)^{-1}X'(X')'\underbrace{((X'X)^{-1})'}_{\text{symmetric}}$$

$$= \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

So $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

What is the distribution of $\mathbf{a}'\widehat{\boldsymbol{\beta}}$ for $\mathbf{a} = (a_1, \ldots, a_p)'$?
What is the distribution of $\mathbf{a}'\widehat{\boldsymbol{\beta}} + \epsilon^*$, where $\epsilon^* \sim N(0, \sigma^2)$ and ϵ^* is independent with $\epsilon_1, \ldots, \epsilon_n$?

known $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$

- $a'\hat{\beta}$ is a linear transformation of normal random variable $\hat{\beta}$

so $a'\hat{\beta}$ is also normal

$$E[a'\hat{\beta}] = a' E[\hat{\beta}] = a'\beta$$

$$V(a'\hat{\beta}) = a' V(\hat{\beta}) (a')'$$

$$= a' \sigma^2 (X'X)^{-1} a$$

$$= \sigma^2 a' (X'X)^{-1} a$$

So $a'\hat{\beta} \sim N(a'\beta, \sigma^2 a'(X'X)^{-1} a)$

- $$\underset{\sim}{a}'\hat{\underset{\sim}{\beta}} \sim N\left(\underset{\sim}{a}'\underset{\sim}{\beta},\ \sigma^2 \underset{\sim}{a}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{a}\right)$$

$$\varepsilon^* \sim N(0, \sigma^2)$$

Noting ε^* is independent with $\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

and $\hat{\underset{\sim}{\beta}} = \underset{\sim}{\beta} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{\varepsilon}$ entirely depends on $\underset{\sim}{\varepsilon}$

So ε^* and $\hat{\underset{\sim}{\beta}}$ are independent

and thus ε^* and $\underset{\sim}{a}'\hat{\underset{\sim}{\beta}}$ are independent

Then $E(\varepsilon^* + \underset{\sim}{a}'\hat{\underset{\sim}{\beta}}) = 0 + \underset{\sim}{a}'\underset{\sim}{\beta}$

$$V(\varepsilon^* + \underset{\sim}{a}'\hat{\underset{\sim}{\beta}}) = V(\varepsilon^*) + V(\underset{\sim}{a}'\hat{\underset{\sim}{\beta}}) + \underbrace{Cov(\varepsilon^*, \underset{\sim}{a}'\hat{\underset{\sim}{\beta}})}_{=0}$$

$$= \sigma^2 + \sigma^2 \underset{\sim}{a}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{a}$$

$$= \sigma^2\{1 + \underset{\sim}{a}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{a}\}.$$

Thus $\underset{\sim}{a}'\hat{\underset{\sim}{\beta}} + \underset{\sim}{\varepsilon}^* \sim N\left(\underset{\sim}{a}'\hat{\underset{\sim}{\beta}},\ \sigma^2\{1 + \underset{\sim}{a}'(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{a}\}\right)$

What is the distribution of $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$? (Recall HW 4 problem 2 part 3).
Prove that $\mathbf{e}$ and $\widehat{\boldsymbol{\beta}}$ are independent.

$$
\begin{aligned}
e = Y - \hat{Y} &= X\beta + \varepsilon - X\hat{\beta} \\
&= X\beta + \varepsilon - X(X'X)^{-1}X'Y \\
&= X\beta + \varepsilon - X(X'X)^{-1}X'(X\beta + \varepsilon) \\
&= \cancel{X\beta} + \varepsilon - \cancel{X\underbrace{(X'X)^{-1}X'X}_{I_n}\beta} - X(X'X)^{-1}X'\varepsilon \\
&= \varepsilon - M\varepsilon = (I_n - M)\varepsilon
\end{aligned}
$$

since $\varepsilon \sim N(0, \sigma^2 I_n)$. $e = (I_n - M)\varepsilon$ is also normal

$$E[e] = (I_n - M)E(\varepsilon) = 0$$

$$
\begin{aligned}
V(e) &= (I_n - M)V(\varepsilon)(I_n - M)' \\
&= \sigma^2 (I_n - M)(I_n - M) \\
&= \sigma^2 (I_n^2 - MI_n - I_nM + M^2) \\
&= \sigma^2 (I_n - M - M + M) \\
&= \sigma^2 (I_n - M)
\end{aligned}
$$

both I and M are symmetric

$M^2 = M$

So $e \sim N(0, \sigma^2 (I_n - M))$

- to show $\underset{\sim}{e}$ and $\hat{Y}$ are independent.

we have $\underset{\sim}{e} = (I_n - M)\underset{\sim}{\varepsilon}$

$$\hat{\underset{\sim}{Y}} = \underset{\sim}{X}\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$$

$$= \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{\varepsilon}$$

$$= \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{M}\underset{\sim}{\varepsilon}$$

both of which are linear transformation of $\underset{\sim}{\varepsilon}$

recall the practice problem on Page 8

it suffices to show $Cov(\underset{\sim}{e}, \hat{\underset{\sim}{Y}}) = 0$

$$Cov(\underset{\sim}{e}, \hat{\underset{\sim}{Y}}) = Cov\left((I_n - M)\underset{\sim}{\varepsilon}, \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{M}\underset{\sim}{\varepsilon}\right)$$

$$= (I_n - \underset{\sim}{M})\, Cov(\underset{\sim}{\varepsilon}, \underset{\sim}{\varepsilon})\, \underset{\sim}{M}'$$

$$= (I_n - \underset{\sim}{M})\, \sigma^2 I_n\, \underset{\sim}{M}$$

$$= \sigma^2 (I_n - \underset{\sim}{M})\underset{\sim}{M}$$

$$= \sigma^2 (\underset{\sim}{M} - \underset{\sim}{M}^2)$$

$$= 0$$

A summary page of last lecture: what we have observed from Example 11.1:

- For two matrices $\mathbf{A}$ and $\mathbf{B}$, generally $\mathbf{AB} \neq \mathbf{BA}$. But

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \quad \text{and} \quad \text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}).$$

  Further

$$\text{tr}(\mathbf{A} - \mathbf{B}) = \text{tr}(\mathbf{A}) - \text{tr}(\mathbf{B}).$$

- $\mathbf{X}'\mathbf{X}$, $(\mathbf{X}'\mathbf{X})^{-1}$, $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ are symmetric matrices.
- If $\mathbf{I}_n$ is the n-dimensional identity matrix, $\mathbf{X}$ is of dimension $n \times p$, then $\text{tr}(\mathbf{I}_n - \mathbf{M}) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{tr}(\mathbf{I}_n) - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_p) = n - p$.
- $\mathbf{M}$ is symmetric; i.e., $\mathbf{M}' = \mathbf{M}$.
- $\mathbf{M}$ is idempotent; i.e., $\mathbf{M}^2 = \mathbf{M}$.
- $\mathbf{MX} = \mathbf{X}$.
- $(\mathbf{I} - \mathbf{M})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$.

<u>We have reviewed that:</u> If $Z \sim N(0,1)$, $W \sim \chi^2(\nu)$, Z and W are independent, what is the distribution of

$$T = \frac{Z}{\sqrt{W/\nu}}$$

Further, we proved that

- $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$
- For any $\mathbf{a} = (a_1, \ldots, a_p)'$, $\mathbf{a}'\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\mathbf{a}'\boldsymbol{\beta}, \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$
- And $\mathbf{a}'\widehat{\boldsymbol{\beta}} + \epsilon^* \sim \mathcal{N}(\mathbf{a}'\boldsymbol{\beta}, \sigma^2\{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\})$, where $\epsilon^* \sim N(0, \sigma^2)$ and ϵ^* is independent with $\epsilon_1, \ldots, \epsilon_n$
- Let $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ be the fitted values. The residual $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$. We have proved that $\mathbf{e} \sim \mathcal{N}(0,)$
- $\mathbf{e}$ and $\widehat{\boldsymbol{\beta}}$ are independent!

How to estimate σ^2? Recall that in

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$. Let $p = k+1$ be the length of $\boldsymbol{\beta}$. The length of Y is n, the dimension of $\mathbf{X}$ is $n \times p$. we assumed ϵ_i's are iid $N(0, \sigma^2)$. Now the question is how to estimate σ^2? Towards this end, we need a new term, called **error (residual) sum of squares**:

$$\text{SSE} = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e}.$$

(1) Prove that

$$\text{SSE} = \mathbf{Y}'(\mathbf{I} - \mathbf{M})\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}.$$

(2) Using the fact, that if $E(\mathbf{Y}) = \boldsymbol{\mu}$, $V(\mathbf{Y}) = \mathbf{V}$, then for any suitable matrix $\mathbf{A}$,

$$E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \text{tr}(\mathbf{A}\mathbf{V}),$$

and $\text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X})$ to prove that

$$E(\text{SSE}) = (n-p)\sigma^2.$$

Then a natural estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{\text{SSE}}{n-p}$$

In fact,

$$\frac{\text{SSE}}{\sigma^2} = \frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p).$$

(3) Find the value of $\widehat{\sigma}^2$ in Example 11.1

(1) $$e = Y - \hat{Y} = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = Y - MY = (I-M)Y$$

$$SSE = e'e = (Y - \hat{Y})'(Y - \hat{Y})$$

$$= \big(\underbrace{(I-M)}_{A}\underbrace{Y}_{B}\big)'(I-M)Y \qquad (AB)' = B'A'$$

$$= Y'\underbrace{(I-M)}_{\text{symmetric}}{}'(I-M)Y$$

$$= Y'(I-M)(I-M)Y \qquad (I-M)^2 = I-M$$

$$= Y'(I-M)Y$$

Further, $\quad SSE = \underset{\sim}{Y}'(I - \underline{M})\underline{Y}$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{Y}'\underline{M}\underset{\sim}{Y}$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - \underbrace{\underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}}_{\hat{\beta}'}\underline{X}'\underline{Y}$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - \hat{\underset{\sim}{\beta}}'\underline{X}'\underset{\sim}{Y}$$

(2) $\quad SSE = \underset{\sim}{Y}'(I_n - \underset{\sim}{M})\underset{\sim}{Y}$

$$E[SSE] = E\left[\underset{\sim}{Y}'(I_n - \underset{\sim}{M})\underset{\sim}{Y}\right]$$

$$\boxed{\begin{array}{l}\text{using the formula. if } E[\underset{\sim}{Y}] = \underset{\sim}{\mu},\ V[\underset{\sim}{Y}] = \underset{\sim}{V} \\ \text{then } E[\underset{\sim}{Y}'\underset{\sim}{A}\underset{\sim}{Y}] = \underset{\sim}{\mu}'\underset{\sim}{A}\underset{\sim}{\mu} + tr[\underset{\sim}{A}\underset{\sim}{V}]\end{array}}$$

take $A = I_n - \underset{\sim}{M}$

since $\underset{\sim}{Y} \sim N(\underset{\sim}{X}\underset{\sim}{\beta} + \sigma^2 I_n)$

$$\underset{\sim}{\mu} = E[\underset{\sim}{Y}] = \underline{X}\underset{\sim}{\beta},\ V = V[\underset{\sim}{Y}] = \sigma^2 I_n$$

So $E[Y'(I_n - M)Y]$

$$= (X\beta)' \underbrace{(I_n - M)X}_{0}\beta + tr[(I_n - M)\sigma^2 I_n]$$

$$= 0 + tr[(I_n - M)\sigma^2]$$

$$= \sigma^2 \, tr(I_n - M)$$

$$= \sigma^2 [tr(I_n) - tr(M)]$$

$$= \sigma^2 (n - tr(\underbrace{X}_{A}\underbrace{(X'X)^{-1}X'}_{B}))$$

$tr(AB) = tr(BA)$

$$= \sigma^2 (n - tr((X'X)^{-1}X'X))$$

note X is n by p, $(X'X)$ is a p by p matrix

So $(X'X)^{-1}X'X = I_p$

$$= \sigma^2 (n - tr(I_p))$$

$$= \sigma^2 (n - p) = E[SSE]$$

(3) in Example 11.1

$$\hat{\beta} = \begin{pmatrix} 1 \\ 0.7 \end{pmatrix} \quad \underset{\sim}{Y} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 3 \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$\hat{\underset{\sim}{Y}} = X\hat{\beta} = \begin{pmatrix} -0.4 \\ 0.3 \\ 1 \\ 1.7 \\ 2.4 \end{pmatrix}$$

$$\underset{\sim}{e} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} = \begin{pmatrix} 0.4 \\ -0.3 \\ 0 \\ -0.7 \\ 0.6 \end{pmatrix}$$

$$SSE = \underset{\sim}{e}'\underset{\sim}{e} = 1.6 + 0.9 + 0 + 4.9 + 3.6$$

$$= 11$$

$$\hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{11}{5-2} = \frac{11}{3} = 3.6667$$

Recall

- that

$$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

  and

$$\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

  are independent;

- further that the quiz problem: If $Z \sim N(0,1)$, $W \sim \chi^2(\nu)$, Z and W are independent,

$$T = \frac{Z}{\sqrt{W/\nu}} \sim T(\nu).$$

These give us an idea to make inference for regression parameters $\boldsymbol{\beta}$.

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$$

Let $$(X'X)^{-1} = \begin{pmatrix} C_{00} & C_{01} & \cdots & C_{0K} \\ C_{10} & C_{11} & \cdots & C_{1K} \\ & \vdots & \ddots & \\ C_{K1} & C_{K2} & \cdots & C_{KK} \end{pmatrix}$$

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}, \sigma^2 \begin{pmatrix} C_{00} & C_{01} & \cdots & C_{0K} \\ C_{10} & C_{11} & \cdots & C_{1K} \\ & \vdots & \ddots & \\ C_{K1} & C_{K2} & \cdots & C_{KK} \end{pmatrix} \right)$$

So for $j = 0, 1, 2, \ldots, K$,

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{jj})$$

- the standard error of $\hat{\beta}_j$ is $\sigma^2 C_{jj}$

where $j = 0, 1, \cdots, k$.

- $$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}} \sim N(0, 1)$$

Note that: $$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

Further, because $\hat{\beta}_j$ is a component of $\hat{\underset{\sim}{\beta}}$

$\hat{\sigma}^2$ is from $\underset{\sim}{e}$

where we have shown $\hat{\underset{\sim}{\beta}}$ and $\underset{\sim}{e}$ are independent

So $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}}$ and $\frac{(n-p)\hat{\sigma}^2}{\sigma^2}$ are independent

Thus
$$\frac{\dfrac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}}}{\sqrt{\dfrac{(n-p)\hat{\sigma}^2}{\sigma^2} \Big/ (n-p)}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \sim T(n-p)$$

- $100 \times (1-\alpha)\%$ CI for β_j

using $\dfrac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} \sim T(n-p)$

we have a $100 \times (1-\alpha)\%$ CI for β_j

is $\hat{\beta}_j \pm t_{n-p, \alpha/2} \sqrt{\hat{\sigma}^2 c_{jj}}$

- Hypothesis test:

two sided: $H_0: \beta_j = \beta_j^*$ vs $H_1: \beta_j \neq \beta_j^*$

$$RR = \left\{ |t| = \left| \frac{\hat{\beta}_j - \beta_j^*}{\sqrt{\hat{\sigma}^2 c_{jj}}} \right| > t_{n-p, \alpha/2} \right\}$$

one-side: $H_0: \beta_j = \beta_j^*$ vs $H_1: \beta_j > \beta_j^*$

$$RR = \left\{ t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} > t_{n-p, \alpha} \right\}$$

one side: $H_0: \beta_j = \beta_j^*$ vs $H_1: \beta_j < \beta_j^*$

$$RR = \left\{ t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} < -t_{n-p, \alpha} \right\}$$

Confidence interval: Similarly, recall that

$$\mathbf{a}'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{a}'\boldsymbol{\beta}, \sigma^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$$

and

$$\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p).$$

$E[Y$ at $x = x^*]$ $= a'\beta$ is a parameter

These tell us **a** $100(1-\alpha)\%$ **confidence interval for** $E(Y)$ when

$$\mathbf{x} = \mathbf{x}^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

is

$$\mathbf{a}'\widehat{\beta} \pm t_{n-p,\alpha/2}\sqrt{\widehat{\sigma}^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

where

$$\mathbf{a} = \begin{pmatrix} 1 \\ x_1^* \\ \vdots \\ x_k^* \end{pmatrix}.$$

Prediction interval: Similarly, recall that

$$\mathbf{a}'\widehat{\boldsymbol{\beta}} + \epsilon^* \sim N(\mathbf{a}'\boldsymbol{\beta}, \sigma^2\{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\}),$$

where ϵ^* is independent with $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$, and

$$\frac{(n-p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p).$$

$Y = a'\beta + \varepsilon$ is a new response with extra error $\varepsilon \sim N(0, \sigma^2)$

These tell us that **a** $100(1-\alpha)\%$ **prediction interval for** Y when

$$\mathbf{x} = \mathbf{x}^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

is

$$\mathbf{a}'\widehat{\beta} \pm t_{n-p,\alpha/2}\sqrt{\widehat{\sigma}^2\{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\}}$$

where

$$\mathbf{a} = \begin{pmatrix} 1 \\ x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

Example 11.1 (continued) Find a 90% confidence interval for $E(Y)$ where $x = 1$. FYI: $t_{2,0.05} = 2.920$, $t_{3,0.05} = 2.353$, $t_{3,0.1} = 1.638$.

at $x=1$, $a = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, (known: $\hat{\beta} = \begin{pmatrix} 1 \\ 0.7 \end{pmatrix}$, $(X'X)^{-1} = \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{pmatrix}$, $\hat{\sigma}^2 = 3.6667$)

$n=5$. $p=2$. $\alpha = 0.1$

an 90% CI for $E(Y)$ at $x=1$ is

$$\underset{\sim}{a}'\hat{\beta} \pm t_{n-p,\alpha/2} \sqrt{\hat{\sigma}^2 \, a'(X'X)^{-1}a}$$

$$= (1,1)\begin{pmatrix} 1 \\ 0.7 \end{pmatrix} \pm t_{3,0.05} \times \sqrt{3.6667 \times (1,1)\begin{bmatrix} \frac{1}{5} & \\ & \frac{1}{10} \end{bmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

$$= 1.7 \pm 2.353 \times \sqrt{3.6667 \times 0.3}$$

Example 11.1 (continued) Find a 90% prediction interval for Y at $x = 2$. FYI: $t_{2,0.05} = 2.920$, $t_{3,0.05} = 2.353$, $t_{3,0.1} = 1.638$.

$$a'\hat{\beta} \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2\{1 + a'(X'X)^{-1}a\}}$$

$$1.7 \pm 2.353 \times \sqrt{3.6667 \times \{1 + 0.3\}}$$

A summary page of last lecture: what we have observed from Example 11.1:

- For two matrices $\mathbf{A}$ and $\mathbf{B}$, generally $\mathbf{AB} \neq \mathbf{BA}$. But

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \quad \text{and} \quad \text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}).$$

  Further

$$\text{tr}(\mathbf{A} - \mathbf{B}) = \text{tr}(\mathbf{A}) - \text{tr}(\mathbf{B}).$$

- $\mathbf{X}'\mathbf{X}$, $(\mathbf{X}'\mathbf{X})^{-1}$, $\mathbf{M} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ are symmetric matrices.
- If $\mathbf{I}_n$ is the n-dimensional identity matrix, $\mathbf{X}$ is of dimension $n \times p$, then $\text{tr}(\mathbf{I}_n - \mathbf{M}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{tr}(\mathbf{I}_n) - \text{tr}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_p) = n - p$.
- $\mathbf{M}$ is symmetric; i.e., $\mathbf{M}' = \mathbf{M}$.
- $\mathbf{M}$ is idempotent; i.e., $\mathbf{M}^2 = \mathbf{M}$.
- $\mathbf{MX} = \mathbf{X}$.
- $(\mathbf{I} - \mathbf{M})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$.

<u>We have reviewed that:</u> If $Z \sim N(0, 1)$, $W \sim \chi^2(\nu)$, Z and W are independent, what is the distribution of

$$T = \frac{Z}{\sqrt{W/\nu}}$$

We have proved that

- $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, where

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \text{and we denote } (\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} c_{00} & c_{01} & c_{02} & \cdots & c_{0k} \\ c_{10} & c_{11} & c_{12} & \cdots & c_{1k} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ c_{k0} & c_{k1} & c_{k2} & \cdots & c_{kk} \end{pmatrix}$$

  Note that means $\widehat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$ for $j = 0, 1, \ldots, k$. The standard error of $\widehat{\beta}_j$ is $\sqrt{c_{jj}\sigma^2}$.

- For any $\mathbf{a} = (a_1, \ldots, a_p)'$, $\mathbf{a}'\widehat{\boldsymbol{\beta}} \sim \mathcal{N}(\mathbf{a}'\boldsymbol{\beta}, \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a})$.
- And $\mathbf{a}'\widehat{\boldsymbol{\beta}} + \epsilon^* \sim \mathcal{N}(\mathbf{a}'\boldsymbol{\beta}, \sigma^2\{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\})$, where $\epsilon^* \sim N(0, \sigma^2)$ and ϵ^* is independent with $\epsilon_1, \ldots, \epsilon_n$
- Let $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$ be the fitted values. The residual $\mathbf{e} = \mathbf{Y} - \widehat{\mathbf{Y}}$. We have proved that $\mathbf{e} \sim \mathcal{N}(0,)$
- $\mathbf{e}$ and $\widehat{\boldsymbol{\beta}}$ are independent!
- An estimator of σ^2 is

$$\widehat{\sigma}^2 = \frac{\text{SSE}}{n - p} = \frac{\mathbf{e}'\mathbf{e}}{n - p} \quad \text{and} \quad \frac{(n - p)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n - p).$$

Further, we have obtained the following results:

- A $100(1-\alpha)\%$ confidence interval of β_j is

$$\widehat{\beta}_j \pm t_{n-p,\alpha/2}\sqrt{c_{jj}\widehat{\sigma}^2}$$

- Reject the test $H_0 : \beta_j = \beta_j^*$ versus $H_1 : \beta_j \neq \beta_j^*$ if

$$|t| = \left| \frac{\widehat{\beta}_j - \beta_j^*}{\sqrt{c_{jj}\widehat{\sigma}^2}} \right| > t_{n-p,\alpha/2}$$

- Reject the test $H_0 : \beta_j = \beta_j^*$ versus $H_1 : \beta_j > \beta_j^*$ if

$$t = \frac{\widehat{\beta}_j - \beta_j^*}{\sqrt{c_{jj}\widehat{\sigma}^2}} > t_{n-p,\alpha}$$

- Reject the test $H_0 : \beta_j = \beta_j^*$ versus $H_1 : \beta_j < \beta_j^*$ if

$$t = \frac{\widehat{\beta}_j - \beta_j^*}{\sqrt{c_{jj}\widehat{\sigma}^2}} < -t_{n-p,\alpha}$$

- A $100(1-\alpha)\%$ confidence interval for $E(Y)$ when

$$\mathbf{x} = \mathbf{x}^* = \begin{pmatrix} x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

  is

$$\mathbf{a}'\widehat{\beta} \pm t_{n-p,\alpha/2}\sqrt{\widehat{\sigma}^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

  where

$$\mathbf{a} = \begin{pmatrix} 1 \\ x_1^* \\ \vdots \\ x_k^* \end{pmatrix}$$

- A $100(1-\alpha)\%$ prediction interval for Y when $\mathbf{x} = \mathbf{x}^*$ is

$$\mathbf{a}'\widehat{\beta} \pm t_{n-p,\alpha/2}\sqrt{\widehat{\sigma}^2\{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\}}.$$

Example 11.2. The taste of matured cheese is related to the concentration of several chemicals in the final product. In a study from the LaTrobe Valley of Victoria, Australia, samples of cheddar cheese were analyzed for their chemical composition and were subjected to taste tests. For each specimen, the taste Y was obtained by combining the scores from several tasters. Data were collected on the following variables:

$$\begin{aligned} Y &= \text{ taste score (}\texttt{TASTE}\text{)} \\ x_1 &= \text{ concentration of acetic acid (}\texttt{ACETIC}\text{)} \\ x_2 &= \text{ concentration of hydrogen sulfide (}\texttt{H2S}\text{)} \\ x_3 &= \text{ concentration of lactic acid (}\texttt{LACTIC}\text{).} \end{aligned}$$

Variables `ACETIC` and `H2S` were both measured on the log scale. The variable `LACTIC` has not been transformed. Table 11.2 contains concentrations of the various chemicals in $n = 30$ specimens of cheddar cheese and the observed taste score.

Specimen	TASTE	ACETIC	H2S	LACTIC	Specimen	TASTE	ACETIC	H2S	LACTIC
1	12.3	4.543	3.135	0.86	16	40.9	6.365	9.588	1.74
2	20.9	5.159	5.043	1.53	17	15.9	4.787	3.912	1.16
3	39.0	5.366	5.438	1.57	18	6.4	5.412	4.700	1.49
4	47.9	5.759	7.496	1.81	19	18.0	5.247	6.174	1.63
5	5.6	4.663	3.807	0.99	20	38.9	5.438	9.064	1.99
6	25.9	5.697	7.601	1.09	21	14.0	4.564	4.949	1.15
7	37.3	5.892	8.726	1.29	22	15.2	5.298	5.220	1.33
8	21.9	6.078	7.966	1.78	23	32.0	5.455	9.242	1.44
9	18.1	4.898	3.850	1.29	24	56.7	5.855	10.20	2.01
10	21.0	5.242	4.174	1.58	25	16.8	5.366	3.664	1.31
11	34.9	5.740	6.142	1.68	26	11.6	6.043	3.219	1.46
12	57.2	6.446	7.908	1.90	27	26.5	6.458	6.962	1.72
13	0.7	4.477	2.996	1.06	28	0.7	5.328	3.912	1.25
14	25.9	5.236	4.942	1.30	29	13.4	5.802	6.685	1.08
15	54.9	6.151	6.752	1.52	30	5.5	6.176	4.787	1.25

Table 11.2: Cheese data. `ACETIC`, `H2S`, and `LACTIC` are independent variables. The response variable is `TASTE`.

REGRESSION MODEL: Suppose the researchers postulate that each of the three chemical composition variables x_1, x_2, and x_3 is important in describing the taste. In this case, they might initially consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i,$$

for $i = 1, 2, ..., 30$. We now use R to fit this model using the method of least squares. Here is the output:

```
> summary(fit)
Call: lm(formula = taste ~ acetic + h2s + lactic)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -28.877     19.735  -1.463  0.15540
acetic         0.328      4.460   0.074  0.94193
h2s            3.912      1.248   3.133  0.00425 **
lactic        19.670      8.629   2.279  0.03109 *

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-squared: 0.6518,     Adjusted R-squared: 0.6116
F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.810e-06
```

OUTPUT: The `Estimate` output gives the values of the least squares estimates:

$$\widehat{\beta}_0 \approx -28.877 \qquad \widehat{\beta}_1 \approx 0.328 \qquad \widehat{\beta}_2 \approx 3.912 \qquad \widehat{\beta}_3 \approx 19.670.$$

Therefore, the fitted least squares regression model is

$$\widehat{Y} = -28.877 + 0.328x_1 + 3.912x_2 + 19.670x_3,$$

or, in other words,

$$\widehat{\texttt{TASTE}} = -28.877 + 0.328\texttt{ACETIC} + 3.912\texttt{H2S} + 19.670\texttt{LACTIC}.$$

The `Std.Error` output gives

$$\begin{aligned}
19.735 &= \widehat{\text{se}}(\widehat{\beta}_0) = \sqrt{c_{00}\widehat{\sigma}^2} = \sqrt{\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}_{00}} \\
4.460 &= \widehat{\text{se}}(\widehat{\beta}_1) = \sqrt{c_{11}\widehat{\sigma}^2} = \sqrt{\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}_{11}} \\
1.248 &= \widehat{\text{se}}(\widehat{\beta}_2) = \sqrt{c_{22}\widehat{\sigma}^2} = \sqrt{\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}_{22}} \\
8.629 &= \widehat{\text{se}}(\widehat{\beta}_3) = \sqrt{c_{33}\widehat{\sigma}^2} = \sqrt{\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}_{33}},
\end{aligned}$$

where

$$\widehat{\sigma}^2 = \frac{\text{SSE}}{30-4} = (10.13)^2 \approx 102.63$$

is the square of the `Residual standard error`. The `t value` output gives the t statistics

$$\begin{aligned}
t = -1.463 &= \frac{\widehat{\beta}_0 - 0}{\sqrt{c_{00}\widehat{\sigma}^2}} \\
t = 0.074 &= \frac{\widehat{\beta}_1 - 0}{\sqrt{c_{11}\widehat{\sigma}^2}} \\
t = 3.133 &= \frac{\widehat{\beta}_2 - 0}{\sqrt{c_{22}\widehat{\sigma}^2}} \\
t = 2.279 &= \frac{\widehat{\beta}_3 - 0}{\sqrt{c_{33}\widehat{\sigma}^2}}.
\end{aligned}$$

These t statistics can be used to test $H_0 : \beta_i = 0$ versus $H_0 : \beta_i \neq 0$, for $i = 0, 1, 2, 3$. Two-sided probability values are in `Pr(>|t|)`. At the $\alpha = 0.05$ level,

- we do not reject $H_0 : \beta_0 = 0$ (p-value $= 0.155$). **Interpretation:** In the model which includes all three independent variables, the intercept term β_0 is not statistically different from zero.

- we do not reject $H_0 : \beta_1 = 0$ (p-value $= 0.942$). **Interpretation:** `ACETIC` does not significantly add to a model that includes `H2S` and `LACTIC`.

- we reject $H_0 : \beta_2 = 0$ (p-value $= 0.004$). **Interpretation:** `H2S` does significantly add to a model that includes `ACETIC` and `LACTIC`.

- we reject $H_0 : \beta_3 = 0$ (p-value $= 0.031$). **Interpretation:** `LACTIC` does significantly add to a model that includes `ACETIC` and `H2S`.

CONFIDENCE INTERVALS: Ninety-five percent confidence intervals for the regression parameters β_0, β_1, β_2, and β_3, respectively, are

$$\begin{aligned}
\widehat{\beta}_0 \pm t_{26,0.025}\widehat{\text{se}}(\widehat{\beta}_0) &\implies -28.877 \pm 2.056(19.735) \implies (-69.45, 11.70) \\
\widehat{\beta}_1 \pm t_{26,0.025}\widehat{\text{se}}(\widehat{\beta}_1) &\implies 0.328 \pm 2.056(4.460) \implies (-8.84, 9.50) \\
\widehat{\beta}_2 \pm t_{26,0.025}\widehat{\text{se}}(\widehat{\beta}_2) &\implies 3.912 \pm 2.056(1.248) \implies (1.35, 6.48) \\
\widehat{\beta}_3 \pm t_{26,0.025}\widehat{\text{se}}(\widehat{\beta}_3) &\implies 19.670 \pm 2.056(8.629) \implies (1.93, 37.41).
\end{aligned}$$

PREDICTION: Suppose that we are interested estimating $E(Y|\mathbf{x}^*)$ and predicting a new Y when `ACETIC` $= 5.5$, `H2S` $= 6.0$, and `LACTIC` $= 1.4$, so that

$$\mathbf{x}^* = \begin{pmatrix} 5.5 \\ 6.0 \\ 1.4 \end{pmatrix}.$$

We use R to compute the following:

```
> predict(fit,data.frame(acetic=5.5,h2s=6.0,lactic=1.4),level=0.95,interval="confidence")
     fit      lwr      upr
23.93552 20.04506 27.82597
> predict(fit,data.frame(acetic=5.5,h2s=6.0,lactic=1.4),level=0.95,interval="prediction")
     fit      lwr      upr
23.93552 2.751379 45.11966
```

- Note that

$$\begin{aligned}
\widehat{E(Y|\mathbf{x}^*)} = \widehat{Y}^* &= \widehat{\beta}_0 + \widehat{\beta}_1 x_1^* + \widehat{\beta}_2 x_2^* + \widehat{\beta}_3 x_3^* \\
&= -28.877 + 0.328(5.5) + 3.912(6.0) + 19.670(1.4) \approx 23.936.
\end{aligned}$$

- A 95 percent **confidence interval** for $E(Y|\mathbf{x}^*)$ is $(20.05, 27.83)$. When `ACETIC` $=$ 5.5, `H2S` $= 6.0$, and `LACTIC` $= 1.4$, we are 95 percent confident that the mean taste rating is between 20.05 and 27.83.

- A 95 percent **prediction interval** for Y^*, when $\mathbf{x} = \mathbf{x}^*$, is $(2.75, 45.12)$. When `ACETIC` $= 5.5$, `H2S` $= 6.0$, and `LACTIC` $= 1.4$, we are 95 percent confident that the taste rating for a new cheese specimen will be between 2.75 and 45.12.

Completed R codes for Example 11.2 (follow this to answer HW problems).

```
##########################################################################
# Reading data
taste=c(12.3,20.9,39,47.9,5.6,25.9,37.3,21.9,18.1,21,34.9,57.2,0.7,25.9,54.9,40.9,
15.9,6.4,18,38.9,14,15.2,32,56.7,16.8,11.6,26.5,0.7,13.4,5.5)
acetic=c(4.543,5.159,5.366,5.759,4.663,5.697,5.892,6.078,4.898,5.242,5.74,6.446,
4.477,5.236,6.151,6.365,4.787,5.412,5.247,5.438,4.564,5.298,5.455,5.855,5.366,
6.043,6.458,5.328,5.802,6.176)
h2s=c(3.135,5.043,5.438,7.496,3.807,7.601,8.726,7.966,3.85,4.174,6.142,7.908,
2.996,4.942,6.752,9.588,3.912,4.7,6.174,9.064,4.949,5.22,9.242,10.199,3.664,
3.219,6.962,3.912,6.685,4.787)
lactic=c(0.86,1.53,1.57,1.81,0.99,1.09,1.29,1.78,1.29,1.58,1.68,1.9,1.06,1.3,1.52,
1.74,1.16,1.49,1.63,1.99,1.15,1.33,1.44,2.01,1.31,1.46,1.72,1.25,1.08,1.25)
##########################################################################
# Fit linear regression
fit=lm(taste~acetic+h2s+lactic)
summary(fit)
##########################################################################
# 95\% confidence interval for E(Y) at acetic=5.5,h2s=6.0,lactic=1.4#
predict(fit,data.frame(acetic=5.5,h2s=6.0,lactic=1.4),level=0.95,interval="confidence")
##########################################################################
# 95\% prediction interval for E(Y) at acetic=5.5,h2s=6.0,lactic=1.4
predict(fit,data.frame(acetic=5.5,h2s=6.0,lactic=1.4),level=0.95,interval="prediction")
```

The analysis of variance (ANOVA) for linear regression

ANOVA TABLE: The general form of an ANOVA table for linear regression (simple or multiple) is given below:

Source	df	SS	MS	F
Regression	$p-1$	SSR	$\text{MSR} = \frac{\text{SSR}}{p-1}$	$F = \frac{\text{MSR}}{\text{MSE}}$
Error	$n-p$	SSE	$\text{MSE} = \frac{\text{SSE}}{n-p}$	
Total	$n-1$	SST		

where

- we have

$$\underbrace{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}_{\text{SSE}}.$$

  - SST is the **corrected total** sum of squares
  - SSR is the **corrected regression (model)** sum of squares
  - SSE is the **error (residual)** sum of squares.

- The column labeled "df" gives the **degrees of freedom** for each.

- The column labeled "MS" contains the **mean squares**

$$\text{MSR} = \frac{\text{SSR}}{p-1}$$

$$\text{MSE} = \frac{\text{SSE}}{n-p} = \widehat{\sigma}^2$$

- Since SST=SSR+SSE, the proportion of the total variation in the data explained by the linear model is

$$R^2 = \frac{\text{SSR}}{\text{SST}},$$

  typically called the **coefficient of determination**. Interpretation: the larger the R^2, the more variation that is being explained by the regression model.

- The ANOVA table F statistic

$$F = \frac{\text{MSE}}{\text{MSE}} = \frac{\text{SSR}/(p-1)}{\text{SSE}/(n-p)}$$

  is used to test

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \text{ versus } H_1 : \text{ at least one of the } \beta_j\text{'s is nonzero}$$

  Rejection region: $\text{RR} = \{F : F > F_{p-1,n-p,\alpha}\}$.

Example 11.2 (continued) Constructing the ANOVA table

Source	df	SS	MS	F
Regression	$p-1$	SSR	$\text{MSR} = \frac{\text{SSR}}{p-1}$	$F = \frac{\text{MSR}}{\text{MSE}}$
Error	$n-p$	SSE	$\text{MSE} = \frac{\text{SSE}}{n-p}$	
Total	$n-1$	SST		

based on the following R output:

```
fit=lm(taste~acetic+h2s+lactic)
summary(fit)

Call:
lm(formula = taste ~ acetic + h2s + lactic)

Residuals:
    Min      1Q  Median      3Q     Max
-17.390  -6.612  -1.009   4.908  25.449

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.8768    19.7354  -1.463  0.15540
acetic        0.3277     4.4598   0.073  0.94198
h2s           3.9118     1.2484   3.133  0.00425 **
lactic       19.6705     8.6291   2.280  0.03108 *
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-squared:  0.6518,Adjusted R-squared:  0.6116
F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

Reduced versus full model testing Recall the multiple linear regression model is

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, \ldots, n,$$

where we believe all the k covariates are useful to explain the variance in the response. However, out of k, the number of truly useful covariates might be only $g < k$. This motivates us to think which of the following is true?

- **the completed model**

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_g x_{ig} + \beta_{g+1} x_{i,g+1} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, \ldots, n,$$

- or **a reduced model**

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_g x_{ig} + \epsilon_i, \quad i = 1, \ldots, n,$$

Equivalently, we are testing

$$H_0 : \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0 \text{ versus } H_1 : \text{ at least one of } \beta_{g+1}, \ldots, \beta_k \text{ is nonzero}$$

Test Statistics:

$$F = \frac{(\mathrm{SSE}_R - \mathrm{SSE}_C)/(k-g)}{\mathrm{SSE}_C/(n-p)}$$

Rejection Region

$$\mathrm{RR} = \{F : F > F_{k-g,n-p,\alpha}\}$$

where $p = k + 1$,

- SSE_C: the error sum of squares under the completed model
- SSE_R: the error sum of squared under the reduced model

Example 11.2 (continued) Y: taste, x_1: acetic, x_2: h2s, x_3: lactic; $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}, i = 1, \ldots, 30.$

```
summary(lm(taste~acetic+h2s+lactic))

Call:
lm(formula = taste ~ acetic + h2s + lactic)

Residuals:
    Min      1Q  Median      3Q     Max
-17.390  -6.612  -1.009   4.908  25.449

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.8768    19.7354  -1.463  0.15540
acetic        0.3277     4.4598   0.073  0.94198
h2s           3.9118     1.2484   3.133  0.00425 **
lactic       19.6705     8.6291   2.280  0.03108 *
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-squared:  0.6518,Adjusted R-squared:  0.6116
F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

(1) Combining the following R results, conduct the below test at significance level $\alpha = 0.05$

$$H_0 : \beta_1 = \beta_3 = 0 \text{ versus } H_1 : \text{ at least one of } \beta_1, \beta_3 \text{ is nonzero}$$

```
#########################################
# R codes:
anova(lm(taste~h2s),lm(taste~acetic+h2s+lactic))
#########################################
Analysis of Variance Table

Model 1: taste ~ h2s
Model 2: taste ~ acetic + h2s + lactic
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     28 3286.1
2     26 2668.4  2    617.73 3.0095 0.06674 .
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

(2) What test are the following codes for?

```
anova(lm(taste~acetic+lactic),lm(taste~acetic+h2s+lactic))
Analysis of Variance Table

Model 1: taste ~ acetic + lactic
Model 2: taste ~ acetic + h2s + lactic
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     27 3676.1
2     26 2668.4  1    1007.7 9.8182 0.004247 **
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```