

© Copyright 2010–2018 by Jenny A. Baglivo. All Rights Reserved.

3	MATH448001 Notebook 3	3
3.1	One Way Layout	3
3.1.1	Analysis of Variance for One Way Layouts	3
3.1.2	Multiple Comparisons in the Parametric Setting	12
3.1.3	Analysis of Linear Contrasts	14
3.1.4	Kruskal-Wallis Statistic, Distribution and Methods	17
3.1.5	Multiple Comparisons in the Nonparametric Setting	19
3.1.6	Permutation f Test	20
3.1.7	Comparison of Tests	20
3.1.8	Permutation Example: Biodiversity Analysis	21
3.2	Blocked Design	23
3.2.1	Analysis of Variance for Blocked Designs	24
3.2.2	Multiple Comparisons in the Parametric Setting	31
3.2.3	Friedman Statistic, Distribution and Methods	32
3.2.4	Multiple Comparisons in the Nonparametric Setting	34
3.2.5	Permutation f Test	36
3.3	Balanced Two Way Layout	37
3.3.1	Analysis of Variance for Balanced Two Way Layouts	38
3.3.2	Multiple Comparisons in the Parametric Setting	46

3 MATH448001 Notebook 3

This notebook includes methods for analyzing more than two samples using classical, permutation and nonparametric approaches. Classical methods assume that the data are the values of independent normal distributions with a common variance. The notes include topics from Chapter 12 (the analysis of variance) of the Rice textbook; these topics are natural generalizations of topics from Chapter 11 (comparing two samples) of the Rice textbook.

3.1 One Way Layout

This section considers methods for I samples (displayed as lists of numbers)

$$\{x_{1,1}, x_{1,2}, \dots, x_{1,n_1}\}, \{x_{2,1}, x_{2,2}, \dots, x_{2,n_2}\}, \dots, \{x_{I,1}, x_{I,2}, \dots, x_{I,n_I}\}$$

where n_i is the number of observations in the i^{th} sample, and $x_{i,j}$ is the j^{th} observation in the i^{th} sample, for $i = 1, 2, \dots, I$.

Let $N = \sum_{i=1}^I n_i$ be the total number of observations.

For example, we might be interested in considering the following three samples of sizes 5, 6 and 4, respectively; the total number of observations is 15.

		<i>Sample Size:</i>
<i>Sample 1:</i>	12.3, 13.5, 13.9, 12.4, 14.9	$n_1 = 5$
<i>Sample 2:</i>	14.5, 15.9, 11.7, 14.4, 13.1, 16.8	$n_2 = 6$
<i>Sample 3:</i>	10.7, 16.6, 15.4, 13.7	$n_3 = 4$
		$N = 15$

3.1.1 Analysis of Variance for One Way Layouts

The data are assumed to be the values of I independent random samples

$$X_{i,1}, X_{i,2}, \dots, X_{i,n_i}, \text{ for } i = 1, 2, \dots, I,$$

from normal distributions with a common unknown standard deviation σ .

1. The i^{th} random sample is called the i^{th} *group*.
2. The mean of i^{th} group, $\mu_i = E(X_{i,j})$, $j = 1, 2, \dots, n_i$, is called the i^{th} *group mean*.

Of interest is a test of

$$H_O : \mu_1 = \mu_2 = \dots = \mu_I \text{ versus } H_A : \text{not all } \mu_i \text{'s are equal.}$$

The general assumptions imply that $X_{i,j}$ can be written in the following linear form:

$$X_{i,j} = \mu + \alpha_i + \epsilon_{i,j} \text{ for } j = 1, 2, \dots, n_i, i = 1, 2, \dots, I, \text{ where}$$

1. $\mu = E \left(\frac{1}{N} \sum_{i=1}^I \sum_{j=1}^{n_i} X_{i,j} \right)$ is the *overall mean*,
2. $\alpha_i = \mu_i - \mu$ is the i^{th} *group effect* for $i = 1, 2, \dots, I$, and
3. The collection of *error terms*, $\{\epsilon_{i,j}\}$, is a random sample of size N from a normal distribution with mean 0 and standard deviation σ .

Since $\mu_i = \mu + \alpha_i$, the hypotheses of interest are equivalent to

$$H_O : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0 \text{ versus } H_A : \text{not all } \alpha_i \text{'s are zero.}$$

Exercise. Demonstrate that $\sum_{i=1}^I n_i \alpha_i = 0$.

Maximum likelihood estimation. Maximum likelihood (ML) estimators for the parameters of the linear model are given in the following theorem:

Theorem (Parameter Estimation). Given the assumptions and definitions above, the following are ML estimators of the parameters in the linear model:

1. Overall mean:

$$\hat{\mu} = \bar{X}_{..} = \frac{1}{N} \left(\sum_{i=1}^I \sum_{j=1}^{n_i} X_{i,j} \right)$$

2. Group effects:

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} = \bar{X}_{i.} - \bar{X}_{..} = \frac{1}{n_i} \left(\sum_{j=1}^{n_i} X_{i,j} \right) - \bar{X}_{..} \text{ for each } i$$

3. Error terms:

$$\hat{\epsilon}_{i,j} = X_{i,j} - \hat{\mu}_i = X_{i,j} - \bar{X}_{i.} \text{ for each } i, j$$

Each is an unbiased estimator.

Further, the *pooled estimator* of the common variance

$$S_p^2 = \frac{1}{N-I} \sum_{i=1}^I \sum_{j=1}^{n_i} \hat{\epsilon}_{i,j}^2 = \frac{1}{N-I} \sum_{i=1}^I \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_{i.})^2$$

is an unbiased estimator of σ^2 .

Note on estimating σ^2 :

Let S_i^2 be the i^{th} sample variance, for $i = 1, 2, \dots, I$. Then the pooled estimator becomes

$$S_p^2 = \frac{1}{N-I} \sum_{i=1}^I \left(\sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_{i.})^2 \right) = \frac{1}{N-I} \sum_{i=1}^I (n_i - 1) S_i^2.$$

For example, data summaries for the example on page 3 are as follows:

	Size:	Mean:	Variance:	Group Effect:
Sample 1:	$n_1 = 5$	$\bar{x}_{1.} = \frac{1}{5}(67.0) = 13.4$	$s_1^2 \approx 1.086^2$	$\bar{x}_{1.} - \bar{x}_{..} \approx -0.587$
Sample 2:	$n_2 = 6$	$\bar{x}_{2.} = \frac{1}{6}(86.4) = 14.4$	$s_2^2 \approx 1.844^2$	$\bar{x}_{2.} - \bar{x}_{..} \approx 0.413$
Sample 3:	$n_3 = 4$	$\bar{x}_{3.} = \frac{1}{4}(56.4) = 14.1$	$s_3^2 \approx 2.560^2$	$\bar{x}_{3.} - \bar{x}_{..} \approx 0.113$
	$N = 15$	$\bar{x}_{..} \approx 13.987$	$s_p^2 \approx 1.857^2$	

Exercise. Let S_p^2 be the pooled estimator of the common variance. Use the theorem on sampling from normal distributions, and the fact that the sum of independent chi-square random variables is chi-square, to demonstrate that

(a) S_p^2 is an unbiased estimator of σ^2 .

(b) $V = \frac{(N-I)}{\sigma^2} S_p^2$ has a chi-square distribution with $N - I$ df.

Initial analysis of variance table. Computations needed for an analysis of variance for a one way layout are often organized into an *analysis of variance table*:

	<i>df</i>	<i>Sum of Squares</i>	<i>Mean Square</i>
<i>Group:</i>	$I - 1$	$SS_g = \sum_{i=1}^I n_i (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$	$MS_g = \frac{1}{I-1} SS_g$
<i>Error:</i>	$N - I$	$SS_e = \sum_{i=1}^I \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_{i\cdot})^2$	$MS_e = \frac{1}{N-I} SS_e$
<i>Total:</i>	$N - 1$	$SS_t = \sum_{i=1}^I \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_{\cdot\cdot})^2$	

1. The total sum of squares (SS_t) is equal to the sum of the group sum of squares (SS_g) and the error sum of squares (SS_e): $SS_t = SS_g + SS_e$.
2. The group mean square (MS_g) is a random variable with mean

$$E(MS_g) = \sigma^2 + \frac{1}{I-1} \sum_{i=1}^I n_i (\mu_i - \mu)^2 = \sigma^2 + \frac{1}{I-1} \sum_{i=1}^I n_i \alpha_i^2.$$

If the null hypothesis that the group means are equal is true, then MS_g is an unbiased estimator of σ^2 ; otherwise, the mean is larger than σ^2 .

3. The error mean square (MS_e) is the same as the pooled estimator of the common variance. Thus, MS_e is an unbiased estimator of σ^2 , and

$$V = \frac{N-I}{\sigma^2} MS_e \text{ has a chi-square distribution with } N - I \text{ degrees of freedom.}$$

Note: Between and Within Sums of Squares

The group sum of squares (SS_g) is often called the *between-group sum of squares*, and the error sum of squares (SS_e) is often called the *within-group sum of squares*.

To conduct an analysis of variance, we work with the ratio $F = MS_g/MS_e$. Large values of F favor the general alternative hypothesis that at least two means differ. The following theorem, proven by Fisher, tells us the sampling distribution of F under the null hypothesis:

Theorem (Distribution Theorem). Under the general assumptions of this section and if the null hypothesis of equality of group means is true, then the ratio

$$F = MS_g/MS_e$$

has an f ratio distribution with $(I - 1)$ and $(N - I)$ degrees of freedom.

Exercise. Demonstrate that the total sum of squares (SS_t) is equal to the sum of the group sum of squares (SS_g) and the error sum of squares (SS_e): $SS_t = SS_g + SS_e$.

Hint: As a first step, observe that $X_{i,j} - \bar{X}_{..} = (X_{i,j} - \bar{X}_{i\cdot}) + (\bar{X}_{i\cdot} - \bar{X}_{..})$.

Exercise. Demonstrate that the expected value of the group mean square is

$$E(MS_g) = \sigma^2 + \frac{1}{I-1} \sum_{i=1}^I n_i (\mu_i - \mu)^2 = \sigma^2 + \frac{1}{I-1} \sum_{i=1}^I n_i \alpha_i^2.$$

Hint: As a first step, notice that we can rewrite the group sum of squares as follows:

$$SS_g = \sum_{i=1}^I n_i \left(\bar{X}_{i\cdot}^2 - 2\bar{X}_{i\cdot}\bar{X}_{\cdot\cdot} + \bar{X}_{\cdot\cdot}^2 \right) = \sum_{i=1}^I n_i \bar{X}_{i\cdot}^2 - 2\bar{X}_{\cdot\cdot} \sum_{i=1}^I n_i \bar{X}_{i\cdot} + \bar{X}_{\cdot\cdot}^2 \sum_{i=1}^I n_i,$$

and that this expansion can be simplified.

Completed ANOVA table. In applications, an analysis of variance table is expanded to include the observed value of the statistic (f_{obs}), and the observed significance level

$$p \text{ Value} = P(F \text{ Ratio RV} \geq f_{\text{obs}}),$$

where the p value is based on the f ratio distribution with $(I - 1)$ and $(N - I)$ df.

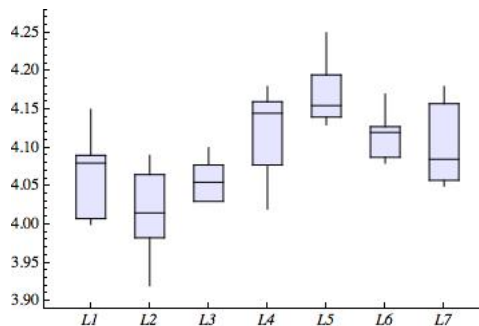
Example. Assume the data presented on page 3 are the values of independent random samples from normal distributions with equal variances, and that we wish to test the null hypothesis that the means are equal versus the general alternative hypothesis that not all means are equal at the 5% significance level. The following is an analysis of variance table for the test:

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p value</i>
<i>Group:</i>	2	2.797	1.399	0.4056	0.675
<i>Error:</i>	12	41.380	3.448		
<i>Total:</i>	14	44.177			

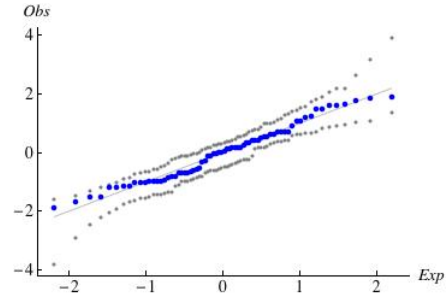
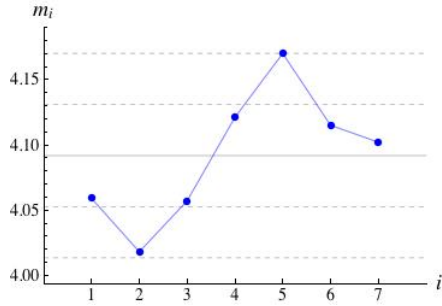
Since the p value is greater than 0.05, we accept the null hypothesis that the means are equal. There is insufficient evidence to conclude otherwise.

Example (Rice textbook, Chapter 12). “Kirchhoefer (1979) reports on a study of the use of a semi-automated method for measuring the amount of chlorpheniramine in tablets. For each of four manufacturers, composites were prepared by grinding and mixing together tablets that had nominal dosage levels of 4 milligrams (mg). Seven labs were asked to make 10 determinations on each composite; each determination was made on a portion of composite whose weight was equivalent to that of one tablet. The purpose of the experiment was to determine the consistency between labs and the variability of the measurement process. This example uses information for one manufacturer.”

Side-by-side box plots of the sample data are shown to the right. Additional summaries are as follows:



	<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>	<i>Lab 4</i>	<i>Lab 5</i>	<i>Lab 6</i>	<i>Lab 7</i>	
<i>Size, n_i:</i>	10	10	10	10	10	10	10	$N = 70$
<i>Mean, $\bar{x}_{i\cdot}$:</i>	4.059	4.018	4.057	4.121	4.170	4.115	4.102	$\bar{x}_{\cdot\cdot} = 4.092$
<i>SD, s_i:</i>	0.051	0.058	0.027	0.054	0.043	0.029	0.052	$s_p = 0.046$



1. *Left Plot:* The left plot shows the observed group means (labeled m_i); a solid horizontal line is drawn at the observed overall mean.
2. *Right Plot:* The right plot is an enhanced normal probability plot of standardized errors,

$$r_{i,j} = \frac{(x_{i,j} - \bar{x}_{i\cdot})}{s_i} \text{ for } i = 1, 2, \dots, I, j = 1, \dots, n_i.$$

The plot suggests that normal theory methods are reasonable in this case.

Assume that the tablets data are the values of independent random samples from normal distributions with equal variances, and that we wish to test the null hypothesis that the means of the distributions are equal versus the general alternative hypothesis that at least two means differ at the 5% significance level. The following is an analysis of variance table for the test:

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p value</i>
<i>Group:</i>	6	0.1534	0.0256	11.9020	0.00000007
<i>Error:</i>	63	0.1354	0.0021		
<i>Total:</i>	69	0.2888			

Comments (please complete):

Note on the Max/Min ratio rule:

The general rule-of-thumb for equality of variances is that the ratio of the maximum to the minimum standard deviation should be no more than about 2.0. But, "...with balanced designs (where all sample sizes are equal) and with small sample sizes (of 10 or less), we can proceed with larger ratios (perhaps up to 4.0)."

In the example above, the ratio is about 2.15.

3.1.2 Multiple Comparisons in the Parametric Setting

If the null hypothesis of equality of means is rejected, then it is natural to try to determine which means differ; this will require doing multiple comparisons.

There are several methods for multiple comparisons: *Tukey's method* (see Rice textbook) can be used with balanced designs only. The more general *Bonferroni method* will be used here.

Bonferroni method. A simple observation, based on an inequality from probability theory, allows us to conduct multiple comparisons while maintaining an α significance level.

Theorem (Bonferroni Method I). If m hypotheses are tested simultaneously, each at the α/m significance level, then the overall level will be at most α .

Outline of Proof: Assume that all m null hypotheses are true. Let E_i be the event that the i^{th} null hypothesis is rejected and assume that $P(E_i) = \alpha/m$. Then, by induction, the probability that at least one null hypothesis is rejected is

$$P(E_1 \cup E_2 \cup \dots \cup E_m) \leq \sum_{i=1}^m P(E_i) = m(\alpha/m) = \alpha.$$

When comparing group means in the analysis of variance for one way layouts setting, we work with the following distribution theorem:

Theorem (Sampling Distribution). Under the general conditions of this section, the statistic

$$T_{i,k} = \frac{(\bar{X}_{i\cdot} - \bar{X}_{k\cdot}) - (\mu_i - \mu_k)}{\sqrt{S_p^2 \left(\frac{1}{n_i} + \frac{1}{n_k} \right)}}, \text{ where } S_p^2 \text{ is the pooled estimator of } \sigma^2,$$

has a Student t distribution with $(N - I)$ df, where $i, k \in \{1, 2, \dots, I\}$ and $i \neq k$.

This theorem can be used to compare any two of the I samples.

To conduct all $m = \binom{J}{2}$ two sided tests of equality of means,

$$\mu_i = \mu_k \text{ versus } \mu_i \neq \mu_k, \text{ for } i < k,$$

we use pooled t statistics of the form

$$T_{i,k} = \frac{\bar{X}_{i\cdot} - \bar{X}_{k\cdot}}{\sqrt{S_p^2 \left(\frac{1}{n_i} + \frac{1}{n_k} \right)}}, \text{ and conduct each test at the } \alpha/m \text{ level.}$$

Continuing with the example on the amount of active ingredient in measurements made at seven laboratories, begun on page 10, a Bonferroni analysis will be conducted at the 5% significance level. A total of $21 = \binom{7}{2}$ comparisons will be made, and the rejection region for each test is

$$|T| \geq t_{63}((0.05/21)/2) = 3.166.$$

The following table gives the 21 observed t statistics:

	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
$i = 1$	1.978	0.096	-2.991	-5.355	-2.701	-2.074
$i = 2$		-1.881	-4.969	-7.333	-4.679	-4.052
$i = 3$			-3.087	-5.451	-2.798	-2.171
$i = 4$				-2.364	0.289	0.917
$i = 5$					2.653	3.280
$i = 6$						0.627

These results suggest (please complete)

3.1.3 Analysis of Linear Contrasts

In some situations, we may be interested in making only certain predetermined comparisons among the means. For example, we may be interested in determining if the average of means in the first two groups equals the average in the next three groups:

$$\frac{1}{2}(\mu_1 + \mu_2) = \frac{1}{3}(\mu_3 + \mu_4 + \mu_5) \quad \text{or} \quad \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 - \frac{1}{3}\mu_5 = 0.$$

The expression $\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \frac{1}{3}\mu_3 - \frac{1}{3}\mu_4 - \frac{1}{3}\mu_5$ is known as a linear contrast of the group means.

Formally, a *linear contrast* of the group means is an expression of the form

$$C = \sum_{i=1}^I a_i \mu_i, \quad \text{where} \quad \sum_{i=1}^I a_i = 0.$$

Contrasts arise naturally in tests of sub-hypotheses. Maximum likelihood (ML) estimation in the analysis of variance for one way layouts setting is straightforward.

Theorem (Parameter Estimation). Under the general assumptions of analysis of variance for one way layouts,

$$\hat{C} = \sum_{i=1}^I a_i \bar{X}_i \quad \text{is the ML estimator of} \quad C = \sum_{i=1}^I a_i \mu_i.$$

Exercise. Find the mean and variance of the ML estimator of C .

Example (Dandruff Study). A study was conducted to compare three treatments for dandruff with a placebo (with no active ingredient). The treatments were

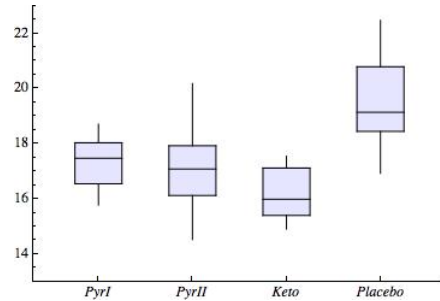
1. A 1% pyrrithione zinc shampoo,
2. The same shampoo, but with instructions to shampoo twice, and
3. A 2% ketoconazole shampoo.

Thus, a natural comparison to make is between the mean for placebo and the average of the means for the three treatments.

A total of 80 participants were randomly assigned to the four study groups. After 6 weeks of treatment, eight sections of the scalp of each participant were examined and measurements in the $[0, 10]$ interval were assigned to each section.

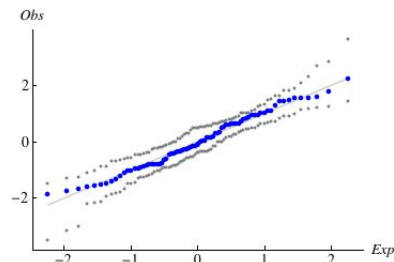
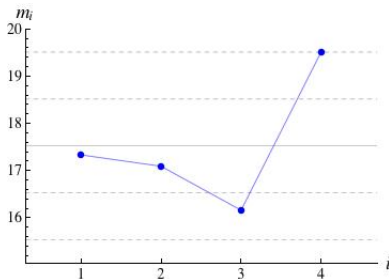
The response variable for each participant was the sum of the eight measurements. The higher the score, the more dandruff was present.

Side-by-side box plots of the dandruff scores are shown to the right, and numerical summaries are presented below.



	1. <i>Pyr I</i>	2. <i>Pyr II</i>	3. <i>Keto</i>	4. <i>Placebo</i>	
<i>Size, n_i:</i>	20	20	20	20	$N = 80$
<i>Mean, $\bar{x}_{i\cdot}$:</i>	17.33	17.09	16.16	19.51	$\bar{x}_{\cdot\cdot} = 17.52$
<i>SD, s_i:</i>	0.88	1.39	0.89	1.67	$s_p = 1.25$

A plot of group means is shown on the *left* below, and an enhanced normal probability plot of standardized errors is shown on the *right*.



Since the ratio of the maximum to the minimum standard deviation is about 1.9, and since the points in the normal probability plot are reasonable close to the line $y = x$, the use of normal-theory methods to analyze these data seems reasonable.

The following table gives the results of a one way analysis of variance analysis of the null hypothesis that the group means are equal versus the general alternative that they are not.

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p value</i>
<i>Group:</i>	3	120.49	40.16	25.55	≈ 0
<i>Error:</i>	76	119.48	1.57		
<i>Total:</i>	79	239.97			

Since the p value is virtually zero, the results suggest that the means are not all equal. The means for the ketoconazole and placebo groups likely differ the most.

Exercise. Let $C = \mu_4 - \frac{1}{3}(\mu_1 + \mu_2 + \mu_3)$ be the contrast reflecting the comparison of most interest to the researchers.

Construct and interpret a 95% CI for the contrast C .

3.1.4 Kruskal-Wallis Statistic, Distribution and Methods

In the 1950's, Kruskal and Wallis developed a nonparametric version of the analysis of variance test for one way layouts appropriate in the following situations:

1. *Population model.* The data are the values of I independent random samples. The null hypothesis is that the distributions from which the data were drawn are equal. Alternative hypotheses of interest are that at least one distribution is stochastically larger or smaller than another.
2. *Randomization model.* The data are measurements on N individuals in distinguishable groups of sizes n_1, n_2, \dots, n_I . The null hypothesis is that observed differences in the groups are due to chance alone. Alternative hypotheses of interest are that the values in at least one group tend to be larger or smaller than those in another.

Let $R_{i,j}$ be the rank of observation $x_{i,j}$ in the combined sample, and $\bar{R}_{i\cdot}$ be the average rank of observations in the i^{th} sample, and assume that all N observations are distinct.

Then the Kruskal-Wallis statistic, K , is defined as follows:

$$K = \frac{12}{N(N+1)} \sum_{i=1}^I n_i (\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2 = \frac{12}{N(N+1)} \sum_{i=1}^I n_i \left(\bar{R}_{i\cdot} - \frac{N+1}{2} \right)^2.$$

Thus, the Kruskal-Wallis statistic is a weighted sum of squared deviations of average group ranks from the overall average rank.

For example, the following table shows the small data example from page 3, corresponding ranks, and observed averages by group ($\bar{r}_{i\cdot}$) and for all 15 observations ($\bar{r}_{\cdot\cdot} = 8$).

	Values:	Ranks:	Rank Means:
Sample 1:	12.3, 13.5, 13.9, 12.4, 14.9	3, 6, 8, 4, 11	$\bar{r}_{1\cdot} = 32/5 = 6.4$
Sample 2:	14.5, 15.9, 11.7, 14.4, 13.1, 16.8	10, 13, 2, 9, 5, 15	$\bar{r}_{2\cdot} = 54/6 = 9.0$
Sample 3:	10.7, 16.6, 15.4, 13.7	1, 14, 12, 7	$\bar{r}_{3\cdot} = 34/4 = 8.5$
		1, 2, ..., 15	$\bar{r}_{\cdot\cdot} = 8.0$

The observed value of the statistic for these data is

$$k_{\text{obs}} = \frac{12}{(15)(16)} (5(6.4 - 8)^2 + 6(9 - 8)^2 + 4(8.5 - 8)^2) = 0.99.$$

Question: Can you see why $\bar{R}_{\cdot\cdot} = (N + 1)/2$?

The sampling distribution of K is obtained by computing its value for each partition of the N ranks into distinguishable groups of sizes n_1, n_2, \dots, n_I . There are a total of $\binom{N}{n_1, n_2, \dots, n_I}$ partitions to consider.

The following theorem gives a large sample approximation to the distribution.

Theorem (Kruskal-Wallis Distribution). If there are no ties in the observations, then under the null hypothesis of randomness and when N is large, the distribution of K is approximately chi-square with $(I - 1)$ df.

Large values of K support the alternative hypothesis that the values in at least one sample tend to be larger or smaller than those in another. The observed significance level is

$$p \text{ Value} = P(K \geq k_{\text{obs}}),$$

where k_{obs} is the observed value of the Kruskal-Wallis statistic.

Note on large sample approximations:

In most practical situations, the chi-square approximation to the sampling distribution of K is used to compute p values.

A “rule of thumb” is that the chi-square approximation is adequate when there are 3 or more samples, each sample has 5 or more observations, and all observations are distinct.

Handling equal observations. Average ranks (or *midranks*) replace ranks when there are ties in the data.

Now, the Kruskal-Wallis statistic becomes

$$K = \sum_{i=1}^I w_i (\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2 = \sum_{i=1}^I w_i \left(\bar{R}_{i\cdot} - \frac{N+1}{2} \right)^2.$$

where the weights, w_i , in the weighted sum are chosen to make the approximate sampling distribution of K under the null hypothesis as close to the chi-square distribution as possible.

Most computer programs automatically use the appropriate weights.

3.1.5 Multiple Comparisons in the Nonparametric Setting

If the Kruskal-Wallis test suggests that differences are not due to chance alone, then the next step is to try to determine which groups differ.

Bonferroni approach to multiple comparisons. In the Bonferroni approach to multiple comparisons, we conduct $m = \binom{J}{2}$ two sided Wilcoxon rank sum tests. If each test is conducted at the $\frac{\alpha}{m}$ level, then the overall type I error will be at most α .

Example. Consider testing the null hypothesis of randomness at the 5% significance level, using the dandruff study from page 15. The following table summarizes the analysis:

<i>Kruskal-Wallis Statistic</i>	<i>p Value</i>	<i>Sampling Distribution</i>
36.9681	≈ 0	Chi-Square, 3 df

Since there are significant differences, the Bonferroni method will be used next.

A total of $\binom{4}{2} = 6$ two sided rank sum tests will be conducted at the $\frac{0.05}{6} = 0.0083$ level. The following table gives the p values:

$(1,2)$	$(1,3)$	$(1,4)$	$(2,3)$	$(2,4)$	$(3,4)$
0.425	0.000234	≈ 0	0.0249	0.0000249	≈ 0

These results suggest (please complete)

3.1.6 Permutation f Test

A permutation version of the analysis of variance f test for one way layouts is appropriate in one of the following situations:

1. *Population model.* The observed data are the values of independent random samples from distributions differing in mean only. The null hypothesis is that the distributions (equivalently, the means) are equal versus the general alternative that at least two distributions (equivalently, means) differ.
2. *Randomization model.* The observed data are measurements taken on N individuals in distinguishable groups of sizes n_1, n_2, \dots, n_I . The null hypothesis is that observed differences in means are due to chance alone versus the alternative that at least one sample has values that tend to be larger or smaller (but not more variable) than the values in another sample.

The sampling distribution of F is obtained by computing its value for each partition of the N observations into distinguishable groups of sizes n_1, n_2, \dots, n_I . There are a total of $\binom{N}{n_1, n_2, \dots, n_I}$ partitions to consider. Since this number can be quite large, Monte Carlo analysis is generally used to estimate a p value.

For example, consider again the example on the amount of active ingredient in measurements made at seven laboratories, begun on page 10. To determine if observed differences in means is due to chance, a permutation f test will be conducted at the 5% significance level.

In a Monte Carlo analysis using 5000 random partitions (including the observed partition of the measurements), 0.02% (1/5000) of f values were greater than or equal to the observed. Further, a 99% confidence interval for the permutation p value was [0.000001, 0.001485].

Thus, there is strong evidence that the means differ. In particular, measurements produced by Lab 5 appear to be significantly greater than those produced by Lab 2.

3.1.7 Comparison of Tests

A useful comparison of the approaches for one way layouts is as follows:

1. *Analysis of variance versus permutation f tests.*

In situations where both the analysis of variance f test and the permutation f test are appropriate, the analysis of variance f test is preferred.

However, it is interesting to note that the tests give similar results.

2. *Permutation f versus Kruskal-Wallis tests.*

In situations where both the permutation f test and the Kruskal-Wallis test are appropriate: if the samples are highly skewed or have extreme outliers, then the Kruskal-Wallis test is preferred; otherwise, the permutation f test is preferred.

In practice, the Kruskal-Wallis test is used almost exclusively since it is easy to implement.

3.1.8 Permutation Example: Biodiversity Analysis

Permutation analyses for I samples generalize analyses for 2 samples. They are appropriate in the following situations:

1. *Population Model:* The data are the values of independent random samples. The null hypothesis of interest is that the distributions from which the data were drawn are equal.
2. *Randomization Model:* A total of $N = n_1 + n_2 + \cdots + n_I$ individuals are randomly assigned to distinguishable groups of sizes n_1, n_2, \dots, n_I , respectively. The data are observations taken on individuals in each group.

Under the null hypothesis of randomness, the N observations can be randomly partitioned into distinguishable groups of sizes n_1, n_2, \dots, n_I .

Application: Simpson's Biodiversity Index. Researchers use biological diversity indices as measures of the ecological quality of natural sites such as tropical rain forests, coral reefs, and temperate woodlands. A commonly used biodiversity index is the Simpson index.

Let S be the total number of different species observed at the site, n_i be the observed number of individuals of species type i , and $N = n_1 + n_2 + \cdots + n_S$ be the total number of individuals observed. The *Simpson diversity index*, D , is defined as follows:

$$D = \sum_{i=1}^S \frac{n_i(n_i - 1)}{N(N - 1)}.$$

The diversity index D is an estimate of the probability that two individuals chosen without replacement will be of the same species. Sites with high diversity have values of D near zero; sites with low diversity have values of D near one.

For example, suppose that 4 different species, and a total of 60 individuals, are observed.

- If $n_1 = n_2 = n_3 = n_4 = 15$, then the value of D is (please complete)

- If $n_1 = n_2 = n_3 = 1, n_4 = 57$, then the value of D is (please complete)

Example (Source: Magurran, 1988). This example uses data on the number of fish of different species collected in three sections of the Upper Region of Black Creek, Mississippi.

A total of 11 different species (labeled 1 through 11) were observed at the three sites. For each site, the following table gives the number of fish of each species, the total number of fish, the total number of different species, and the observed value of the diversity index.

	1	2	3	4	5	6	7	8	9	10	11	N	S	
Site 1:	1	0	1	4	25	79	2	7	26	4	1	150	10	$d_1 = 0.335$
Site 2:	0	11	12	24	57	76	0	0	0	0	0	180	5	$d_2 = 0.301$
Site 3:	0	0	1	56	9	57	1	0	42	3	1	170	8	$d_3 = 0.281$
	1	11	14	84	91	212	3	7	68	7	2	500	11	$d = 0.260$

The diversity index for the combination of three sites is given on the bottom row, $d = 0.260$.

A natural statistic to use to determine if the biodiversity is the same across I study sites is the weighted sum of squared deviations of the individual indices from the combined index,

$$T = \sum_{i=1}^I N_i (D_i - D)^2,$$

where N_i is the number of individuals, and D_i is the diversity index, for the i^{th} study site, and D is the overall diversity index. The null hypothesis of equal biodiversity would be rejected for large values of T .

To determine if there are differences among the sites in this example, a permutation test will be conducted at the 5% significance level using the T statistic. For these data, $t_{\text{obs}} = 1.21455$.

In a Monte Carlo analysis using 2000 random partitions (including the observed partition of the sample data), 0.25% (5/2000) test statistics were as larger or larger than the observed. Further, a 99% confidence interval for the permutation p value is [0.0005,0.0071].

Thus (please complete),

3.2 Blocked Design

This section considers methods for matrices of observations of the form:

$$\{\{x_{1,1}, x_{1,2}, \dots, x_{1,J}\}, \{x_{2,1}, x_{2,2}, \dots, x_{2,J}\}, \dots, \{x_{I,1}, x_{I,2}, \dots, x_{I,J}\}\}$$

where

- the i^{th} *sample* is the list $\{x_{i,j} : j = 1, 2, \dots, J\}$, and
- the j^{th} *block* is the list $\{x_{i,j} : i = 1, 2, \dots, I\}$.

Let $N = IJ$ be the total number of observations.

For example, we might be interested in considering the following 4-by-5 matrix of numbers, where each row corresponds to a sample, and each column to a block.

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$i = 1$	4.3	9.9	13.9	8.5	9.2
$i = 2$	19.1	6.7	6.4	11.8	12.2
$i = 3$	12.1	11.7	11.8	18.3	12.8
$i = 4$	11.5	20.1	20.5	14.1	19.9

There are $I = 4$ samples, $J = 5$ blocks, and $N = 20$ observations.

Experimental setting: randomized blocks. Blocked samples arise in experiments known as *randomized block experiments*:

1. There are I treatments under study and J blocks of I individuals each, where individuals in a given block are matched on important factors.
2. The individuals within each block are randomly assigned to the I treatments. The individual receiving the i^{th} treatment is in the i^{th} experimental *group*.

For example,

- Age, gender and general health measures are typical blocking factors in medical studies.
- Time is a typical blocking factor in agricultural studies.

Note: Blocked Samples are Used to Reduce Variability

Blocked designs generalize paired samples designs.

Researchers use blocked designs to reduce the variability of the results since individuals within each block are expected to respond similarly to treatment, while those in different blocks are expected to respond differently to treatment.

3.2.1 Analysis of Variance for Blocked Designs

The data are assumed to be the values of $N = IJ$ independent normal random variables satisfying the following linear model:

$$X_{i,j} = \mu + \alpha_i + \beta_j + \epsilon_{i,j} \text{ for } i = 1, 2, \dots, I, j = 1, 2, \dots, J, \text{ where}$$

1. $\mu = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^J E(X_{i,j}) = \frac{1}{N} \sum_{i=1}^I \sum_{j=1}^J \mu_{i,j}$ is the *overall mean*.
2. $\alpha_i = \frac{1}{J} \sum_{j=1}^J \mu_{i,j} - \mu = \mu_{i\cdot} - \mu$ is the i^{th} *group effect* for $i = 1, 2, \dots, I$.
3. $\beta_j = \frac{1}{I} \sum_{i=1}^I \mu_{i,j} - \mu = \mu_{\cdot j} - \mu$ is the j^{th} *block effect* for $j = 1, 2, \dots, J$.
4. The collection of *error terms*, $\{\epsilon_{i,j}\}$, is a random sample of size N from a normal distribution with mean 0 and standard deviation σ .

Of primary interest is a test of

$$H_O : \alpha_1 = \alpha_2 = \dots = \alpha_I = 0 \text{ versus } H_A : \text{At least one } \alpha_i \text{ is not zero.}$$

The null hypothesis that the block effects are identically zero can also be tested.

Note: The definitions above imply that

$$\sum_{i=1}^I \alpha_i = 0 \text{ and } \sum_{j=1}^J \beta_j = 0.$$

Question: Can you see why this is true?

Exercise. Assume the linear model above holds.

Find the mean and variance of the difference between

- (1) Two random variables in the same block: $X_{i,j} - X_{k,j}$ when $i \neq k$.
- (2) Two random variables in the same group: $X_{i,j} - X_{i,k}$ when $j \neq k$.

Maximum likelihood estimation. Maximum likelihood (ML) estimators for the parameters of the linear model are given in the following theorem:

Theorem (Parameter Estimation). Given the assumptions and definitions above, the following are ML estimators of the parameters in the linear model:

1. Overall mean:

$$\hat{\mu} = \bar{X}_{..} = \frac{1}{N} \left(\sum_{i=1}^I \sum_{j=1}^J X_{i,j} \right).$$

2. Group effects:

$$\hat{\alpha}_i = \hat{\mu}_{i\cdot} - \hat{\mu} = \bar{X}_{i\cdot} - \bar{X}_{..} = \frac{1}{J} \left(\sum_{j=1}^J X_{i,j} \right) - \bar{X}_{..} \text{ for all } i.$$

3. Block effects:

$$\hat{\beta}_j = \hat{\mu}_{\cdot j} - \hat{\mu} = \bar{X}_{\cdot j} - \bar{X}_{..} = \frac{1}{I} \left(\sum_{i=1}^I X_{i,j} \right) - \bar{X}_{..} \text{ for all } j.$$

4. Error terms:

$$\hat{\epsilon}_{i,j} = X_{i,j} - \hat{\mu}_{i,j} = X_{i,j} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j) = X_{i,j} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{..} \text{ for all } i, j$$

Each is an unbiased estimator.

Further, the *pooled estimator* of the common variance,

$$S_p^2 = \frac{1}{(I-1)(J-1)} \sum_{i=1}^I \sum_{j=1}^J \hat{\epsilon}_{i,j}^2 = \frac{1}{(I-1)(J-1)} \sum_{i=1}^I \sum_{j=1}^J (X_{i,j} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{..})^2,$$

is an unbiased estimator of σ^2 .

Note on degrees of freedom:

IJ means are estimated using $1 + (I - 1) + (J - 1)$ parameters (the overall mean, the group effects, and the block block effects); the remaining degrees of freedom,

$$df = IJ - (1 + (I - 1) + (J - 1)) = (I - 1)(J - 1), \text{ are used to estimate } \sigma^2.$$

To illustrate the computations using the 4-by-5 data example on page 23:

The estimated overall mean is 12.74, and the pooled estimate of the common variance is 4.517². The following table shows the residuals (estimated errors), as well as the estimated group effects and estimated block effects.

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	
$i = 1$	$\hat{\epsilon}_{1,1}$	1.380	4.330	-1.095	-0.745	$\hat{\alpha}_1 = -3.58$
$i = 2$	8.850	-3.90	-5.250	0.125	0.175	$\hat{\alpha}_2 = -1.50$
$i = 3$	-0.250	-1.000	-1.950	4.525	-1.325	$\hat{\alpha}_3 = 0.60$
$i = 4$	-4.730	3.520	2.870	-3.555	1.895	$\hat{\alpha}_4 = 4.48$
	$\hat{\beta}_1 = -0.990$	$\hat{\beta}_2 = -0.640$	$\hat{\beta}_3 = 0.410$	$\hat{\beta}_4 = 0.435$	$\hat{\beta}_5 = 0.785$	

Finally, $\hat{\epsilon}_{1,1} = \underline{\hspace{15em}}$.

Initial analysis of variance table. Computations needed for an analysis of variance for a blocked design are often organized into an *analysis of variance table*:

	df	<i>Sum of Squares</i>	<i>Mean Square</i>
<i>Group</i>	$I - 1$	$SS_g = \sum_{i=1}^I J (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$	$MS_g = \frac{1}{I-1} SS_g$
<i>Block</i>	$J - 1$	$SS_b = \sum_{j=1}^J I (\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})^2$	$MS_b = \frac{1}{J-1} SS_b$
<i>Error</i>	$(I - 1)(J - 1)$	$SS_e = \sum_{i=1}^I \sum_{j=1}^J (X_{i,j} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2$	$MS_e = \frac{1}{(I-1)(J-1)} SS_e$
<i>Total</i>	$IJ - 1$	$SS_t = \sum_{i=1}^I \sum_{j=1}^J (X_{i,j} - \bar{X}_{\cdot\cdot})^2$	

1. The total sum of squares (SS_t) is equal to the sum of the group sum of squares (SS_g), the block sum of squares (SS_b), and the error sum of squares (SS_e):

$$SS_t = SS_g + SS_b + SS_e.$$

2. The group mean square (MS_g) is a random variable with mean

$$E(MS_g) = \sigma^2 + \frac{J}{I-1} \sum_{i=1}^I (\mu_{i\cdot} - \mu)^2 = \sigma^2 + \frac{J}{I-1} \sum_{i=1}^I \alpha_i^2.$$

Under the null hypothesis that the group effects are identically zero, the mean is σ^2 ; otherwise, the mean is greater than σ^2 .

3. The block mean square (MS_b) is a random variable with mean

$$E(MS_b) = \sigma^2 + \frac{I}{J-1} \sum_{j=1}^J (\mu_{\cdot j} - \mu)^2 = \sigma^2 + \frac{I}{J-1} \sum_{j=1}^J \beta_j^2.$$

Under the null hypothesis that the block effects are identically zero, the mean is σ^2 ; otherwise, the mean is greater than σ^2 .

4. The error mean square (MS_e) is the same as the pooled estimator of σ^2 .

Under the general assumptions of the linear model for blocked designs, the error mean square is an unbiased estimator of σ^2 , and

$$V = \frac{(I-1)(J-1)}{\sigma^2} MS_e \text{ has a chi-square distribution with } (I-1)(J-1) \text{ df.}$$

To conduct an analysis of variance for group effects, we compute the ratio $F = MS_g/MS_e$. Large value of F favor the general alternative that not all group effects are zero. And, we can conduct a similar analysis for block effects.

The following distribution theorem, proven by Fisher, gives us sampling distributions under both null hypotheses:

Theorem (Distribution Theorem). Under the general assumptions of the linear model in this section,

1. If the null hypothesis that the group effects are identically zero is true, then

$$F = MS_g/MS_e$$

has an f ratio distribution with $(I-1)$ and $(I-1)(J-1)$ df.

2. If the null hypothesis that the block effects are identically zero is true, then

$$F = MS_b/MS_e$$

has an f ratio distribution with $(J-1)$ and $(I-1)(J-1)$ df.

Completed ANOVA table. In applications, an analysis of variance table is expanded to include columns for observed f statistics (f_{obs}), and observed significance levels,

$$p \text{ Value} = P(F \text{ Ratio RV} \geq f_{\text{obs}}),$$

where each p value is based on the appropriate f ratio distribution.

Example. Assume the data presented on page 23 are the values of independent normal random variables satisfying the linear model of this section, and that we are interested in testing both hypotheses at the 5% significance level.

The following is an analysis of variance table for these data:

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p value</i>
<i>Group:</i>	3	177.484	59.161	2.899	0.079
<i>Block:</i>	4	9.453	2.363	0.116	0.974
<i>Error:</i>	12	244.851	20.404		
<i>Total:</i>	19	431.788			

Since both p values are greater than 0.05, we accept the null hypotheses that the group effects are identically zero and that the block effects are identically zero. There is insufficient evidence to conclude otherwise.

Example (Source: Rice textbook, Chapter 12). “Five drugs [for relieving itching] were compared to a placebo and no drug with 10 volunteer male subjects aged 20-30. . . Each volunteer underwent one treatment per day, and the time-order was randomized. Thus, individuals were “blocks.” The subjects were given a drug (or placebo) intravenously, and then itching was induced on their forearms with cowage, an effective itch stimulus. The subjects recorded the duration of the itching [in seconds].”

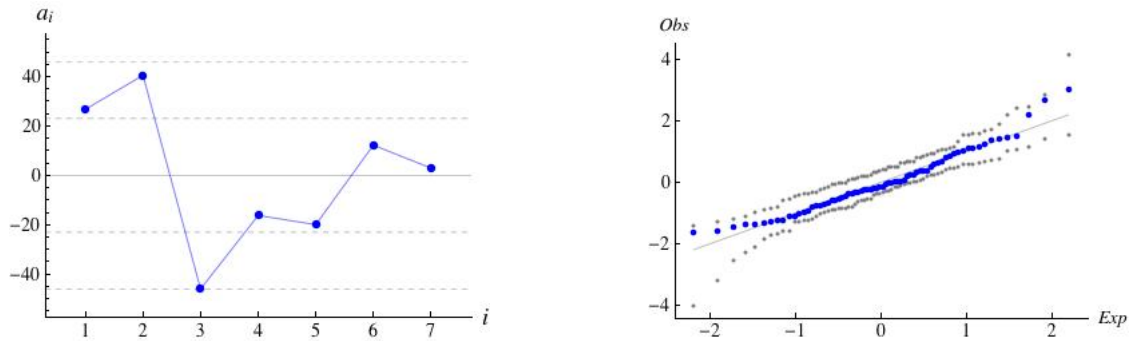
Observations for the 10 subjects (labeled 1 through 10) and the 7 treatments are given in the following table, along with observed group means (in seconds).

	1	2	3	4	5	6	7	8	9	10	<i>Group Mean:</i>
<i>No Drug:</i>	174	224	260	255	165	237	191	100	115	189	191.0
<i>Placebo:</i>	263	213	231	291	168	121	137	102	89	433	204.8
<i>Papaverine:</i>	105	103	145	103	144	94	35	133	83	237	118.2
<i>Morphine:</i>	199	143	113	225	176	144	87	120	100	173	148.0
<i>Amnophylline:</i>	141	168	78	164	127	114	96	222	165	168	144.3
<i>Pentobarbital:</i>	108	341	159	135	239	136	140	134	185	188	176.5
<i>Tripelennamine:</i>	141	184	125	227	194	155	121	129	79	317	167.2

For reference, here is a table of mean times (in seconds) for each subject:

1	2	3	4	5	6	7	8	9	10
161.6	196.6	158.7	200.0	173.3	143.0	115.3	134.3	116.6	243.6

The *left plot* below is a plot of estimated group effects (labeled a_i); a horizontal solid line is drawn at 0, which is the average of the estimated effects. The *right plot* is an enhanced normal probability plot of standardized errors.



There is a very slight curvature, but it is not bad.

Assume these data are the values of independent normal random variables satisfying the general assumptions of this section, and that we are interested in testing the null hypothesis that the group effects are identically zero at the 5% significance level.

The following is an analysis of variance table including results for both significance tests:

	df	SS	MS	F	p value
<i>Group:</i>	6	53012.9	8835.48	2.855	0.01730
<i>Block:</i>	9	103280.0	11475.50	3.708	0.00122
<i>Error:</i>	54	167130.0	3094.99		
<i>Total:</i>	69	323422.0			

Thus (please complete),

3.2.2 Multiple Comparisons in the Parametric Setting

If the null hypothesis that the group effects are identically zero is rejected, then it is natural to try to determine which groups differ; multiple comparisons are needed.

When comparing group means in the analysis of variance for blocked designs setting, we work with the following distribution theorem:

Theorem (Sampling Distribution). Under the general conditions of this section, the statistic

$$T_{i,k} = \frac{(\bar{X}_{i\cdot} - \bar{X}_{k\cdot}) - (\mu_{i\cdot} - \mu_{k\cdot})}{\sqrt{2S_p^2/J}}, \text{ where } S_p^2 \text{ is the pooled estimator of } \sigma^2,$$

has a Student t distribution with $(I-1)(J-1)$ df. In this statement, i and k are distinct integers satisfying $i, k \in \{1, 2, \dots, I\}$.

This theorem can be used when comparing any two samples.

Bonferroni method for multiple comparisons. Using the Bonferroni approach, to conduct all $m = \binom{I}{2}$ two sided tests of equality of group means,

$$\mu_{i\cdot} = \mu_{k\cdot} \text{ versus } \mu_{i\cdot} \neq \mu_{k\cdot} \text{ for } i < k,$$

we use paired t statistics of the form

$$T_{i,k} = \frac{\bar{X}_{i\cdot} - \bar{X}_{k\cdot}}{\sqrt{2S_p^2/J}}, \text{ and conduct each test at the } \alpha/m \text{ level.}$$

Exercise. Continuing with the study on time to relief from itching, begun on page 29, a Bonferroni analysis will be conducted at the 5% significance level.

A total of $21 = \binom{7}{2}$ comparisons will be made.

The rejection region for each test is $|T| \geq t_{54}((0.05/21)/2) = 3.188$.

The 21 observed t statistics are given in the following table:

	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
$i = 1$	-0.555	2.926	1.728	1.877	0.583	0.957
$i = 2$		3.481	2.283	2.432	1.137	1.511
$i = 3$			-1.198	-1.049	-2.343	-1.969
$i = 4$				0.149	-1.146	-0.772
$i = 5$					-1.294	-0.920
$i = 6$						0.374

There is only one significant difference at the 5% level, namely the difference between the mean time to relief for papaverine and the mean time to relief for placebo.

Construct and interpret a 95% confidence interval for comparing papaverine and placebo.

3.2.3 Friedman Statistic, Distribution and Methods

In the 1930's, Milton Friedman developed a nonparametric test for comparing groups in blocked designs appropriate in the following situations:

1. *Population model.* The data are the values of $N = IJ$ independent random variables. The null hypothesis is that all I distributions within each block are equal. Alternatives of interest are that for at least one block, at least one distribution is stochastically larger or smaller than another.
2. *Randomization model.* The blocked data are measurements on J individuals (or J blocks of I individuals each). The null hypothesis is that observed differences in measurements within each block are due to chance alone. Alternatives of interest are that at least one group has values that tend to be larger or smaller than the values in another.

Let $R_{i,j}$ be the rank of observation $x_{i,j}$ in the j^{th} block, and $\bar{R}_{i\cdot}$ be the average rank of observations in the i^{th} sample, and assume that observations are distinct within blocks.

Then the Friedman statistic, Q , is defined as follows:

$$Q = \frac{12J}{I(I+1)} \sum_{i=1}^I (\bar{R}_{i\cdot} - \bar{R}_{..})^2 = \frac{12J}{I(I+1)} \sum_{i=1}^I \left(\bar{R}_{i\cdot} - \frac{I+1}{2} \right)^2.$$

Thus, the Friedman statistic is a weighted average of squared deviations of average group ranks from the overall average rank.

For example, the following table shows the ranks for the small data example from page 23, and observed averages by group ($\bar{r}_{i\cdot}$) and for all 20 observations ($\bar{r}_{\cdot\cdot} = 2.5$).

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	<i>Group Mean:</i>
$i = 1$	1	2	3	1	1	$\bar{r}_{1\cdot} = (8/5) = 1.6$
$i = 2$	4	1	1	2	2	$\bar{r}_{2\cdot} = (10/5) = 2.0$
$i = 3$	3	3	2	4	3	$\bar{r}_{3\cdot} = (15/5) = 3.0$
$i = 4$	2	4	4	3	4	$\bar{r}_{4\cdot} = (17/5) = 3.4$

The observed value of the statistic for these data is

$$q_{\text{obs}} = \frac{\binom{12}{4}\binom{5}{5}}{\binom{4}{5}} ((1.6 - 2.5)^2 + (2.0 - 2.5)^2 + (3.0 - 2.5)^2 + (3.4 - 2.5)^2) = 6.36,$$

Question: Can you see why $\bar{R}_{\cdot\cdot} = (I + 1)/2$?

The sampling distribution of Q is obtained by computing its value for each matching of the I ranks in each block to the I groups. There are a total of $(I!)^J$ matchings to consider.

The following theorem gives a large sample approximation to the distribution.

Theorem (Friedman Distribution). If there are no ties within blocks, then under the null hypothesis of randomness and when N is large, the distribution of Q is approximately chi-square with $(I - 1)$ df.

Large values of Q support the alternative hypothesis that the values in at least one group tend to be larger or smaller than those in another. The observed significance level is

$$P(Q \geq q_{\text{obs}}),$$

where q_{obs} is the observed value of the Friedman statistic.

Note on large sample approximations:

In most practical situations, the chi-square approximation to the sampling distribution of Q is used to compute p values.

A “rule of thumb” is that the chi-square approximation is adequate when there are 3 or more groups, the total sample size is 30 or more, and observations are distinct within each block.

Handling equal observations. Average ranks (or *midranks*) replace ranks when there are ties in a given block.

Now, the Friedman statistic becomes

$$Q = \sum_{i=1}^I w_i (\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2 = \sum_{i=1}^I w_i \left(\bar{R}_{i\cdot} - \frac{I+1}{2} \right)^2,$$

where the weights, w_i , in the weighted sum are chosen to make the approximate sampling distribution of Q under the null hypothesis as close to the chi-square distribution as possible.

Most computer programs automatically use the appropriate weights.

3.2.4 Multiple Comparisons in the Nonparametric Setting

If the Friedman test suggests that group differences are not due to chance alone, then the next step is to try to determine which groups differ.

Bonferroni approach to multiple comparisons. In the Bonferroni approach to multiple comparisons, we conduct $m = \binom{I}{2}$ two sided Wilcoxon signed ranks tests. If each test is conducted at the $\frac{\alpha}{m}$ level, then the overall type I error will be at most α .

Example (Source: Wackerley, et al, 2002). Corrosion of metals is a problem in many mechanical devices. Three sealants used to help retard corrosion of metals were tested to see whether there were any differences among them. Samples from ten different ingots (bars) of the same metal composition were treated with each of the three sealants, with location-order randomized, and the amount of corrosion was measured after exposure to the same environmental conditions; the higher the corrosion score, the greater the corrosion.

The first table shows the original corrosion scores, and the second table shows the ranks within blocks, for each sealant and each ingot (labeled 1 through 10):

	1	2	3	4	5	6	7	8	9	10	Group Mean
<i>Sealant 1:</i>	4.6	7.2	3.4	6.2	8.4	5.6	3.7	6.1	4.9	5.2	5.53
<i>Sealant 2:</i>	4.2	6.4	3.5	5.3	6.8	4.8	3.7	6.2	4.1	5.0	5.00
<i>Sealant 3:</i>	4.9	7.0	3.4	5.9	7.8	5.7	4.1	6.4	4.2	5.1	5.45

<i>Sealant 1:</i>	2	3	1.5	3	3	2	1.5	1	3	3	2.30
<i>Sealant 2:</i>	1	1	3	1	1	1	1.5	2	1	1	1.35
<i>Sealant 3:</i>	3	2	1.5	2	2	3	3	3	2	2	2.35

Group means have been added to each table.

Consider testing the null hypothesis of randomness at the 5% significance level. The following table summarizes the analysis:

<i>Friedman Statistic</i>	<i>p Value</i>	<i>Sampling Distribution</i>
6.684	0.0354	Chi-Square, 2 df

Since there are significant differences, the Bonferroni method will be used next.

A total of $3 = \binom{3}{2}$ two sided signed rank tests will be conducted at the $\frac{0.05}{3} = 0.016\bar{6}$ level. The following table gives the p values:

<i>(1,2)</i>	<i>(1,3)</i>	<i>(2,3)</i>
0.0234375	0.753906	0.00585938

The results suggest (please complete)

3.2.5 Permutation f Test

A permutation version of the analysis of variance f test for blocked designs is appropriate in one of the following situations:

1. *Population model.* The observed data are the values of IJ independent random variables from distributions differing in mean only. The null hypothesis is that for each j , the distributions of

$$X_{i,j}, \quad i = 1, 2, \dots, I$$

are equal (the means are equal within each block) versus the general alternative that at least two means differ in some block.

2. *Randomization model.* The blocked data are measurements on J individuals (or J blocks of I individuals). The null hypothesis is that observed differences in measurements within each block are due to chance alone versus the alternative that at least one group has values that tend to be larger or smaller (but not more variable) than the values in another.

The sampling distribution of $F = MS_g/MS_e$ is obtained by computing its value for each matching of the I observations in a given block to the I treatments. There are a total of $(I!)^J$ matchings to consider. Since this number can be quite large, Monte Carlo analysis is generally used to estimate a p value.

Example. Consider again the study on relief from itching begun on page 29. To determine if there are significant differences among treatments, a permutation f test will be conducted at the 5% significance level.

In a Monte Carlo analysis with 10000 random permutations (including the observed permutation of the 70 duration times), 1.92% (192/10000) of F values were greater than or equal to $f_{\text{obs}} = 2.855$. Further, a 99% confidence interval for the p value is [0.0158, 0.0230].

The results suggest that the group effects are not identically zero. Further, papaverine appears to have the minimum average time to relief.

3.3 Balanced Two Way Layout

This section considers methods for matrices of samples of the form:

$$\{\{s_{1,1}, s_{1,2}, \dots, s_{1,J}\}, \{s_{2,1}, s_{2,2}, \dots, s_{2,J}\}, \dots, \{s_{I,1}, s_{I,2}, \dots, s_{I,J}\}\}$$

where each $s_{i,j}$ is a list of K observations:

$$s_{i,j} = \{x_{i,j,1}, x_{i,j,2}, \dots, x_{i,j,K}\}.$$

- i is used to index the I levels of the first factor of interest (the *row* factor), and
- j is used to index the J levels of the second factor of interest (the *column* factor).
- There are a total of IJ samples (or *groups*), and
- a total of $N = IJK$ observations.

For example, we might be interested in considering the following table of 60 numbers:

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$i = 1$	1.52	2.76	5.35	11.46	14.02
	2.40	14.30	13.50	18.74	18.09
	12.31	12.88	8.69	11.39	18.40
$i = 2$	9.29	16.46	13.11	12.16	18.18
	3.56	18.25	11.11	20.50	17.25
	16.83	15.08	27.45	16.65	14.75
$i = 3$	10.11	17.93	18.16	22.44	18.48
	1.48	12.27	9.18	19.55	5.12
	5.73	6.53	13.52	20.55	16.47
$i = 4$	9.97	17.40	24.97	20.32	26.66
	1.13	14.62	16.92	21.36	25.07
	14.94	9.85	18.89	26.00	20.87

The table is structured as a 4-by-5 matrix of samples of size 3.

Experimental setting: randomized factorial designs. Balanced two way layouts arise from *randomized factorial designs*, where researchers are interested in determining if each factor separately affects the outcome and if there are certain combinations of factors that are especially good (or especially bad).

For example, when studying weight loss, the first factor might correspond to different diet programs and the second factor to different exercise programs. Researchers would be interested in knowing

- if the diet programs differ,
- if the exercise programs differ, and
- if particular diet-exercise programs are especially effective (or especially detrimental).

The last item suggests an *interaction* between diet and exercise.

3.3.1 Analysis of Variance for Balanced Two Way Layouts

The data are assumed to be the values of IJ independent random samples, each of size K , from normal distributions with a common unknown standard deviation σ .

Let $\mu_{i,j}$ be the mean of observations in the (i, j) -sample:

$$\mu_{i,j} = E(X_{i,j,k}), \quad k = 1, 2, \dots, K.$$

The general assumptions imply that each $X_{i,j,k}$ can be written in the following linear form:

$$X_{i,j,k} = \mu + \alpha_i + \beta_j + \delta_{i,j} + \epsilon_{i,j,k} \quad \text{for all } i, j, k$$

where

1. $\mu = E\left(\frac{1}{N} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K X_{i,j,k}\right) = \frac{1}{IJ} \left(\sum_{i=1}^I \sum_{j=1}^J \mu_{i,j}\right)$ is the *overall mean*.
2. $\alpha_i = \mu_{i\cdot} - \mu = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K \mu_{i,j} - \mu = \frac{1}{J} \sum_{j=1}^J \mu_{i,j} - \mu$ is the i^{th} *row effect* for all i .
3. $\beta_j = \mu_{\cdot j} - \mu = \frac{1}{IK} \sum_{i=1}^I \sum_{k=1}^K \mu_{i,j} - \mu = \frac{1}{I} \sum_{i=1}^I \mu_{i,j} - \mu$ is the j^{th} *column effect* for all j .
4. $\delta_{i,j} = \mu_{i,j} - (\mu + \alpha_i + \beta_j)$ is the $(i, j)^{\text{th}}$ *interaction term*.
5. The collection of *error terms*, $\{\epsilon_{i,j,k}\}$, is a random sample of size N from a normal distribution with mean 0 and standard deviation σ .

Three null hypotheses are of interest:

1. the differential effects of the first factor are identically zero,
2. the differential effects of the second factor are identically zero, and
3. the interaction terms are identically zero.

In each case, the alternative hypothesis is that at least one effect is not zero.

In this model,

- The sum of row effects is zero: $\sum_{i=1}^I \alpha_i = 0$.
- The sum of column effects is zero: $\sum_{j=1}^J \beta_j = 0$.
- The interaction effects satisfy

$$\sum_{j=1}^J \delta_{i,j} = 0 \text{ for each } i, \text{ and } \sum_{i=1}^I \delta_{i,j} = 0 \text{ for each } j.$$

- Group means have the form $\mu_{i,j} = \mu + \alpha_i + \beta_j + \delta_{i,j}$ for all i, j .

Note: You should be able to figure out why each of the items above is true.

Simple additive model. If the interaction terms are identically zero, then the means satisfy the simple additive model:

$$\mu_{i,j} = \mu + \alpha_i + \beta_j.$$

For example, in the weight loss scenario, where the row factor corresponds to different diet programs and the column factor corresponds to different exercise programs,

- if a particular diet-exercise combination is extremely effective (or extremely detrimental), then there will be some nonzero interaction terms;
- otherwise, the simple additive model would hold.

Maximum likelihood estimation. Maximum likelihood (ML) estimators for the parameters in the linear model are given in the following theorem.

Theorem (Parameter Estimation). Given the assumptions and definitions above, the following are ML estimators of the parameters in the linear model:

1. Overall mean:

$$\hat{\mu} = \bar{X}\dots = \frac{1}{N} \left(\sum_{i,j,k} X_{i,j,k} \right)$$

2. Row effects:

$$\hat{\alpha}_i = \widehat{\mu_{i\cdot}} - \hat{\mu} = \frac{1}{JK} \left(\sum_{j,k} X_{i,j,k} \right) - \hat{\mu} = \bar{X}_{i\cdot\cdot} - \bar{X}\dots \text{ for all } i$$

3. Column effects:

$$\hat{\beta}_j = \widehat{\mu_{\cdot j}} - \hat{\mu} = \frac{1}{IK} \left(\sum_{i,k} X_{i,j,k} \right) - \hat{\mu} = \bar{X}_{\cdot j\cdot} - \bar{X}\dots \text{ for all } j$$

4. Interactions (row-by-column effects):

$$\begin{aligned} \widehat{\delta_{i,j}} &= \widehat{\mu_{i,j}} - \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \right) \\ &= \frac{1}{K} \left(\sum_k X_{i,j,k} \right) - \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \right) \\ &= \bar{X}_{i,j\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}\dots \text{ for all } i, j \end{aligned}$$

5. Error terms:

$$\widehat{\epsilon_{i,j,k}} = X_{i,j,k} - \widehat{\mu_{i,j}} = X_{i,j,k} - \bar{X}_{i,j\cdot} \text{ for all } i, j, k$$

Each is an unbiased estimator.

Further, the *pooled estimator* of the common variance

$$S_p^2 = \frac{1}{N - IJ} \sum_{i,j,k} \widehat{\epsilon_{i,j,k}}^2 = \frac{1}{N - IJ} \sum_{i,j,k} (X_{i,j,k} - \bar{X}_{i,j\cdot})^2$$

is an unbiased estimator of σ^2 .

Note on estimating σ^2 :

Let $S_{i,j}^2$ be the sample variance of the (i, j) -sample.

Since the design is balanced, the pooled estimator of the common variance is

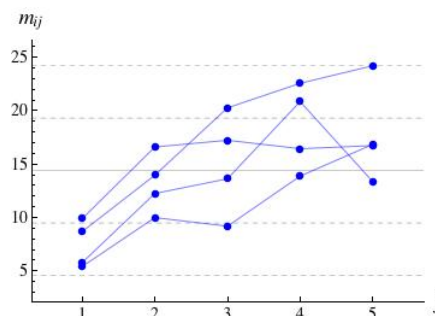
$$S_p^2 = \frac{1}{N - IJ} \sum_{i,j,k} (X_{i,j,k} - \bar{X}_{i,j\cdot})^2 = \frac{1}{IJ(K-1)} \sum_{i,j} (K-1)S_{i,j}^2 = \frac{1}{IJ} \sum_{i,j} S_{i,j}^2$$

To illustrate the computations using the data given on page 37:

The estimated overall mean for the 60 numbers is 14.382, and the pooled estimate of the common variance is 4.937^2 .

The table below gives the 20 estimated group means, and the plot displays the estimated group means (labeled m_{ij}) as separate line plots for each level of the first factor.

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$i = 1$	5.410	9.980	9.180	13.863	16.837
$i = 2$	9.893	16.597	17.223	16.437	16.727
$i = 3$	5.773	12.243	13.620	20.847	13.357
$i = 4$	8.680	13.957	20.260	22.560	24.200



Interaction plot. The plot above is an example of an *interaction plot*: If the simple additive model holds, then the separate line plots will be roughly parallel; significant departures from parallel suggest that the row and column factors interact.

Effects table. Researchers often display the estimated interaction terms ($\hat{\delta}_{i,j}$) in an I -by- J table, along with the estimated row effects ($\hat{\alpha}_i$) and the estimated column ($\hat{\beta}_j$).

For example, the data from our simple example from page 37 can be displayed as follows:

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	
$i = 1$	$\hat{\delta}_{1,1}$	0.114	-2.563	-1.235	2.385	$\hat{\alpha}_1 = -3.328$
$i = 2$	1.461	2.409	1.159	-2.983	-2.047	$\hat{\alpha}_2 = 0.993$
$i = 3$	-0.452	0.263	-0.237	3.634	-3.209	$\hat{\alpha}_3 = -1.214$
$i = 4$	-2.308	-2.787	1.640	0.584	2.871	$\hat{\alpha}_4 = 3.549$
	$\hat{\beta}_1 = -6.943$	$\hat{\beta}_2 = -1.188$	$\hat{\beta}_3 = 0.689$	$\hat{\beta}_4 = 4.045$	$\hat{\beta}_5 = 3.398$	

The missing value is $\hat{\delta}_{1,1} = \underline{\hspace{10cm}}$.

Initial analysis of variance table. Computations needed for an analysis of variance for a balanced two way layout are often organized into an *analysis of variance table*:

	<i>df</i>	<i>Sum of Squares</i>	<i>Mean Square</i>
<i>Row</i>	$I - 1$	$SS_r = \sum_i JK (\bar{X}_{i..} - \bar{X}...)^2$	$MS_r = \frac{1}{I-1} SS_r$
<i>Col</i>	$J - 1$	$SS_c = \sum_j IK (\bar{X}_{.j.} - \bar{X}...)^2$	$MS_c = \frac{1}{J-1} SS_c$
<i>R-by-C</i>	$(I - 1)(J - 1)$	$SS_{r \times c} = \sum_{i,j} K (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}...)^2$	$MS_{r \times c} = \frac{1}{(I-1)(J-1)} SS_{r \times c}$
<i>Error</i>	$IJ(K - 1)$	$SS_e = \sum_{i,j,k} (X_{i,j,k} - \bar{X}_{ij.})^2$	$MS_e = \frac{1}{IJ(K-1)} SS_e$
<i>Total</i>	$N - 1$	$SS_t = \sum_{i,j,k} (X_{i,j,k} - \bar{X}...)^2$	

In this table,

1. The total sum of squares (SS_t) equals the sum of the row sum of squares (SS_r), the column sum of squares (SS_c), the interaction (or row-by-column) sum of squares ($SS_{r \times c}$), and the error sum of squares (SS_e):

$$SS_t = SS_r + SS_c + SS_{r \times c} + SS_e.$$

2. The row mean square (MS_r) is a random variable with mean

$$E(MS_r) = \sigma^2 + \frac{JK}{I-1} \sum_i (\mu_{i.} - \mu)^2 = \sigma^2 + \frac{JK}{I-1} \sum_i \alpha_i^2.$$

If the null hypothesis that the row effects are identically zero is true, then the expected value of MS_r is σ^2 ; otherwise, values of MS_r will tend to be larger than σ^2 .

3. The column mean square (MS_c) is a random variable with mean

$$E(MS_c) = \sigma^2 + \frac{IK}{J-1} \sum_j (\mu_{.j} - \mu)^2 = \sigma^2 + \frac{IK}{J-1} \sum_j \beta_j^2.$$

If the null hypothesis that the column effects are identically zero is true, then the expected value of MS_c is σ^2 ; otherwise, values of MS_c will tend to be larger than σ^2 .

4. The interaction mean square ($MS_{r \times c}$) is a random variable with mean

$$E(MS_{r \times c}) = \sigma^2 + \frac{K}{(I-1)(J-1)} \sum_{i,j} (\mu_{ij} - (\mu + \alpha_i + \beta_j))^2 = \sigma^2 + \frac{K}{(I-1)(J-1)} \sum_{i,j} \delta_{i,j}^2.$$

If the null hypothesis that the interaction effects are identically zero is true, then the expected value of $MS_{r \times c}$ is σ^2 ; otherwise, the values will tend to be larger than σ^2 .

5. The error mean square (MS_e) is the same as the pooled estimator σ^2 .

Thus, under the general assumptions of the linear model, the error mean square is an unbiased estimator of σ^2 , and

$$V = \frac{IJ(K-1)}{\sigma^2} MS_e \text{ has a chi-square distribution with } IJ(K-1) \text{ df.}$$

To conduct an analysis of variance for any one of the three hypotheses, we compute the ratio of the appropriate mean square to the error mean square. Large value of the ratio favor the general alternative that not all effects are zero.

The following distribution theorem, proven by Fisher, gives us sampling distributions for all three null hypotheses.

Theorem (Distribution Theorem). Under the general assumptions of the linear model in this section:

1. If the row effects are identically zero, then the ratio

$$F = MS_r/MS_e$$

has an f ratio distribution with $(I - 1)$ and $IJ(K - 1)$ degrees of freedom.

2. If the column effects are identically zero, then the ratio

$$F = MS_c/MS_e$$

has an f ratio distribution with $(J - 1)$ and $IJ(K - 1)$ degrees of freedom.

3. If the interaction effects are identically zero, then the ratio

$$F = MS_{r \times c}/MS_e$$

has an f ratio distribution with $(I - 1)(J - 1)$ and $IJ(K - 1)$ degrees of freedom.

Completed ANOVA table. In applications, an analysis of variance table is expanded to include columns for observed f statistics (f_{obs}), and observed significance levels,

$$p \text{ Value} = P(F \text{ Ratio RV} \geq f_{\text{obs}}),$$

where each p value is based on the appropriate f ratio distribution.

Example. Assume the data presented on page 37 are the values of independent normal random variables satisfying the linear model of this section, and that we are interested in testing each null hypothesis at the 5% significance level.

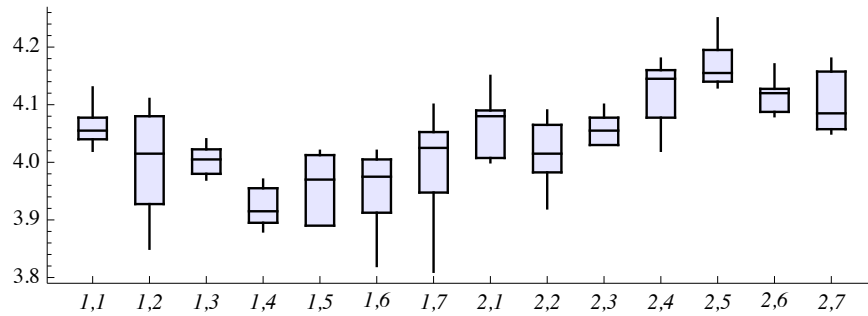
The following is an analysis of variance table for these data:

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p value</i>
<i>Row:</i>	3	392.008	130.669	5.361	0.00337
<i>Col:</i>	4	935.929	233.982	9.600	0.00002
<i>Row-by-Col:</i>	12	258.160	21.513	0.883	0.57100
<i>Error:</i>	40	974.930	24.373		
<i>Total:</i>	59	2561.030			

The results suggest significant row and column effects at the 5% significance level. Further, since the interaction terms are not significantly different from zero at the 5% level, the results suggest that a simple linear model holds.

Example (Source: Rice textbook, Chapter 12). “Kirchhoefer (1979) reports on a study of the use of a semiautomated method for measuring the amount of chlorpheniramine in tablets. For each of four manufacturers, composites were prepared by grinding and mixing together tablets that had nominal dosage levels of 4 milligrams (mg). Seven labs were asked to make 10 determinations on each composite; each determination was made on a portion of composite whose weight was equivalent to that of one tablet. The purpose of the experiment was to determine the consistency between labs and the variability of the measurement process.”

This example extends the example from page 10 to data on two manufacturers. The plot below shows side-by-side box plots for the two manufacturers:

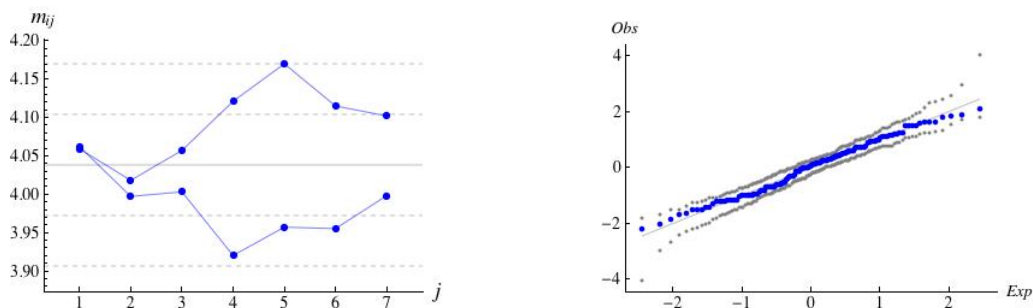


The following table gives estimated means for each group, with estimated standard deviations in parentheses, and other summary information:

	<i>Lab 1</i>	<i>Lab 2</i>	<i>Lab 3</i>	<i>Lab 4</i>	<i>Lab 5</i>	<i>Lab 6</i>	<i>Lab 7</i>	<i>Mfr Mean:</i>
<i>Mfr 1</i>	4.062 (0.033)	3.997 (0.09)	4.003 (0.023)	3.920 (0.033)	3.957 (0.057)	3.955 (0.067)	3.998 (0.085)	3.985
<i>Mfr 2</i>	4.059 (0.051)	4.018 (0.058)	4.057 (0.027)	4.121 (0.054)	4.170 (0.043)	4.115 (0.029)	4.102 (0.052)	4.092
<i>Lab Mean:</i>	4.061	4.008	4.030	4.021	4.064	4.035	4.050	

The estimated overall mean is 4.038, and the pooled estimate of the common variance is 0.054^2 . The ratio of the maximum to the minimum sample standard deviation is about 3.913.

The *left* plot is a plot of estimated group means by manufacturer, and the *right* plot is an enhanced normal probability plot of estimated standardized errors:



Assume that the tablets data are the values of independent random samples from normal distributions satisfying the general assumptions of this section, and that we wish to test all three null hypotheses at the 5% significance level.

The following is an analysis of variance table for these tests:

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p value</i>
<i>Row:</i>	1	0.4017	0.4017	138.033	≈ 0
<i>Col:</i>	6	0.0522	0.0087	2.989	0.0092
<i>Row-by-Col:</i>	6	0.2260	0.0377	12.939	≈ 0
<i>Error:</i>	126	0.3668	0.0029		
<i>Total:</i>	139	1.0467			

(Please complete this analysis)

3.3.2 Multiple Comparisons in the Parametric Setting

Multiple comparisons in the two way analysis of variance setting are conducted using methods for the analysis of IJ samples for one way analysis of variance.

Bonferroni method for multiple comparisons. Using the Bonferroni approach, to conduct all $m = \binom{IJ}{2}$ two sided tests of equality of means,

$$\mu_i = \mu_k \text{ versus } \mu_i \neq \mu_k, \text{ for } i < k,$$

we use pooled t statistics of the form

$$T_{i,k} = \frac{\bar{X}_{i\cdot} - \bar{X}_{k\cdot}}{\sqrt{S_p^2 \left(\frac{1}{n_i} + \frac{1}{n_k} \right)}}, \text{ where } S_p^2 \text{ is the pooled estimator of } \sigma^2,$$

and conduct each test at the α/m level.

The single indexing needed for Bonferroni analysis is usually implemented as first row samples,

$$1 \equiv (1, 1), 2 \equiv (1, 2), \dots, I \equiv (1, I),$$

followed by second row samples,

$$I + 1 \equiv (2, 1), I + 2 \equiv (2, 2), \dots, 2I \equiv (2, I),$$

and so forth.

Example. Continuing with the last example, there are $91 = \binom{14}{2}$ comparisons to consider.

A Bonferroni analysis was conducted at the 5% significance level; there were 40 significant results. In the following table, plus/minus signs are used to indicate significant mean differences: the plus when an observed difference was positive; and minus otherwise.

(i, k)	2	3	4	5	6	7	8	9	10	11	12	13	14
1			+	+	+						-		
2											-	-	-
3											-	-	-
4							-	-	-	-	-	-	-
5							-		-	-	-	-	-
6							-		-	-	-	-	-
7										-	-	-	-
8											-		
9										-	-	-	
10											-		
11													
12													
13													

These results suggest (please complete)