CHAPTER 19
Two Factor Studies, Balanced case

We'll consider experimental studies first. both factors are experimental

There are pros and cons to including two factors in the same study, rather than taking the "one factor at a time approach" (OFAAT).

One pro is efficiency; you get two experiments in one.

A con is that it's a more complicated approach, therefore more prone to errors and to missing data. Missing data cause more problems with the analysis than in the one-factor design.

Example of OFAAT approach Gas mileage. Factors: Gas, Hi or Low octane; Tire pressure, Normal or Low

In the OFAAT approach, you need to run two separate experiments, such as the following:

pressure is not studied here

- Experiment 1. Normal tire pressure throughout. Randomly assign Hi, Low octane gas to experimental units (time periods of driving car). Suppose we find no difference between Hi and Low octane gas in effect on gas mileage from Exp. 1.

Gas type is not studied in Exp. 2

- Experiment 2. Fix Gas at low octane. Randomly assign Normal and Low tire pressure to units.

In summary, the OFAAT approach is to carry out a separate experiment for each factor, making decisions sequentially about which levels of the already-studied factors to use in successive experiments.

Two-way ANOVA: Meaning of Model Elements

Ex. Corn Yield

A field was divided into 20 equal-size plots, each with the same base fertility level, and same exposure to sun and rain; each was planted wth the same amount of corn seed in the same manner.

Fertilizer and manure were applied to each plot. At the end of the growing season, the corn was harvested and the yields of the twenty plots were measured.

Goal: Assess how corn yield depends on level of nitrogen-based fertilizer and level of manure used.

In this experiment, two levels of each factor were used (low, high).

$Y =$ yield of corn (metric tons)

Factor 1: Fertilizer (low is 45 kg per hectare, high is 135 kg per hectare)

Factor 2: Manure (low is 84 kg per hectare, high is 168 kg per hectare)

Example: Corn Yield $n = 5$ plots per “cell” (A “cell” is a particular combination of the levels of the two factors, e.g. “low, low” means “low fertilizer” and “low manure”). The summary statistics are given in the two tables below.

manure fertilizer

Cell Means

column factor

Common data layout

	Fertilizer	
Manure	Low	High
Low	11.3	13.9
High	14.0	15.1

row factor

Cell SDs

	Fertilizer	
Manure	Low	High
Low	1.9	1.7
High	1.8	1.3

General framework for balanced two-way ANOVA

Our approach to analysis will allow us to easily accommodate more than two levels per factor.

To start with, for discussion of the model parameters, suppose we know the true mean yields for every treatment (factor-level combination, or "cell" of the design).

Two-way ANOVA Model

Table of cell/treatment means, factor level means, and main effects:

	Factor B		Row	Main (Row)
Factor A	$j=1$	$j=2$	Average	Effects
$i=1$	μ_{11}	μ_{12}	$\mu_{1.}$	$\mu_{1.}-\mu_{..}$ = α_1
$i=2$	μ_{21}	μ_{22}	$\mu_{2.}$	$\mu_{2.}-\mu_{..}$ = α_2
Column average	$\mu_{.1}$	$\mu_{.2}$	$\mu_{..}$	
Main (col) effects	β_1 = $\mu_{.1}-\mu_{..}$	$\mu_{.2}-\mu_{..}$ = β_2		

The mean response for a given treatment in a two-factor study is denoted by μ_{ij}, where i is the level of factor A ($i=1,\ldots,a$) and j is the level of factor B ($j=1,\ldots,b$).

Row averages, column averages, and overall average:

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

The main row effects are denoted $\alpha_i,\ i=1,\ldots,a$

The main column effects are denoted $\beta_j,\ j=1,\ldots,b$

Comment: The reason for defining the main effects this way, is so the definition can be extended to three or more levels of the factor.
If there are only two levels of a factor (“Low” and “High”), a natural definition of effect is the difference between the averages for High vs. Low.

Overall average:

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

Example 1 Find the factor level means and main effects for the following set of **true** cell means:

μ_{ij}'s – hypothetical population parameters

	Fertilizer			
Manure	Low	High	Row Means	Main row effect
Low	11.0	14.0	12.5	12.5 − 13.5 = −1.0
High	14.0	15.0	14.5	14.5 − 13.5 = 1.0
Column means	12.5	14.5	Grand mean = 13.5	
Main col. effects	−1.0	1.0		

Example 1. Table of cell means

	Fertilizer	
Manure	Low	High
Low	11.0	14.0
High	14.0	15.0

Important concept: There is *interaction* in this model because $\mu_{21} - \mu_{11} = 14 - 11 = 3$, while $\mu_{22} - \mu_{12} = 15 - 14 = 1$.

That is, the effect of HIgh vs. Low Manure is not the same at both Low and High levels of Fertilizer. We say that the two factors, Manure and Fertilizer, interact.

Exercise Are the effects of HIgh vs. Low Fertilizer the same at Low and High Manure? Use the model notation to explain your answer.

Example 2 Suppose the true cell means were the following:

	Fertilizer		Row	Main row
Manure	Low	High	Average	effect
Low	12.0	14.0	13.0	-2.0
High	16.0	18.0	17.0	2.0
Column average	14.0	16.0	15.0	
Main col. effect	-1.0	1.0		

An additive model holds.

In this special case, the factor effects are **additive.** This means, for instance, that each cell mean can be found by adding the respective row and column effects to the overall mean $\mu_{..}$.

$$\mu_{11} = \mu_{..} + \alpha_1 + \beta_1 = 15 + (-2) + (-1)$$
$$\mu_{12} = \mu_{..} + \alpha_1 + \beta_2 = 15 + (-2) + 1$$
$$\vdots$$

This also means that $\mu_{ij} = \mu_{i.} + \mu_{.j} - \mu_{..}$.

This is an example of an **additive** model. We also say there is **no interaction** between the two factors.