Applied Statistical Methods II

Chapter #19 in KNNL

Two-Factor Studies with Equal Sample Sizes

Looking forward:

- Ch 19:
  - Start two-way ANOVA with equal group sizes.
  - Consider model and inference with interaction.
  - Consider inference without interaction.

Two-Way ANOVA Models

- We will start to consider two-factor studies.
- Also known as two-way ANOVA models.
- Consider a study where we want to know how shelf height (high, middle, low) and shelf width (regular, wide) are associated with sales of bread.
  - 12 stores are participating.
  - each treatment combination has two stores.
- Often called a 2×3 factorial study.
- Presents one level more of complexity than one-way ANOVA.
  - Do the two treatments affect one-another?
  - In the regression context: question of interactions.

Some Notation

- Y_{ijk} is the k^{th} observation with first factor A at level i and second factor B at level j.
  - $i = 1, \ldots, a$ and $j = 1, \ldots, b$
  - $\mu_{ij} = E(Y_{ijk})$
  - $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$
  - When might we worry about identifiability of μ_{ij}?
  - n is the number of subjects per treatment pair. For now, we assume balanced design.

$$\begin{aligned}
\mu_{\cdot j} &= \frac{\sum_i \mu_{ij}}{a} \\
\mu_{i\cdot} &= \frac{\sum_j \mu_{ij}}{b} \\
\mu_{\cdot\cdot} &= \frac{\sum_i \sum_j \mu_{ij}}{a \times b}
\end{aligned}$$

These are all estimable quantities.

Some Notation

- Y_{ijk} is the k^{th} observation with first factor A at level i and second factor B at level j.
  - $i = 1, \ldots, a$ and $j = 1, \ldots, b$
  - $\mu_{ij} = E(Y_{ijk})$
  - $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$
  - When might we worry about identifiability of μ_{ij}?
  - n is the number of subjects per treatment pair. For now, we assume balanced design.

$$\begin{aligned}
\mu_{\cdot j} &= \frac{\sum_i \mu_{ij}}{a} \\
\mu_{i\cdot} &= \frac{\sum_j \mu_{ij}}{b} \\
\mu_{\cdot\cdot} &= \frac{\sum_i \sum_j \mu_{ij}}{a \times b}
\end{aligned}$$

These are all estimable quantities.

More Notation

- $Y_{ij\cdot} = \sum_k Y_{ijk}$
- $\overline{Y}_{ij\cdot} = \frac{Y_{ij\cdot}}{n}$
- $Y_{i\cdot\cdot} = \sum_j Y_{ij\cdot}$
- $\overline{Y}_{i\cdot\cdot} = \frac{Y_{i\cdot\cdot}}{bn}$
- Similarly for $Y_{\cdot j\cdot}$, $\overline{Y}_{\cdot j\cdot}$, $Y_{\cdot\cdot\cdot}$ and $\overline{Y}_{\cdot\cdot\cdot}$

Additive Model

- Often the science dictates that there should be an additive effect between the two factors.
- Consider a study of HIV viral counts.
  - Factor A is drug 1 or placebo.
  - Factor B is drug 2 or placebo.
  - Drug 1 stops the virus from binding with cells.
  - Drug 2 allows the immune system to recognize and destroy the virus.
- These are called either additive models or models with no interaction.

The Additive Two-Way ANOVA Model

- $\mu_{ij} = \mu_0 + \alpha_i + \beta_j$
- Are parameters identifiable?
- Convenient restriction: $\sum_i \alpha_i = \sum_j \beta_j = 0$
- What is interpretation of μ_0? α_i? β_j?
- $\mu_{..} = \mu_0 + \frac{1}{ab} \sum_i \sum_j (\alpha_i + \beta_j) = \mu_0$
- Main effect for the i^{th} level of A:
  $\alpha_i = \mu_{i.} - \mu_{..}$
- Main effect for the j^{th} level of B:
  $\beta_j = \mu_{.j} - \mu_{..}$
- Under additive model, we find:
  - $\hat{\mu}_{..} = \overline{Y}_{...}$
  - $\hat{\mu}_{i.} = \overline{Y}_{i..}$
  - $\hat{\mu}_{.j} = \overline{Y}_{.j.}$
  - $\hat{Y}_{ijk} = \hat{\mu}_{ij} = \overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}$

The Additive Two-Way ANOVA Model

- $\mu_{ij} = \mu_0 + \alpha_i + \beta_j$
- Are parameters identifiable?
- Convenient restriction: $\sum_i \alpha_i = \sum_j \beta_j = 0$
- What is interpretation of μ_0? α_i? β_j?
- $\mu_{..} = \mu_0 + \frac{1}{ab} \sum_i \sum_j (\alpha_i + \beta_j) = \mu_0$
- Main effect for the i^{th} level of A:
  $\alpha_i = \mu_{i.} - \mu_{..}$
- Main effect for the j^{th} level of B:
  $\beta_j = \mu_{.j} - \mu_{..}$
- Under additive model, we find:
  - $\hat{\mu}_{..} = \overline{Y}_{...}$
  - $\hat{\mu}_{i.} = \overline{Y}_{i..}$
  - $\hat{\mu}_{.j} = \overline{Y}_{.j.}$
  - $\hat{Y}_{ijk} = \hat{\mu}_{ij} = \overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}$

The Additive Two-Way ANOVA Model

- $\mu_{ij} = \mu_0 + \alpha_i + \beta_j$
- Are parameters identifiable?
- Convenient restriction: $\sum_i \alpha_i = \sum_j \beta_j = 0$
- What is interpretation of μ_0? α_i? β_j?
- $\mu_{\cdot\cdot} = \mu_0 + \frac{1}{ab} \sum_i \sum_j \left(\alpha_i + \beta_j\right) = \mu_0$
- Main effect for the i^{th} level of A:
  $\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot}$
- Main effect for the j^{th} level of B:
  $\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot}$
- Under additive model, we find:
  - $\hat{\mu}_{\cdot\cdot} = \overline{Y}_{\cdot\cdot\cdot}$
  - $\hat{\mu}_{i\cdot} = \overline{Y}_{i\cdot\cdot}$
  - $\hat{\mu}_{\cdot j} = \overline{Y}_{\cdot j\cdot}$
  - $\hat{Y}_{ijk} = \hat{\mu}_{ij} = \overline{Y}_{i\cdot\cdot} + \overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdot\cdot\cdot}$

Two-Way ANOVA Model with Interaction

- Often the two factors can affect one-another.
- Or testing this possible interaction is of importance.
- Consider a study of two drugs to battle ovarian cancer.
  - A 2×2 factorial design.
  - First drug destroys the cellular structure of a tumor.
  - The second drug attacks the blood supply to the tumor.
  - What if the first drug also destroys the veins into the tumor?

The Two-Way ANOVA Model with Interaction

- Main effect for the i^{th} level of A:
  $\alpha_i = \mu_{i.} - \mu_{..}$
- Main effect for the j^{th} level of B:
  $\beta_j = \mu_{.j} - \mu_{..}$
- Would like to write $\mu_{ij} = \underbrace{\mu_{..}}_{\text{intercept}} + \underbrace{\alpha_i + \beta_j}_{\text{main effects}} + \underbrace{(\alpha\beta)_{ij}}_{\text{interaction}}$
- Identifiability issues?
- Restrictions:
  - $\sum_i \alpha_i = \sum_j \beta_j = 0$
  - $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$
- $(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{i.} + \mu_{.j} - \mu_{..})$
  - Intuition: $(\alpha\beta)_{ij}$ is the different between μ_{ij} and the additive effects

The Two-Way ANOVA Model with Interaction

- Main effect for the i^{th} level of A:
  $\alpha_i = \mu_{i.} - \mu_{..}$
- Main effect for the j^{th} level of B:
  $\beta_j = \mu_{.j} - \mu_{..}$
- Would like to write $\mu_{ij} = \underbrace{\mu_{..}}_{\text{intercept}} + \underbrace{\alpha_i + \beta_j}_{\text{main effects}} + \underbrace{(\alpha\beta)_{ij}}_{\text{interaction}}$
- Identifiability issues?
- Restrictions:
  - $\sum_i \alpha_i = \sum_j \beta_j = 0$
  - $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$
- $(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{i.} + \mu_{.j} - \mu_{..})$
  - Intuition: $(\alpha\beta)_{ij}$ is the different between μ_{ij} and the additive effects

ANOVA Decomposition

- Two-way ANOVA with interaction, we find:
  - $\hat{\mu}_{..} = \overline{Y}_{...}$
  - $\hat{\mu}_{i.} = \overline{Y}_{i..}$
  - $\hat{\mu}_{.j} = \overline{Y}_{.j.}$
  - $\hat{Y}_{ijk} = \hat{\mu}_{ij} = \overline{Y}_{ij.}$
- The ANOVA decomposition $SSTO = SSTR + SSE$
  - $SSTO = \sum_{ijk} \left(Y_{ijk} - \overline{Y}_{...}\right)^2$ has $n \times a \times b - 1$ df.
  - $SSTR = n \sum_{ij} \left(\overline{Y}_{ij.} - \overline{Y}_{...}\right)^2$ has $a \times b - 1$ df.
  - $SSE = \sum_{ijk} \left(Y_{ijk} - \overline{Y}_{ij.}\right)^2$ has $a \times b \times (n-1)$ df.

Further Decomposition

- $SSTR = \underbrace{SSA}_{\text{main effect}} + \underbrace{SSB}_{\text{main effect}} + \underbrace{SSAB}_{\text{interaction}}$.
- $SSA = nb\sum_i \left(\overline{Y}_{i..} - \overline{Y}_{...}\right)^2$ has a-1 df.
- $SSB = na\sum_j \left(\overline{Y}_{.j.} - \overline{Y}_{...}\right)^2$ has b-1 df.
- $SSAB = n\sum_{ij} \left(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)^2$ has (a-1)(b-1) df.
  - The sum of the df must equal df from SSTR ab-1.
  - Know df for SSA and SSB.
  - Subtract to get df from SSAB.
- Define the mean squared errors by dividing the sums of squares errors by their degrees of freedom.

Quadratic Forms

- Let **Y** be the vector of all of the data.
- We can find matrices A, B, C, E such that:
  - $SSA = \mathbf{Y}^T A \mathbf{Y}$ where $rk(A) = a - 1$
  - $SSB = \mathbf{Y}^T B \mathbf{Y}$ where $rk(B) = b - 1$
  - $SSAB = \mathbf{Y}^T C \mathbf{Y}$ where $rk(C) = (a - 1)(b - 1)$
  - $SSE = \mathbf{Y}^T E \mathbf{Y}$ where $rk(E) = a \times b \times (n - 1)$
- For **balanced designs**,
  - $AB = 0$, $AC = 0$, $AE = 0$, $BC = 0$, $BE = 0$, $CE = 0$
  - Each of these sums-of-squares is independent χ^2 with some non-centrality parameter.
- This only holds for the two-way ANOVA under balanced design.

Expected Values of MS

- Recall: If A is idempotent and $\mathbf{Y} \sim \mathbf{N}(\mu, \mathbf{V})$ then $E\boldsymbol{Y}^T A\boldsymbol{Y} = tr(AV) + \mu^T A\mu$
- This can be used to find that:
  - $E[MSA] = \sigma^2 + bn\frac{\sum_i(\mu_{i.} - \mu_{..})^2}{a-1}$
  - $E[MSB] = \sigma^2 + an\frac{\sum_j(\mu_{.j} - \mu_{..})^2}{b-1}$
  - $E[MSAB] = \sigma^2 + n\frac{\sum_i\sum_j\{\mu_{ij} - (\mu_{i.} + \mu_{.j} - \mu_{..})\}^2}{(a-1)(b-1)}$
  - $E[MSE] = \sigma^2$

Testing

- Assuming normality, F-tests allow us to test for:
  1. Need for interactions $H_0 : (\alpha\beta)_{ij} = 0$ for all i, j
     - $F^* = \frac{MSAB}{MSE} \sim F_{(a-1)(b-1),(n-1)ab}$ under H_0.
  2. Level averages for factor A are equal: $H_0 : \mu_{1\cdot} = \cdots = \mu_{a\cdot}$
     - $F^* = \frac{MSA}{MSE} \sim F_{(a-1),(n-1)ab}$ under H_0.
  3. Level averages for factor B are equal $H_0 : \mu_{\cdot 1} = \cdots = \mu_{\cdot b}$
     - $F^* = \frac{MSB}{MSE} \sim F_{(b-1),(n-1)ab}$ under H_0.
- Care must be taken in interpreting the last two tests when there is interaction.
- Let's look at an example.

Recall the Additive Two-Way ANOVA Model

- Main effect for the i^{th} level of A:
  $\alpha_i = \mu_{i.} - \mu_{..}$
- Main effect for the j^{th} level of B:
  $\beta_j = \mu_{.j} - \mu_{..}$
- $\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$
- Under additive model, we find:
  - $\hat{\mu}_{..} = \overline{Y}_{...}$
  - $\hat{\mu}_{i.} = \overline{Y}_{i..}$
  - $\hat{\mu}_{.j} = \overline{Y}_{.j.}$
  - $\hat{Y}_{ijk} = \hat{\mu}_{ij} = \overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}$

ANOVA Decomposition for the Additive Model

$SSTO = SSTR + SSE_p$

- $SSTO = \sum_{ijk} \left(Y_{ijk} - \overline{Y}_{...}\right)^2$ has $n \times a \times b - 1$ df.
- $SSTR_* = n \sum_{ij} \left(\overline{Y}_{i..} + \overline{Y}_{.j.} - 2\overline{Y}_{...}\right)^2$ has $(a-1) + (b-1)$ df.
- $SSE_p = \sum_{ijk} \left(Y_{ijk} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)^2$ has $a \times b \times n - a - b + 1$ df.

Can see that $SSTR_* = SSA + SSB$.

Under the additive model, the SS for the interaction is absorbed into the SS for error.

The text calls SSE_p the pooled SSE.

Testing in the Additive Model

Assuming normality, F-tests allow us to test for:

1. Factor level A has any affect: $H_0 : \mu_{1\cdot} = \cdots = \mu_{a\cdot}$
   - $F^* = \frac{MSA}{MSE_p} \sim F_{(a-1),abn-a-b+1}$ under H_0.
2. Factor level B has any affect: $H_0 : \mu_{\cdot 1} = \cdots = \mu_{\cdot b}$
   - $F^* = \frac{MSB}{MSE_p} \sim F_{(b-1),abn-a-b+1}$ under H_0.

Let's do an example.

Analysis of Factor Means with No Interaction

- The analysis of the main effects is usually only of interest for the additive model.
- Consider several different scenarios:
  - Want to know the pairwise comparisons $D = \mu_{i\cdot} - \mu_{i'\cdot}$ for one factor.
    - Can use Tukey.
  - We want to know a few linear combinations that are known a priori.
    - Bonferroni
  - We want to know a lot of linear combinations of one factor.
    - Sheffé
  - We want to compare linear combinations of both factors.
    - Modify the procedures for a single factor.

Pairwise Comparisons for Factor A (Additive Model)

- We want confidence intervals for each difference $D_{ii'} = \mu_{i\cdot} - \mu_{i'\cdot}$.
- Tukey's procedure has optimal Type I error.
- Based on the distribution of: $\frac{\max(\overline{Y}_{i\cdot\cdot}) - \min(\overline{Y}_{i\cdot\cdot})}{s(\hat{D})} \sim q_{a,abn-a-b+1}$
- $\hat{D}_{ii'} \pm Ts\left(\hat{D}_{ii'}\right)$
- $\hat{D}_{ii'} = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{i'\cdot\cdot}$
- $s^2(\hat{D}_{ii'}) = 2\frac{MSE}{bn}$
- $T = \frac{1}{\sqrt{2}} q_{a,abn-a-b+1}(1-\alpha)$
- Can get tests by inverting the confidence interval.

Sheffé for Factor A

- We want to be able to test any liner combination $L = \sum_i c_i \mu_{i\cdot}$.
- Sheffé's confidence intervals are

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

- $S^2 = (a-1)F_{a-1,abn-a-b+1}(1-\alpha)$
- If we are interested in only a couple of comparisons.
    - or a couple of linear combinations
    - just apply Bonferroni.
- Let's do an example.

Interest in Two Factors

- You do not want to look at just pairwise difference of one factor but of both factors.
- You can use a Bonferroni adjustment to the two separate Tukey intervals.
- How would you do that?