STAT 212: Principles of Statistics II

Lecture Notes: Chapter 2 (Part B)

Two-Way ANOVA + Experimental Design

Sharmistha Guha

Dept. of Statistics

Texas A&M University

Fall 2022

Multifactor ANOVA

Often one wishes to do an experiment to assess the effects of *two or more* factors on a response.

Example (Two-factor or two-way ANOVA):

- Response = orange yield (in lbs.)
- Factor A: *orange variety*
- Factor B: *type of pesticide used*

Important questions:

1. Do either or both factors have an effect on yield?
2. Is there an **interaction** between *orange variety* and *pesticide type*?

Example of an interaction:

> *Average yield of variety 1 oranges is the same regardless of which pesticide is used, but average yield of variety 2 oranges is higher with pesticide 1 than pesticide 2.*

An interaction can be paraphrased as follows:

> *The type of effect that Factor A has on the response depends upon which level of Factor B is being used.*

Suppose there are a levels of factor A, b levels of factor B, and we obtain n observations for each of the ab combinations of factor levels.

The most basic model for a two-factor experiment is:

$$X_{ijk} = \mu_{ij} + \epsilon_{ijk},$$

with $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n$.

- X_{ijk} = response of the kth experimental unit treated with level i of Factor A and level j of Factor B.

- ϵ_{ijk}s are independent random variables each having a $N(0, \sigma^2)$ distribution.

- μ_{ij} = mean response of a population of experimental units treated with level i of Factor A and level j of Factor B.

Additive model: no interaction

Suppose that

$$\mu_{ij} = \mu + \alpha_i + \beta_j,$$

where $\sum_{i=1}^{a} \alpha_i = 0 = \sum_{j=1}^{b} \beta_j$.

In the most general model, there are ab parameters, i.e., means, to estimate, whereas in the additive model there are only $a + b - 1$ mean parameters.

When $a > 1$ and $b > 1$, $a + b - 1 < ab$.

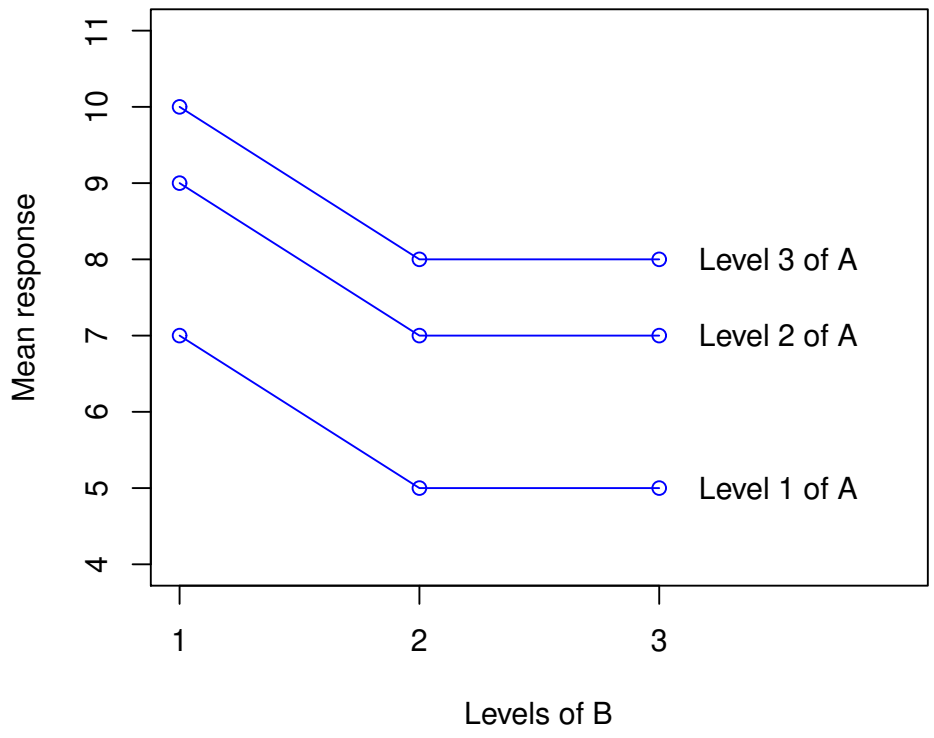
This simplicity comes with a price: *the additive model does not allow interactions*.

Fix factor B at level j and compare levels i and ℓ of factor A. In the additive model

$$\begin{aligned} \mu_{ij} - \mu_{\ell j} &= (\mu + \alpha_i + \beta_j) - (\mu + \alpha_\ell + \beta_j) \\ &= \alpha_i - \alpha_\ell. \end{aligned}$$

Notice that *the difference does not depend on* j.

In other words, *whatever effect Factor* A *has on the response, it is consistently the same across all levels of* B. Graphically this looks as follows:



Note: The lines are parallel, indicating no interaction.

Two-factor ANOVA with interaction

As before assume that

$$\begin{aligned} X_{ijk} &= \mu_{ij} + \epsilon_{ijk} \\ &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \end{aligned}$$

where

$$\mu = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij}.$$

Note that

$$\gamma_{ij} = \mu_{ij} - \left(\mu + \alpha_i + \beta_j \right).$$

A given γ_{ij} represents how much different μ_{ij} is from what it would be if the additive model held.

No interaction exists if and only if all the γ_{ij}s equal 0.

The two-factor model allowing for the possibility of interaction is

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

with $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n$.

The same assumptions as before hold for the error terms ϵ_{ijk}.

The first step in analyzing data from a two-factor ANOVA is to test whether there is an interaction between the two factors.

A two-factor experiment can be summarized with an ANOVA table.

First we define means:

$$\bar{X}_{ij\cdot} = \frac{1}{n}\sum_{k=1}^{n} X_{ijk}, \quad \bar{X}_{i\cdot\cdot} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} X_{ijk}$$

$$\bar{X}_{\cdot j\cdot} = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} X_{ijk} \quad \text{and}$$

$$\bar{X}_{\cdots} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} X_{ijk} \ .$$

The sums of squares are

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (X_{ijk} - \bar{X}_{\cdots})^2,$$

$$SSA = bn\sum_{i=1}^{a} (\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdots})^2,$$

$$SSB = an\sum_{j=1}^{b} (\bar{X}_{\cdot j\cdot} - \bar{X}_{\cdots})^2,$$

$$SSAB = n \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2$$

and

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (X_{ijk} - \bar{X}_{ij.})^2.$$

Important relationship:

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

ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Factor A	$a-1$	SSA	MSA	F_A
Factor B	$b-1$	SSB	MSB	F_B
Interaction	$(a-1)(b-1)$	$SSAB$	$MSAB$	F_{AB}
Error	$ab(n-1)$	SSE	MSE	
Total	$abn-1$	SST		

Each mean square is the corresponding sum of squares divided by its degrees of freedom. For example,

$$MSA = \frac{SSA}{a-1}.$$

Each F-statistic is the corresponding mean square divided by MSE. For example,

$$F_{AB} = \frac{MSAB}{MSE}.$$

Hypotheses that may be tested in the two-factor experiment are

$$H_{0AB} : \gamma_{ij} = 0 \quad \text{for all } i, j,$$

$$H_{0A} : \alpha_i = 0 \quad i = 1, \ldots, a$$

and

$$H_{0B} : \beta_j = 0 \quad j = 1, \ldots, b.$$

H_{0AB} is tested with F_{AB}, which has the F-distribution with degrees of freedom $(a-1)(b-1)$ and $ab(n-1)$ when there is no interaction.

H_{0AB} is rejected at level of significance α if

$$F_{AB} \geq F_{(a-1)(b-1),ab(n-1);\alpha}.$$

Recall the definition of $SSAB$. The term

$$\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdots}$$

equals

$$\bar{X}_{ij\cdot} - [\bar{X}_{\cdots} + (\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdots}) + (\bar{X}_{\cdot j\cdot} - \bar{X}_{\cdots})],$$

which estimates the interaction term

$$\mu_{ij} - (\mu + \alpha_i + \beta_j).$$

A "large" value of $SSAB$ is thus evidence of an interaction.

Strategy for testing hypotheses

1. Test H_{0AB} as described on p. 151N.

2. If H_{0AB} is rejected, try to explain the nature of the interaction. Looking at a plot of means is a useful device in this regard. *Do not test the hypotheses H_{0A} and H_{0B} since doing so is potentially misleading.*

3. If H_{0AB} is not rejected, then test H_{0A} and H_{0B} using the F-statistics F_A and F_B.

Using R to do two-factor ANOVA

- The data file should have three columns: one column is all the values of X, one column indicates which level of factor A each X belongs to, and one indicates which level of factor B X belongs to.

- Suppose your data is in an R dataframe called `Data`, whose columns are called

  `Response`, `FactorA` and `FactorB`.

  Use the following commands:

```
fit = aov(Response ~ FactorA * FactorB,
          data = Data)
anova(fit)
```

Example 13: *Effect of planting date and fertilizer type on soybean yield*

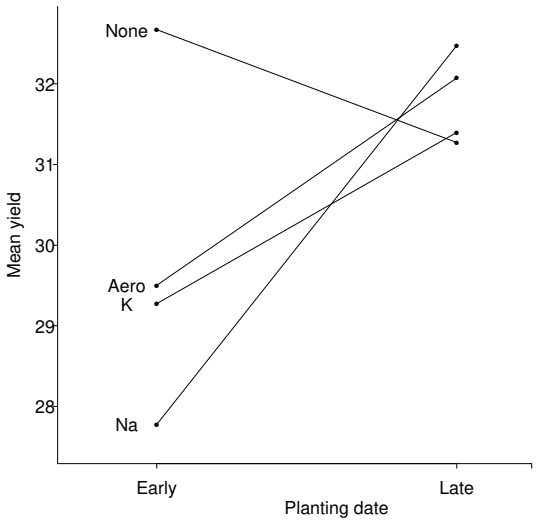
Factor A: Planting date (early and late)

Factor B: Fertilizer (none, aero, sodium and potassium)

For each combination of planting date and fertilizer, four plots were randomly selected and planted with soybeans.

Response: Soybean yield in lbs.

$$a = 2, \quad b = 4 \quad n = 4$$



ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Planting date	1	32	32	10.414
Fertilizer	3	16.40	5.47	1.78
Interaction	3	38.39	12.80	4.17
Error	24	73.75	3.07	
Total	31	160.54		

Test for interaction:

$$F_{AB} = 4.17 > 3.01 = F_{3,24;0.05}$$

So, it's reasonable to conclude that there is an interaction. Looking at the means plot on p. 154N, the most obvious explanation for the interaction is as follows:

When planting late, there don't seem to be big differences in yield between any of the fertilizers. However, when planting early, using no fertilizer results in higher average yield than any of the fertilizers.

We may do *quantitative* comparisons using Tukey's procedure. When we find a significant interaction, we may compare the difference between any two *treatment* means with

$$Q_{\alpha,ab,ab(n-1)}\sqrt{\frac{MSE}{n}}.$$

A *treatment* mean is $\bar{X}_{ij\cdot}$, i.e., the average response for a given combination of factor A and B levels.

In general there are ab treatment means.

In our example,

$$\begin{aligned} Q_{.05,2(4),2(4)(4-1)}\sqrt{\frac{MSE}{n}} &= 4.68\sqrt{\frac{3.0727}{4}} \\ &= 4.10. \end{aligned}$$

It turns out that Early/Sodium results in significantly smaller mean yield than any of Early/None, Late/Na and Late/Aero. However, no other pair of treatments differ significantly with respect to mean yield.

In R we may obtain the results of Tukey's procedure using either one of the following commands:

```
TukeyHSD(aov(Response~FactorA*FactorB))
```

or

```
plot(TukeyHSD(aov(Response~FactorA*FactorB)))
```.

Suppose we fail to reject the hypothesis of no interaction.

Then test both H_{0A} and H_{0B} using F_A and F_B, respectively.

H_{0A} is rejected if

$$F_A \geq F_{a-1,ab(n-1);\alpha}$$

and H_{0B} is rejected if

$$F_B \geq F_{b-1,ab(n-1);\alpha}.$$

If either of these hypotheses is rejected, then one may use multiple comparisons to compare the levels of a given factor.

For example, if we reject H_{0A}, then we may use Tukey's procedure as follows. Conclude that levels i and ℓ of Factor A are significantly different if

$$|\bar{X}_{i\cdot\cdot} - \bar{X}_{\ell\cdot\cdot}| \geq Q_{\alpha,a,ab(n-1)}\sqrt{\frac{MSE}{bn}}.$$

If H_{0B} is rejected, then conclude that levels j and m of Factor B are significantly different if

$$|\bar{X}_{\cdot j \cdot} - \bar{X}_{\cdot m \cdot}| \geq Q_{\alpha,b,ab(n-1)} \sqrt{\frac{MSE}{an}}.$$

In R we may obtain Tukey's procedure for Factor A as follows:

```
fit=aov(Response~FactorA*FactorB)
TukeyHSD(fit,''FactorA'')
```

Analogously we could do Tukey's procedure for Factor B.

Experimental design

Often in doing an experiment one has choices as to how treatments are assigned to experimental subjects. *Design of experiment* refers to how the assignment is done.

- Completely randomized design (CRD)

- Randomized block design (RBD)

The CRD is the design (implicitly) assumed in our discussion of one- and two-factor experiments.

Suppose we have N experimental units, t treatments, and we want to administer treatment i to n_i experimental units.

Note: $n_1 + n_2 + \cdots + n_t = N$

In a CRD, we randomly assign n_1 units to treatment 1, n_2 units to treatment 2, and so forth.

In our one-factor experiment,

$$t = k.$$

Two-factor experiment:

$$t = ab, \quad n_1 = n_2 = \cdots = n_t = n$$

$$\text{and} \quad N = abn.$$

Randomized block design

Differences between experimental units usually lead to variance, or "noise," among responses. This makes it more difficult to detect when two treatments are different.

Suppose we want to see if a treatment is better than no treatment. Consider the following two scenarios:

1. We have 10 pairs of identical twins. Administer treatment to one twin in a pair and give the other a placebo. Do this for all 10 pairs.

2. We have 20 people. Randomly select 10 and give them the treatment. Give the placebo to the other ten.

Intuitively, the first seems better! Why?

The second scenario is a CRD and the first is a *block design*, where each pair of twins is a block.

In a RBD, similar experimental units are combined to form a *block*. All treatments of interest are then assigned randomly to the units within a block.

Our previous twins example would be a *randomized* block design if a coin is tossed to decide which of the two twins in a pair is to receive the treatment.

The motivation for a RBD is to, effectively, reduce variation in responses due to differences between experimental units. *By applying treatments to similar experimental units, differences between treatments become easier to detect.*

Analysis of data from RBD

Let X_{ij} be the response to treatment i in block j, where $i = 1, \ldots, k$ and $j = 1, \ldots, n$.

(So, we have k treatments and n blocks.)

The following statistical model is assumed:

$$X_{ij} = \mu + \alpha_i + b_j + \epsilon_{ij}, \quad \begin{array}{l} i = 1, \ldots, k, \\ j = 1, \ldots, n. \end{array}$$

- α_i: effect due to treatment i

- b_j: effect due to block j

Thinking back to the twins example,

- α_i accounts for a difference between treatment and placebo,

- b_j accounts for differences from one pair of twins to the next, and

- ϵ_{ij} accounts for any slight differences between twins in the same pair.

Treatment means:

$$\bar{X}_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}, \quad i = 1, \ldots, k$$

Block means:

$$\bar{X}_{\cdot j} = \frac{1}{k}\sum_{i=1}^{k} X_{ij}, \quad j = 1, \ldots, n$$

Grand mean:

$$\bar{X}_{\cdot\cdot} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}$$

Sums of squares:

$$SSTr = n\sum_{i=1}^{k}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$$

$$SSB = k \sum_{j=1}^{n} (\bar{X}_{.j} - \bar{X}_{..})^2$$

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - \bar{X}_{..})^2$$

$$SSE = SST - SSTr - SSB$$

ANOVA table

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	$k-1$	$SSTr$	$MSTr$	F_{Tr}
Blocks	$n-1$	SSB	MSB	F_B
Error	$(k-1)(n-1)$	SSE	MSE	
Total	$kn-1$	SST		

As always, each mean square is the corresponding sum of squares divided by its degrees of freedom.

We test $H_0 : \alpha_1 = \cdots = \alpha_k$ using the statistic $F_{Tr} = MSTr/MSE$. H_0 is rejected if

$$F_{Tr} \geq F_{k-1,(k-1)(n-1);\alpha}.$$

Can also use F_B to test for a block effect, although this is not of primary importance.

Example 14: *Analyzing data from an RBD*

An experiment was conducted to assess the effect of caffeine on the endurance of athletes. Nine well-conditioned cyclists were available for the study, and four caffeine dosages were considered: 0, 5, 9 and 13 mg.

The experiment was conducted over four days and each cyclist ended up receiving all four caffeine doses, one dose per day. The order in which the doses were administered was randomly determined for each cyclist.

The response variable was number of minutes of cycling between receiving the caffeine and exhaustion setting in.

Treatments: Caffeine dosages

Blocks: Cyclists

ANOVA table

Source of variation	SS	df	MS	F
Caffeine	933.122	3	311.041	5.917
Cyclists	5557.994	8	694.749	13.216
Error	1261.657	24	52.569	
Total	7752.773	35		

Since $5.917 > 4.72 = F_{3,24;0.01}$, it is reasonable to conclude that not all caffeine doses produce the same average time to exhaustion.

As usual, we may use Tukey's procedure to determine what pattern of differences exist among the doses.

Treatment means $\bar{X}_{i\cdot}$ and $\bar{X}_{\ell\cdot}$ are significantly different if

$$|\bar{X}_{i\cdot} - \bar{X}_{\ell\cdot}| \geq Q_{\alpha,k,(k-1)(n-1)}\sqrt{\frac{MSE}{n}}.$$

Caffeine dose	Mean time to exhaustion
0	46.4
5	57.7
13	58.1
9	58.7

$$\begin{aligned} Q_{.05,4,24}\sqrt{\frac{52.57}{9}} &= 3.90(2.417) \\ &= 9.43 \end{aligned}$$

- No caffeine leads to a lower average time to exhaustion than any of the other three doses.

- We can't reject the hypothesis that doses 5, 9 and 13 have the same average time to exhaustion.

In R, suppose that the response, treatment and block vectors are called `Response`, `Treatment` and `Block`, respectively. Then the following commands produces the ANOVA table and Tukey's procedure for an RBD:

```
fit=aov(Response~Treatment+Block)
anova(fit)
TukeyHSD(fit,``Treatment'')
```