Applied Statistical Methods II

Chapter #17

Analysis of Factor Level Means

A Single Estimate

- Assume that you are interested in the estimation/testing of one linear combination of the factor means.
  - μ_i for some i.
  - $\mu_i - \mu_j$ for some i, j.
  - $\frac{\mu_1+\mu_2}{2} - \frac{\mu_3+\mu_4}{2}$.
- Before we undertake the analysis, we know that we are interested in these combinations.
- Assuming normality, these are easy to estimate:
  - Use normality to get the sampling distribution of a test statistic.
  - Invert it to get a confidence interval.
  - Only interested in one combination, no need to worry about multiple testing.

Inference for μ_i

- Recall $\hat{\mu}_i = \overline{Y}_{i.}$
- $var(\overline{Y}_{i.}) = \sigma^2/n_i$
- $s^2(\overline{Y}_{i.}) = MSE/n_i$
- Under normality, $\frac{\overline{Y}_{i.} - \mu_i}{s(\overline{Y}_{i.})} \sim t_{n_T - r}$
    - $SSE/\sigma^2 \sim \chi^2_{n_T - r}$ is independent of $\overline{Y}_{i.}$.
- Easy to do tests and get confidence intervals:
    - Reject $H_0 : \mu_i = c$ vs $H_0 : \mu_i \neq c$ when

$$\left| \frac{\overline{Y}_{i.} - c}{s(\overline{Y}_{i.})} \right| > t_{n_T - r}(1 - \alpha/2).$$

    - $\overline{Y}_{i.} \pm t_{n_T - r}(1 - \alpha/2) s(\overline{Y}_{i.})$

Inference for $\mu_i - \mu_j$

- Let $D = \mu_i - \mu_j$ and $\hat{D} = \overline{Y}_{i.} - \overline{Y}_{j.}$
- $E\hat{D} = D$ follows directly from
  - unbiasedness of $\overline{Y}_{i.}$ and $\overline{Y}_{j.}$
- $Var(\hat{D}) = Var(\overline{Y}_{i.}) + Var(\overline{Y}_{j.}) = \sigma^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)$ follows from
  - independence of $\overline{Y}_{i.}$ and $\overline{Y}_{j.}$
- $s^2(\hat{D}) = s^2(\overline{Y}_{i.}) + s^2(\overline{Y}_{j.}) = MSE \left(\frac{1}{n_i} + \frac{1}{n_j}\right)$
- $\frac{\hat{D}-D}{s(\hat{D})} \sim t_{n_T - r}$.
- Use this distribution to get confidence intervals and tests.

Inference for $\sum_i c_i \mu_i$

- Assume you want to draw inference on a linear combination $L = \sum_i c_i \mu_i$.
  - Want to estimate $\mu_i - \mu_j$.
  - Want to compare $\frac{\mu_1+\mu_2}{2}$ to $\frac{\mu_3+\mu_4}{2}$.
- Can formulate the testing problem as $H_0 : L\mu = C$.
  - In second bullet above: $L = [\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}][\mu_1, \ldots, \mu_4]'$ and $C = 0$.
- Let $\hat{L} = \sum_i c_i \overline{Y}_{i\cdot}$.
  - Unbiased and normal.
- $Var(\hat{L}) = \sum_i c_i^2 Var(\overline{Y}_{i\cdot}) = \sigma^2 \sum c_i^2 / n_i$
- $s^2(\hat{L}) = MSE \sum c_i^2 / n_i$.
- $\frac{\hat{L}-L}{s(\hat{L})} \sim t_{n_T - r}$.

The Cereal Example

- Estimate μ_2 and its confidence interval.
- Is there a difference between μ_1 and μ_2?
- Do designs 3 and 4 collectively sell more than designs 1 and 2 on average?

Recall the One-Way ANOVA Model

The one-way ANOVA model:

- $Y_{ij} = \mu_i + \epsilon_{ij}$
- $i = 1, \ldots, r$
- $j = 1, \ldots, n_i$
- $E\epsilon_{ij} = 0$, $\text{Var}\epsilon_{ij} = \sigma^2$.
- $EY_{ij} = \mu_i$

Now we will focus on:

- inference on $g = 1, \ldots, G$ linear combinations of the group means $L_g = \sum_{i=1}^{r} c_{gi}\mu_i$.

Controlling Type I Error

- Main consideration:
  - Controlling type I error rate.
- Computing multiple tests at level α will inflate the *family-wise* Type I error.
  - We do three different level $\alpha = 0.05$ pairwise comparisons.
  - If tests are independent, probability of rejecting any of the 3 tests when H_0 is true is $1 - (.95)^3 = 0.143$
  - Dealt with these topics before.
  - Has a similar consequence on confidence intervals.

Controlling Type I Error

- Main consideration:
  - Controlling type I error rate.
- Computing multiple tests at level α will inflate the *family-wise* Type I error.
  - We do three different level $\alpha = 0.05$ pairwise comparisons.
  - If tests are independent, probability of rejecting any of the 3 tests when H_0 is true is $1 - (.95)^3 = 0.143$
  - Dealt with these topics before.
  - Has a similar consequence on confidence intervals.

Consider three situations

- You want to consider all pairwise comparisons. For r levels, that's $\frac{r(r+1)}{2}$ tests.
  - Which group means can you consider equal?
  - Will use Tukey's procedure for this.
- You want to explore any linear combination of factor means. i.e. $q^T \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix}$ for all unit vectors $q \in \mathbb{R}^r$.
  - Do a complete data exploration.
  - Will use Scheffé.
- Have a finite set of comparisons you want to look at.
  - If the number of contrasts is equal to or smaller than the number of factor levels.
  - Use Bonferroni.

Consider three situations

- You want to consider all pairwise comparisons. For r levels, that's $\frac{r(r+1)}{2}$ tests.
  - Which group means can you consider equal?
  - Will use Tukey's procedure for this.
- You want to explore any linear combination of factor means. i.e. $q^T \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix}$ for all unit vectors $q \in \mathbb{R}^r$.
  - Do a complete data exploration.
  - Will use Scheffé.
- Have a finite set of comparisons you want to look at.
  - If the number of contrasts is equal to or smaller than the number of factor levels.
  - Use Bonferroni.

Consider three situations

- You want to consider all pairwise comparisons. For r levels, that's $\frac{r(r+1)}{2}$ tests.
    - Which group means can you consider equal?
    - Will use Tukey's procedure for this.
- You want to explore any linear combination of factor means. i.e. $q^T \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix}$ for all unit vectors $q \in \mathbb{R}^r$.
    - Do a complete data exploration.
    - Will use Scheffé.
- Have a finite set of comparisons you want to look at.
    - If the number of contrasts is equal to or smaller than the number of factor levels.
    - Use Bonferroni.

Tukey's Multiple Comparisons

- You want to test, for all $1 \leq i < j \leq r$: $H_0(i,j) : \mu_i - \mu_j = 0$ vs. $H_a(i,j) : \mu_i - \mu_j \neq 0$.
- Tukey's multiple comparison method for testing:
  - has level exactly α for balanced design.
  - is conservative for unbalanced designs. Type I error < α.
- The resulting confidence intervals, if family of tests are repeated for multiple replications:
  - all confidence intervals for the family will be correct $(1-\alpha)$% of the time for equal sample sizes.
  - all confidence intervals for the family will be correct $> (1-\alpha)$% of the time for unequal sample sizes.
- Procedure is often called Tukey-Kramer for unequal sample sizes.

Idea Behind Tukey

- Consider $Y_1, \ldots, Y_r$ as independent normal random variables with variance σ^2.
- Studentized range is $\frac{\max(Y_i) - \min(Y_i)}{\sqrt{MSE}}$.
- Recall that MSE, estimate of σ^2, depends on degrees of freedom ν.
- Tukey computed the distribution $\frac{\max(Y_i) - \min(Y_i)}{\sqrt{MSE}} \sim q_{r,\nu}$.
- To get simultaneous confidence intervals for $D_{ij} = \mu_i - \mu_j$
  - $\hat{D}_{ij} \pm Ts(\hat{D}_{ij})$
  - $\hat{D}_{ij} = \overline{Y}_{i.} - \overline{Y}_{j.}$
  - $s^2(\hat{D}) = MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)$
  - $T = \frac{1}{\sqrt{2}} q_{r, n_T - r}(1 - \alpha)$
  - Adjust the multiple of the estimated standard deviation.

- Note that we use $1 - \alpha$ rather than $1 - \alpha/2$.
- To get the p-value of pairwise tests:
  - $q_{ij}^* = \frac{\sqrt{2}|\hat{D}_{ij}|}{s(\hat{D}_{ij})}$
  - p-val = $\Pr(q > q_{ij}^* | q \sim q_{r,n_T - r})$.
- Note that for unequal sample sizes, $s(\hat{D})$ will vary.

Scheffé

- Same principle used to get Working-Hotelling confidence bands can be used to get Scheffé adjusted tests.
- Contains all estimable comparisons with $(1-\alpha)\%$ confidence.
- Viewing ANOVA as regression, regression function is over a discrete space.
  - Confidence band can be seen as covering all possible linear combinations $L = \sum c_i \mu_i$.
  - Let's derive the confidence band.
- Adjust the multiple of the standard deviation:
  - $\hat{L} \pm Ss(\hat{L})$
  - $s^2(\hat{L}) = MSE \sum c_i^2/n_i$
  - $S^2 = (r-1)F_{r-1,n_T-r}(1-\alpha)$
- Tests based on $F^* = \frac{\hat{L}^2}{(r-1)s^2(\hat{L})} \sim F_{r-1,n_T-r}$.

Scheffé

- Same principle used to get Working-Hotelling confidence bands can be used to get Scheffé adjusted tests.
- Contains all estimable comparisons with $(1-\alpha)\%$ confidence.
- Viewing ANOVA as regression, regression function is over a discrete space.
  - Confidence band can be seen as covering all possible linear combinations $L = \sum c_i \mu_i$.
  - Let's derive the confidence band.
- Adjust the multiple of the standard deviation:
  - $\hat{L} \pm S s(\hat{L})$
  - $s^2(\hat{L}) = MSE \sum c_i^2 / n_i$
  - $S^2 = (r-1) F_{r-1, n_T - r}(1-\alpha)$
- Tests based on $F^* = \frac{\hat{L}^2}{(r-1)s^2(\hat{L})} \sim F_{r-1, n_T - r}$.

Bonferroni

- If have a few linear comparisons that we want to make.
- If the number of comparisons, G, is equal to or smaller than the number of factor levels.
- Deflate the level of a single test from α to α/G.
- What is the problem with using Bonferroni with many tests?
- Let's go back to simulation example.

Bonferroni

- If have a few linear comparisons that we want to make.
- If the number of comparisons, G, is equal to or smaller than the number of factor levels.
- Deflate the level of a single test from α to α/G.
- What is the problem with using Bonferroni with many tests?
- Let's go back to simulation example.

- If our goal is to identify different pairs, use Tukey.
- In R:
  - Fit an ANOVA object using `aov`
  - Call internal function `TukeyHSD(x = <aov object>)`
  - Try cereal example
- Using Tukey's multiple comparison procedure we would conclude at the level $\alpha = 0.05$ that there is no statistical difference between items sold with designs 1,2 and 3 but that design 4 sold more items then either designs 1,2, or 3.
- To get CI for multiple contrasts, have to do some work.
  - Want to know if the average number of sales with designs 1 and 2 is equal to design 3.
  - Want to know if the average number of sales with designs 1 and 2 is equal to design 4.

- If our goal is to identify different pairs, use Tukey.
- In R:
    - Fit an ANOVA object using `aov`
    - Call internal function `TukeyHSD(x = <aov object>)`
    - Try cereal example
- Using Tukey's multiple comparison procedure we would conclude at the level $\alpha = 0.05$ that there is no statistical difference between items sold with designs 1,2 and 3 but that design 4 sold more items then either designs 1,2, or 3.
- To get CI for multiple contrasts, have to do some work.
    - Want to know if the average number of sales with designs 1 and 2 is equal to design 3.
    - Want to know if the average number of sales with designs 1 and 2 is equal to design 4.