Parametric and Nonparametric Factorial Designs

Benjamin Chiao

Apr 20, 2009 (updated on Apr 20, 2009) Lecture Notes for Graduate Level Empirical Microeconomics, Peking University Guanghua School of Management

Agenda

- Two-factor Designs (Parametric)
- Two-factor Designs (Non-Parametric)

Two-Factor Designs

- The section of parametric two-factor designs largely follows D. Montgomery (2003). *Design and Analysis of Experiments*. John Wiley & Sons.
- The section of non-parametric counterpart largely follows Akritas, M.G., Arnold S.F. and Brunner E. (1997). "Nonparametric Hypothesis and Rank Statistics for Unbalanced Factorial Designs," *Journal of the American Statistical Association*, 92, 258-265.

Two-Factor Designs (Parametric)

- Dependent variable is continuous
  - "The only well-accepted multivariable procedures for ordinal dependent variables are the ones that can be used as non-parametric equivalents to certain ANOVA designs." – Richard Riegelman (2004). *Studying a study and testing a test*. Lippincott Williams & Wilkins. p. 357.
- Independent variables are called the main factors and interaction

Two-Factor Designs (Parametric)

- Two factors/treatments $(A \ \& \ B)$,assumed to be fixed (not random) effects
- A has possible levels (values) indexed by i from 1 to a ; B has possible levels (values) indexed by j from 1 to b;
- Each pair of (i,j) is called a cell
- Each cell contains many observations indexed by k
- $(\alpha\beta)_{ij}$ is called an interaction of the cell (i,j)
- The following is called an effects model:
  - The kth observation for the cell (i,j) :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{jik}; i = 1, ..., a; j = 1, ..., b \quad (1)$$

  - $\epsilon_{jik} \sim niid(0, \sigma^2)$ and unrelated to all other variables

Example:

- The kth observation for the cell (i, j) :

$$\text{Fight}_{ijk} = \mu + \text{Husband's Temperament}_i + \text{Wife's Temperament}_j + \text{Husband's Temperament}_i \times \text{Wife's Temperament}_j + \epsilon_{jik}$$

  - where
    - $i = 1, ..., a; j = 1, ..., b;$
    - Temperament runs from "Like Sheep" to "Like Tigers"
    - We unrealistically assume one to one relationships so no single individuals appear in more than one cell :)
- Is it more likely for a fight to occur when two tiger types are matched, or a tiger type is matched with a sheep type?

Beyond Two-Factor Designs

- If there are more than $k > 2$ factors, we call it a 2^k design.
- The number of cells quickly goes out of bound as k increases, so in practice we could run fractional factorial designs which usually cut the number of treatments run to half
- These are not the subject for today as the vast majority of factorial designs involve only two factors

- The kth observation for the cell (i, j) :

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{jik}; i = 1, ..., a; j = 1, ..., b \quad (2)$$

- Exercise: Show that there are $1 + a + ab$ parameters, $1 + a + b + ab$ normal equations, $1 + a + b$ linearly dependencies, and the following $1 + a + b$ restrictions allows exact identification

$$\sum_{i=1}^{a} \alpha_i = 0; \sum_{j=1}^{b} \beta_j = 0, \sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0 \quad (3)$$

- That is we impose that the effects are deviations from means:
- See Montgomery (2003) pp 185-189 for answers

Agenda

- Two-factor Designs (Parametric)
  - Model
  - $\rightarrow$Hypothesis Testing
  - Model Adequacy Checking
  - Fixing Inadequacy
  - Parameter Estimations
- Two-factor Designs (Non-Parametric)

Hypotheses

$$H_0(\alpha) : \alpha_1 = ... = \alpha_a = 0 \quad (4)$$

$$H_a(\alpha) : \text{ at least one } \alpha_i \neq 0 \quad (5)$$

$$H_0(\beta) : \beta_1 = ... = \beta_b = 0 \quad (6)$$

$$H_a(\beta) : \text{ at least one } \beta_j \neq 0 \quad (7)$$

$$H_0(\alpha\beta) : (\alpha\beta)_{11} = ... = (\alpha\beta)_{ab} = 0 \quad (8)$$

$$H_a(\alpha\beta) : \text{ at least one } (\alpha\beta)_{ij} \neq 0 \quad (9)$$

Notations

- $y_{i..}$ denotes the total of all observations under the i^{th} level of factor A
- $y_{.j.}$ denotes the total of all observations under the j^{th} level of factor B
- $y_{ij.}$ denotes the total of all observations under the ij^{th} cell
- $y_{...}$ denotes the total of all observations
- use a bar to denote mean values
- Total corrected sum of squares (SS_T) is

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 \tag{10}$$

- By straighforward algebraic expansion, SS_T can be written as

$$SS_T = bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2$$
$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$
$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2 \qquad (11)$$

or

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \qquad (12)$$

Degrees of Freedom

Effects	Degrees of Freedom
A	a-1
B	b-1
AB	(a-1)(b-1)
Error	ab(n-1)
Total	abn-1

- A and B have a and b levels respectively, so df are a-1 and b-1
- The interaction df is the df of cells (which is ab-1) minus the df's of a and b. So ab-1-(a-1)-(b-1)=(a-1)(b-1)
- In each ab cell, there are n-1 df. So there are ab(n-1) df for errors

Expected Mean Square

- Mean Square is defined by dividing the sum of squares by its degree of freedom. Taking expected values, we have:

$$E(MS_A) = E(\frac{SS_A}{a-1}) \tag{13}$$

$$E(MS_B) = E(\frac{SS_B}{b-1}) \tag{14}$$

$$E(MS_{AB}) = E(\frac{SS_{AB}}{(a-1)(b-1)}) \tag{15}$$

$$E(MS_E) = E(\frac{SS_E}{ab(n-1)}) \tag{16}$$

- Recall that $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}[\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij\cdot})^2]$.
- Dividing the squared bracket terms by n-1 gives

$$\sum_{i=1}^{a}\sum_{j=1}^{b}[\sum_{k=1}^{n}\frac{(y_{ijk} - \bar{y}_{ij\cdot})^2}{n-1}] = \sum_{i=1}^{a}\sum_{j=1}^{b} s_{ij}^2 \tag{17}$$

  - where s_{ij}^2 is the sample variance for cell (i, j).
- Since $\epsilon_{jik} \sim niid(0, \sigma^2)$, $Es_{ij}^2 = \sigma^2$.
- So $E(MS_E) = E(\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}[\sum_{k=1}^{n}\frac{(y_{ijk}-\bar{y}_{ij\cdot})^2}{n-1}]}{ab}) = E(\frac{\sum_{i=1}^{a}\sum_{j=1}^{b} s_{ij}^2}{ab}) = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} Es_{ij}^2}{ab} = \sigma^2$
- Similarly, we can find $E(MS_A), E(MS_B), E(MS_{AB})$ on the next page

- $$E(MS_A) = \sigma^2 + \frac{bn\sum_{i=1}^{a}\alpha_i}{a-1} \tag{18}$$
$$E(MS_B) = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j}{b-1} \tag{19}$$
$$E(MS_{AB}) = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}}{(a-1)(b-1)} \tag{20}$$
$$E(MS_E) = \sigma^2 \tag{21}$$

- So if $H_0(\alpha)$, $H_0(\beta)$,and $H_0(\alpha\beta)$ are all true. $MS_A, MS_B, MS_{AB}, MS_E$ are all unbiased estimates of σ^2
- But if only $H_0(\alpha)$ is not true, then $E(MS_A) > E(MS_E)$.

Hypothesis Testing Using Mean Squares

- So $\frac{MS_A}{MS_E}$ measures if $H_0(\alpha)$ is true. Turned out $\frac{MS_A}{MS_E} \sim F$ with $a-1$ df.
- Similarly, $\frac{MS_B}{MS_E}$ and $\frac{MS_{AB}}{MS_E}$ also $\sim F$ with $b-1$ and $(a-1)(b-1)$ df's.

Hypothesis Testing of Pairwise Means Difference

- Of N cells, there are $\binom{N}{2}$ possible combinations, and thus pairwise testings possible.
- Consider a 3 by 3 example. There are three levels for factor A and three levels for factor B.
- There are $\binom{9}{2} = 36$ total pairwise testings possible.
- Consider that if we are only interested to find the means difference when a level of B is fixed.
- Then we can use the Tukey test.

Tukey Test: Hypothesis Testing of Pairwise Means Difference

- The Tukey Test Statistics is the absolute difference between the two means, denoted by T.
- In a two tailed test, we reject the null of no difference if $T > T_\alpha = q_\alpha(p, df_{MS_E})\sqrt{\frac{MS_E}{n}}$
  - where
    - n is the no. of observations in a cell which is the same across cells by assumption
    - p is number of means, here it is three (the three levels of factor A)
    - df_{MS_E} is the degree of freedom of MS_E
    - $q_\alpha(p, df_{MS_E})$ can then be looked up in a statistical table or software

Tukey Test: Hypothesis Testing of Pairwise Means Difference

- Tukey (1953) shows that the overall significance level is exactly α when the sample sizes are equal (and at most α when the sample sizes are not equal but we will not consider this case below). This α is not the fixed effect, its the significance level. Now you see sometimes we use τ_i instead of α_i for the fixed effects.

Choice of Sample Size

- To find the appropriate sample size for a two-factor design, one can use the operating characteristic curves in Appendix Chart V in Montgomery. See also Montgomery Ch 5, pp 189.

Agenda

- Two-factor Designs (Parametric)
  - Model
  - Hypothesis Testing
  - $\longrightarrow$Model Adequacy Checking
  - Fixing Inadequacy
  - Parameter Estimations
- Two-factor Designs (Non-Parametric)

Model Adequacy Checking

- To $\epsilon_{jik} \sim niid(0, \sigma^2)$ is reasonable, we can
  - do some visual inspection of plots of residuals
  - do more formal tests

First Plots: Does it shape like normal distribution?

- We first check if $\epsilon_{jik} \sim niid(0, \sigma^2)$ is reasonable by doing a residual analysis
- The residuals for a two-factor factorial model are

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} \tag{22}$$

  - where $\hat{y}_{ijk}$ is the predicted value of y_{ijk}
  - but $\hat{y}_{ijk} = \bar{y}_{ij.}$ because $\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + (\widehat{\alpha\beta})_{ij} = \bar{y}_{ijk}$ (check it yourself) so

$$e_{ijk} = y_{ijk} - \bar{y}_{ijk} \tag{23}$$

- If $niid(0, \sigma^2)$ is satisfied, a plot of histogram of the e_{ijk} would give a normal distribution shape with mean zero.

Second Plots: Are There Outliers?

- You can also do a normal probability plot, which will be a straight line if $niid(0, \sigma^2)$ is satisfied. Outliers can be detected using this plot.
- We can do some a rough statistical check. A standardized residual is defined as

$$d_{ij} = \frac{e_{ijk}}{\sqrt{MS_E}} \tag{24}$$

  - A rough check is to say that 68%, 95%, 99% of d_{ij} should be $\pm 1, \pm 2, \pm 3$.
  - At times, you will find that even the largest d_{ij} is 1.83 so there is no obvious indicators of outliers.
  - See details in Barnett and Lewis (1994), John and Prescott (1975) and Stefansky (1972)

Third Plots: Residuals in Time Sequence

-
  - There should be no correlation between residuals
  - The variance should be constant over time

Forth Plots: Residuals v. Fitted Values

-
    - The residuals should be unrelated to any other variables including predicted response
    - Some measuring instruments cause nonconstant variance because errors could be percentage of the size of observation

Formal Tests of Nonconstant Variance

- When normality and independence are satisfied, use the Bartlett's test

$$H_0(\sigma) : \sigma_1 = ... = \sigma_a \quad (25)$$

$$H_a(\sigma) : \text{ at least one } \sigma_i \text{ differs} \quad (26)$$

  - The test statistic is $\chi_0^2 = 2.3026\frac{q}{c}$ with $a - 1$ df where (for a one factor design)
    - $q = (N - a) \log S_p^2 - \sum_{i=1}^{a} (n_i - 1) \log S_i^2$
    - $c = 1 + \frac{1}{3(a-1)} (\sum_{i=1}^{a} (n_i - 1)^{-1} - (N - a)^{-1})$
    - $S_p^2 = \frac{\sum_{i=1}^{a} (n_i - 1) S_i^2}{N - a}$
- When normality is not satisfied, use the modified Levene test (Levene [1960] and Conover et al [1981])

Agenda

- Two-factor Designs (Parametric)
    - Model
    - Hypothesis Testing
    - Model Adequacy Checking
    - $\longrightarrow$Fixing Inadequacy
    - Parameter Estimations
- Two-factor Designs (Non-Parametric)

Fixing Inadequacy

- If homogenety of variance is violated,
  - F test is only slightly affected in the balanced (equal no. of observations in each cell) fixed effect models.
  - If the cells having the larger variances also have smaller (larger) sample sizes, the actual type 1 error is larger (smaller) than anticipated.
  - So use equal sample size whenever possible
  - For random effects models, this violation is severe even for balanced design
  - See details on p.80 in Montgomery

- The usual approach is to apply a variance-stabilizing transformation and the analysis on the transformed data
- If the observations follow:
  - Poisson distribution, then use $y^*_{ijk} = \sqrt{y_{ijk}}$ or $\sqrt{1 + y_{ijk}}$
  - Lognormal distribution, then use $y^*_{ijk} = \log y_{ijk}$
  - Binomial distribution, then use $y^*_{ijk} = \arcsin \sqrt{y_{ijk}}$
- There is a large literature on this, see Ch 3(p.80) and Ch 14 in Montgomery

Agenda

- Two-factor Designs (Parametric)
  - Model
  - Hypothesis Testing
  - Model Adequacy Checking
  - Fixing Inadequacy
  - $\longrightarrow$Parameter Estimations
- Two-factor Designs (Non-Parametric)

Parameter Estimation

- See Montgomery (2003) pp. 185-189 for details

Agenda

- Two-factor Designs (Parametric)
  - Model
  - Hypothesis Testing
  - Model Adequacy Checking
  - Fixing Inadequacy
  - Parameter Estimations
- $\rightarrow$Two-factor Designs (Non-Parametric)

Non-parametric Equivalence?

- The extension of parametric approach to testing higher order interactions is well established.
- However, the nonparametric alternatives are uncommon for testing interactions, especially for three factors or above
- More commonly, however, we deal with only two factors. This paper studies this:
  - Akritas, M.G., Arnold S.F. and Brunner E. (1997). "Nonparametric Hypothesis and Rank Statistics for Unbalanced Factorial Designs," *Journal of the American Statistical Association*, 92, 258-265.

- For a three factor model and the statistical procedure used in SAS See
  - G. Bakeerathan and S. Samita. A Non-parametric Approach in Testing Higher Order Interactions.

Agenda

- Two-factor Designs (Parametric)
- Two-factor Designs (Non-Parametric)
  - $\longrightarrow$Model
  - Hypothesis Testing
  - Other Considerations

Two-Factor Designs (Non-parametric) with Independent Observations

- For ordinal dependent variables
- A denotes the row-factor with a levels
- B denotes the column factor with b levels
- Y_{ijk} is the k^{th} observation for the cell (i, j)
- $N = \sum_i \sum_j n_{ij}$
- Y_{ijk} is assumed to be independent with
  - $P(Y_{ijk} \leq x) = F_{ij}^{+}(x), i = 1, ..., a; j = 1, ..., b$
  - We do not require a continuous distribution, we just required that it is right continuous (that's why we add the $+$ sign to the F_{ij}).
  - We will ignore this and simply write $Y_{ijk} \sim F_{ij}$
- It also does not require homoskedasticity

- Akritas et al. (1997) show that F_{ij} can be decomposed into:

$$F_{ij}(y) = M(y) + A_i(y) + B_j(y) + C_{ij}(y) \tag{27}$$

- $A_i(y)$ is the main factor A effects
- $B_j(y)$ is the main factor B effects
- $C_{ij}(y)$ is the interaction effects
- $\sum_{i=1} A_i = 0; \sum_{j=1} B_j = 0, \sum_i C_{ij} = 0, \sum_j C_{ij} = 0$.Thus, denoting an index by a dot (.) when averaging over all values of the index, we have:
    - $M = F_{..}$,
    - $A_i = F_{i.} - M$
    - $B_j = F_{.j} - M$
    - $C_{ij} = F_{ij} - F_{i.} - F_{.j} + M$

Agenda

- Two-factor Designs (Parametric)
- Two-factor Designs (Non-Parametric)
  - Model
  - $\longrightarrow$Hypothesis Testing
  - Other Considerations

Hypotheses

- No main factor A effects: $H_0(A) : A_i = 0 \forall i = 1, ..., a$
- No main factor B effects: $H_0(B) : B_j = 0 \forall j = 1, ..., b$
- No interaction: $H_0(AB) : C_{ij} = 0, \forall i = 1, ..., a, j = 1, ..., b$
- No simple factor A effect (i.e. A has no effect on the response either through the main effect or the interaction) $H_0(A|B) : A_i + C_{ij} = 0, \forall i = 1, ..., a, j = 1, ..., b$
- No simple factor B effect: $H_0(B|A) : B_j + C_{ij} = 0, \forall i = 1, ..., a, j = 1, ..., b$

- The hypotheses can equivalently be written as:
  - $H_0(A) : C_A F = 0$
  - $H_0(B) : C_B F = 0$
  - $H_0(AB) : C_{AB} F = 0$
  - $H_0(A|B) : C_{A|B} F = 0$
  - $H_0(B|A) : C_{B|A} F = 0$

- where
  - 1_d is the $d \times 1$ column vector of one's
  - $J_d = 1_d 1_d'$
  - I_d is the d-dimentional identity matrix
  - $M_d = (1_{d-1} | I_{d-1})$
  - $F = (F_{11}, ... F_{1b}, ..., F_{a1}, ..., F_{ab})'$
  - $C_A = M_a \otimes b^{-1} 1_b'$; $C_B = a^{-1} 1_a' \otimes M_b$
  - $C_{AB} = M_a \otimes M_b$
  - $C_{A|B} = M_a \otimes I_b$; $C_{B|A} = I_a \otimes M_b$

- Writing in this form is useful for conducting tests
- Under the null hypothesis of the form $CF = 0$, the test statistic is

$$Q(C) = N\hat{T}_C'(C\hat{V}C')^{-1}\hat{T}_C' \qquad (28)$$

  - which is χ^2 distributed for N large, with degree of freedom equals to the number of rows of C.
  - where
    - $\hat{T}_C$ is an estimate of $T_C = C\int\sum\sum\lim_{N\to\infty}(n_{ij}/N)F_{ij}dF$
    - $\hat{V}$ is a consistent estimate of the variance matrix given in Akritas et al (2007).

Agenda

- Two-factor Designs (Parametric)
- Two-factor Designs (Non-Parametric)
  - Model
  - Hypothesis Testing
  - $\longrightarrow$Other Considerations

Other Considerations

- Two-factor designs (Non-Parametric) in related samples

- Recall that we already covered in the previous lecture about the Friedman Two-Way ANOVA by Ranks
- It actually tests two factor models (without the interaction term). Latest research have worked extending the Friedman test to test the interaction effects, but these are outside our scope for now.
- Consider again this dataset this test applies:
  - Raw scores of three age groups under four treatments (teaching methods)

Groups	*I*	*II*	*III*	*IV*
A (age below 5)	9	4	1	7
B (age between 5 & 9)	6	5	2	8
C (age above 9)	9	1	2	6

- One factor is age group, the other is treatments but observations across treatments within each age group are not independent. Say, younger kids might learn faster

- But it is too restrictive in that we need the number of treatments equals to the number of individuals. It is for complete block designs.
- If the design is a balanced incomplete block design, that is
  - the no. of students in each group equals the number of treatments (even though some treatments might not have any students)
  - each treatment is available in each group
  - the total no. of students exposed to each treatment are the same
- then we can use the Durbin (1951) test
- Example:
  -

Groups	I	II	III	IV
A (age below 5)	9	4, 2		7
B (age between 5 & 9)	6, 2		2	8
C (age between 9 & 10)		1	2, 3	5
D (age above 10)	1	5	5	5