CHAPTER 15. INTRODUCTION TO DESIGN

- Experiments vs. Observational Studies, Section 15.1
- Experimental design principles; Section 15.2
- Overview of designs for experimental studies; Section 15.3

Section 15.1: Experimental and Observational Studies

We need just a little more detail about experimental and observational studies.

In the background, we have a specific population, and we wish to compare two or more "treatments."

A *treatment* is a specific set of conditions that in principle can be applied to the population (e.g. administer vitamin C to an individual child, apply a new fertilizer to plots of land, use a new method for teaching statistics in a class, etc.)

We will talk about *observational studies* and *experiments*.

Observational study A study in which we observe individuals and measure variables, but don't interfere in any way.

Example: Method 1 in the vitamin C example

Experiment A study in which we specifically impose a "treatment" on individuals in order to observe their responses to it. Purpose is to try to determine whether the treatment *causes* a change in the response.

Example: Method 2 in the vitamin C example (the actual experiment)

Advantage of well-designed experiment over observational study

From observational studies, it is difficult (or impossible) to infer causation. From a well-designed experiment, we *can* conclude causation.

Usefulness of Randomization in Experiments

Randomization makes the two groups similar with respect to all factors (other than that one gets the new treatment and the other gets the old treatment). This includes factors we might have predicted would affect the response *and* factors we had not ever even thought about. Thus the two groups differ only in that one got the new treatment and the other got the old.

Another way to put this:

Randomization eliminates confounding *Confounding* means a difference between the treatment and control groups–other than the treatment—which affects the responses being studied. A *confounder* is a third variable, associated with the treatment factor and with the response.

In a controlled, randomized experiment, the difference in response between the two groups is caused either by a genuine difference in the effects of the two treatments *or* by chance.

If we can essentially exclude the possibility that chance alone caused the difference in the responses, then we have established causality.

If by use of randomization, we have eliminated the possibility of confounding in our experiment, then the result of a statistical hypothesis test is valuable information.

Statistical Significance

When the observed difference in responses between the experimental and control groups is so large that it cannot be attributed to just a chance outcome, the difference is called “statistically significant.”

If we have randomized subjects to experimental and control groups, then there are just two possible explanations for an observed difference in the mean responses for the two groups:

1. There is a real population difference between effects of treatment and placebo, or
2. The observed difference is not ”real,” but occurred just by chance.

If the P-value of our two-sample t test (or other appropriate procedure) is very small, this eliminates Possibility 2, and proves the ”alternative hypothesis” (Possibility 1).

Example. Vitamin C (See p. 4.)

Method 1 The people taking Vit C every day may be older and more likely to catch colds because of weaker immune system than the people not taking Vit C. This would produce a bias against Vitamin C. What is the confounder here?

Or, there could be a bias in favor of Vitamin C. Some offered by students in class: Healthier people take Vitamin C as part of their health regimen. Even without the Vitamin C, they get fewer colds because they eat better food than average, stay away from sick people, etc.

Method 2 An older person is equally likely to be assigned to the placebo group, as to the Vitamin C group. Thus there is no bias either for or against Vitamin C (at least, no bias, probabilistically speaking).

What to Do if We Have Observational Studies

If you can, do a controlled (i.e. randomized) study.

If you have to do with an observational study, then you can and should try to "control for confounding factors." This means to subdivide the population into smaller but more homogeneous subgroups, within which confounding factors do not exist.

Example:

Caution

Section 15.2: Basic Concepts in Experimental Studies

- Factor: an explanatory variable, with *levels* or values that it can take on
  - Experimental factor is one that is assigned at random to units
  - Observational factor is intrinsic to the unit and cannot be assigned at random

- Multifactor studies have more than one factor. *Ex.*: Effect of three levels of temperature and two levels of concentration of solvent on yield of a chemical process.

Example. Quick bread volume Effect of baking temperature (low, medium, high, very high) on volume of bread was studied by assigning five package mixes at random to each condition.

Design Vocabulary

Experimental factor: Vitamin C in Colds experiment, temperature in Quick Bread Volume.

Observational factor: Vitamin C in hypothetical Method 1 on Vit C and colds.

In regression, factors are usually called predictors or explanatory variables. Note that if we have an observational factor in an experiment, we still have an experiment wrt the primary factor of interest. However, we can't infer causality wrt the observational factor.

Steps in designing an actual experiment

Proper design requires consideration of what factors and how many factors to include, number and value of levels of each included factor, range of levels for quantitative factors, what control treatment to include.

Section 15.2 continued: Crossed vs. nested factors

- In multifactor studies, factors can be *crossed* or *nested*. *Ex. Yield of chemical process* If all 3×2 possible treatments (factor combinations) are included in the experiment, then the factors are *crossed*.

- The levels of a *nested* factor are unique to a particular level of another factor. Ex. of study with *nested* factor: Effects of operators on production yield in three manufacturing plants. The operators are different people in each of the three plants.

- *Experimental unit*: the smallest unit to which a treatment can be assigned. Thus the method of randomization determines what is the experimental unit. *Example* Quick Bread Volume. The twenty package mixes are the experimental units.

- Sample size and replication. *Ex* Quick bread; sample size is 20, and there are five (5) replicates of the basic experiment.

. **Experimental Units**

Ex. Vit C/Colds The child is the experimental unit.

Ex. Educational methods study Three high schools participated in a study to evaluate the effectiveness of a new computer-based math curriculum. In each school, four $24-$student sections of freshman algebra were available for the study. The two types of instruction (standard curriculum, computer-based curriculum) were randomly assigned to the four sections in each of the three schools. At the end of the term, a standard math achievement test was given to each of the 24 students in each section.

If the treatment is applied to all the students in the classroom, then the class is the unit, not the individual student.

- You do collect data (test scores) on each individual student separately. The unit on which you collect data may not be the same as the experiment unit.
- The **experimental unit** is that entity to which the treatment can be applied at random.

How to randomize (conceptually) We will use a "random number generator" on the computer (in R) in practice.

Ex. Vit C/Colds Say there are 20 children in the experiment, for simplicity.

- Number the 20 children, 1 to 20.
- Generate "random permutation" of these numbers.
- Assign the children with the first 10 id numbers to Vitamin C, and the remaining children to Placebo.

Constrained randomization, or blocking

Eg. Say you want to control the number of males and females in the experiment, because you know gender is associated with frequency of colds.

Steps in a blocked experiment (one with constrained randomization):

- Select 434 males and 434 females for the experiment.
- Randomize the 434 males to Vit C and Placebo groups (as described above)
- Randomize the 434 females to Vit C and Placebo groups (as described above)

Section 15.3: EXPERIMENTAL DESIGNS

- Completely randomized design (CRD)
  To compare two treatments, so far our experimental design has been a CRD:
  > All experimental subjects are allocated at random among the two treatments.

  More generally, the Quick Bread Volume experiment is an example of a CRD with four treatments, provided the twenty package mixes are randomly allocated, five apiece to each of the four treatments.
- Completely randomized factorial design
  Uses the same randomization scheme as the CRD. There are two or more crossed factors which define the set of treatments.
  Example. Chemical yield p. 43 Experimental unit might be a single run of the chemical process, using standard company equipment. In this 3×2 factorial design, if there are 18 runs assigned at random to the six treatments, this is a CR-factorial.

- Randomized complete block design. In a blocked design, first the experimental units are divided into homogenous groups, each with the same number of units as there are treatments. Then the treatments are assigned at random within each of the blocks separately.

  Example Quick Bread Volume , where you have eight mixes, but four mixes were produced in Plant A and four mixes were produced in Plant B. You decide to run the experiment in two blocks of size four:

  Note: The treatments can be defined by crossing of two or more factors in a factorial design; this would be called a factorial experiment run in blocks.

- Repeated measures design
  This is a variation on randomized block design.
  A repeated measures design is a randomized blocks experiment, where the subject is the block. The order in which a subject receives the treatments is randomized.

  Example Compare three cereal formulations using $n = 12$ consumer volunteer taste-testers. Compare CRD and repeated measures design.

- Split-plot design
  Example Does perceived wholesomeness affect rating? Six consumers are told that the cereals are formulations of a new organic, nutritious cereal; the other six consumers are told just that the formulations are new formulations of a cereal.

How to randomize an experiment in R.

Illustration of R function `sample`:

```
committee <- c("Al", "Bob", "Cathy", "Don", "Ed", "Gail")
set.seed(1)
subcommittee <- sample(x=committee, size=3, replace=F)
subcommittee
[1] "Bob"   "Gail"  "Cathy"
```

From the committee, select 3 people at random and w/o replacement.

```
individuals <- 1:20
set.seed(1); sample(x=individuals, size=10, replace=F)
[1]  6  8 11 16  4 14 15  9 19  1
```

This randomly splits the individuals into two groups, each of size 10.

How to randomize an experiment in R (Completely randomized design)

Example Obtain a randomization of units to treatments for a CRD with four treatments and two replicates. Use R function `sample`

Solution:

Label the eight units: 1, 2, 3, 4, 5, 6, 7, 8

List the treatments: T1, T1, T2, T2, T3, T3, T4, T4

Get a random permutation of the units, and assign this permutation of the units to the treatments as listed above:

```
> set.seed(12)
> sample(1:8,replace=FALSE)
[1] 1 6 7 2 8 4 3 5
```

Randomization scheme:

One-way Analysis of Variance (ANOVA)

The basic setup is simple: We have several (≥ 3) groups. This is like the two-sample problem, but with at least one additional group. The groups are *independent* (separate and unrelated or unpaired).

A grouping factor can be *experimental*, perhaps a control group and two or more treatment groups, e.g. two different doses of a medicine, or two different medicines.

Or, a grouping factor can be *observational*, not assigned by the experimenter but a feature of the data.

Example 1. Dose of medicine What is the best dosage level of a particular medicine? In order to compare effectiveness of three dosage levels, first recruit 30 patients with the medical problem to participate in a pilot study. Assign each patient at random to one of the three drug dosage levels, such that exactly ten patients receive each dosage level. One of the groups is a placebo group (dosage is zero, and a fake medicine is given). This is a *balanced* completely randomized one-way experimental design.

Example 2. Absorptive Properties of Paper Towels In an experiment to compare paper towel brands, five sheets of paper were selected from each of Bounty, Scott, Viva, and Generic. Twenty 6-ounce beakers of water were prepared, and the twenty paper towel sheets were randomly assigned to the beakers. This is a balanced CR one-way design with one factor at four levels.

In both of the examples, we commonly view the predictor variable as a single, qualitative predictor variable.

The resulting model can be viewed as a regression model with several dummy variable predictors.

For r levels of the factor, we need $r - 1$ dummy variables. Spell this out:

Choice between ANOVA model and regression model

Example 1. Effectiveness of Medicine Dose The goal is to compare effectiveness of various doses of a certain medicine in reducing blood pressure.

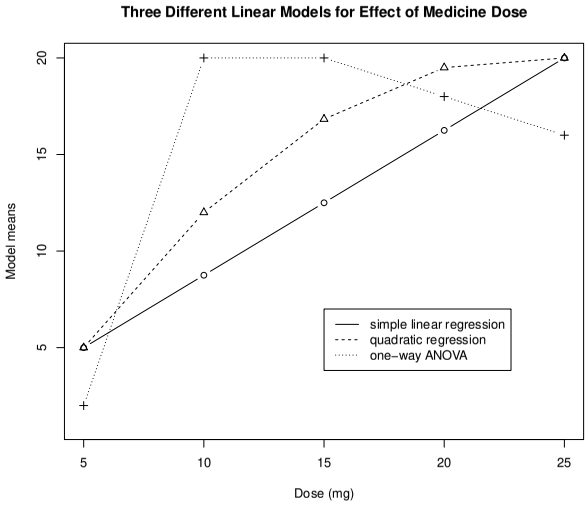
- Experimental units: 30 subjects
- Response variable: Y is change in blood pressure (pre - post)
- One factor, dose of medicine. Let's say there are six levels. (We need multiple doses for the purposes of this discussion.)
- Design: Randomly assign the 30 subjects to the six groups. This is a CRD, with one factor at six levels.
- Analysis: You can use one-way analysis of variance. OR: use regression analysis.

Because the predictor variable X is quantitative, there are two possible approaches to the analysis.

The reason this is important to think about is that for dose to be treated as numerical, you'd need a regression model to be appropriate.

Ex. 1 Effectiveness of Medicine Dose, con.

Compare the following models for this situation: simple linear regression, quadratic regression, and one-way ANOVA:



Model 1, simple linear regression: $EY_i = \beta_0 + \beta_1 x_i$ Key assumption: The slope is constant over the whole range of doses. This might not hold.

Model 2, quadratic regression: $EY_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2$ (we would center the x's first to do quadratic regression right) This model allows the means to increase less as doses increase and is sometimes a realistic model.

Model 3, one-way ANOVA: This is most general model of the three: allows arbitrarily different means for each dose. This figure has only one out of a huge number of possible patterns of the means.

Example for Illustration of Statistical Analysis

Example. Kenton Food Company (Section 16.4, p. 691) The company wants to compare four different package designs for a new breakfast cereal in terms of sales.

- Experimental units: 20 stores [2]
- Data: A fire occurred in one store, which had to be dropped from the study. Response variable Y was sales, in number of cases.
- Design: Completely randomized design (CRD) with package design as the single factor, with four levels.
- Further points: Shelf space, level of advertising, and other factors which might affect sales were kept constant among the stores in the study.

Key point about the model The model says there could be arbitrarily different means μ_A, μ_B, μ_C, μ_D in the four groups.

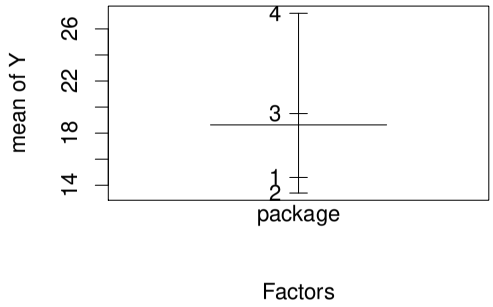
[2] comparable in sales volume and location

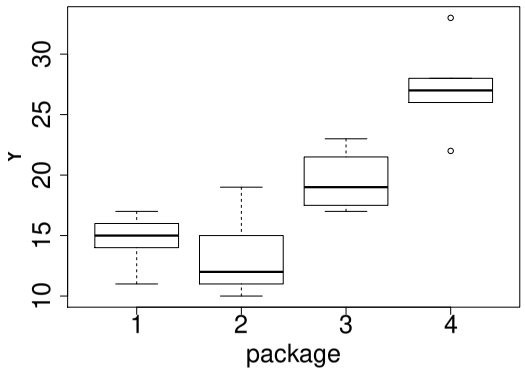
Example. Kenton Food Company. Recall: Four package designs for a breakfast cereal are being tested in five stores each (except for one package design which was tested in four stores).

Data:

	Store (j)				
Package Design	1	2	3	4	5
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}
1	11	17	16	14	15
2	12	10	15	19	11
3	23	20	18	17	
4	27	33	22	26	28

Plot the data:





There are two main parts in the analysis:

1. An overall test to see if there is statistical evidence that there exist *any* differences.
2. A more detailed *follow-up* analysis to decide which of the populations differ, and to estimate how large the differences are.

Let $r =$ number of levels of the explanatory variable (number of treatment groups for example). Let $n_i =$ number of cases (experimental units) in i^{th} group, and let $n_T = \sum_{i=1}^{r} n_i$ be the total number of observations.

Let the population mean parameters be μ_i, $i = 1, \ldots, r$.

Hypotheses to be tested:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_r$ vs.

H_a: at least two of the means are not equal.

Assumptions:

- Each of the r population distributions is normal.
- The r standard deviations are all equal.
- All n_T observations are taken independently.

Cell-means model

Let Y_{ij} denote the response of the j^{th} unit in the i^{th} group (i^{th} level of the factor).

$$Y_{ij} = \mu_i + \epsilon_{ij},\ \text{i} = 1, \ldots, r;\ \text{j} = 1, \ldots, n_i$$

where

- the μ_i are parameters,
- the errors, ϵ_{ij}, are independent $\mathcal{N}(0, \sigma^2)$

Another way to state the model:

Y_{ij} are independent $\mathcal{N}(\mu_i, \sigma^2),\ \text{i} = 1, \ldots, r;\ \text{j} = 1, \ldots, n_i$

Example Kenton Food Company

Suppose we know $\mu_1 = 15$, $\mu_2 = 16$, $\mu_3 = 20$, $\mu_4 = 28$, $\sigma = 1.5$

Two of the observations are:

Y	package	id
11	1	1
19	2	4

In this hypothetical situation, find the error terms for these two observations.

Cell means model is a linear model

Illustrate why, using a case involving $r = 3$ treatments and two replicates per treatment.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{pmatrix}.$$

Notation for cell and grand means

Let $Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}$, be the sum of the observations in Group i, for $i = 1, \ldots, r$.

Then the mean of the i^{th} group is:

$$\overline{Y}_{i\cdot} = \frac{1}{n_i} Y_{i\cdot}$$

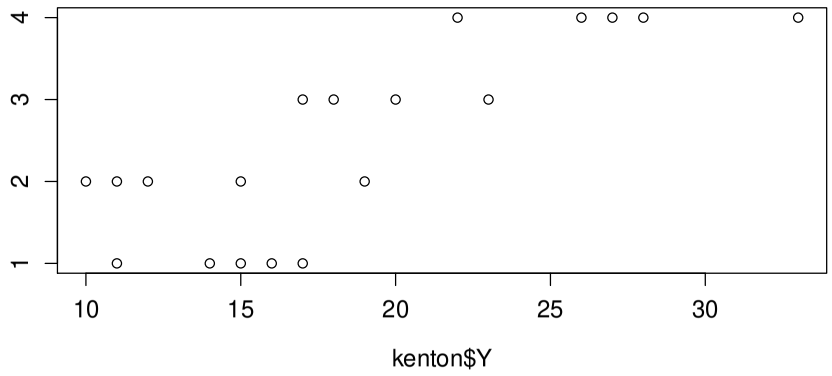
The group means are also called the *cell means*.

Let $Y_{\cdot\cdot} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij}$ be the total of all the observations.

Then the *grand mean* is:

$$\overline{Y}_{\cdot\cdot} = \frac{1}{n_T} Y_{\cdot\cdot}$$

Fitting the one-way ANOVA model: Least Squares



- For each possible set of parameter values $(\mu_1, \mu_2, \mu_3, \mu_4)$, calculate the sum of the squared distances (SS) between the y-values and their means
- Find the values of $(\mu_1, \mu_2, \mu_3, \mu_4)$ that minimize SS; these are the "least squares" estimates.

Least Squares Estimators

The least-squares estimates of the $\mu_i, i = 1, \ldots, r$ are the values $\hat{\mu}_i$ which minimize the sum of squared deviations of Y's from their expected values. Call the quantity to be minimized Q; then Q is given by

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2$$

To minimize Q, you can minimize each of the r terms above separately, since each term involves exactly one parameter. The answer for term i is $\hat{\mu}_i = \overline{Y}_{i.}$

A simple explanation, not requiring calculus, can be given in this case. Drop the index i for simplicity.

We want to minimize $\sum_{j=1}^{n} (Y_j - \mu)^2$ w.r.t. μ.

We want to minimize $Q = \sum_{j=1}^{n} \left(Y_j - \mu\right)^2$ w.r.t. μ.

First, add and subtract $\overline{Y}$ to each term in the sum; then expand the quadratic. We have

$$Q = \sum_{j=1}^{n} \left(Y_j - \overline{Y} + \overline{Y} - \mu\right)^2 = \sum_{j=1}^{n} \left(Y_j - \overline{Y}\right)^2 + \sum_{j=1}^{n} 2(Y_j - \overline{Y})(\overline{Y} - \mu) + \sum_{j=1}^{n} (\overline{Y} - \mu)^2$$

The second term on the right is 0.

So we have

$$Q = \sum_{j=1}^{n} \left(Y_j - \overline{Y}\right)^2 + \sum_{j=1}^{n} (\overline{Y} - \mu)^2$$

The first term is constant wrt μ, and the second term is obviously minimized when we take $\mu = \overline{Y}$. Therefore, $\hat{\mu} = \overline{Y}$.

QED.

Fitted values and residuals

Recall the model equation for the cell-means model:

$$Y_{ij} = \mu_i + \epsilon_{ij},\ \mathrm{i} = 1, \ldots, r;\ \mathrm{j} = 1, \ldots, n_i$$

The fitted values $\hat{Y}_{ij}$ are defined to be:

$$\hat{Y}_{ij} = \overline{Y}_{i.},\ \mathrm{i} = 1, \ldots, r;\ \mathrm{j} = 1, \ldots, n_i$$

The residuals:

$e_{ij} = Y_{ij} - \overline{Y}_{i.}$

Fitted equation:

$$Y_{ij} = \overline{Y}_{i.} + e_{ij},\ \mathrm{i} = 1, \ldots, r;\ \mathrm{j} = 1, \ldots, n_i$$

Property of residuals in one-way ANOVA is that within every group, $i = 1, \ldots r$, the residuals sum to zero:

$$\sum_{j=1}^{n_i} e_{ij} = 0$$

Ex. Kenton Food Company

	Store (j)					Number of			
Package Design	1	2	3	4	5	Stores	Total	Mean	SD
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	n_i	$Y_{i.}$	$\bar{Y}_{i.}$	s_i
1	11	17	16	14	15	5	73		
2	12	10	15	19	11	5	67		
3	23	20	18	17		4	78		
4	27	33	22	26	28	5	136		

$Y_{..} = 354$, $\bar{Y}_{..} = 18.63$, $n_T = 19$

Residuals

$e_{11} = Y_{11} - \hat{Y}_{11} = Y_{11} - \bar{Y}_{1.} =$

$e_{34} =$

Analysis of variance Basis is decomposition of observations, sums of squares and df, like in regression.

Decomposition of observations:

$$Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.}) \tag{1}$$

Decomposition of sums of squares and degrees of freedom:

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

$$n_T - 1 = (r-1) + (n_T - r)$$

The decomposition of sums of squares holds because when you square the right-hand side of Equation (1), the cross-product term is zero.

Explanation of why cross-product term is zero:

Decomposition of observations:

$$Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.})$$

When you square the right-hand side of the above equation, the cross-product term is:

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i} 2(\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.}) = \sum_{i=1}^{r} 2(\overline{Y}_{i.} - \overline{Y}_{..}) \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.}) \qquad = 0$$

Decomposition of sum of squares:

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

$$\text{SSTO} = \text{SSTR} + \text{SSE}$$

Ex. Kenton Food Company

Package Design	Number of Stores	Mean	SD
i	n_i	$\bar{Y}_{i.}$	s_i
1	5	14.6	2.302
2	5	13.4	3.647
3	4	19.5	2.646
4	5	27.2	3.962

Grand mean: $\bar{Y}_{..} = (5(14.6) + 5(13.4) + 4(19.5) + 5(27.2))/19 = 18.632$

$$\begin{aligned}\text{SSTR} &= 5(14.6 - 18.632)^2 + 5(13.4 - 18.682)^2 + 4(19.5 - 18.632)^2 \\ &\quad + 5(19.5 - 18.632)^2 = 588.2211 \\ \text{SSE} &= 4(2.302^2) + 4(3.647^2) + 3(2.646^2) + 4(3.962^2) = 158.2\end{aligned}$$

Mean Square: a sum of squares, divided by its degrees of freedom. Note that a *mean square* is a statistic and has a sampling distribution.

$$\text{MSTR} = \frac{\text{SSTR}}{r-1}$$
$$\text{MSE} = \frac{\text{SSE}}{n_T - r}$$

Define $\mu_.$ to be the weighted mean of the $\mu_i, i = 1, \ldots, r$ in the cell-means model:

$$\mu_. = \frac{\sum_{i=1}^{r} n_i \mu_i}{n_T}$$

Expected Mean Square: the mean of the sampling distribution of a mean square.

$$\text{E(MSTR)} = \sigma^2 + \frac{\sum n_i (\mu_i - \mu_.)^2}{r-1}$$
$$\text{E(MSE)} = \sigma^2$$

ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	Expected Mean Square
Treatment	SSTR	$r-1$	$\frac{SSTR}{r-1}$	
Error	SSE	$n_T - r$	$\frac{SSE}{n_T - r}$	
Total	SSTO	$n_T - 1$		

Kenton Food Company

Source of Variation	Sum of Squares	df	Mean Square	F	P
Treatment	588.22	3	196.07		
Error	158.2	15	10.55		
Total	746.42	18			

F Test for Equality of Factor Level Means

Hypotheses to be tested:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_r$ vs.

H_a: at least two of the means are not equal.

The F Statistic

$$F_{\text{obs}} = \frac{MSTR}{MSE}$$

For fixed r and $n_1, n_2, \ldots, n_r$, if H_0 is true then the distribution of F is completely determined.

$$F_{\text{obs}} = \frac{\frac{\text{SSTR}}{r-1}}{\frac{\text{SSE}}{n_T - r}}$$

If H_0 is true, the distribution of F_{obs} is $F_{r-1, n_T - r}$. The two parameters are the numerator degrees of freedom and denominator degrees of freedom ($r - 1$, $n_T - r$). [3]

If H_0 is not true, the numerator MSTR will tend to be larger than what would be expected if H_0 was true. To see this, compare E(MSTR) when H_0 is true, to E(MSTR) when H_0 is false.

So, under H_a, F_{obs} will also be larger than expected under H_0. So the test consists of comparing the ratio with the 95% point of the F distribution with $r - 1$ and $n_T - r$ degrees of freedom.

[3] The numerator is a χ^2 r.v. divided by its degrees of freedom, and the denominator is another χ^2 r.v. divided by its degrees of freedom. The numerator and denominator are independent of each other by Cochran's Thm.

ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	Expected Mean Square
Treatment	SSTR	$r-1$	$\frac{\text{SSTR}}{r-1}$	$\sigma^2 + \frac{\sum n_i(\mu_i - \mu_.)^2}{r-1}$
Error	SSE	$n_T - r$	$\frac{\text{SSE}}{n_T - r}$	σ^2
Total	SSTO	$n_T - 1$		

Kenton Food Company

Source of Variation	Sum of Squares	df	Mean Square	F	P
Treatment	588.22	3	196.07		
Error	158.2	15	10.55		
Total	746.42	18			

P-value of the F test is area to the right of the observed statistic F_{obs}.

Ex: Kenton Food Company P-value $= P(F_{3,15} > 18.591)$

```
> 1 - pf(18.591,df1=3,df2=15)
[1] 2.585007e-05
```

Since $P < .05$, reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ at level $\alpha = .05$. Conclude that at least two of the package designs differ in number of packages sold on average.

F Test with pre-set Type I error probability α

For $\alpha = .05$ first find the $.95$ quantile of the appropriate F distribution.

```
> qf(.95,df1=3,df2=15)
[1] 3.287382
```

Decision Rule: For a level $\alpha = .05$ test, we will reject H_0 if $F_{\text{obs}} > 3.288$

Factor Effects Model (Section 16.7)

$$Y_{ij} = \mu_\cdot + \tau_i + \epsilon_{ij},\ \mathrm{i} = 1, \ldots, r;\ \mathrm{j} = 1, \ldots, n_i$$

where:

- $\mu_\cdot$ is the "overall mean," more accurately described as a "constant component common to all observations"
- τ_i is the effect of the i^{th} factor level
- the errors, ϵ_{ij}, are independent $\mathcal{N}(0, \sigma^2)$

Usually take $\mu_\cdot$ to be the unweighted mean of the factor level means:

$$\mu_\cdot = \frac{\sum_{i=1}^{r} \mu_i}{r}$$

This leads to the constraint on the τ_i's:

$$\sum_{i=1}^{r} \tau_i = 0$$

Linear model approach to factor effects model *Example* $r = 4$, $n = 2$ observations per group

Parameter vector:

$$\boldsymbol{\beta} = \begin{pmatrix} \mu_\cdot \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{pmatrix}$$

Design matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The design matrix is **not full rank**

Linear model approach to factor effects model

We will take the approach of solving for one of the τ's in terms of the others. In the example, take $\tau_4 = -\tau_1 - \tau_2 - \tau_3$

Parameter vector:

$$\boldsymbol{\beta} = \begin{pmatrix} \mu_\cdot \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}$$

Design matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix}$$

Chapter 17 Analysis of Factor Level Means

It is convenient to use the cell-means formulation of the model for this discussion.

Topics:

- Individual inferences, e.g. for a single μ_i or for a difference of means $\mu_i - \mu_j$.
- Contrasts
- Simultaneous inference

Individual inferences

Ex: Kenton Food Company (See p. 60)

Package Design	Characteristics
1	3 colors, with cartoons
2	3 colors, without cartoons
3	5 colors, with cartoons
4	5 colors, without cartoons

Confidence Interval for a factor level mean

The goal is as in the one-sample situation, but since we have data from several groups, we use all the data to estimate the variance σ^2, with $\hat{\sigma}^2 = \text{MSE} = \text{SSE}/(n_T - r)$.

Basic Facts

1. $E(\hat{\mu}_i) = \mu_i$.
2. $\text{SD}(\hat{\mu}_i) = \dfrac{\sigma}{\sqrt{n_i}}$
3. $\hat{\mu}_i$ has a normal distribution.
4. $(\hat{\mu}_i - \mu_i) \Big/ \left(\dfrac{\sigma}{\sqrt{n_i}}\right) \sim \mathcal{N}(0,1)$.
5. $(\hat{\mu}_i - \mu_i) \Big/ \left(\dfrac{\hat{\sigma}}{\sqrt{n_i}}\right) \sim t_{n_T - r}$.

It follows from point 5 that a $100(1-\alpha)\%$ CI for μ_i is:

Ex. Kenton Food Company

- Estimate the mean sales for package design 4 with a 95% CI:

- Estimate the mean difference in sales for package design 2 vs. package design 1.

- Let's compare mean sales for designs with and without cartoons; estimate the following contrast.

$$L = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$$

Need for Simultaneous Inference in One-Way ANOVA

With r groups, we are interested in up to $\binom{r}{2}$ comparisons of means, and may be interested in other *contrasts* as well.

Methods are needed to control the family-wise confidence level or significance level. In experimental studies, this makes conclusions more interpretable. In observational studies, these methods make *data snooping* possible.

We will consider three multiple comparison methods:

1. Tukey's studentized range, for all possible pairwise comparisons. Useful in practice. Has to be tweaked to handle unequal group sizes.

2. Scheffé method, for all possible contrasts. Amazing that such a method exists, but the intervals tend to be very wide.

3. Bonferroni method. Useful, flexible, easy. Requires that you specify contrasts to be estimated in advance of getting the data.

Tukey Studentized Range Intervals

Consider the usual one-way ANOVA model with r treatments, and suppose we want to do $\binom{r}{2}$ tests to compare pairs μ_i, μ_j for all $i \neq j$, and also to get CI's for all possible differences $\mu_i - \mu_j$.

Suppose the sample sizes are all equal, and let n denote the common value.

Studentized range distribution

Suppose:

- $Y_1, Y_2, \ldots, Y_r$ are independent and normally distributed with mean μ and standard deviation σ.
- The statistic s^2 is an unbiased estimate of σ^2, is independent of the Y_i's, and the quantity νs^2 has a χ^2 distribution with ν d.f..

Let $R = \max(Y_i) - \min(Y_i)$ be the range of the Y's.

The studentized range is

$$q = \frac{R}{s}$$

The distribution of q depends on the parameters r and ν, is available in tables and in software.

Example For $r = 7$, $\nu = 21$, the $.95$ quantile of the studentized range distribution is:

```
> qtukey(.95,r=7,df=21)
[1] 4.597302
```


Pairwise comparisons of means

r groups with n observations per group, MSE estimates σ^2.

We want to give a confidence interval formula which used the standard error of the difference of means. Recall:

$$\mathrm{SE}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}) = \sqrt{\mathrm{MSE}}\sqrt{\frac{2}{n}}$$

The Tukey studentized range method for all pairwise comparisons of means forms $\binom{r}{2}$ intervals as follows. For estimation of $\mu_i - \mu_j$, take $\hat{D} = \overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}$.

$$\hat{D} \pm \frac{1}{\sqrt{2}} q(1-\alpha, r, n_T - r)\mathrm{SE}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})$$

Example 1: Rust inhibitors. Four brands of rust inhibitor were tested (A, B, C, D). Ten experimental units were assigned at random to each brand. Sample size is 40, number of replicates is 10. Response Y is a coded value for which higher values indicate less rust.

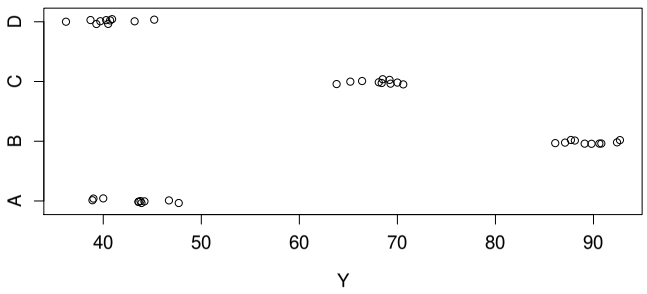
This is an experimental study with one categorical predictor variable. The correct model to use is the one-way ANOVA model; the single factor has four levels.

Read the dataset into R, rename variables, and make the group variable a `factor` object:

```
> rust <- read.table("CH17TA02.txt", header=FALSE)
> str(rust)
'data.frame':        40 obs. of  3 variables:
 $ V1: num  43.9 39 46.7 43.8 44.2 47.7 43.6 38.9 43.6 40 ...
 $ V2: int  1 1 1 1 1 1 1 1 1 1 ...
 $ V3: int  1 2 3 4 5 6 7 8 9 10 ...
> names(rust) <- c("Y","brand","obs")
> rust$brand <- factor(rust$brand,labels=c("A","B","C","D"))
```


ALWAYS PLOT YOUR DATA

Dotplots of rust inhibition scores for each of the four brands:



Do the means appear different?

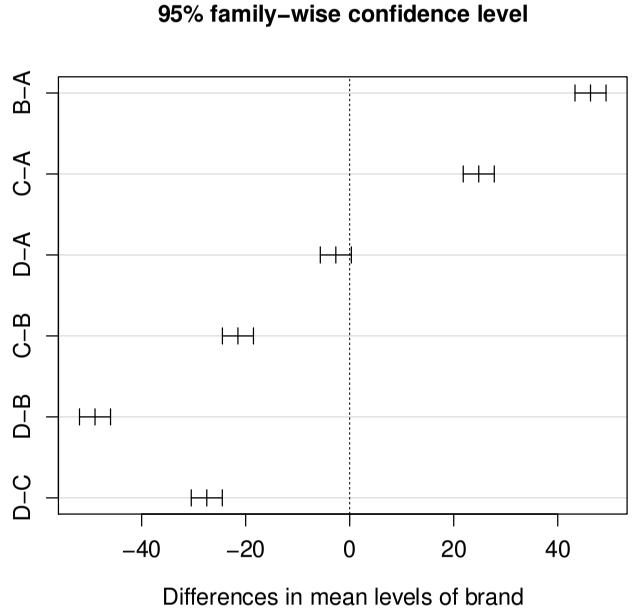
Are the variances within groups roughly equal? (informal check of equal-variances assumption)

```
> rust.aov <- aov(Y~brand, data=rust)
> anova(rust.aov)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
brand      3  15954  5317.8  866.12 < 2.2e-16 ***
Residuals 36    221     6.1
```

98

```
> plot(TukeyHSD(rust.aov))
```



99

Interpretation of family-wise confidence level in the example:

We are 95% confident that all of the six confident intervals are correct (contain the parameter being esimated).

In infinitely many repeat experiments like this one, with the same number of groups and sample sizes, and where we form the six intervals by Tukey's method, the proportion of experiments with all intervals correct would be 95%.

Simultaneous Testing

Scheffé multiple comparison method

The Scheffé method allows simultaneous confidence intervals (or tests) for all possible contrasts of the cell means,

$$L = \sum_{i=1}^{r} c_i \mu_i, \text{ where } \sum c_i = 0$$

Procedure:

The estimate of L is:

$$\hat{L} = \sum c_i \overline{Y}_{i.}$$

Estimate the variance of $\hat{L}$ by

$$s^2(\hat{L}) = \text{MSE} \sum \frac{c_i^2}{n_i}$$

Let $S^2 = (r-1)F(1-\alpha; r-1, n_T - r)$ (Here the capital "S" refers to critical constant for Scheffé method)

Scheffé critical constant:

Let $S^2 = (r-1)F(1-\alpha; r-1, n_T - r)$ (Here the capital "S" refers to critical constant for Scheffé method)

The Scheffé confidence intervals are:

$$\hat{L} \pm Ss(\hat{L})$$

Ex Rust, all pairwise comparisons

Purpose of this example is to compare Scheffé method to Tukey's studentized range method, in the case where both apply.

Tukey:

$$\hat{D} \pm \frac{1}{\sqrt{2}} q(1-\alpha, r, n_T - r)\text{SE}(\overline{Y}_{i.} - \overline{Y}_{j.})$$

Scheffé:

$$\hat{D} \pm \sqrt{(r-1)F(1-\alpha; r-1, n_T - r)}\text{SE}(\overline{Y}_{i.} - \overline{Y}_{j.})$$

Rust example Find the critical constants for: individual 95% t CI, Tukey's method, and Scheffé method

```
> qt(.975,36)
[1] 2.028094
> qtukey(.95,nmeans=4,df=36)/sqrt(2)
[1] 2.693227
> S <- sqrt(3*qf(.95,3,36)); S
[1] 2.93237
```

Scheffé method can handle any contrast, not just differences of two means.

Ex Suppose you wanted to compare Brands A and D to Brands B and C, so you decide to estimate the contrast

$$L = \frac{\mu_A + \mu_D}{2} - \frac{\mu_B + \mu_C}{2}$$

The estimate is:

$$\hat{L} = \frac{\overline{Y}_{\text{A.}} + \overline{Y}_{\text{D.}}}{2} - \frac{\overline{Y}_{\text{B.}} + \overline{Y}_{\text{C.}}}{2}$$

and the estimate of the variance is

$$\hat{\text{Var}}(\hat{L}) = \frac{1}{4}\left(4 \times \frac{\text{MSE}}{n}\right)$$

Bonferroni method

If we want to just make k comparisons, where k is small, then we can use the (trivial) "Bonferroni method": We simply use the individual t confidence interval method, adjusting the significance level (or error rate of confidence intervals) to α/k.

E.g., suppose want to make 3 comparisons. Let

$A_1 = \{$CI #1 does not contain the true value$\}$

$A_2 = \{$CI #2 does not contain the true value$\}$

$A_3 = \{$CI #3 does not contain the true value$\}$

If $P(A_1) = P(A_2) = P(A_3) = \alpha/3$, then

$P(A_1 \cup A_2 \cup A_3) \leq \alpha/3 + \alpha/3 + \alpha/3 = \alpha.$

So need to make 98.33% confidence intervals.