

Analysis of Variance (cont.)

Presenters 10/14/2016

7:50-8:00	Group 16
8:00 -8:10	Group 7
8:10-8:20	Group 3
8:20-8:30	Group 19

16	Wang, Yiren	Li, Yanjin	Wen, Litong	Lu, Yicheng	Sun, Xuechun	Wang, Yue	Zhao, Fei
----	-------------	------------	-------------	-------------	--------------	-----------	-----------

7	Zhang, Yunyi	Qin, Yunlin	Liu, Haojiang	Fei, Yang	Kim, Hayoung	Cho, Younhyuk	Song, Hyounghook	Fan, Yang	Wang, Weitong
---	--------------	-------------	---------------	-----------	--------------	---------------	------------------	-----------	---------------

3	Bai, Silvia	Cai, Weipan	Dai, Di	Han, Siqi	Xie, Tianzhao	Yang, Mengting	Zeng, Cen	Zhuang, Shiyu	Zhang, Yifan	Huang, Rui
---	-------------	-------------	---------	-----------	---------------	----------------	-----------	---------------	--------------	------------

19	Xiao, Han	Cui, Han	Wang, Danmo	Feng, Jingjing	Sheng, Tian	Zhao, Ran	Zhang, Yimin	Zhi, Chi	Xin, Xieke	Xue, Lifu
----	-----------	----------	-------------	----------------	-------------	-----------	--------------	----------	------------	-----------

One-Way ANOVA

- **Data:** $Y_{ij}, i = 1, \dots, I; j = 1, \dots, n_i$

Y_{ij} is distributed normally with mean μ_i , and constant variance

Wish to test:

$$H_0 : \mu_1 = \dots = \mu_I$$

against the alternative $H_1 : \mu_i \neq \mu_j$ for at least one pair $(i, j), i \neq j$.

One approach to model the data is to use the following formulation,

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

One-Way ANOVA Table

Source of Variation	df	Sum of Squares	Mean Squares
Treatment	I-1	SSTrt	SSTrt/(I-1)
Error	N-I	SSE	SSE/N-I
Total	N-1	SST	

A test statistic for H_o may be constructed based on the ratio

$$F = MSTrt/MSE$$

which, under H_o and model assumptions, has an $F_{I-1, N-I}$ distribution, where $N = \sum_i n_i$.

Two-Way ANOVA

ANOVA model is given by:

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

where $i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, n_{ij}$,
and ϵ_{ijk} are typically assumed to be i.i.d. $N(0, \sigma^2)$.

OLS estimator:

$$\hat{\mu}_{ij} = \bar{Y}_{ij}.$$

It is often more convenient to use the following
alternative formulation

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

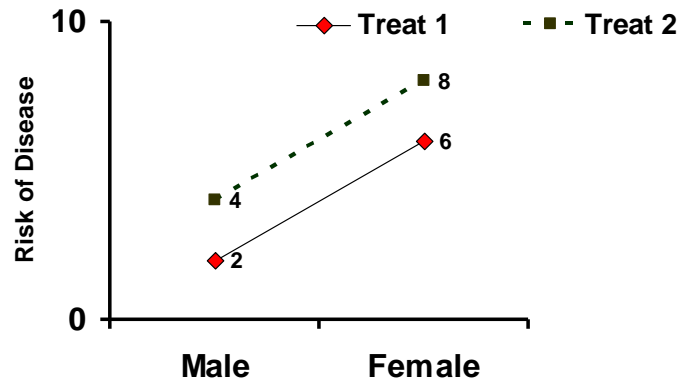
Two-Way ANOVA Table: Balanced Design

Source	df	SS	MS
A	I-1	SSA	$MSA=SSA/(I-1)$
B	J-1	SSB	$MSB=SSB/(J-1)$
AB	$(I-1)(J-1)$	SSAB	$MSAB/(I-1)(J-1)$
Error	$IJ(n-1)$	SSE	$SSE/IJ(n-1)$
Total	$IJN-1$	SST	

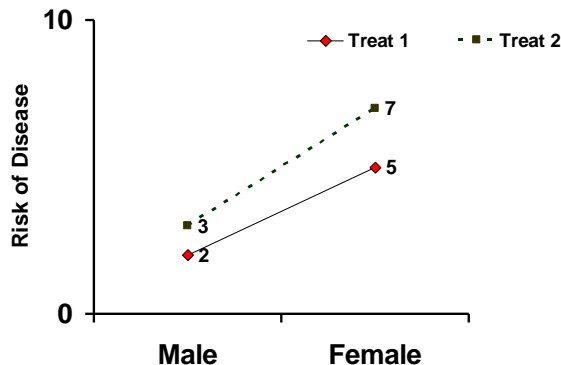
When the interaction term is significant:

- Evaluate the nature and strength of the interaction.

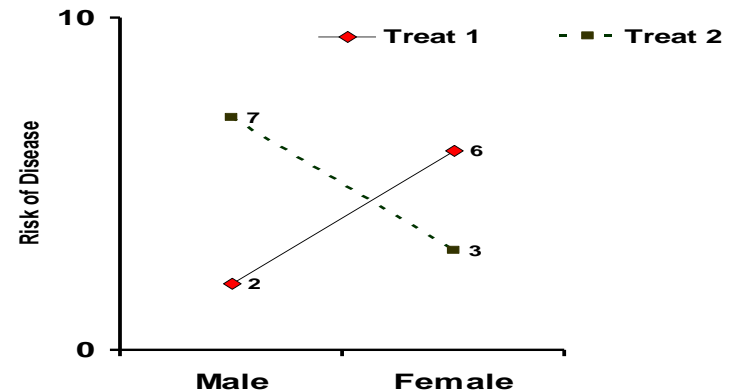
No Interaction



Quantitative Interaction



Qualitative Interaction



Suppose X_{ij} is a covariate of interest, and consider the ANOCVA model

$$Y_{ij} = \mu_i + \beta X_{ij} + \epsilon_{ij}$$

Note that

$$\bar{Y}_{i.} \approx \hat{\mu}_i + \beta \bar{X}_{i.}$$

Thus, comparing μ_i and μ_k based on $\bar{Y}_{i.} - \bar{Y}_{k.}$ would be inappropriate unless $\bar{X}_{i.} = \bar{X}_{k.}$. So , comparison is generally performed at a common value of X, say $\bar{X}_{..}$. For convenience, let

$$Y_{ij} = \mu_i + \beta(X_{ij} - \bar{X}_{..}) + \epsilon_{ij}$$

Adjusted mean (a.k.a. LS Mean)

$$\hat{\mu}_i = \hat{Y}_{i.} - \hat{\beta}(\bar{X}_{i.} - \bar{X}_{..})$$

In the above

$$\hat{\beta} = \frac{\sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{Y}_{i.})(X_{ij} - \bar{X}_{i.})}{\sum_{i=1}^I \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2}$$

$$\hat{\sigma}^2 = \frac{1}{N - I - 1} \sum_{ij} (Y_{ij} - \hat{\mu}_i - \hat{\beta}(\bar{X}_{i.} - \bar{X}_{..}))^2$$

$$var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^I \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2}$$

Other ANOVA Models

Example 1. Study on effects of teaching methods on student performance. All 5 teachers in a given school (i.e., 5 different methods) included in a study, each assigned 10 students at random. At the end of a training period, scores on a standardized test recorded.

Example 2. In another school there are 100 teachers. 5 teachers chosen at random (i.e., 5 different methods), and each assigned 10 students at random. At the end of a training period, scores on a standardized test recorded.

NB: The first school corresponds to Fixed Effects ANOVA (Model I), since all the levels of the factor “Teacher” are in the study.

The levels of the factor “Teacher” in the second case are random.
Corresponds to Random Effects ANOVA (Model II)

One Way Random Effects ANOVA Model (Model II)

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

where μ_i are iid $N(\mu, \sigma_A^2)$, ϵ_{ij} are also iid $N(0, \sigma_e)$, and independent of μ_i . Note that if all teachers teach the same way, $\mu_i = \mu$, and hence $\sigma_A^2 = 0$.

A test for treatment difference may then be formulated in term sof

$$H_0 : \sigma_A^2 = 0$$

vs

$$H_1 : \sigma_A^2 > 0$$

Total sum of squares (SST) decomposed:

- Sum of squares due to treatment (SSA)+
- Error sum of squares (SSE), where

$$SSA = \sum_i n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$SSE = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2$$

Assuming all $n_i = n$, it can also be shown that,

$$E[MSA] = E[SSA/(I - 1)] = n\sigma_A^2 + \sigma_e^2.$$

$$E(MSE) = \sigma_e^2.$$

To test

$$H_0 : \sigma_A^2 = 0$$

vs

$$H_1 : \sigma_A^2 > 0$$

$$F = \frac{MSA}{MSE}$$

which under H_0 has an $F_{I-1, N-I}$ distribution

Remark: In the One-way case, the test similar
to that of Fixed Effects model

Reading assignment: Confidence intervals for σ_A^2 and $\frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}$

Two-Factor Models: Model II

$$Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \epsilon_{ijk}$$

where $i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, n_{ij};$
 $a_i, b_j, (ab)_{ij}, \epsilon_{ijk}$ are mutually independent normal random variables, with mean 0, and respective variances: $\sigma_A^2, \sigma_B^2, \sigma_{AB}^2, \sigma_e^2$.

When the design is balanced, we have

$$SST = SSA + SSB + SSAB + SSE$$

where

$$SSA = nJ \sum_i (\bar{Y}_{i..} - \bar{Y}_{..})^2$$

$$SSB = nI \sum_j (\bar{Y}_{.j.} - \bar{Y}_{..})^2$$

$$SSAB = n \sum_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{..})^2$$

$$SSE = \sum_{ijk} (Y_{ijk} - \bar{Y}_{ij.})^2$$

and

$$SST = \sum_{ijk} (Y_{ijk} - \bar{Y}_{..})^2$$

$$E(MSA) = \sigma_e^2 + n\sigma_{AB}^2 + nJ\sigma_A^2$$

$$E(MSB) = \sigma_e^2 + n\sigma_{AB}^2 + NI\sigma_B^2$$

$$E(MSAB) = \sigma_e^2 + n\sigma_{AB}^2$$

and

$$E(MSE) = \sigma_e^2$$

Two-Way Random Effects Model ANOVA Table

Source	df	SS	MS	EMS
A	I-1	SSA	SSA/(I-1)	$\sigma_e^2 + n\sigma_{AB}^2 + nJ\sigma_A^2$
B	J-1	SSB	SSB/(J-1)	$\sigma_e^2 + n\sigma_{AB}^2 + nI\sigma_B^2$
AB	(I-1)(J-1)	SSAB	SSAB/(I-1)(J-1)	$\sigma_e^2 + n\sigma_{AB}^2$
Error	(n-1)IJ	SSE	SSE / (n-1)IJ	σ_e^2
TOTAL	N-1	SST		

Mixed Effects Models: Model III

Example: Suppose three treatments are to be compared.

- 5 hospitals selected at random from a district of 100 hospitals
- 5 patients are randomly assigned to each treatment in each hospital
- The factor "treatment" is fixed, since all the levels are included in the study.
- "Hospital" is random, since the levels are a random sample.

Two-way Mixed Effects Model

$$Y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + \epsilon_{ijk}$$

where μ and α_i are constant,

$$b_j \sim N(\mu, \sigma_b^2)$$

$$(\alpha b)_{ij} \sim N(0, \frac{(I-1)}{I} \sigma_{AB}^2),$$

ϵ_{ijk} is $N(0, \sigma_e^2)$, and all the random quantities are mutually independent.

Two -Way Mixed Effects ANOVA Table

Source	df	SS	MS	EMS
A (Fixed)	I-1	SSA	SSA/(I-1)	$\sigma_e^2 + \frac{nJ}{J-1} \sum_i \alpha_i^2 + n\sigma_{AB}^2$
B (Random)	J-1	SSB	SSB/(J-1)	$\sigma_e^2 + nI\sigma_B^2$
AB (Random)	(I-1)(J-1)	SSAB	SSAB/(I-1)(J-1)	$\sigma_e^2 + n\sigma_{AB}^2$
Error (Random)	(n-1)IJ	SSE	SSE / (n-1)IJ	σ_e^2
TOTAL	N-1	SST		

Two -Way Mixed Effects ANOVA Table: Balanced Design

Source	df	SS	MS	EMS
A (Fixed)	I-1	SSA	SSA/(I-1)	$\sigma_e^2 + \frac{nJ}{I-1} \sum_i \alpha_i^2 + n\sigma_{AB}^2$
B (Random)	J-1	SSB	SSB/(J-1)	$\sigma_e^2 + nI\sigma_B^2$
AB (Random)	(I-1)(J-1)	SSAB	SSAB/(I-1)(J-1)	$\sigma_e^2 + n\sigma_{AB}^2$
Error (Random)	(n-1)IJ	SSE	SSE / (n-1)IJ	σ_e^2
TOTAL	N-1	SST		

Inference about the fixed effect (A)

$$H_0 : \alpha_i = 0$$

$$F = \frac{MSA}{MSAB}$$

which under H_0 has an $F_{I-1, (I-1)(J-1)}$ distribution

If the interaction term is not significant, then an appropriate test statistic, based on the reduced model, is

$$F = \frac{MSA}{MSAB + MSE}$$

Null distribution?

Inference about the random effect (B)

$$H_0 : \beta_j = 0$$

$$F = \frac{MSB}{MSE}$$

which under H_o has an $F_{I-1,(n-1)IJ}$ distribution

Two -Way Mixed Effects ANOVA Table: Balanced Design

Source	df	SS	MS	EMS
A (Fixed)	I-1	SSA	SSA/(I-1)	$\sigma_e^2 + \frac{nJ}{I-1} \sum_i \alpha_i^2 + n\sigma_{AB}^2$
B (Random)	J-1	SSB	SSB/(J-1)	$\sigma_e^2 + nI\sigma_B^2$
AB (Random)	(I-1)(J-1)	SSAB	SSAB/((I-1)(J-1))	$\sigma_e^2 + n\sigma_{AB}^2$
Error (Random)	(n-1)IJ	SSE	SSE /((n-1)IJ)	σ_e^2
TOTAL	N-1	SST		

Example. Consider a study comparing 3 teaching methods.

- A random sample of 5 schools selected
- From each school 3 teachers randomly chosen.
 - Each teacher was then assigned a teaching method at random and asked to apply it in their class of about 20 students each.
- The scores (Y_{ij}) of each student were then recorded at the end of the semester.

In this example, each level of the factor "teacher" occurs with only one level of "school", and each of the 15 levels is meaningful only given the level of "school". The factor "teacher" is said to be *nested* within "school".

Two-Way Nested Designs

When both factors are random,

$$Y_{ijk} = \mu + a_i + b_{j(i)} + \epsilon_{ijk}$$

where $b_{j(i)}$ denotes the effect of B at the j 'th level, when A is at the i 'th level. Further a_i , $b_{j(i)}$ and ϵ_{ijk} are mutually independent normal

random variables, with mean 0 and, respective variances:

$$\sigma_A^2, \sigma_{B(A)}^2 \text{ and } \sigma_e^2$$

For the balanced case, the sums of squares are given by:

$$SSA = Jn \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$SSB(A) = n \sum_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i..})^2$$

$$SSE = \sum_{ijk} (\bar{Y}_{ijk} - \bar{Y}_{ij.})^2$$

Two -Way Nested Random Effects ANOVA Table

Source	df	SS	MS	EMS
A (Random)	I-1	SSA	SSA/(I-1)	$\sigma_e^2 + nJ\sigma_A^2 + n\sigma_{B(A)}^2$
B(A) (Random)	I(J-1)	SSB(A)	SSB(A)/I(J-1)	$\sigma_e^2 + n\sigma_{B(A)}^2$
Error (Random)	(n-1)IJ	SSE	SSE /(n-1)IJ	σ_e^2
TOTAL	N-1	SST		

When A is fixed and B is random, and B(A):

Source	df	SS	MS	EMS
A (Fixed)	I-1	SSA	SSA/(I-1)	$\sigma_e^2 + nJ \frac{J_n}{I-1} \sum_i \alpha_i^2$
B(A) (Random)	I(J-1)	SSB(A)	SSB(A)/I(J-1)	$\sigma_e^2 + \frac{n}{I(n-1)} \sum_{ij} \beta_{j(i)}^2$
Error (Random)	(n-1)IJ	SSE	SSE / (n-1)IJ	σ_e^2
TOTAL	N-1	SST		

Repeated Measures Design

Example. Consider a clinical trial comparing three treatment groups. Subjects were randomized to each treatment, and measurements were taken at weekly.

Observation taken over time on the same subject may be correlated.

- The usual ANOVA will not be applicable to this case.

Repeated Measures Design

Let Y_{ijk} denote the measurement on the k 'th subject, assigned to treatment i , and taken at time j .

$$Y_{ijk} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + S(\alpha)_{k(i)} + \epsilon_{ijk}$$

where α_i is the i 'th treatment effect, τ is time effect, and $S(\alpha)$ stands for subject nested in treatment.

Since the error terms may be correlated, several correlation structures may be possible:

- Compound symmetry (i.e., equal correlations)

$$\text{Corr}(Y_{ijk}, Y_{ilk}) = \rho, \quad \forall j \neq l$$

- AR(1)
- Unstructured

The following is a decomposition of the Total Sum of Squares (SST)

$$SSTreatment = J \sum_i n_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$SSTime = n_+ \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

$$SSTreat * Time = \sum_{i,j} n_{ij} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$$SSS(Treat) = J \sum_{i,k} n_{i.} (\bar{Y}_{i.k} - \bar{Y}_{i..})^2$$

$$SSE = \sum_{i,j,k} (Y_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{ij.} + \bar{Y}_{i..})^2$$

$$SST = \sum_{i,j,k} (Y_{ijk} - \bar{Y}_{i..})^2$$

Repeated Measures ANOVA Table

Source	df	SS	F Statistic
Treatment	I-1	SS Treat	$\frac{MSTreat}{MS\ S(Treat)}$
Time	J-1	SS Time	$\frac{MSTime}{MSE}$
Treat*Time	(I-1)(J-1)	SS Time*Time	$\frac{MSTreat*Time}{MSE}$
S (Treat)	$\sum_i (n_i - 1) = n_+ - I$	SS S(Treat)	$\frac{MSS(Treat)}{MSE}$
Error	$(\sum_i n_i - J)(J - 1)$	SSE	
Total	SST		

Remarks:

- $\frac{MSTime}{MSE}$ and $\frac{MSTreat*Time}{MSE}$ may not have an F distribution with the usual degrees of freedom. Indeed, actual significance may be less strong than given by table.

Under certain conditions (Huynh & Feldt), the distributions are F (e.g., when the correlation structure is independent or exchangeable or AR(1)).

- More generally, models that take into account the correlation structure must be used (SAS PROC MIXED).

Example: Repeated measures

Subjects randomized to either Group 1, 2 or 3.

For each subject, response measured at Time 1, 2 and 3,
following randomization and treatment.

ID Grp T Y

1 1 1 15

1 1 2 29

1 1 3 25

2 1 1 11

2 1 2 28

2 1 3 27

3 2 1 14

3 2 2 12

3 2 3 16

4 2 1 11

4 2 2 10

4 2 3 13

5 3 1 21

5 3 2 22

5 3 3 19

6 3 1 14

6 3 2 18

6 3 3 16

7 3 1 13

7 3 2 10

7 3 3 11

Grp = Group. T = time.

```

proc glm data=repeat;
class ID Group Time;
model Y=Group ID(Group)Time group*time/ss3;
test h=group e=id(group);
run;

```

Source	DF	SS3	MS	F	Pr > F
Group	2	303	152	50.50	<.0001
ID(Group)	4	143	35.7	11.90	0.0019
Time	2	105	52.6	17.53	0.0012
Group*Time	4	211.6	52.9	17.63	0.0005

Tests of Hypotheses Using the Type III MS for ID(Group) as an Error Term

Source	DF	Type III SS	MS	F Value	Pr > F
Group	2	303	152	4.24	0.1027

Reading Assignment :

SAS' PROC GLM gives Type I - Type III SS.

Example: Model $Y = A + B + A * B$

Type I SS: Order-dependent (hierarchical, sequential). Each effect is adjusted for all other effects that appear earlier (to the left) in the model, but not for any effects that appear later in the model.

Type II SS are the reduction in the SSE due to adding the effect to a model that contains all other effects except effects that contain the effect being tested

Types III SS are each adjusted for all other effects in the model, regardless of order.

```
proc glm data=repeat;
class ID Group Time;
model Y=Group ID(Group)Time group*time/ss1;
test h=group e=id(group);
run;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Group	2	302.9761905	151.4880952	50.50	<.0001
ID(Group)	4	142.8333333	35.7083333	11.90	0.0019
Time	2	80.3809524	40.1904762	13.40	0.0028
Group*Time	4	211.6190476	52.9047619	17.63	0.0005

Tests of Hypotheses Using the Type I MS for ID(Group) as an Error Term

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Group	2	302.9761905	151.4880952	4.24	0.1027

Reading assignment:

Compare the above results with the results obtained using the R function

aov(Y ~ Group*Time+Error(ID))

```

proc mixed data=repeat;
class ID Group Time;
model Y=Group Time group*time;
repeated/type=cs subject=ID;
run;

```

Type 3 Tests of Fixed Effects

	Num	Den		
Effect	DF	DF	F Value	Pr > F
Group	2	4	4.24	0.1027
Time	2	8	17.53	0.0012
Group*Time	4	8	17.63	0.0005

Problem Set

Consider the *ChickWeight* data in R. The body weights of the chicks were measured at birth (i.e., time=0) and every second day thereafter until day 20. They were also measured on day 21. There were four groups of chicks on different protein diets.

Perform an appropriate repeated measures ANOVA to determine whether there is a significant difference in the mean weights of the four groups using the measurements on Days 4, 8, 12, 16 and 20.

1. Do the analyses assuming compound symmetry, unstructured and AR(1) covariance structures and compare the results.
2. In each case determine whether it might be appropriate to adjust for Birth Weight
3. Check the validity of your assumptions

	Name: Last, First	Name: Last, First	Name: Last, First	Name: Last, First	Name: Last, First	Name: Last, First	Name: Last, First	Name: Last, First	Name: Last, First	Name: Last, First
1	Wang, Yicheng	Xue, Yichao	He, Hanming	Zhang, Tian	Yi, Ziyun	Tao, Mengyuan	Tian, Jiani	Wang, Shiyu		
2	Li, Jingwei	Liu, Ao	Wu, Xiangyu	Yang, Yutong	Aoyuan, Liao	Yi, Jian	Zhang, Wanyi	Wang, Jia	Gao, Duanhong	Shi, Hanqing
3	Bai, Silvia	Cai, Weipan	Dai, Di	Han, Siqi	Xie, Tianzhao	Yang, Mengting	Zeng, Cen	Zhuang, Shiyu	Zhang, Yifan	Huang, Rui
4	Zuo, Nian Yao	Chen, Jiayang	Huang, Xian gkai	Tang, Mingz hen	He, Jin	Li, Weihang	Gao, Fei	Qin, Liwen	Zuo, Zhaoy u	
5	Yang, Chuhan	Duan, Ziyang	Chen, Jinglin	Huang, Yirui	Zhu, Ming	Liu, Zhaoze	Yin, Qing	Zhang, Baizheng	Yue, Wenshu	Zeng, Neng
6	You, Guanzhong	Wang, Lu	Zhou, Xingyu	Luan, Sitao	Bao, Wenhang	Liu, Chang	Yang, Tianmeng	Zhu, Feiran	Chen, Jie	
7	Zhang, Yunyi	Qin, Yunlin	Liu, Haojiang	Fei, Yang	Kim, Hayoung	Cho, Younhyuk	Song, Hyoungmok	Fan, Yang	Wang, Weitong	
8	Jin, Chengzhe	Liu, Youzhu	Yu, Xingzao	Zhu, Ying	You, Jiwen	Li, Linna	Lyu, Yihua	Ye, Hexiu	Lin, Chi-Heng	Jiang, Bo
9	Wang, Suling	Cheng, Tianyuan	Li, Cheng	Wang, Han	Shang, Renfei	Yao, Wei	Yu, Zhao			
10	Chen, Haoyang	lin	Zhou, Longwei	Chen, Zachary	Huang, Biyue	Qian, Quan				
11	Nian, Yiqun	An, Huilong	Lin, Zida	Zhang, Xiaohan	Hu, Yifei	Qin, Yu	Dai, Peijun	Gu, Kexin		
12	Gao, Chenying	Yao, Weichi								
13	Sun, Yuhan	Zhou, Jingying	Jiang, Chencheng	Teng, Yueying	Wang, Yanran	Gu, Xinghao	Chen, Ying	Meng, Ziwei		
14	Jin, Zhaoyan	Ji, Chenlu	Wang, Jiayi	Lu, Ke	Zhang, Xuan	Zhang, Chi	Lang, Yifei	Yu, Tianying		
15	Chen, Geer	Wang, Bo	Zhao, Edward	Zhang, Keren	Bao, Yu	Ren, Ruoxi	Zhang, Yunyi	Wang, Xiaoxiao	Zhou, Xiaoyu	Sun, Sirui

