Design of Experiments (Math 556)

DL MWF 8:00am-9:30am, (or 4:00pm-5:30pm ?)
Office: WH 132
Office hours: M 7-8pm, T 4:00-5:00pm
Textbook: Statistics for Experimenters (2nd ed.)
by George Box, J Stuart Hunter and William G. Hunter

Quiz: Once a week at a random day,
right after quiz, take a picture of your answer, output as a pdf file and email me.
quiz problems: formulas for Math 447-448 (see my website)

Midterm: Oct. 19 (M), Final: Dec. 8-10
Each is allowed to bring a piece of paper with anything you prefer on it.
Homework assigned during a week is due next Wednesday.
Email me at qyu@math.binghamton.edu before 4:00pm on Wednesday.

It is on my website: http://www.math.binghamton.edu/qyu/qyu
personal note and note2 are updated one,
Grading Policy: 50% hw and quizzes +50% exams,
B = 70 ±

Chapter 1. Introduction

All the concepts of this chapter have been introduced in 501, except autocorrelation.
Recall

X and Y are random variables, with observations \((X_1, Y_1), \ldots, (X_n, Y_n)\).
Population covariance and correlation:
\[
\text{Cov}(X, Y) = E(XY) - E(X)E(Y),
\]
\[
\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y},
\]

Sample Covariance
\[
\hat{\text{Cov}}(X, Y) = \overline{XY} - \overline{X} \cdot \overline{Y},
\]
Sample correlation \(\hat{\rho} = r = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\hat{\sigma}_X \hat{\sigma}_Y},\)
where \(\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2,\)
\(S^2\) is also denoted by \(s^2\) in the textbook. Which is a better notation ?

Definition. The lag-\(k\) sample autocorrelation coefficient of \(Y_i\)'s is
\[
r_k = \frac{\sum_{i=k+1}^{n} (Y_i - \overline{Y})(Y_{i-k} - \overline{Y})}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}, \quad k = 1, 2, \ldots
\]
It measures the serial dependence of the data in time.
If \(r_k \neq 0,\) are the data i.i.d.? 

If \(r_k > 0\) significantly, are the data i.i.d.? 

Theorem 1. If \(X_1, \ldots, X_n\) are i.i.d. from \(N(\mu, \sigma^2),\) then
(a) \(\overline{X} \perp S^2;\)
(b) \(\overline{X} \sim N(\mu, \sigma^2/n);\)
(c) \((n-1)S^2/\sigma^2 \sim \chi^2(n-1).\)

Chapter 2. Basic

Chapter 3. Comparing Two Entities
3.1. Consider the test for the difference of the means of two random samples $X_i$’s and $Y_j$’s.

$H_0$: $\mu_Y - \mu_X = \delta$ v.s. $H_1$: $\mu_Y - \mu_X > \delta$.

**Two-samples test:** Under the assumption that (1) two samples are independent,
(2) $X_i$’s are from $N(\mu_X, \sigma^2)$ and (3) $Y_j$’s are $N(\mu_Y, \sigma^2)$, then a common test is

$$\phi = 1(t > t_{a,n_Y+n_X-2}), \text{ where}$$

$$t = \frac{\overline{Y} - \overline{X} - \delta}{s_p \sqrt{1/n_X + 1/n_Y}} \text{ and } s_p^2 = \frac{\sum_{i=1}^{n_X} (X_i - \overline{X})^2 + \sum_{j=1}^{n_Y} (Y_j - \overline{Y})^2}{n_Y + n_X - 2}.$$  \( (1) \)

This is due to
(a) $T = \frac{N(0,1)}{\sqrt{\chi^2(\nu)}} \chi^2(\nu)$, where $N(0, 1) \perp \chi^2(\nu)$
(b) $t = \frac{\overline{X} - \overline{Y} - \delta}{s_p \sqrt{1/n_X + 1/n_Y}} \chi^2(\nu)$
(c) $\frac{\sum_{i=1}^{n_X} (X_i - \overline{X})^2}{\sigma^2} \chi^2(n_X - 1)$, $\frac{\sum_{j=1}^{n_Y} (Y_j - \overline{Y})^2}{\sigma^2} \chi^2(n_Y - 1)$
(d) $\frac{\sum_{i=1}^{n_X} (X_i - \overline{X})^2}{\sigma^2} + \frac{\sum_{j=1}^{n_Y} (Y_j - \overline{Y})^2}{\sigma^2} \chi^2(\nu)$

**The paired t-test:** Under the paired random sample of size $n$ from $N(\mu_Y, \sigma^2)$ and $N(\mu_X, \sigma^2)$, then a common test is

$$\phi_p = 1(t > t_{a,n-1}), \text{ where}$$

$$t = \frac{\overline{X} - \overline{Y} - \delta}{s_p \sqrt{1/n}} \text{ and } s_p^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i - \overline{Y} - \overline{X})^2.$$  \( (2) \)

**Importance of the independent normally distributed assumptions in both tests.**

**Chemical Example in Table 3.2.** An experiment was performed on a factory by making in sequence 10 batches of chemical using a standard production method (A) followed by 10 batches of a chemical using a modified method (B). The data are

A: 89.7, 81.4, ..., 84.5
B: 84.7, 86.1, ..., 88.5

See Table 3.1 on page 69.

Summary:

$n_A = n_B = 10,$
$\overline{X}_A = 84.24,$
$\overline{Y}_B = 85.54,$
$s_p^2 = 10.8727,$

$H_0$: $\mu_B - \mu_A = 0$, v.s. $H_1$: $\mu_B - \mu_A > 0.$
$\overline{Y}_B - \overline{X}_A = 1.3.$

Is it significant? **What does it mean?**

We need to
(1) set $\alpha (=0.05)$, and
(2) compute $P(\overline{Y}_B - \overline{X}_A \geq 1.3) = ?$ what is it called?

Then conclude that if $P(\overline{Y}_B - \overline{X}_A \geq 1.3) < \alpha$ ...

One often uses the two-sample t-test in Eq. (1), then the P-value is 19%.

Is it significant?
Do we reject $H_0$?

**Can we use paired t-test?**

Does the SD become larger or smaller if we use it?

$$\sigma^2/(2n - 2) \text{ v.s. } \sigma^2/(n - 1).$$

$$s_p^2 = \frac{\sum_{i=1}^{n_X} (X_i - \overline{X})^2 + \sum_{j=1}^{n_Y} (Y_j - \overline{Y})^2}{n_Y + n_X - 2} \text{ v.s. } S_{Y_B-Y_A}$$

What is the conclusion if we use it? $P\text{value}=P(T \geq \frac{\overline{Y}_B-\overline{X}_A}{S_{SE}})$
Introduce two alternative approaches next.

**External Reference Distribution.**

Old data. 210 batches of the chemical products recorded in time order before the 20 data: 

\[ x_1, \ldots, x_{210} \]

The old data (see p.120) provide an external reference distribution.

Under \( H_0 \), the 20 data can be viewed as a sample from the population of the 210 data.

Compute

\[
D_t = \sum_{i=t-10}^{t+19} x_i / 10 - \sum_{i=t-9}^{t-1} x_i / 10, \quad t = 1, \ldots, 191.
\]

See the histogram Figure 3.3 on page 70.

\[ P(\overline{y}_B - \overline{y}_A \geq 1.3) = 9/191 \approx 0.047. \text{ Is it significant?} \]

Recall that if one uses t-test, the P-value is 19%. **Anything wrong?**

1. The lag-1 sample auto-correlation of the data is \( r_1 = \hat{\rho}_1 = -0.29 \).

The data are not independent.

   If one pretends independence, it leads to incorrect conclusion.

2. Normal assumption may not be valid (do we need to check it?)

**Internal Reference distribution.** Random sampling distribution.

A randomized design in the comparison of standard and modified fertilizer mixtures for tomato plants. 11 plants in a row. 5 with standard (A), 6 with modified (B).

One way is to apply A to the first 5 and B to the next 6 in a row. There are correlation between locations and it is not a good idea without randomization.

Randomizing the order in the row (sample(1:11,5) = ?) resulting

<table>
<thead>
<tr>
<th>location</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>fertilizer</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>yield</td>
<td>29.2</td>
<td>11.4</td>
<td>26.6</td>
<td>23.7</td>
<td>25.3</td>
<td>28.5</td>
<td>14.2</td>
<td>17.9</td>
<td>16.5</td>
<td>21.1</td>
<td>24.3</td>
</tr>
</tbody>
</table>

**Remark:** Role of a statistician:

1. randomization before an experiment (DOE);
2. make inferences after the experiment.

\[ x = c(29.2,11.4,26.6,23.7,25.3,28.5,14.2,17.9,16.5,21.1,24.3) \]

\[ z = c(3,4,6,7,8,11) \]

\[ \text{mean}(x[z]) - \text{mean}(x[\text{subset}=-z]) \]

# results in \( \overline{y}_B - \overline{y}_A \approx 1.69 \).

To test \( H_0: \mu_B - \mu_A = 0 \) against \( H_1: \mu_B - \mu_A > 0 \).

Need to compute \( P(\overline{y}_B - \overline{y}_A \geq 1.69) = ? \)

Rather than using t-test, which needs normal assumption, and equal variance, we make use of the

**Permutation distribution.**

Table (1) is one combination of selecting 5 out of 11.

\[
\begin{array}{cccccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
29.2 & 11.4 & 26.6 & 23.7 & 25.3 & 28.5 & 14.2 & 17.9 & 16.5 & 21.1 & 24.3 \\
\end{array}
\]

is another combination under \( H_0: \mu_B - \mu_A = 0 \).

> \text{mean}(x[6:11]) - \text{mean}(x[1:5]) \]

# results in \(-2.82\)

Eq.(2) yields \( \overline{y}_B - \overline{y}_A \approx -2.82 \); while

Eq.(1) yields \( \overline{y}_B - \overline{y}_A \approx 1.69 \).

There are \( \binom{11}{5} = \frac{11!}{5!6!} = 11 \cdot 3 \cdot 2 \cdot 7 = 462 \) such combinations.

> P=combn(1:11,6)
Thus these 462 combinations yield 462 \( \bar{y}_B - \bar{y}_A \) values. These 462 values form a (discrete) distribution called the permutation distribution.

```r
x=c(29.2,11.4,26.6,23.7,25.3,28.5,14.2,17.9,16.5,21.1,24.3)
N=choose(11,6) # =462
y=1:N
P=combn(1:11,6)
for(i in 1:N)
  y[i]=mean(x[P[,i]])-mean(x[-P[,i]])
length(y[y>1.69])/N # result is 0.3203463
```

Or without loop:
```r
y=x[P]
dim(y)=c(6,462)
B=apply(y,2,sum)
y=B/6-(sum(x)-B)/5
length(y[y>1.69])/N # result is 0.3203463
```

What is the conclusion of the test?
The permutation distribution can also be simulated by the R code as follows.
```r
x=c(29.2,11.4,26.6,23.7,25.3,28.5,14.2,17.9,16.5,21.1,24.3)
N=10000
y=rep(0,N)
for(i in 1:N){
  u=sample(x)
  y[i]=mean(u[1:6])-mean(u[7:11])
}
length(y[y>1.69])/N # result is 0.3209
```

![Histogram of y](image1)

![Histogram of y](image2)

Figure 3.1. Histograms of permutation distribution v.s. simulation one
Should we use simulation here?

Remark. The two-samples t-test $P(t_{n_A+n_B-2} > \frac{1.69}{\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}) \approx 0.34)$ for the current data.

If the normal assumption is not valid, the t-test is not applicable (though it happens to be close to 0.32.

The permutation distribution is based on a different sample space from the sample space where the data come from. But if $n_A + n_B$ is large, the permutation distribution of $Y_B - Y_A$ is very close to $t_{n_A+n_B-2}$, whereas the two-sample t-test may not have the $t_{n_A+n_B-2}$ distribution (e.g. if $X_1 = \cdots = X_{n_A} = Y_1 = \cdots = Y_{n_B}$).

Can we say $n_A + n_B$ is large here?

Remark. In the fertilizer example, the data are resulted from randomization, whereas in the previous chemical example, the data are in sequence.

A A A A A A A A A B B B B B B B B

If they had done

sample(1:20,10)

[1] 10 4 15 12 13 2 11 6 20 8

for the order of 10 batches of chemical using method A, then the permutation distribution would be valid.

Is it appropriate to apply randomization distribution in the chemical example?
3.2. Randomized paired comparison design: Boys shoes example. The shoe soles were made of two different materials, A and B. Ten boys were chosen randomly to compare the shoe wear. Each boy wore a special pair of shoes. The decision as to whether the left or right sole was made with A or B was determined by

(1) convenience,
(2) by flipping a coin (or rbinom(n,1,0.5)).

Which result in a random sample? (took 2 steps in DOE).

The randomization results

\[
\text{boy: } 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10
\]

\[
\text{material A: } L \quad L \quad R \quad L \quad R \quad L \quad L \quad R \quad L
\]

The experiment results in

\[
x = (0.8, 0.6, 0.3, -0.1, 1.1, -0.2, 0.3, 0.5, 0.5, 0.3)
\]

Then 10 \( \bar{y}_B - \bar{y}_A \)'s yield

\[
\text{mean}(x) \quad # \bar{y}_B - \bar{y}_A = 0.41
\]

Should we use two-sample t-test or paired t-test?

What assumptions do we need in order to use one of them?

Another way to compute P-value for \( \bar{y}_B - \bar{y}_A \geq 0 \).

Under \( H_0: \mu_B - \mu_A = 0 \), a combination could be (R L R L R L L L R L)

\[
\begin{pmatrix}
R & L & R & L & R & L & L & L & R & L
\end{pmatrix}
\]

Then the data become

\[
x = (-0.8, 0.6, 0.3, -0.1, 1.1, -0.2, 0.3, 0.5, 0.5, 0.3)
\]

Compare to the real data:

\[
x = (0.8, 0.6, 0.3, -0.1, 1.1, -0.2, 0.3, 0.5, 0.5, 0.3)
\]

The randomized reference distribution under \( H_0: \mu_A = \mu_B \) can be obtained as follows.

\[
x = c(0.8, 0.6, 0.3, -0.1, 1.1, -0.2, 0.3, 0.5, 0.5, 0.3)
\]

\[
\text{sum}(x) \quad # \text{result}=4.1
\]

\[
y = 1:1024 \quad # \text{initialize } y
\]

for(i1 in 0:1)
for(i2 in 0:1)
for(i3 in 0:1)
for(i4 in 0:1)
for(i5 in 0:1)
for(i6 in 0:1)
for(i7 in 0:1)
for(i8 in 0:1)
for(i9 in 0:1)
for(i10 in 0:1)

\[
i = c(i1,i2,i3,i4,i5,i6,i7,i8,i9,i10)
\]

\[
h = 0:9
\]

\[
y[i %% %%(2 * h) +1] = \text{sum}(x*((-1)**i))
\]

# i1 * 2^0 + i2 * 2^1 + i3 * 2^2 + ... + i10 * 2^9,
# Examples:
# binary number 1110 = 2^3 + 2^2 + 2^1 + 0 * 2^0 = 14
# ternary number 2101 = 2 * 3^3 + 1 * 3^2 + 0 * 3^1 + 1 * 3^0 = 64
# decimal number 2101 = 2 * 10^3 + 1 * 10^2 + 0 * 10^1 + 1 * 10^0
\]

\[
\text{length}(y[y>= 4.1])/1024 \quad # \text{result} = 0.0068
\]

\[
z = \text{seq}(-6, 6, 0.1)
\]

\[
\text{hist}(y)
\]

\[
\text{lines}(z, dt(z, 9))
\]

The randomized reference distribution under \( H_0: \mu_A = \mu_B \) can be approximated
by simulation as follows.

```r
N=10000
y=rep(0,N)
x=c(0.8,0.6,0.3,-0.1,1.1,-0.2,0.3,0.5,0.5,0.3)
for (i in 1:N) {
  s=rbinom(10,1,0.5)
z=(-1)**s
y[i]=sum(x*z)
}
length(y[y > 4.1])/N  # 0.0063
hist(y, xlim=c(-6,6), breaks=12, freq=F)
```

One can see from Figure 3.2 that the simulation distribution is very close to the true permutation distribution.
The density of $t_9$ is displayed at the top of Fig. 3.2. The P-value using the paired t-test is 0.4%.

![Histogram of y](image1)

**Figure 3.2. Histograms of permutation distribution v.s. simulation one**

Any thing wrong with the solution?
Is it one sided test or two-sided test?
10.1. Main assumption:

\[ Y = \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon, \]

or

\[ E(Y|X) = \beta_1 X_1 + \cdots + \beta_p X_p, \]

where

- \( \epsilon \) is an unobservable random variable with \( E(\epsilon|X) = 0 \) (no assumption on \( V(\epsilon|X) \) yet),
- \( \beta_i \)'s are parameters,
- \( X_i \)'s and \( Y \) are observable.

Given (independent) observations \( (Y_i, x_{i1}, \ldots, x_{ip}), i = 1, \ldots, n, \)
we shall make inference about \( \beta_i \)'s.

**Remark.** A special case of the linear regression model is

\[ Y = \alpha + \beta X + \epsilon. \]

**Least squares estimator (LSE) minimizes**

\[ S(\beta) = \sum_{i=1}^{n} (Y_i - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 \]

where \( \beta = (\beta_1, \ldots, \beta_p)' \).

Notice that \( S(\beta) \) can be written as a matrix form

\[ S(\beta) = (Y - X\beta)'(Y - X\beta) \]

where \( Y' = (Y_1, \ldots, Y_n) \),

\[ X = (x_{ij})_{n \times p} = \begin{pmatrix} x_{i1} & \cdots & x_{ip} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \]

The LSE can be obtained by solving the normal equation

\[ \frac{\partial S}{\partial \beta} = 0, \text{ a } p \text{ dimensional zero vector.} \]

That is,

\[ X'(Y - X\beta) = 0. \quad \text{(Why not } (Y - X\beta)'X = 0?) \]

The LSE has the form

\[ \hat{\beta} = (XX')^{-1}X'Y \text{ if } X'X \text{ is invertible,} \]

otherwise, the solution to LSE is not unique, and one often imposes further constraints to get a unique solution.

If \( \epsilon \) is normal, then \( \hat{\beta} \) is the MLE. Otherwise, it is a semi-parametric estimator.

**Fitted value** \( \hat{y}_i = (x_{i1}, \ldots, x_{ip})\hat{\beta}. \) (\( E(Y|x) \))

**Residuals** \( y_i - \hat{y}_i, i = 1, \ldots, n. \)

If one further assumes that \( V(\epsilon_i) = \sigma^2 \forall i, \) then

\[ \hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

is an unbiased estimator of \( \sigma^2, \)

and conditional on \( X \) (if one assumes \( X \) is random),

\[ V(\hat{\beta}) = \sigma^2(X'X)^{-1} \text{ or } V(\hat{\beta}|X) = \sigma^2(X'X)^{-1} \]

(are they both correct?)

**Is \( V(\hat{\beta}) \) variance or covariance matrix?**

SE of \( \hat{\beta}_j = \sqrt{v}, \) where \( v \) is obtained by the \( j \)-th diagonal element of \( \hat{\sigma}^2(X'X)^{-1} \)

(why not \( \sigma^2(X'X)^{-1} \)? \( \text{SD = SE? Are they r.v.\'s?} \)

**Under NID** \( a (1 - \alpha)100\% \) CI of \( \beta_i \) is \( \hat{\beta}_i \pm t_{n-p, \alpha/2}SE \)

**Example 0:** Suppose that \( Y_i = \begin{cases} -\gamma + W_i & \text{if } i \in \{1, \ldots, n_+\} \\ \gamma + W_i & \text{if } i \in \{n_+ + 1, \ldots, n\} \end{cases} \quad n_- > 1, \)

and \( W_1, \ldots, W_n \) are i.i.d. from the exponential distribution and \( E(W_i) = 1. \) \( \gamma \) is unknown, \( Y_i \)'s are observations, but \( W_i \)'s are not observable, though we know \( W_i \sim Exp(1). \) Derive the LSE and the MLE of \( \gamma \) based on the regression data.

**Discussion.** Let the model be \( Y_i = \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i = X_i'\beta \) with \( E(\epsilon_i) = 0. \)

\( p = \)?

Do we observe \( Y_i \)?

Do we observe \( (X_{i1}, \ldots, X_{ip}) \)?

\( W_i = \epsilon_i \)?

Do we know \( \beta \)? or \( (\beta_1, \ldots, \beta_p) \)?

If we rewrite the model as \( Y_i = \alpha + \gamma X_i + \epsilon_i, \) then \( \alpha = \)?
Homework 10.1. Find the MLE and the LSE of \( \beta \) under the assumptions above.

Polynomial model: \( Y_i = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k + \epsilon_i, \ i = 1, \ldots, n \).

\( k \) can be as large as \( n - 1 \) if \( x_i \)'s are all distinct.

Example 1. Data: \( (X_i, Y_i) \): (1,2), (3,4). The LSE \( \hat{\beta} = (X'X)^{-1}X'Y \) under the models:

\[
Y = \beta_0 + \epsilon, \quad X = ?
\]
\[
Y = \beta_1 x + \epsilon, \quad X = ?
\]
\[
Y = \beta_0 + \beta_1 x + \epsilon, \quad X = ?
\]

If one fits model \( Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \). Then

\[
Y = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ : & : & : & : \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}
\]

\[
\text{rank of } X'X = 2. \ X'X \text{ is not invertible. The LSE is not uniquely determined.}
\]

We say that the parameter is not identifiable.

Possible modification: Add a constraint to \( \beta_i \)'s, e.g. \( \beta_0 = 1 \) or \( \beta_1 = \beta_2 \), etc.

Remark. In Example 1, there are 3 ways to intereprete \( X_i \)'s or \( X \):

<table>
<thead>
<tr>
<th>models</th>
<th>( X ) type: ( \text{original} )</th>
<th>( \text{in model} )</th>
<th>( X ) in LSE formula</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = \beta_0 + \epsilon )</td>
<td>( 1, 3 )</td>
<td>( 1, 1 )</td>
<td>( (1,1)' )</td>
<td>( \beta_0 )</td>
</tr>
<tr>
<td>( Y = \beta_1 x + \epsilon )</td>
<td>( 1, 3 )</td>
<td>( 1, 3 )</td>
<td>( (1,3)' )</td>
<td>( \beta_1 )</td>
</tr>
<tr>
<td>( Y = \beta_0 + \beta_1 x + \epsilon )</td>
<td>( 1, 3 )</td>
<td>( (1,1), (1,3) )</td>
<td>( \begin{pmatrix} 1 &amp; 1 \ 1 &amp; 3 \end{pmatrix} )</td>
<td>( \beta_0, \beta_1 )'</td>
</tr>
<tr>
<td>( Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon )</td>
<td>( 1, 3 )</td>
<td>( (1,1,1), (1,3,9) )</td>
<td>?</td>
<td>( \beta_0, \beta_1, \beta_2 )'</td>
</tr>
<tr>
<td>( Y - 1 = \beta_1 x + \beta_2 x^2 + \epsilon )</td>
<td>( 1, 3 )</td>
<td>( (1,1), (3,9) )</td>
<td>?</td>
<td>( \beta_1, \beta_2 )'</td>
</tr>
<tr>
<td>( Y = \beta_0 + \beta_1 (x + x^2) + \epsilon )</td>
<td>( 1, 3 )</td>
<td>( (1,2), (1,12) )</td>
<td>?</td>
<td>( \beta_0, \beta_1 )'</td>
</tr>
</tbody>
</table>

Example 2. One way anova table \( Y_{ij} = \mu + \alpha_j + \epsilon_{ij}, \ i = 1, \ldots, 4, \) and \( j = 1, \ldots, 3 \).

Consider an example that there are three treatments \( A, B \) and \( C \). There are \( I (=4) \) groups, each consists of 3 patients. Total of 12 patients. In each group, the 3 patients receive 3 different treatments separately. The result for the \( j \)th patient in the \( i \)th group is \( Y_{ij} \).

Is it a linear regression model?

\[ \beta = ? \]

LSE = \( (X'X)^{-1}X'Y \).

\[ X = ? \]

One possibility is based on \( Y = X\beta + e \),

\[
Y_{11} \quad Y_{12} \quad Y_{13} \quad Y_{14} \quad Y_{21} \quad Y_{22} \quad Y_{23} \quad Y_{24} \quad Y_{31} \quad Y_{32} \quad Y_{33} \quad Y_{34} \quad Y_{41} \quad Y_{42} \quad Y_{43} \quad Y_{44}
\]

\[
= \begin{pmatrix} 1 & X_{11} & X_{12} & X_{13} \\ : & : & : & : \\ 1 & X_{n1} & X_{n2} & X_{n3} \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + e = X\beta + e, \quad e = \epsilon_1, \ldots, \epsilon_{12} \)' |

\( X_{11} = 1 \) (treatment = \( A \) for the \( i \)-th patient).
\( X_{12} = 1 \) (treatment = \( B \) for the \( i \)-th patient).
\( X_{13} = 1 \) (treatment = \( C \) for the \( i \)-th patient).

Notice that \( X_{11} + X_{12} + X_{13} = 1 \).

Thus the LSE for \( Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i \) is not unique.

(We say that the parameters are not identifiable).

\( X'X \) is not invertible as
\[
X = \begin{pmatrix}
1 & X_{11} & X_{12} & X_{13} \\
\vdots & \vdots & \vdots & \vdots \\
1 & X_{n1} & X_{n2} & X_{n3} \\
\end{pmatrix}
\]

is of rank at most 3, not 4,

as
\[
\begin{pmatrix}
X_{11} + X_{12} + X_{13} \\
\vdots \\
X_{n1} + X_{n2} + X_{n3}
\end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \end{pmatrix}
\]

why?

**Three modifications:**

1. Revise the model. Let \( Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \) with \( \mu = 0 \),

\[
\beta = (\alpha_1, \alpha_2, \alpha_3)', \ X = \begin{pmatrix} X_{11} & X_{12} & X_{13} \\
\vdots & \vdots & \vdots \\
X_{n1} & X_{n2} & X_{n3} \end{pmatrix}, \hat{\beta} = (X'X)^{-1}X'Y \text{ works}.
\]

\( \text{lm}(Y \sim X_1 + X_2 + X_3 - 1) \)

2. Impose a constraint \( \alpha_1 = 0 \) for the model

\[
Y_i = \mu + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \epsilon_i.
\]

Then \( \beta_1 = \mu + \alpha_1, \ i = 1, 2, 3, \) in \( Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i \) with \( \mu = 0 \),

\( \text{e.g., if } X_{i1} = 1, \) then \( Y_i = \beta_1 + \epsilon_i = \mu + \alpha_1, \) where \( \alpha_1 = 0, \ ... \)

\( \text{options(contrasts =c("contr.treatment", "contr.poly")}) \)

\( \text{lm}(Y \sim X_1 + X_2 + X_3) \).

3. Impose another constraint \( \sum_i \alpha_i = 0 \) \( \alpha_3 = -\alpha_1 - \alpha_2 \) for the model

\[
Y_i = \mu + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \epsilon_i.
\]

then \( \mu + \alpha_i = \beta_i, \ i = 1, 2, 3. \)

\( \text{options(contrasts =c("contr.sum", "contr.poly")}) \)

\( \text{lm}(Y \sim X_1 + X_2 + X_3) \)

**Example 3** (a simulation study on the Two way anova table).

\[
Y_{ij} = \mu + \alpha_i + b_j + \epsilon_{ij}, \ i \in \{1, \ldots, 4\}, \ j \in \{1, \ldots, 6\}
\]

\( \text{> y=norm(24)} \)

\( \text{> a=gl(4,6,24)} \)

\( \text{> b=gl(6,1,24)} \)

\( \text{> a} \)

\( \text{[1] 1 1 1 1 1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 Levels: 1 2 3 4 } \)

\( \text{> b} \)

\( \text{[1] 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 Levels: 1 2 3 4 5 6 } \)

\( \text{> lm(y~a+b-1) #1.} \)

\( \text{a1 a2 a3 a4 a5 a6} \)

\( \text{0.266 0.235 0.246 -0.379 -0.102 -0.319 0.913 -0.227 -0.125} \)

\( \mu \text{ and b1 ?} \)

\( \text{> lm(y~a+b) #2} \)

\( \# \text{ or} \)

\( \# \text{ lm(y~a+b, contrasts =c("contr.treatment", "contr.poly")}) \)

\( \text{Intercept a1 a2 a3 a4 a5 a6} \)

\( \text{0.268 -0.031 -0.020 -0.645 -0.102 -0.319 0.913 -0.227 -0.125} \)

\( a1, b1 ? \)

\( \text{> options(contrasts =c("contr.sum", "contr.poly")}) #3.} \)

\( \text{> lm(y~a+b)} \)

\( \text{Intercept a1 a2 a3 a4 a5 a6} \)

\( \text{0.115 0.174 0.143 0.154 -0.023 -0.125 -0.342 0.890 -0.250} \)

\( a4, b6 ? \)

**Relation between these three ?**
\[ \hat{E}(Y_{ij}) = \text{intercept} + ai + bj \text{ same?} \]

<table>
<thead>
<tr>
<th>(i)</th>
<th>(int.) = 0</th>
<th>(b1 = 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(a1)</td>
<td>(a2)</td>
</tr>
<tr>
<td>0</td>
<td>0.27</td>
<td>0.24</td>
</tr>
<tr>
<td>2</td>
<td>(a1 = 0)</td>
<td>(b1 = 0)</td>
</tr>
<tr>
<td>3</td>
<td>(a4 =?)</td>
<td>(b6 =?)</td>
</tr>
</tbody>
</table>

\[ \hat{E}(Y_{11}) = \begin{cases} 0 + 0.266 + 0 & \text{from #1} \\ 0.268 + 0 + 0 & \text{from #2} \\ 0.115 + 0.174 - 0.023 = 0.266 & \text{from #3.} \end{cases} \]

Are they the same?

What is \(X\), \(\beta\) and \(\hat{\beta}\) in the model \(Y = X'\beta + \epsilon\) for \(\text{lm}(y \sim a + b)\) in Example 3?

\[
X = \begin{pmatrix}
1 \\
1(a=1) \\
1(a=2) \\
1(a=3) \\
1(a=4) \\
1(b=1) \\
1(b=2) \\
1(b=3) \\
1(b=4) \\
1(b=5) \\
1(b=6)
\end{pmatrix}
\]

The sample size is \(n = 24\), \(\hat{y} = X'\hat{\beta} = 0.27 + 01(a = 1) - 0.031(a = 2) - 0.021(a = 3) + \cdots + 01(b = 1) + \cdots - 0.231(b = 5) - 0.131(b = 6)\).

What is \(\beta\) and \(X\) here?

\[
\beta = (int, a2, a3, a4, b2, \ldots, b6)' \text{ and } X = \begin{pmatrix}
a2 & a3 & a4 & b2 & \cdots & b6 \\
1 & 0 & 0 & 0 & 0 & \cdots & 0 & 2nd coordi \\
1 & 0 & 0 & 0 & 1 & \cdots & 0 & 3rd coordi \\
\vdots \\
1 & 0 & 0 & 1 & 0 & \cdots & 1 & last coordi
\end{pmatrix}
\]

> a
> [1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 3 4 4 4 4
> b
> [1] 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

Homework. 10.2. What is \(X\) and \(\beta\) for \(\hat{\beta} = (X'X)^{-1}X'Y\) in \(\text{lm}(y \sim a + b - 1)\) in Example 3?

Example 3 (continued).

Another way to generate the same type of data:
> y=rnorm(24)
> a=rep(1,6)
> a=c(a,a+1,a+2,a+3)
> b=rep(1:6,4)
> lm(y~a+b) # #A
(Output)
(Intercept) a b
> a=factor(a)
> b=factor(b)
> lm(y~a+b) # #B
(Output)
(Intercept) a2 a3 a4 b2 b3 b4 b5 b6

What is \( X \) and \( \beta \) for \( \hat{\beta} = (X'X)^{-1}X'Y \) in \# A?

What is \( X \) and \( \beta \) for \( \hat{\beta} = (X'X)^{-1}X'Y \) in \# B?

What is the difference between outcomes \# A and \# B?

> lm(y~a+b-1) #1.
> lm(y~a+b) #2
> options(contrasts =c("contr.sum", "contr.poly")) #3.
> lm(y~a+b)

Which is the way as in Example 3? \#1 or \#2?

What do you expect the estimates before seeing output?

> summary(lm(y~a+b)) # justify the answer to the question

Homework 10.3.
1. Repeat Example 3 once youself and answer the questions there.
2. Mimic Example 3 (continued) by inserting \( y=1+2a+y \) before \( \text{lm}(y\sim a+b) \).
   Then ask youself relevant questions and answer them.

Announcement:
1. Please put your name on top of your pdf files.
2. List the 5 elements whenever do a test.
3. About the reason of the correction of 502 exam.
4. About answer to 2.2(h)

p124. \# 1. 1. A civil engineer tested two different types (A and B) of a special reinforced concrete beam. He made nine test beams (five A’s and 4 B’s) and mea- sured the strength of each. From the following strength data (in coded units) he wants to decide whether there is any real difference between the two types. What assumptions does he need to draw conclusions? What might he conclude? Give reasons for your answers.
   Type A: 67 80 106 83 89
   Type B: 45 71 87 53