Lecture 5: Two-Factor Studies with One Case per Treatment

STA 106: Analysis of Variance

Xiner Zhou
Department of Statistics, University of California, Davis

Example

(The Insurance Study)

<u>The objective of the study:</u>

City size (factor A)

Geographic region (factor B)

Amount of premium for automobile insurance charged by an insurance company (for a specific type of coverage)

<u>The study setup:</u>

6 cities were selected to represent different regions of the state and different sizes of cities

(a) Premiums for Automobile Insurance Policy (in dollars)

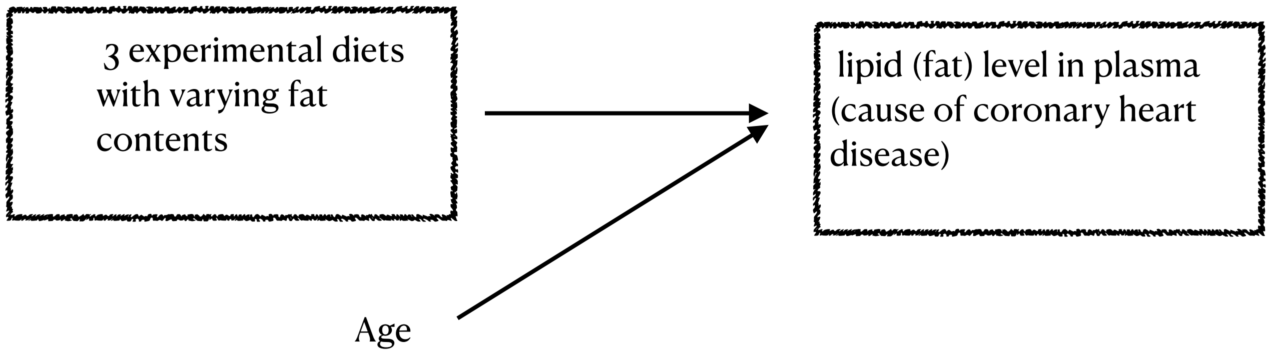
	Region (factor B)		
Size of City (factor A)	East ($j = 1$)	West ($j = 2$)	Average
Small ($i = 1$)	140	100	120
Medium ($i = 2$)	210	180	195
Large ($i = 3$)	220	200	210
Average	190	160	175

☑ Two-factor observational study

Example

(The Fat-in-Diet Study)

<u>The objective of the study:</u>



<u>The study setup:</u>

Within each block, 3 experimental diets were randomly assigned to the 3 subjects

Reduction in lipid level after some a certain period of time were recorded as the outcome

Block i		Fat Content of Diet $j = 1$ Extremely Low	 $j = 2$ Fairly Low	 $j = 3$ Moderately Low
1	Ages 15–24	.73	.67	.15
2	Ages 25–34	.86	.75	.21
3	Ages 35–44	.94	.81	.26
4	Ages 45–54	1.40	1.32	.75
5	Ages 55–64	1.62	1.41	.78

☑ Randomized Complete Block Design

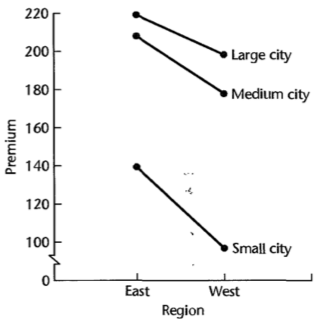
Two-Factor Studies with One Case per Treatment

Why and When would such cases occur?

- Constraints on cost, time, materials. …… severely limit the number of observations can be obtained
  - Only 1 subject (such as patient with certain characteristic) is available
- Outcome of interest is a single aggregated measure, there is no way to get more than 1 replicates for each treatment
  - Only aggregated measure at some large geographic regions, such as state, city…
  - Only aggregated measure at institutional level, such as hospitals, schools …
- Two very important kinds of studies
  - Randomized Complete Block Design
  - Observational studies with matched pairs

Two-Factor Studies with One Case per Treatment

What can go wrong with the two-factor factor effects model with interaction?



$$Y_{ij1} = \mu_{..} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij1}$$

$\rightarrow$ $\hat{Y}_{ij1} = \bar{Y}_{ij\cdot} = Y_{ij1}$

$\rightarrow$ $e_{ij1} = 0$

$\rightarrow$ $MSE = 0$

With only one case per treatment, there is no way to estimate variability within treatments

there is no way to estimate error variance σ^2 by MSE, which is one parameter in the ANOVA model and the key to any inference .

Issue: over-parameterization

Two-Factor Studies with One Case per Treatment

- Two-Way ANOVA Model without Interaction
- Analysis of Variance
- Test for Factor A and B Main Effects
- Multiple Comparison Procedures
- Tukey's test
- Randomized Complete Block Designs

Two-Way ANOVA Model without Interaction

Assume factor A and B do not interact, i.e. all the interaction $\gamma_{ij} = 0$

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + \varepsilon_{ij}$$ (Subscript k=1 dropped)

- $\mu_{..} = \dfrac{\sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij}}{ab}$: overall mean

- $\alpha_i = \mu_{i.} - \mu_{..}$ main effect of factor A at ith level

  Subject to constraint $\sum \alpha_i = 0$

- $\beta_j = \mu_{.j} - \mu_{..}$ main effect of factor B at jth level

  Subject to constraint $\sum \beta_j = 0$

- $\gamma_{ij} = \mu_{ij} - \alpha_i - \beta_j = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}$ interaction effect of factor A at ith level with factor B at jth level

  Subject to a+b-1 constraints

  $$\sum_i \gamma_{ij} = 0 \;\; j = 1, \ldots, b$$
  $$\sum_j \gamma_{ij} = 0 \;\; i = 1, \ldots, a$$

- ε_{ij} are independent $N\left(0, \sigma^2\right)$ for $i = 1, \ldots, a;\ j = 1, \ldots, b$

Fitting the Two-Way ANOVA Model without interaction

Least squares estimates for parameters in factor effects parameterization:

$$\hat{\mu}_{\cdot\cdot} = \frac{\sum_i \sum_j \hat{\mu}_{ij}}{ab} = \frac{\sum_i \sum_j Y_{ij}}{ab} = \bar{Y}_{\cdot\cdot}$$

$$\hat{\alpha}_i = \hat{\mu}_{i\cdot} - \hat{\mu}_{\cdot\cdot} = \frac{\sum_j Y_{ij}}{b} - \bar{Y}_{\cdot\cdot} = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$$

$$\hat{\beta}_j = \hat{\mu}_{\cdot j} - \hat{\mu}_{\cdot\cdot} = \frac{\sum_i Y_{ij}}{a} - \bar{Y}_{\cdot\cdot} = \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}$$

Where:

$$\bar{Y}_{\cdot\cdot} = \frac{\sum_i \sum_j Y_{ij}}{ab}$$

$$\bar{Y}_{i\cdot} = \frac{\sum_j Y_{ij}}{b}$$

$$\bar{Y}_{\cdot j} = \frac{\sum_i Y_{ij}}{a}$$

- fitted value for an observation Y_{ij}

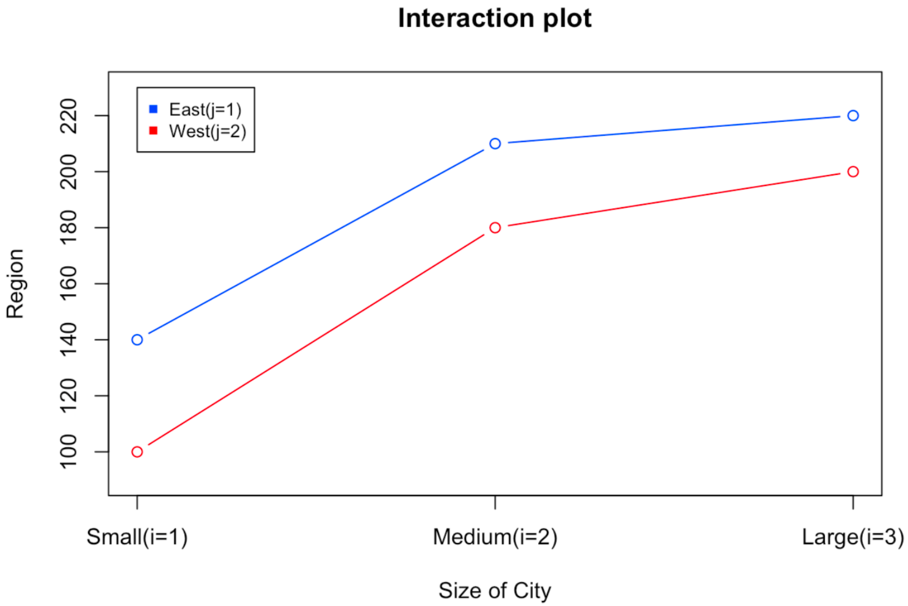
  ANOVA model's "best guess" or "best prediction"

$$\hat{Y}_{ij} = \hat{\mu}_{\cdot\cdot} + \hat{\alpha}_i + \hat{\beta}_j = \overline{Y_{\cdot\cdot}} + \overline{Y_{i\cdot}} - \overline{Y_{\cdot\cdot}} + \overline{Y_{\cdot j}} - \overline{Y_{\cdot\cdot}} = \overline{Y_{i\cdot}} + \overline{Y_{\cdot j}} - \overline{Y_{\cdot\cdot}}$$

Example

Size of City (factor A)	Region (factor B) East ($j = 1$)	West ($j = 2$)	Average
Small ($i = 1$)	140	100	120
Medium ($i = 2$)	210	180	195
Large ($i = 3$)	220	200	210
Average	190	160	175

Plot the data. Does it appear that interaction effects are present? Does it appear that factor A and factor B main effects are present? Discuss.



It appears that there could be a slight interaction between region and size of city in their effects on the premium. However, the lack of parallelism in the response lines could simply be the result of randomness.

It appears that factor A size of city do have effects, as we can see the premium increases as size increases.

It appears that factor B region also have effects, with east has higher premium than the west.