# Contents



# 4 Analysis of Variance (ANOVA)

## 4.1 Introduction

The goal is to extend our previous results from two samples to more than two samples. In direct extension of the two-sample case, we can imagine that we collect samples from $T \geq 2$ populations.

Another scenario that fits in this framework is the case where a single sample of size $n$ is randomly assigned to $T$ treatments. In this case, we imagine $T$ *hypothetical* populations. The $t$th population is *hypothetically* subjected to treatment $t$. The samples randomly assigned to treatment $t$ are like a random sample from the $t$th population.

### 4.1.1 Terminology

**Definitions/Terminology**

Definition: *experimental unit*

> The experimental unit of a study is the object on which measurements are taken.

Experimental units can be people, computers, animals, classes, universities, *etc*. Let's consider some examples.

- Rats are randomly fed different iron supplements in their food, and iron retention in their bodies is measured 3 days later. The experimental unit is the rat.

- Rats live in cages, and all rats in the same cage eat from a communal food bowl. Food bowls were randomly dosed with different iron supplements, and all rats measured for iron retention 3 days later. The experimental unit is the cage, because rats within cages cannot be assigned to different treatments.
- A scientist wants to determine whether childhood exposure to TV commercials impacts obesity. They enroll random families in the study, recording TV usage of the household periodically over several years. During the study, they also record height and weight of children in the family. The experimental unit is the collection of all children in the household, since TV usage of individual children cannot be determined.
- When testing chemicals for teratogenesis (the propensity to cause birth defects), pregnant rats are subject to high doses and the pups are scored for birth defects. Because the individual pups *cannot* be assigned different treatments, the litter is an experimental unit.
- A rabbit is sacrificed and the retinas are used for scientific experiments on neurons (horrible thing, but true). To avoid waste, both retinas are used in separate, independent experiments. The experimental units are the retinas, but one should worry about the foced blocking on rabbit (see discussion of blocking later). If within each retina, individual amacrine cells are located and independently tested, then the experimental unit is the amacrine cell. There is blocking at two hierarchical levels: the rabbit, and the retina within the rabbit.

Definition: *factor*

> A factor is a variable that is controlled in an experiment. Distinct values of the factor are called *levels*.

Some examples of factors are: (1) time exposed to teratogenic chemical, (2) dose of teratogenic chemical, (3) temperature of greenhouse, (4) initial number of infectious agents in an agent-based model of an epidemic, *etc*.

Definition: *treatment*

> A treatment is a specific combination of factors to which experimental units may be exposed in an experiment.

An example is the dose and time of exposure to a teratogenic chemical.

ANOVA analyses can be classified in several ways. A **one-way** ANOVA considers a set of treatments caused by varying one factor. A **two-way** ANOVA consideres a set of treatments caused by varying two factors simultaneously. For example, if one designed a teratogenic study that varied the dose and exposure time to the chemical, then you are varying two factors, dose and time. One can expand to multi-way ANOVA by including additional factors.

If the populations (or treatments) included in the study are selected by the experimenter and inferences are to be made *only* about those populations, then the model is called a **fixed effects model**. If instead the populations are *representative* of a large collection of populations, many of which are not sampled, and the experimenter wishes to infer properties of *all* populations, then the model is called a **random effects model**. For example, if three antidepressant drugs and a control are applied to four random groups of patients in order to determine which of these drugs can reduce depression symptoms, then the mean treatment responses are fixed effects associated with these three treatments only. In contrast, if three, representative antidepressant drugs are applied to three random groups of patients to determine the side effects of taking antidepressants in general, then the observed means for the three drugs are random variables representative of the kinds of effects caused by any antidepressant drug, even those not included in the study.

We will focus on fixed effects ANOVA in these notes.

### 4.1.2 Data

The data associated with ANOVA might be summarized in a table of the following form:

| Treatment | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | $\cdots$ | $k$ |
| $y_{11}$ | $y_{21}$ | $y_{31}$ | $\cdots$ | $y_{k1}$ |
| $y_{12}$ | $y_{22}$ | $y_{32}$ | $\cdots$ | $y_{k2}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $\vdots$ | $y_{2n_2}$ | $y_{3n_3}$ | $\cdots$ | $\vdots$ |
| $y_{1n_1}$ | | | $\cdots$ | $y_{kn_k}$ |

**Example: heights of singers in a choir**

Suppose, for example, that you are studying the heights of singers in a choir. Your data table is below.

| Soprano | Alto | Tenor | Bass |
|---|---|---|---|
| 64 | 65 | 69 | 72 |
| 62 | 62 | 72 | 70 |
| 66 | 68 | 71 | 72 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 63 | 66 | 66 | 68 |
| 65 | 66 | 68 | 70 |
| 62 | 66 | 67 | 75 |
| 65 | 62 | 64 | 68 |
| 66 | 70 | | 71 |
| 62 | 65 | | 70 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 65 | 67 | | 70 |
| 66 | 66 | | 75 |
| 65 | 68 | | 72 |
| 62 | | | 66 |
| | | | 72 |
| | | | 70 |
| | | | 69 |

[Find the original data at the Data & Story Library.]

The "treatments" are actual populations in this case, i.e. the different singer types (soprano, alto, tenor, bass). In working with this data, you might be interested in determining whether the mean heights of all treatments are the same. Specifically, you might expect a significant difference in the mean heights of basses and sopranos, because most, if not all, of the former are male, and the latter are female.

**First Step**

The first step in any data analysis is to plot the data. Dr. Zhu introduced you to the `R` functions `boxplot()` and `stripchart()` particularly useful in this context.

**Statistics**

We now introduce some notation and common statistics that are computed from ANOVA-type data. The sample treatment mean is

$$\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

The overall sample mean is obtained as

$$\bar{Y}_{\cdot\cdot} = \frac{1}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}$$

Intuitively, it should be clear that large differences in the sample treatment means may indicate that the treatments affect the quantity being measured. Thus, it should come as no surprise, that variation in sample means is exactly the signal used to reject the null hypothesis of no treatment effects. The details are below.

## 4.2 One-Way ANOVA

### 4.2.1 The Model

**Cell Means Model**

$$Y_{ij} = \theta_i + \epsilon_{ij} \qquad\qquad i = 1, 2, \ldots, k, \qquad j = 1, 2, \ldots, n_i$$

where $\theta_i$ are unknown population parameters, $\epsilon_{ij}$ are random errors, $k$ is the number of distinct populations, and $n_i$ is the sample size in the $i$th population. Note, the sample sizes may not be equal.

Note, if we assume $E[\epsilon_{ij}] = 0$, then the expected value of data is

$$E[Y_{ij}] = \theta_i, j = 1, 2, \ldots, n_i.$$

Thus, the mean of the data depends only on the treatment $i$. In particular, we conclude that the parameter, $\theta_i$, is the population mean of population $i$.

Since we focus on **fixed effect** models, this collection of $k$ populations are the only ones of interest. Therefore, $\theta_i$ are viewed as unknown constants.

**Alternative Parameterization**

Often, you will see another parameterization of the one-way ANOVA model.

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Then,

$$E[Y_{ij}] = \mu + \alpha_i$$

where $\mu$ is the *grand mean* and $\alpha_i$ is the unique effect of treatment $i$. Notice the expected difference between two measurements is given as

$$E[Y_{ij} - Y_{kl}] = \mu + \alpha_i - (\mu + \alpha_k) = \alpha_i - \alpha_k,$$

a difference in effects. Note, there are $k + 1$ parameters in this model formulation and this leads to identifiability problems, discussed in the next section.

### 4.2.2 Identifiability

Recall our overall framework. We have a population and associated with it are unknown population parameter(s) $\theta$. We assume there is some probability model that describes data $X$ sampled from this population. The probability model defines the pdf $f_\theta(x)$ (or pmf for discrete outcomes) for the data.

**Definition:** identifiable

A population parameter $\theta$ is *identifiable* if distinct $\theta$ correspond to distinct pdfs (or pmfs for discrete random variables). That is, if $\theta \neq \theta'$, then the pdf of the data $f_\theta(x) \neq f_{\theta'}(x)$ are distinct functions.

For example, if $\mu_1 \neq \mu_2$, then the corresponding normal pdfs are not the same:

$$f_{\mu_1}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu_1)}{2\sigma}\right] \neq \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu_1)}{2\sigma}\right] = f_{\mu_2}(x)$$

indicating that population mean of normally distributed random variables is identifiable.

1. identifiability is a property of the model (not the estimates of the population parameter), so solving identifability problems involves changing the model
2. if a model is not identifiable, then estimation of or inference on its population parameters is not possible

**Alternative Parameterization is Overparameterized**

In the alternative formulation, there are $k+1$ parameters and $k$ sample means available from the data. The extra degree of freedom in the data indicates that the model is *unidentifiable*. More than one choice of $(\mu, \alpha_1, \ldots, \alpha_k)$ can lead to the same pdf. One restriction on the parameters must be added to make the model identifiable. There are multiple choices for that restriction that change the way the parameters are interpretted.

- $\sum \alpha_i = 0$ means that we can interpret the $\alpha_i$ as deviations from the overall mean attributable to each population $i$.
- $\alpha_1 = 0$ might be useful if population 1 is the control group and we want to interpret the $\alpha_i, i > 1$ as deviations from no treatment.

### 4.2.3 Model Assumptions

**Assumptions**

1. $E[\epsilon_{ij}] = 0$, Var$(\epsilon_{ij}) = \sigma_i^2 < \infty$ for all $i, j$, Cov$(\epsilon_{ij}, \epsilon_{kl}) = 0$ for all $i, j, k, l$ with $i \neq j$ or $k \neq l$.
2. $\epsilon_{ij} \sim N(0, \sigma_i^2)$ independent.
3. *Homoscedasticity*: $\sigma_i^2 = \sigma^2$

Comments:

- Assumption 2 is required for hypothesis testing and confidence intervals.
- Without assumption 2, we are limited to do estimation. With assumption 1 about variance, we can find the estimate with minimum variance.
- Non-normality can lead to difficulties, but there are solutions for other kinds of distributions. We will not discuss much here.
- We can use CLT to get normality on population means if $n_i$ is large enough and the real distribution is fairly symmetric.
- Robustness to violations of 3 is lessened if $n_i \approx n$ constant for all treatments $i$.
- Robustness to violations of 2 depends on the extent to which 3 is true. For this reason, people will often transform the $Y$ random variables to achieve 3 so that they do not need to worry so much about normality of their data.

### 4.2.4 Inference

**Estimating $\mu$ and $\alpha_i$**

First, we address the problem of estimating the parameters $\mu$ and $\alpha_i$ (we address estimation of $\sigma^2$ in the context of hypothesis testing for ANOVA). It should be intuitively clear that

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

and

$$\hat{\alpha}_i = \bar{Y}_{i.} - \hat{\mu}.$$

are good estimators. You can also show these are the maximum likelihood estimates of the parameters given the model assumptions defined in the last section. We will discuss estimation of $\sigma^2$ when we discuss inference.

**Classic ANOVA Hypothesis**

$$\begin{aligned} H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \quad &\text{or} \quad H_0 : \alpha_1 = \cdots = \alpha_k = 0 \\ H_A : \theta_i \neq \theta_j \quad \text{for some } i \neq j \quad &\text{or} \quad H_A : \alpha_i \neq 0 \quad \text{for some } i \end{aligned}$$

This hypothesis is not often so interesting. Take the example of comparing several treatments. One may often include a control as a treatment to make sure that the experiment runs as planned. One knows before even collecting data that the control should have a different outcome compared to the rest, which means this classic $H_0$ will always be rejected. We might still like to know if $\theta_2 \neq \theta_3$. We will come back to this problem later.

**Partitioning Variance**

Often, ANOVA is presented as a way of partitioning the variance. The total variability can be summarized as the total sum of squares

$$SS_{\text{tot}} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$$

Note, this is just $N - 1$ times the combined sample variance, where $N = \sum_{i=1}^{k} n_i$.

By adding and subtracting the sample means $\bar{Y}_{i.}$, we can partition the total variance into parts

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left[ (\bar{Y}_{i.} - \bar{Y}_{..})^2 + (Y_{ij} - \bar{Y}_{i.})^2 \right]$$

Expand the quadratic and recognize the cross-term becomes 0 because

$$\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.}) = n_i \bar{Y}_{i.} - n_i \bar{Y}_{i.} = 0$$

to find

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

Interpretting each part of this sum, we have

$$SS_{\text{tot}} = SS_B + SS_W$$

where $SS_B$ is the sum-of-squares due to variation *between* treatments and $SS_W$ is the sum-of-squares due to error *within* treatments. There are $N$ observations, so there are $N-1$ d.f. for $SS_{\text{tot}}$. There are $k$ treatments, so there are $k-1$ d.f. for $SS_B$. Within the $i$th treatment, there are $n_i - 1$ d.f. for a total of $\sum_i (n_i - 1) = N - k$ d.f. within treatments for the $SS_W$.

**Estimating $\sigma^2$**

We will now show how all sums-of-squares, $SS_{\text{tot}}$, $SS_B$, and $SS_W$, are estimates of population variance $\sigma^2$ under the ANOVA null hypothesis. This fact will also allows us to propose statistics with sampling distributions for testing the null hypothesis.

**Lemma 13.** *Suppose $X_i$ independent are random variables with $E[X_i] = \mu_i$ and Var$(X_i) = \sigma^2$, $i = 1, \ldots, n$. Then,*

$$E\left[(X_i - \bar{X})^2\right] = (\mu_i - \bar{\mu})^2 + \frac{n-1}{n}\sigma^2,$$

*where $\bar{\mu} = \frac{1}{n}\sum_{i=1}^n \mu_i$.*

*Proof.* Recall for any random variable $Z$, that $E[Z^2] = (E[Z])^2 + \text{Var}(Z)$ by definition of variance, and $E\left[X_i - \bar{X}\right] = \mu_i - \bar{X}$ by linearity of expectation. The only part missing is

$$\begin{aligned}
\text{Var}(X_i - \bar{X}) &= \text{Var}(X_i) + \text{Var}(\bar{X}) - 2\text{Cov}(X_i, \bar{X}) \\
&= \sigma^2 + \frac{\sigma^2}{n} - 2\text{Cov}\left(X_i, \frac{1}{n}\sum_j X_j\right) \\
&= \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n}
\end{aligned}$$

Putting the parts back into the formula for $E[Z^2]$, the lemma is proved. □

**Theorem 14.** *Given the ANOVA assumptions, and assuming $n_i = n$ for all $i$,*

$$\begin{aligned}
E[SS_W] &= k(n-1)\sigma^2 \\
E[SS_B] &= n\sum_{i=1}^k \alpha_i^2 + (k-1)\sigma^2
\end{aligned}$$

*Proof.*

$$\begin{aligned}
E\left[SS_W\right] &= \textstyle\sum_{i=1}^k E\left[\sum_{j=1}^n (Y_{ij} - \bar{Y}_{i\cdot})^2\right] \\
&= \textstyle\sum_{i=1}^k E[(n-1)S_i^2] && \text{definition of sample variance} \\
&= \textstyle\sum_{i=1}^k (n-1)\sigma^2 && \text{constant variance assumption \& sample variance unbiased} \\
&= k(n-1)\sigma^2.
\end{aligned}$$

The second result uses the lemma.

$$\begin{aligned}
E[SS_B] &= n\sum_{i=1}^k E\left[(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right] \\
&= n\sum_{i=1}^k \left[\alpha_i^2 + \frac{k-1}{kn}\sigma^2\right] \\
&= n\sum_{i=1}^k \alpha_i^2 + (k-1)\sigma^2
\end{aligned}$$

where the second step results because $E[\bar{Y}_{i\cdot}] = E[Y_{ij}] = \mu + \alpha_i$, Var$(\bar{Y}_{i\cdot}) = \frac{\sigma^2}{n}$, $\bar{\mu} = \frac{1}{k}\sum_{i=1}^k (\mu + \alpha_i) = \mu$, and $\mu_i - \bar{\mu} = \alpha_i$. □

The expectations just derived suggest two ways to estimate the population variance $\sigma^2$. Most naturally, we define the pooled sample variance

$$S_p^2 := \frac{SS_W}{k(n-1)}.$$

Under the ANOVA assumption of constant variance, $S_p^2$ uses all the data to estimate population variance $\sigma^2$. (The denominator becomes $N-k$ for unequal sample sizes.) This is the multi-sample extension of the pooled sample variance from the two-sample $t$-test with equal variances.

When the ANOVA hypothesis is true, so $\alpha_i = 0$ for all $i$, then

$$\frac{SS_B}{k-1}$$

also estimates the population variance. By the way, so does

$$\frac{SS_{\text{tot}}}{kn-1}.$$

It is the difference between $E[SS_B]$ and $E[SS_W]$ that forms the basis of a statistical test of the null hypothesis. If $\alpha_i \neq 0$ for some $i$, then $E\left[\frac{SS_B}{k-1}\right] > E\left[\frac{SS_W}{k(n-1)}\right]$. It is this "signal" we'll capture in a statistic. The only trick then, will be to choose the statistic and determine its distribution under the null. To help us toward that goal, we consider the following theorem.

**Theorem 15.** *If* $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$*, then*

$$\frac{SS_W}{\sigma^2} \sim \chi^2_{k(n-1)}$$

*and if* $\alpha_i = 0$ *for all* $i$*, then*

$$\frac{SS_B}{\sigma^2} \sim \chi^2_{k-1}$$

*independent of* $\frac{SS_W}{\sigma^2}$.

*Proof.* We will not prove this result, but the proof follows the same kind of reasoning that gave us the chi-squared distribution in the one- or two-sample case. □

This theorem is a special case of Cochran's Theorem, applied by recognizing that $\frac{SS_{\text{tot}}}{\sigma^2} \sim \chi^2_{kn-1}$ when $H_0$ is true by results from the single sample result. (To be clear, under $H_0$, the multiple samples are all part of one big single sample from the same population.) We state Cochran's theorem for its general applicability to partitioned sums-of-squares.

**Theorem 16** (Cochran's Theorem)**.** *Let* $Z_i \sim N(0,1)$ *for* $i = 1, \ldots, \nu$ *and*

$$\sum_{i=1}^{\nu} Z_i^2 = Q_1 + Q_2 + \cdots + Q_s$$

*with* $s \leq \nu$*. Then,* $Q_1, Q_2, \ldots, Q_s$ *are independent* $\chi^2$ *random variables with* $\nu_1, \nu_2, \ldots, \nu_s$ *d.f., respectively if and only if*

$$\nu = \nu_1 + \nu_2 + \cdots + \nu_s.$$

**F Test for Testing Classical ANOVA Hypothesis**

These estimates of $\sigma^2$ provide the basis of the $F$ test for the classic hypothesis. Define statistic

$$F = \frac{\frac{SS_B}{k-1}}{\frac{SS_W}{k(n-1)}}$$

As already argued, this statistic should be close to 1 if $H_0$ is correct. Otherwise, it should tend to exceed 1.

We take this moment to define a new distribution called the $F$ distribution.

Definition: *F distribution*

Given $U \sim \chi^2_m$ and $V \sim \chi^2_n$ independent chi-square random variables, then

$$W = \frac{\frac{U}{m}}{\frac{V}{n}} \sim F(m,n)$$

is said to have an $F$ distribution with $m$ and $n$ degrees of freedom. The pdf of the $F$ distribution is given by

$$f(w) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \left(1 + \frac{mw}{n}\right)^{-(m+n)/2}$$

Thus, if $H_0$ is correct, then theorem 15 shows us $F \sim F(k-1, k(n-1))$, and in the case with variable samples sizes $F(k-1, N-k)$.

If the alternative hypothesis is correct, then we expect $SS_B/(k-1)$ to *overestimate* the population variance, so large values statistic $F$ will indicate problems with $H_0$, thus rejection is according to a one-tailed test when

$$F > F_{k-1,N-k}(1-\alpha/2)$$

In `R`, the functions `pf()`, `qf()`, and friends are for the $F$ distribution.

**The ANOVA Table**

The one-way ANOVA analysis is summarized in the *ANOVA table*.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | F |
|---|---|---|---|---|
| Between treatments | $SS_B$ | $k-1$ | $MS_B = \frac{SS_B}{k-1}$ | $F = \frac{MS_B}{MS_W}$ |
| Within treatments | $SS_W$ | $N-k$ | $MS_W = \frac{SS_W}{N-k}$ | |
| Total | $SS_{\text{tot}}$ | $N-1$ | | |

**Tukey Method**

If the ANOVA null hypothesis is rejected, then there is some effect $\alpha_i \neq 0$ for some population $i$. It becomes important to figure out which effects are non-zero, or which population means differ significantly.

Recalling $\theta_i = \mu + \alpha_i$, our model assumptions yield

$$\bar{Y}_{i\cdot} \sim N\left(\theta_i, \frac{\sigma^2}{n_i}\right)$$

so we can construct CI using $t$ statistics if we can produce an estimate of $\sigma^2$. Previously, we argued that $S_p^2$ always estimates $\sigma^2$, even when the null hypothesis is not satisfied, so it is natural to form CI for $\theta_i$ of

$$\bar{Y}_{i\cdot} \pm t_{N-k}(1-\alpha/2)S_p$$

The degrees of freedom used are the degrees of freedom of $SS_W$, which is used to compute $S_p^2$. A way to remember the degrees of freedom is to realize there are $N$ observations, but $k$ d.f. are lost to estimate the sample means in order to compute the sum of squared deviations in $SS_W$. The above CI is directly relevant to testing $H_0 : \alpha_i = 0$. You can figure out the test statistic and its sampling distribution. Notice, there are $k$ such tests we might need to run.

As for computing CI for mean differences, e.g. $\theta_i - \theta_j$, we recognize that testing $H_0 : \alpha_i = \alpha_j$ is equivalent. For this case, statistic

$$\frac{\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}}{S_p\sqrt{1/n_i + 1/n_j}} \sim t_{N-k}$$

is useful, but there are $\binom{k}{2}$ pairs of means we could test.

In both cases above, it is unwise to run that many tests without correcting the type $I$ error rate $\alpha$. The objective of Tukey's method is to estimate CI for all pairwise mean differences $\mu_i - \mu_j$ that simultaneously have the desired coverage.

Recall

$$(\bar{Y}_{i\cdot} - \theta_i) \sim N(0, \sigma^2/n)$$

if $\sigma^2$ is constant across sample sizes and the sample size is constant $n$. Define statistic

$$SR = \max_{i,j} \frac{|(\bar{Y}_{i\cdot} - \theta_i) - (\bar{Y}_{j\cdot} - \theta_j)|}{S_p/\sqrt{n}}$$

Under the ANOVA model, $SR$ follows a studentized range distribution with parameters $k$ and $k(n-1)$. Unusually large values of $SR$ suggest that the proposed population means $\theta_i$ are not the true population means. We will not write a formula for the studentized range distribution, but suppose $q_{k,k(n-1)}(1-\alpha)$ is its quantile. Then,

$$P\left[|(\bar{Y}_{i\cdot} - \theta_i) - (\bar{Y}_{j\cdot} - \theta_j)| \leq q_{k,k(n-1)}(1-\alpha)\frac{S_p}{\sqrt{n}}\right] = 1 - \alpha$$

for all $i \neq j$. When we hypothesize $\theta_i = \theta_j = \mu$, the confidence interval for $\alpha_i - \alpha_j$ is

$$\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} \pm q_{k,k(n-1)}(1-\alpha)\frac{S_p}{\sqrt{n}}$$

If the CI does not contain 0, then we reject $H_0 : \alpha_i = \alpha_j$ with $p$-value $< \alpha$.

The key advantage of the Tukey method is that if $H_0 : \alpha_i = \alpha_l$ is also rejected, then the $p$-value for that conclusion is also $< \alpha$. If separate $t$ tests at the $\alpha$ level are used for these analyses, the CI's would be narrower, more nulls would be rejected, and the probability of a type I error for any test would exceed $\alpha$.

Another solution to this problem is to use Bonferroni corrected $\alpha$ values on the separate $t$ tests.

Example:

We consider the following data with sample means computed from 7 treatments, each based on 10 measurements. Suppose we are given the pooled sample variance is $S_p = 0.061$, which we could read off an ANOVA table as the square root of $MS_W$.

| Lab | Mean |
|---|---|
| 1 | 4.062 |
| 2 | 4.003 |
| 3 | 3.998 |
| 4 | 3.997 |
| 5 | 3.957 |
| 6 | 3.955 |
| 7 | 3.920 |

The quantile for the Tukey statistic is given in `R` as `qtukey(0.95, nmeans=7, df=63)`, so $q_{7,63}(0.95)S_p/\sqrt{10} = 0.082$. We can examine all absolute pairwise differences, and any difference that exceeds $0.082$ allows us to reject the corresponding null that there is no difference with $\alpha = 0.05$. We find populations 1 and 4, 1 and 5, 1 and 6, and 3 and 4 have significantly different treatment effects.

If we, incorrectly, performed a two-sample $t$-test, a significant difference is anything larger than

$$t_{63}(0.975)S_p\sqrt{\frac{2}{10}} = 0.054$$

If we perform the two-sample $t$-tests with Bonferroni correction, a significant difference is found for every pairwise distance exceeding

$$t_{63}\left(1 - \frac{0.05}{2 \times 21}\right) S_p\sqrt{\frac{2}{10}} = 0.086.$$

You might also be interested to see the `TukeyHSD()` function in `R`.

### 4.2.5 Checking Model Assumptions

Definition: *residual*

> The residual is the difference between the observation and its model-estimated mean. In this case,
> $$r_{ij} = Y_{ij} - \hat{\mu} - \hat{\alpha}_i = Y_{ij} - \bar{Y}_{i\cdot}$$

By assumption of the model, the residuals should be normally distributed. One can check this assumption with probability plots for $r_{ij}$ or other tests of normality that we have discussed.

The ANOVA model additionally assumes constant variance, and we have not yet discussed methods for checking variance, though some problems can be identified from the boxplots.

#### Testing Common Variance

The $F$ test itself suggests a way to test equal variance in two-sample tests. If we have two independent samples:

$$\begin{aligned} X_1, \ldots, X_{n_x} &\overset{\text{iid}}{\sim} N(\mu_x, \sigma_x^2) \\ Y_1, \ldots, Y_{n_y} &\overset{\text{iid}}{\sim} N(\mu_y, \sigma_y^2) \end{aligned}$$

then we know

$$\begin{aligned} \frac{(n_x - 1)S_x^2}{\sigma^2} &\sim \chi^2_{n_x - 1} \\ \frac{(n_y - 1)S_y^2}{\sigma^2} &\sim \chi^2_{n_y - 1} \end{aligned}$$

are independent, so statistic

$$\frac{S_x^2}{S_y^2} \sim F(n_x - 1, n_y - 1)$$

In this case, the statistic may be unusually small or unusually large, but should be around 1 if the hypothesis

$$H_0 : \sigma_x^2 = \sigma_x^2 = \sigma^2$$

is correct. Thus, a two-tailed test can be used to find samples with significantly different variances.

#### Testing Common Variance: Multiple Samples

To test the null hypothesis that all sample variances are equal across more than two samples, i.e.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$$

we can use Bartlett's test or Levene's test. We will not spend time deriving these tests, but only show you how to use them.

- **Bartlett's Test.** See `bartlett.test()` in `R`. The downside of this test is it relies on the normality assumption.
- **Levene's Test.** Perform a second ANOVA on the absolute residuals, $|r_{ij}|$, testing the classic ANOVA hypothesis of constant means (or no effects).

### 4.2.6 Nonparametric Test

#### Kruskal-Wallis Test

If you find that your data does not satisfy the ANOVA assumptions, there is an alternative test that is related to the rank sum test. Let $R_{ij}$ be the rank of $Y_{ij}$ in the combined sample. Handle ties as for the rank sum test. Define

$$\bar{R}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} R_{ij} \qquad \text{and} \qquad \bar{R}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij} = \frac{N+1}{2}$$

and

$$SS_B = \sum_{i=1}^{k} n_i(\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2.$$

Then, it should be clear that the larger $SS_B$, the more evidence there is against the hypothesis

$$H_0 : \text{same probability distribution for all } k \text{ groups.}$$

As for the rank sum test, the statistic is most sensitive to changes in location of the distributions.

For small samples, you can use `R`'s function `kruskal.test()` to compute the $p$-value. For larger samples, it turns out that

$$K = \frac{12SS_B}{N(N+1)} \sim \chi^2_{k-1}$$

has an asymptotic chi-square distribution. The conditions for good asymptotics are $I = 3, n_i \geq 5$ or $I > 3$ and $n_i \geq 4$.

### 4.2.7 Contrasts

#### Contrast

The following is optional material. It was not discussed in class, but it covers a very common aspect of ANOVA.

Definition: *contrast*

> Let $t = (t_1, \ldots, t_k)$ be a vector of random variables, their realizations, parameters, or statistics.
> Let $a = (a_1, \ldots, a_k)$ be constants, then
>
> $$\sum_{i=1}^{k} a_i t_i$$
>
> is a linear combination of $t_i$'s. If $\sum_i a_i = 0$, then the linear combination is called a *contrast*.

We can write the classical ANOVA hypothesis in terms of contrasts.

**Theorem 17.** *$\theta_1 = \cdots = \theta_k$ if and only if $\sum_i a_i\theta_i = 0$ for all $a \in \mathcal{A}$, where $\mathcal{A} = \{a = (a_1, \ldots, a_k) : \sum_i a_i = 0\}$.*

*Proof.* The forward implication is obvious

$$\sum_i a_i\theta_i = \theta\sum_i a_i = 0$$

The reverse implication is also quite easy. Consider $a^{(1)} = (1, -1, 0, \ldots, 0) \in \mathcal{A}$. This one shows $\theta_1 = \theta_2$. Similarly, $a^{(2)} = (0, 1, -1, 0, \ldots, 0)$ shows $\theta_2 = \theta_3$. In general, the set $a^{(1)}, a^{(2)}, \ldots, a^{(k-1)}$ spans the space $\mathcal{A}$. Therefore, all possible equalities encoded in $\theta_1 = \cdots = \theta_k$ are implied by combining these vectors appropriately. □

**Inference on Contrasts**

Under the ANOVA assumptions, we have

$$Y_{ij} \sim N(\theta_i, \sigma^2)$$

and

$$\bar{Y}_{i\cdot} \sim N(\theta_i, \sigma^2/n_i).$$

Also, for any $a$,

$$\sum_{i=1}^{k} a_i \bar{Y}_{i\cdot} \sim N(\cdot, \cdot)$$

with mean and variance

$$E\left[\sum_i a_i \bar{Y}_{i\cdot}\right] = \sum_i a_i \theta_i \qquad \text{Var}\left(\sum_i a_i \bar{Y}_{i\cdot}\right) = \sigma^2 \sum_{i=1}^{k} \frac{a_i^2}{n_i}$$

**$t$-test for Generic Contrast**

But of course, we don't usually know $\sigma^2$. Instead, we use

$$S_i^2 = \frac{1}{n_i - 1} \sum_{i=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2$$

which is unbiased for $\sigma^2$ ($\sigma_i^2$ with heteroscedasticity) and also has distribution

$$\frac{(n_i - 1)S_i^2}{\sigma^2} \sim \chi^2_{n_i - 1}$$

If assumption 3 of homoscedasticity applies, then we can pool sample variances to get a better estimate of $\sigma^2$. Namely, with $N = \sum_i n_i$, we use the pooled sample variance

$$S_p^2 = \frac{1}{N - k} \sum_{i=1}^{k} (n_i - 1) S_i^2 = \frac{1}{N - k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2$$

Because the $S_i^2$ are independent, we also have

$$\frac{(N - k)S_p^2}{\sigma^2} \sim \chi^2_{N-k}$$

Also, because $S_p^2$ is independent of $\bar{Y}_{i\cdot}$, we have that statistic

$$\frac{\sum_i a_i \bar{Y}_{i\cdot} - \sum_i a_i \theta_i}{S_p \sqrt{\frac{a_i^2}{n_i}}} \sim t_{N-k}$$

which allows confidence intervals of the usual form

$$\sum_i a_i \bar{Y}_{i\cdot} - t_{N-k,\alpha/2} S_p \sqrt{\sum_i \frac{a_i^2}{n_i}} \leq \sum_i a_i \theta_i \leq \sum_i a_i \bar{Y}_{i\cdot} + t_{N-k,\alpha/2} S_p \sqrt{\sum_i \frac{a_i^2}{n_i}}$$

## 4.3 Two-Way ANOVA

### 4.3.1 Model

**Two-Way ANOVA Model**

In the two-way ANOVA, the experimenter simultaneously controls two factors, e.g. dosage level and exposure time. Each combination of factors is a treatment, and forms a *cell* in the two-way layout. Suppose there we observe a constant $K$ observations per cell, $I$ levels of factor one, and $J$ levels of factor two. Then the two-way ANOVA model is

$$\begin{aligned} Y_{ijk} &= \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk} \\ \epsilon_{ijk} &\overset{\text{iid}}{\sim} N(0, \sigma^2) \end{aligned}$$

$\alpha_i$ effect of factor one, level $i$, $\sum_{i=1}^{I} \alpha_i = 0$

$\beta_j$ effect of factor two, level $j$, $\sum_{j=1}^{J} \beta_i = 0$

$\delta_{ij}$ interaction effect of factor one, level $i$, and factor two, level $j$, $\sum_{i=1}^{I} \delta_{ij} = \sum_{j=1}^{J} \delta_{ij} = 0$

### 4.3.2 Parameter Estimation

As we motivated the estimates for the one-way ANOVA, we can use

$$\begin{aligned} \mu_{ijk} = E[Y_{ijk}] &= \mu + \alpha_i + \beta_j + \delta_{ij} \\ \bar{\mu}_{i\cdot k} = \frac{1}{J}\sum_j E[Y_{ijk}] &= \frac{1}{J}\sum_j (\mu + \alpha_i + \beta_j + \delta_{ij}) = \mu + \alpha_i \\ \bar{\mu}_{\cdot jk} = \frac{1}{I}\sum_i E[Y_{ijk}] &= \frac{1}{I}\sum_i (\mu + \alpha_i + \beta_j + \delta_{ij}) = \mu + \beta_j \end{aligned}$$

to justify

$$\begin{aligned} \hat{\mu} &= \bar{Y}_{...} \\ \hat{\alpha}_i &= \bar{Y}_{i..} - \bar{Y}_{...} \\ \hat{\beta}_i &= \bar{Y}_{\cdot j\cdot} - \bar{Y}_{...} \\ \hat{\delta}_{ij} &= \bar{Y}_{ij\cdot} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j) \\ &= \bar{Y}_{ij\cdot} - \bar{Y}_{...} - (\bar{Y}_{i..} - \bar{Y}_{...}) - (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{...}) \\ &= \bar{Y}_{ij\cdot} - \bar{Y}_{i..} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{...} \end{aligned}$$

but it is again possible to show that these are maximum likelihood estimates under the assumption of normally distributed errors and fixed effects. The likelihood is

$$L(Y_{ijk}; \mu, \alpha_i, \beta_j, \delta_{ij}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})^2\right].$$

Because of independence of observations, we have the log likelihood of all the data $Y$ is

$$l(Y; \mu, \alpha_i, \beta_j, \delta_{ij}, \sigma^2) = -\frac{IJK}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i\sum_j\sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})^2$$

Maximizing simultaneously for all the parameters, yields the intuitive estimates above.

Again, we leave estimation of $\sigma^2$ until later, as it is intimately related to hypothesis testing.

### 4.3.3 Hypothesis Testing

**Sums-of-Squares**

As we did for one-way ANOVA, we can break down the total sums-of-squares into components.

$$SS_{\text{tot}} = SS_A + SS_B + SS_{AB} + SS_E$$

where $SS_A$ measures the variation in the factor one means, $SS_B$ measures the variation in the factor two means, $SS_{AB}$ quantifies the strength of the interaction effects, and $SS_E$ is analagous to the within sum-of-squares, measuring the measurement error (hence subscript $E$).

In terms of the data, this partition of the sum-of-squares is

$$\sum_{i,j,k}(Y_{ijk}-\bar{Y}_{...})^2 = JK\sum_i(\bar{Y}_{i..}-\bar{Y}_{...})^2+IK\sum_j(\bar{Y}_{.j.}-\bar{Y}_{...})^2+K\sum_{i,j}(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})^2+\sum_{i,j,k}(Y_{ijk}-\bar{Y}_{ij.})^2$$

which can be proven by expanding

$$Y_{ijk} - \bar{Y}_{...} = (Y_{ijk} - \bar{Y}_{ij.}) + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}),$$

squaring both sides, and dropping cross-terms (because they sum to 0).

As before, we first work out the expectations of each of these partitioned sums-of-squares.

**Theorem 18.** *Under the two-way ANOVA model with* $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$,

$$\begin{aligned} E[SS_A] &= (I-1)\sigma^2 + JK\sum_i \alpha_i^2 \\ E[SS_B] &= (J-1)\sigma^2 + IK\sum_j \beta_j^2 \\ E[SS_{AB}] &= (I-1)(J-1)\sigma^2 + K\sum_{i,j}\delta_{ij}^2 \\ E[SS_E] &= IJ(K-1)\sigma^2 \end{aligned}$$

*Proof.* You can use lemma 13 to prove the result for $SS_A$ and $SS_B$.

For $SS_{AB}$, we apply the lemma to

$$\begin{aligned} E[SS_{\text{tot}}] &= E\left[\sum_{i,j,k}(Y_{ijk} - \bar{Y}_{...})^2\right] \\ &= \sum_{i,j,k}\left[\frac{IJK-1}{IJK}\sigma^2 + (\alpha_i + \beta_j + \delta_{ij})^2\right] \\ &= (IJK-1)\sigma^2 + JK\sum_i \alpha_i^2 + IK\sum_j \beta_j^2 + K\sum_{i,j}\delta_{ij}^2 \end{aligned}$$

which uses all the identities like $\sum_i \alpha_i = 0$ to simplify the result. Then, because $E[SS_{AB}] = E[SS_{\text{tot}}] - E[SS_A] - E[SS_B]$, we are done. □

Next comes distributional information for the sums-of-squares.

**Theorem 19.** *Under the two-way ANOVA model with* $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$,

- $SS_E/\sigma^2 \sim \chi^2_{IJ(K-1)}$.
- *If* $H_A : \alpha_i = 0 \; \forall \; i$, *then* $SS_A/\sigma^2 \sim \chi^2_{I-1}$.
- *If* $H_B : \beta_j = 0 \; \forall \; j$, *then* $SS_B/\sigma^2 \sim \chi^2_{J-1}$.
- *If* $H_{AB} : \delta_{ij} = 0 \; \forall \; i, j$, *then* $SS_{AB}/\sigma^2 \sim \chi^2_{(I-1)(J-1)}$.
- *And* $SS_A$, $SS_B$, $SS_{AB}$, *and* $SS_E$ *are all independent of each other.*

**Estimating $\sigma^2$**

We can see that $SS_E$ can be used to estimate population variance $\sigma^2$, so

$$\hat{\sigma}^2 = \frac{SS_E}{IJ(K-1)} := S_p^2$$

Under appropriate null hypotheses, the other sums-of-squares also estimate $\sigma^2$.

**Testing Hypotheses**

As before, this realization motivates the $F$ tests.

**Theorem 20.** *Under the two-way ANOVA model with $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$,*

- *For testing $H_A$, $F = SS_A/SS_E \sim F(I-1, IJ(K-1))$.*
- *For testing $H_B$, $F = SS_B/SS_E \sim F(J-1, IJ(K-1))$.*
- *For testing $H_{AB}$, $F = SS_{AB}/SS_E \sim F((I-1)(J-1), IJ(K-1))$.*

**ANOVA Table**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | F |
|---|---|---|---|---|
| Factor one | $SS_A$ | $I-1$ | $MS_A = \frac{SS_A}{I-1}$ | $F = \frac{MS_A}{MS_E}$ |
| Factor two | $SS_B$ | $J-1$ | $MS_B = \frac{SS_B}{J-1}$ | $F = \frac{MS_B}{MS_E}$ |
| Interaction | $SS_{AB}$ | $(I-1)(J-1)$ | $MS_{AB} = \frac{SS_{AB}}{(I-1)(J-1)}$ | $F = \frac{MS_{AB}}{MS_E}$ |
| Error | $SS_E$ | $IJ(K-1)$ | $MS_E = \frac{SS_E}{IJ(K-1)}$ | |
| Total | $SS_{\text{tot}}$ | $IJK-1$ | | |

**Reduced Model: No Interaction**

Notice, if interaction effects are assumed $\delta_{ij} = 0$, then the ANOVA table reduces and the degrees of freedom changes.

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | F |
|---|---|---|---|---|
| Factor one | $SS_A$ | $I-1$ | $MS_A = \frac{SS_A}{I-1}$ | $F = \frac{MS_A}{MS_E}$ |
| Factor two | $SS_B$ | $J-1$ | $MS_B = \frac{SS_B}{J-1}$ | $F = \frac{MS_B}{MS_E}$ |
| Error | $SS_E$ | $IJK-I-J+1$ | $MS_E = \frac{SS_E}{IJK-I-J+1}$ | |
| Total | $SS_{\text{tot}}$ | $IJK-1$ | | |

**Confidence Intervals**

Tukey's method can be extended to the two-way ANOVA, but we will focus on uncorrected CI here.

Suppose we want a CI for $\alpha_i - \alpha_{i'}$, then the relevant statistic is

$$\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}.$$

Because the samples are independent, we have

$$\text{Var}(\bar{Y}_{i\cdot\cdot}) = \text{Var}(\bar{Y}_{i'\cdot\cdot}) = \frac{\sigma^2}{JK} \text{ and } \text{Var}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}) = \frac{2\sigma^2}{JK}.$$

and the CI are

$$\bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot} \pm t_{IJ(K-1)}(1-\alpha/2)\sqrt{\frac{2SS_E}{IJ^2K(K-1)}}$$

### 4.3.4 Randomized Block Design

**Experimental Design**

We now take a moment to discuss experimental design because one of the most common experimental designs produces a two-way ANOVA.

Definition: *completely randomized design (CRD)*

> Given $T$ treatments and $n$ experimental units, the completely randomized design results if the EU are randomly divided into $T$ groups with $n_1, \ldots, n_T$ EU in each, such that all EU in group $t$ receive treatment $t$.

As we have discussed, the randomization of the CRD is a good thing because it insures that there are no confounding factors introduced by experimenter in assigning treatments that might also affect the response.

Definition: *randomized block design (RBD)*

> The RBD consists of $B$ blocks of $T$ EU each, with treatments randomly assigned such that each treatment appears exactly once in each block.

The RBD is an extension of the matched pair design. If the experimenter can identify a confounding factor, e.g. weight of subject, computer lab containing computer, etc., that might affect the measured response, then it is a good idea to use a RBD design to block ("pair" in the context of $> 2$ samples) on the confounding factor.

Example:

> Suppose four treatments are to be applied to $8$ experimental units. In a CRD, we would probably randomly choose $n_1 = \cdots = n_4 = 2$ EU per treatment. The problem is that EU will often vary tremendously in their response to even the same treatment. Thus, $n_i = 2$ may not be enough EU to see small treatment effects amongst large subject effects. Suppose the treatments can be applied sequentially to the same EU (i.e. there is nothing irreversible to the treatments, for example no surgeries). Then, a lot of power can be gained by blocking on EU. Each EU is subject to all four treatments, applied in random order (here is where the randomization enters). A random RBD for four EU is shown below, where $T_i$ is treatment $i$.

| Subject | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| $T_2$ | $T_4$ | $T_1$ | $T_1$ |
| $T_1$ | $T_2$ | $T_3$ | $T_4$ |
| $T_4$ | $T_1$ | $T_2$ | $T_3$ |
| $T_3$ | $T_3$ | $T_4$ | $T_2$ |

> The treatment order may have an effect as well. One can also "block" on timing of treatment, so that each temporal sequence of the treatments is observed only once. Efficient designs for this multi-dimensional blocking are *Latin Hypercube Designs*. An example is shown below. Notice the treatment orders are no longer random, but do vary from subject to subject.

| Subject | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| $T_1$ | $T_4$ | $T_3$ | $T_2$ |
| $T_2$ | $T_1$ | $T_4$ | $T_3$ |
| $T_3$ | $T_2$ | $T_1$ | $T_4$ |
| $T_4$ | $T_3$ | $T_2$ | $T_1$ |

**RBD as Two-Way ANOVA**

The RBD is very popular and it leads to the following model

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

where $\alpha_i$ is the treatment effect, $\beta_j$ is the block effect (of little interest), and $\epsilon_{ij}$ are the usual errors. Because $K = 1$, we drop the subscript $k$.

I should note that RBD designs often are mixed effects models, that is where $\alpha_i$ are fixed effects and $\beta_j$ are random effects. Take our example. We probably don't want to just make inference for the blocks (subjects) in our study, but to extrapolate to the population of EU. In this case, $\beta_j$ are random effects. Fortunately, the hypothesis testing for $\alpha_i$ are unchanged from the fixed effects models we've been discussing.

### 4.3.5 Nonparametric Test

**Friedman's Test**

The one-way ANOVA assumptions about the errors also apply to the two-way ANOVA. If these assumptions are suspect for your dataset, then nonparametric methods may be warranted. Friedman's test for the RBD is a generalization of the sign rank test for paired samples.

For each treatment, rank the measurements $Y_{i1}, \ldots, Y_{iB}$ to obtain $R_{i1}, \ldots, R_{iB}$. Then compute

$$SS_A = J \sum_i (\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2,$$

a measure of the difference in ranks across treatments. `R`'s function `friedman.test()` can be used to compute $p$-values using this statistic, but for large samples, the statistic

$$Q = \frac{12SS_A}{I(I+1)} \sim \chi^2_{I-1}.$$