Multi-factor Factorial Experiments

- In the one-way ANOVA, we had a single factor having several different levels.

- Many experiments have multiple factors that may affect the response.

Example: Studying weight gain in puppies

Response (Y) = weight gain in pounds

Factors: Type of Diet (A, B, C)
Exercise Program (None, Medium, Intense)
Amount of Food (oz.) (4, 8, 12, 16)

- Here, 3 factors, each with several levels.

- Levels could be quantitative or qualitative.

- A factorial experiment measures a response for each combination of levels of several factors.

- Example above is a: $3 \times 3 \times 4$ factorial experiment (based on # of levels for each factor)

- We will study the effect on the response of the factors, taken individually and taken together.

Two Types of Effects

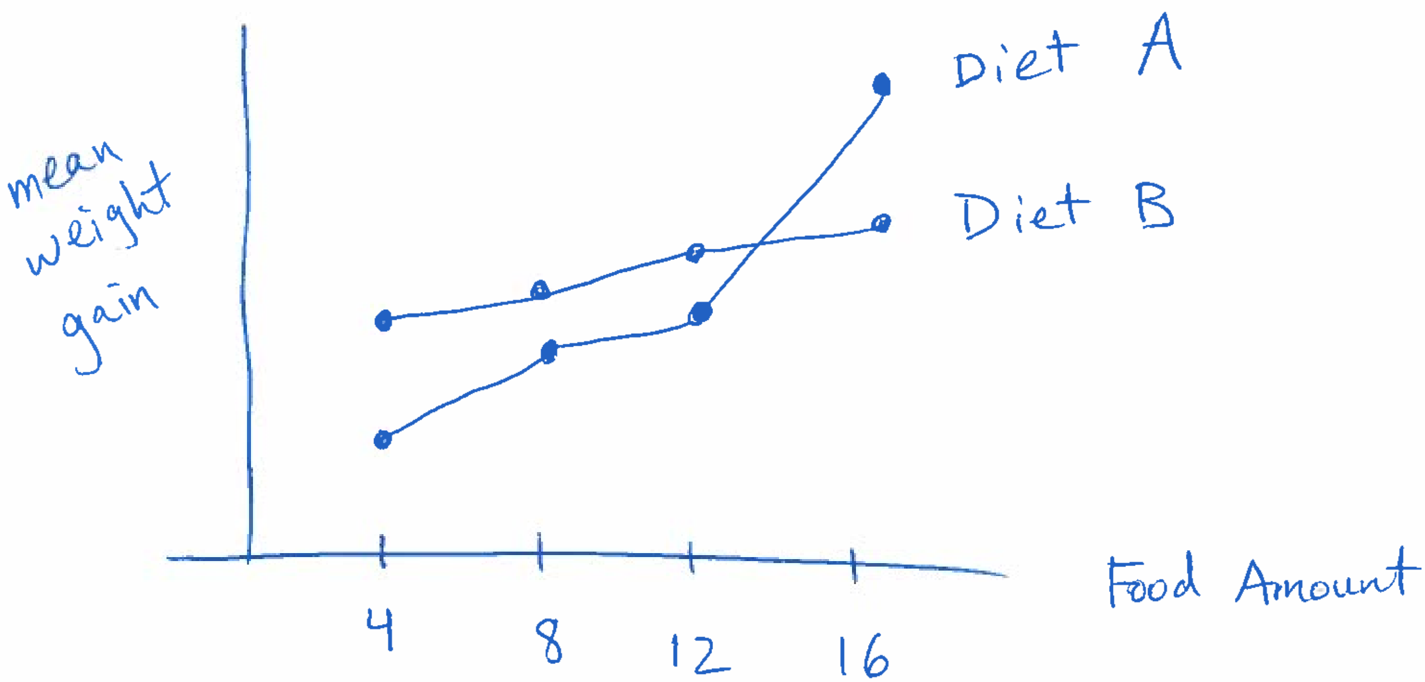
• The main effects of a factor measure the change in mean response across the levels of that factor (taken individually).

• Interaction effects measure how the effect of one factor varies for different levels of another factor.

Example: We may study the main effects of food amount on weight gain.

• But perhaps the effect of food amount is different for each type of diet: Interaction between amount and diet!

Picture: Example:



Two-Factor Factorial Experiments

- Model is more complicated than one-way ANOVA model.

- Assume we have two factors, A and C, with *a* and *c* levels, respectively: ($a \times c$ factorial experiment)

- Assume we have *n* observations at each combination of factor levels.

- Total of acn observations.

Model: $Y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \varepsilon_{ijk}$

$i = 1, \ldots, a \qquad j = 1, \ldots, c \qquad k = 1, \ldots, n$

- Y_{ijk} = *k*-th observed response at level *i* of factor A and level *j* of factor C.
- μ = an overall mean response
- α_i's (main effects of factor A) = difference between mean response for *i*-th level of A and the overall mean response
- γ_j's (main effects of factor C) = difference between mean response for *j*-th level of C and the overall mean response
- $(\alpha\gamma)_{ij}$'s (interaction effects between factors A and C)
- ε_{ijk} = random error component → accounts for the variation among responses <u>at the same combination</u> of factor levels

• **Again, we assume the random error is approximately normal, with mean 0 and variance σ^2.**

• **We also restrict** $\sum_i \alpha_i = \sum_j \gamma_j = \sum_i (\alpha\gamma)_{ij} = \sum_j (\alpha\gamma)_{ij} = 0.$

<u>Example: (Meaning of main effects)</u>

• **Suppose $\alpha_1 = 3.5$ and $\alpha_2 = 2$. What does this mean?**

<u>Case I: (No interaction between A and C)</u>

$\rightarrow (\alpha\gamma)_{ij} = 0$ for all i, j

• **Mean response at level 1 of factor A is:**

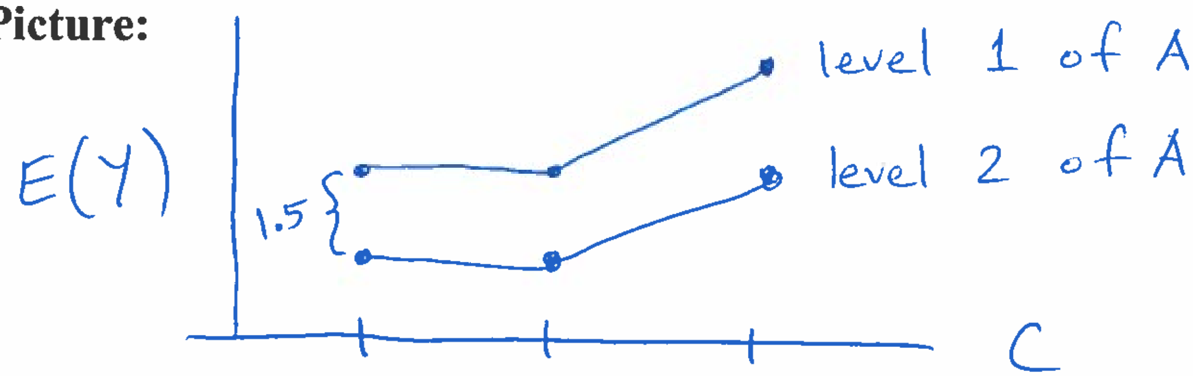
$$E(Y_{1jk}) = \mu + \alpha_1 + \gamma_j$$
$$= \mu + 3.5 + \gamma_j$$

• **Mean response at level 2 of factor A is:**

$$E(Y_{2jk}) = \mu + \alpha_2 + \gamma_j$$
$$= \mu + 2 + \gamma_j$$

• **For <u>any</u> fixed level of C, mean response at level 1 of A** is 1.5 more than mean response at level 2 of A, since $E(Y_{1jk}) - E(Y_{2jk}) = 1.5$

Picture:



Case II: (Interaction between A and C)

- **Mean response at level 1 of factor A is:**

$$E(Y_{1jk}) = \mu + \alpha_1 + \gamma_j + (\alpha\gamma)_{1j}$$
$$= \mu + 3.5 + \gamma_j + (\alpha\gamma)_{1j}$$

- **Mean response at level 2 of factor A is:**

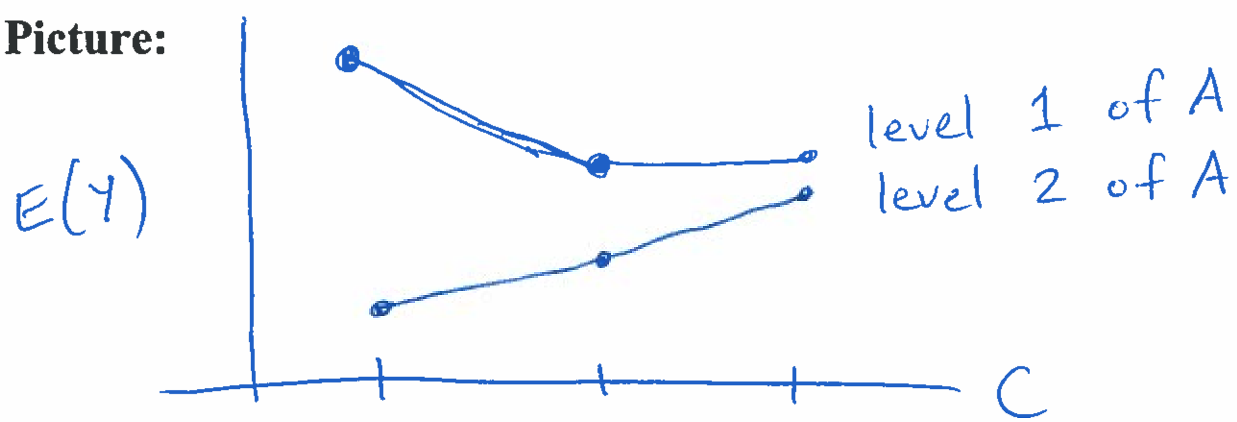
$$E(Y_{2jk}) = \mu + \alpha_2 + \gamma_j + (\alpha\gamma)_{2j}$$
$$= \mu + 2 + \gamma_j + (\alpha\gamma)_{2j}$$

- **Here, the difference in mean responses for levels 1 and 2 of factor A is:**

$$E(Y_{1jk}) - E(Y_{2jk}) = 3.5 - 2 + (\alpha\gamma)_{1j} - (\alpha\gamma)_{2j}$$

- **This difference depends on the level of C!**

Picture:



- **We see that the main effects are not directly interpretable in the presence of interaction.**

• **In a two-factor study, first we will test for interaction:**

$H_0: (\alpha\gamma)_{ij} = 0$ for all i, j

$H_a: (\alpha\gamma)_{ij} \neq 0$ for some i, j

• **If there is no significant interaction, we will test for main effects of each factor:**

$H_0: \alpha_i = 0$ for all i	$H_0: \gamma_j = 0$ for all j
$H_a: \alpha_i \neq 0$ for some i	$H_a: \gamma_j \neq 0$ for some j

Notation for Sample Means:

$\bar{Y}_{ij\bullet}$ **= sample mean of observations for level *i* of A and level *j* of C [This is the (*i*, *j*) cell sample mean]**

$\bar{Y}_{i\bullet\bullet}$ **= sample mean of observations for level *i* of A**

$\bar{Y}_{\bullet j\bullet}$ **= sample mean of observations for level *j* of C**

$\bar{Y}_{\bullet\bullet\bullet}$ **= sample mean of all observations in the study [This is the <u>overall</u> sample mean]**

ANOVA Table for Two-Factor Experiment

- **Partitioning the Variation in Y:**

TSS = $\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$ $\quad df = acn - 1$

→ measures total variation in Y-values

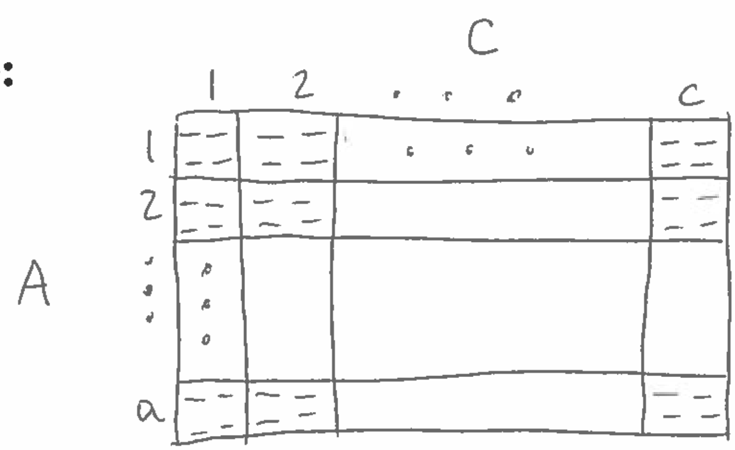
SS(Cells) = $n \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$ $\quad df = ac - 1$

→ measures variation across cell means

SSW = $\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2$ $\quad df = ac(n-1)$

→ measures variation <u>within</u> cells.

Picture:



MS(Cells) = $\dfrac{SS(Cells)}{ac - 1}$ $\qquad$ **MSW =** $\dfrac{SSW}{ac(n-1)}$

• **If MS(Cells) > MSW, the mean response is different across the cells → the ANOVA model is not useless.**

<u>Overall F-test</u>: If F* = MS(Cells) / MSW is greater than $F_\alpha[ac-1,\ ac(n-1)]$, then we conclude there is a difference among the population cell means.

<u>Example (Table 9.5 data)</u>:

Y = Gas mileage (in mpg)

<u>Factors:</u> Engine Type (A): 4-cylinder, 6-cylinder

Motor Oil (C): Standard, Multi, Gasmiser

2×3 factorial, $a=2$, $c=3$, $n=5$

• **Software will calculate:**

TSS = 92.547
SS(Cells) = 66.523
SSW = 26.024

$\Rightarrow$

$$MS(\text{Cells}) = \frac{66.523}{5} = 13.30$$

$$MSW = \frac{26.024}{24} = 1.084$$

$$F^* = \frac{13.30}{1.084} = 12.27 \quad (\text{P-value near } 0)$$

Using α = 0.05: $F_{.05}(5, 24) = 2.62$ (Table A.4.A)

Conclusion: $12.27 > 2.62$, so we reject H_0 and conclude the population cell means are <u>not</u> equal (ANOVA model is not useless)

• **If we reject H_0: "all cell means are equal" with the overall F-test, then we test for (1) interaction and possibly (2) main effects.**

• **Further Partitioning of SS(Cells):**

SSA = $cn\sum_i (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2$ **d.f. =** $a - 1$

→ measures variation due to factor A

SSC = $an\sum_j (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2$ **d.f. =** $c - 1$

→ measures variation due to factor C

SSAC = SS(Cells) – SSA – SSC **d.f. =** $(a-1)(c-1)$

→ measures variation due to interaction of A and C.

Mean Squares:

MSA = $\frac{SSA}{a-1}$ **MSC =** $\frac{SSC}{c-1}$ **MSAC =** $\frac{SSAC}{(a-1)(c-1)}$

ANOVA table

Source	d.f.	SS	MS	F*
Between Cells	$ac-1$	SS(Cells)	MS(Cells)	MS(cells)/MSW
A	$a-1$	SSA	MSA	MSA/MSW
C	$c-1$	SSC	MSC	MSC/MSW
A × C	$(a-1)(c-1)$	SSAC	MSAC	MSAC/MSW
Within Cells (Error)	$ac(n-1)$	SSW	MSW	
Total	$acn-1$	TSS		

• **We will usually calculate the ANOVA table quantities using software.**

Useful F-tests in Two-Factor ANOVA

Testing for Significant Interaction: We reject

$$H_0: (\alpha\gamma)_{ij} = 0 \text{ for all } i, j$$

if: $F^* = \frac{MSAC}{MSW} > F_\alpha\left[(a-1)(c-1), ac(n-1)\right]$

Example: $SSAC = 20.328$, $MSAC = \frac{20.328}{2} = 10.164$

$F^* = \frac{10.164}{1.084} = 9.37$ and $F_{.05}(2,24) = 3.40$

$9.37 > 3.40$, so we reject H_0 and conclude there is significant interaction between engine type and motor oil. (P-value $\approx .001$)

Note: If (and only if) the interaction is NOT significant, we test for significant main effects of factor A and of factor C:

• **For factor A: We reject H_0: $\alpha_i = 0$ for all i**
if:

$$F^* = \frac{MSA}{MSW} > F_\alpha\left[a-1, ac(n-1)\right]$$

• **For factor C: We reject H_0: $\gamma_j = 0$ for all j**
if:

$$F^* = \frac{MSC}{MSW} > F_\alpha\left[c-1, ac(n-1)\right]$$