STAT 5000

Statistical Methods I

Week 6

Fall 2024

Dr. Danica Ommen

Unit 2

ANOVA: Diagnostics & Remedies

- ANOVA assumes:

$$\epsilon_{ij} \text{ are i.i.d. } N(0, \sigma^2)$$

  - Independence of groups and observations
  - Homogeneous (equal) variance:
    $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_r^2 = \sigma^2$
  - Normal Distribution:
    Random error terms are normally distributed

Model Diagnostics

- Many results from two-sample model diagnostics apply.
  - Independence: critical aspect
  - Equal Variances: important
  - Normality: only a concern for small sample sizes, or very skewed distributions
  - Outliers: results not robust
- Use residuals to assess model assumptions

$$e_{ij} = Y_{ij} - \overline{Y}_{i.}$$

Independence Assumption:

- Data Collection
  - Random sample(s) from multiple populations
  - Observations from multiple independent groups
- Study designed to produce independent responses

Equal Variance Assumption:

- Construct histograms of residuals for each groups
- Construct boxplots of residuals for each groups
- Plot residuals versus predicted values
  (there should be no trend)
  - Beware of interpretation if n_i's are very unequal
  - Expect larger range of ϵ_{ij} if n_i is larger
- Study ratio of sample standard deviations

$$\frac{\max\{S_i\}}{\min\{S_i\}}$$

Equal Variance Assumption:

- Tests for equality of variances
  - Brown-Forsythe test
  - Levene's test
  - etc.
- Consequences of unequal variances on *F*-test:
  - Minor if sample sizes are the same.
  - Large distortion of α level if very unequal sample sizes
  - Decreased power

Normality Assumption:

- Histogram of *residuals*
- Normal probability plot of residuals
- Numerical summaries - skewness and kurtosis
- Tests for Normality
    - Shapiro-Wilk
    - Kolmogorov-Smirnov
    - Cramer-von Mises
    - Anderson-Darling

- Assumptions
  - Independence
  - Homogeneous Variances
  - Normal Distribution
- What if the homogeneous variances and/or normal distribution assumptions are violated to the point where p-values and confidence levels cannot be trusted?
  - Transform data and check whether the homogeneous variances and normal distribution assumptions are appropriate for transformed data
  - Non-parametric Tests

- One-Way ANOVA on Ranks
- Assumptions
  - Independence
- Null hypothesis: r populations have the same distribution
  - Distribution is not required to be normal
  - Implies equal medians, percentiles, means and variances

- Combine the data into a single data set
- Order the N observations from smallest to largest
- Assign ranks R_{ij}
  - The smallest observation gets rank=1,
    the second smallest gets rank=2, etc...
  - For tied observations, average the ranks
- Calculate $\bar{R}_{i.}$ = the mean rank of observations in group i
- The test statistic is:

$$H = (N-1)\frac{\sum_{i=1}^{r} n_i(\bar{R}_{i.} - \bar{R})^2}{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(R_{ij} - \bar{R})^2}$$

where $\overline{R} = (N+1)/2$ = the average of all ranks 1 through N

- If H_0 is true, H will have an approximate χ^2 distribution with $r-1$ degrees of freedom
- Approximation is best when $n_i \geq 5$ for all i
- p-value $= P(\chi^2_{r-1} > H)$

Unit 2

ANOVA: Contrasts

Donut Example:

- Decision: We rejected the null hypothesis of an equal mean amount of oil absorbed for the four cooking oils.
- Interpretation: At least some of the means for the four cooking oils are different.
- Question: Which ones and by how much?

Additional Analyses:

- Inference for a single population mean
- Linear combinations of means, including contrasts
- Pairwise comparisons

Inference for Single Population Mean

- $100(1-\alpha)\%$ confidence interval for a single group mean

$$\bar{Y}_{i.} \pm t_{N-r,1-\alpha/2}\sqrt{\frac{MS_{error}}{n_i}}$$

- Note: MS_{error} is the estimate of the population variance σ^2
- Note: *df* for *t* distribution is $N-r$
- Valid for a single population mean
(not used for comparison between means)

Contrast

Linear combination of the population means with $\sum_{i=1}^{r} c_i = 0$:

$$\gamma = \sum_i c_i \mu_i$$

Examples:

- Difference between group 1 mean and mean of groups 2 & 3:

$$\gamma = \mu_1 - \frac{\mu_2 + \mu_3}{2}$$

  ($c_1 = 1$, $c_2 = -0.5$, $c_3 = -0.5$, and $c_4 = 0$)

- Difference between two group means:

$$\gamma = \mu_i - \mu_k$$

  ($c_i = 1, c_k = -1$, all other c's = 0)

- Point estimate: $\hat{\gamma} = \sum_i c_i \bar{Y}_{i.}$
- Standard error assuming σ^2 is known:

$$\sigma_{\hat{\gamma}} = \sqrt{\sigma^2 \sum_i (c_i^2 / n_i)}$$

- Standard error when σ^2 is NOT known:

$$S_{\hat{\gamma}} = \sqrt{MS_{error} \sum_i (c_i^2 / n_i)}$$

- $100(1-\alpha)\%$ confidence intervals:

$$\hat{\gamma} \pm t_{N-r,1-\alpha/2} S_{\hat{\gamma}}$$

Hypothesis Test

- Test $H_0 : \gamma = \sum_i c_i \mu_i = 0$
  - Using t distribution:

$$t = \frac{\hat{\gamma} - 0}{S_{\hat{\gamma}}} \text{ has } N - r \text{ d.f.}$$

  - Using F distribution:

$$F = \frac{SS_\gamma}{MS_{error}} \text{ has } (1, N - r) \text{ d.f.}$$

where $SS_\gamma = \dfrac{\hat{\gamma}^2}{(\sum_i c_i^2 / n_i)}$

Orthogonal Contrasts

Two contrasts $\gamma_1 = \sum_i c_i\mu_i$ and $\gamma_2 = \sum_i b_i\mu_i$ are *orthogonal* if

$$\sum_i b_i c_i / n_i = 0$$

Example:

$$\gamma_1 = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) \text{ and } \gamma_2 = \mu_2 - \mu_3$$

- $c_1 = 1, c_2 = -0.5, c_3 = -0.5$ and $b_1 = 0, b_2 = 1, b_3 = -1$
- $\sum_i b_i c_i / n_i = 0(1)/n_1 + 1(-0.5)/n_2 + -1(-0.5)/n_3 = 0$

If γ_1 and γ_2 are orthogonal contrasts, then

- They represent statistically unrelated pieces of information
- One contrast conveys no information about the other
  - Estimates $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are *uncorrelated*
  - Hypothesis tests for γ_1 and γ_2 are independent, i.e. results of one test do not affect results of other
  - Confidence intervals for γ_1 and γ_2 are independent i.e. results of one do not affect results of other

Properties of Orthogonal Contrasts

- A set of contrasts are orthogonal if all pairs are orthogonal
- For r means, there are at most $r-1$ mutually orthogonal contrasts in a set
- For r means, there are many possible sets of $r-1$ mutually orthogonal contrasts
- SS_{model} can be decomposed by $r-1$ mutually orthogonal contrasts
- Example: Let $r=4$ as in the donut example, and let $\gamma_1, \gamma_2, \gamma_3$ be mutually orthogonal contrasts. Then

$$SS_{\text{model}} = SS_{\gamma_1} + SS_{\gamma_2} + SS_{\gamma_3}$$

Orthogonal Polynomial Contrasts

- Analyze trends for quantitative factors or ordered treatments
- Assume equal spacing of levels and equal sample sizes
- For a factor with three equally spaced levels:

Trend	μ_1	μ_2	μ_3
Linear	-1	0	1
Quadratic	-1	2	-1

$$SS_{\text{model}} = SS_{\text{linear}} + SS_{\text{quad}}$$

Orthogonal Polynomial Contrasts

- For a factor with five equally spaced levels:

Trend	μ_1	μ_2	μ_3	μ_4	μ_5
Linear	-2	-1	0	1	2
Quadratic	-2	1	2	1	-2
Cubic	-1	2	0	-2	1
Quartic	1	-4	6	-4	1

$$SS_{\text{model}} = SS_{\text{linear}} + SS_{\text{quad}} + SS_{\text{cubic}} + SS_{\text{quartic}}$$

Why are orthogonal contrasts useful?

- *F*-test from the ANOVA table
  - Tests whether all groups have the same mean
  - We don't always care about the *F*-test ...
    Contrasts focus attention on specific questions
  - Researcher must specify the questions
- Independence of test results means we can interpret tests for contrasts individually
- Motivate partitioning of SS into "interesting" and "everything else" parts

Why are orthogonal contrasts useful?

- Researchers specify one question: "Did type1 oil have a different mean from the other three types?"
  answered by contrast with c=(-1,1/3, 1/3, 1/3)
- Does this contrast explain all differences among means?
  SS_{model} = $SS_1 + SS_2 + SS_3 = SS_1$+ rest

Source	SS	d.f.
Model	SS_{model}	3
Type1 vs Others	SS_c	1
rest	$SS_{model} - SS_c$	3-1
Error	SS_{error}	df_{error}

Unit 2

ANOVA: Multiple Comparisons

Compare means for each pair of treatments:

- Each comparison is a contrast: $\mu_i - \mu_k$ for all $i \neq k$
- There are $\binom{r}{2}$ possible pairwise comparisons
- Set of $\binom{r}{2}$ comparisons are NOT orthogonal
    - Example: $\mu_1 - \mu_2$ and $\mu_1 - \mu_3$

Contrast	μ_1	μ_2	μ_3
1	1	-1	0
2	1	0	-1

Inference for Pairwise Comparisons

Using MS_{error} as the estimate of the common variance:

- $(1-\alpha) \times 100\%$ confidence intervals for difference in two means:

$$(\bar{Y}_i - \bar{Y}_k) \pm t_{N-r,1-\alpha/2}\sqrt{MS_{error}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}$$

- Hypothesis test to compare two-means:

$$t = \frac{\bar{Y}_i - \bar{Y}_k}{\sqrt{MS_{error}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}} \quad \text{with } N - r \text{ df}$$

- Each pairwise comparison has Type I error level α or confidence level $100(1-\alpha)\%$
- We do $\binom{r}{2}$ such comparisons!
- If r is large, some significant differences are expected by chance even if all of the means are the same

- Known as the multiple comparisons problem
- When many comparisons are made, how should one interpret the *p*-value for a single comparison?
- Reminder: traditional *p*-value interpretation is derived from P(observe more extreme result $\mid H_o$ is true).
  small p-value $\Rightarrow$ observed statistic unlikely if H_o is true
  i.e. reject H_o if observed result is "unusual"
- Doesn't have the same interpretation when many comparisons are made

Example:

Experiment with 10 treatments: $\binom{10}{2} = 45$ possible tests

Case 1: Pre-specified Contrast

Case 2: Post-hoc Testing

Example:

Experiment with 10 treatments: $\binom{10}{2} = 45$ possible tests

Case 1: Pre-specified Contrast

- Test # 10 is the only test you want to do
- Result: p-value $= 0.032$
- Conclusion: p-value for test # 10 has the usual interpretation $\Rightarrow$ significant evidence of difference between two means since p-value < 0.05.

Case 2: Post-hoc Testing

Example:

Experiment with 10 treatments: $\binom{10}{2} = 45$ possible tests

Case 1: Pre-specified Contrast

Case 2: Post-hoc Testing

- Test # 10 has the smallest p-value (p-value $= 0.032$)
- With 45 *independent* tests, one would expect (45)(0.05) = 2.25 of the p-values to be smaller than 0.05 if all H_0's are true
- A p-value of 0.032 is no longer unusual!

- Comparison-wise type I error rate:

$$P(\text{ reject } H_0 \text{ for } \textbf{one} \text{ test } | H_0 \text{ is true for that test })$$

- Experiment-wise type I error rate:

$$P(\text{ reject at least one of the } H_0\text{'s } | \text{ all } H_0\text{'s are true })$$

- Multiple comparisons adjustment:
  - Avoid too many *false* significant findings
  - Make experiment-wise Type I error rate reasonably small
  - Equivalent to simultaneous confidence intervals, i.e. all confidence intervals in a set include their individual targets with a specified probability

Basic Approach

- Adjust the $t_{N-r,1-\alpha/2}$ critical value used in individual $100(1-\alpha)\%$ confidence intervals or individual α-level t-tests
- Cost is lower power: less likely to detect a non-zero effect
- Benefit is that the experiment-wise Type I error rate is no larger than the specified α

- Comparison-wise Type I error rate
  - Least Significant Difference (LSD)
- Experiment-wise Type I error rate
  - Tukey-Kramer Honest Significant Difference (HSD)
  - Scheffe's
  - Bonferroni

SAS Code:

```
proc glm data=donut;
  class oil;
  model y = oil / p;
  estimate 'o4-(o1+o2+o3)/3' oil -1 -1 -1 3 / divisor=3;
  estimate 'o2-(o1+o3)/2' oil -0.5 1 -0.5 0;
  estimate 'o1-o3' oil 1 0 -1 0 ;
  contrast 'o4-(o1+o2+o3)/3' oil -1 -1 -1 3 ;
  contrast 'o2-(o1+o3)/2' oil -0.5 1 -0.5 0;
  contrast 'o1-o3' oil 1 0 -1 0 ;
  means oil / alpha=.05 bon lsd scheffe tukey snk;
  output out=set2 residual=r predicted=yhat;
run;
```

Least Significant Difference (LSD)

- Conduct overall F-test of $H_0 : \mu_1 = ... = \mu_r$ at the α level
- If H_0 is not rejected then declare all means the same (chance of any false declarations of significant differences is less than α)
- If H_0 is rejected then calculate confidence intervals or conduct hypothesis tests
- Commonly used, but substantial loss of power when only a few groups have different means

LSD: Donut Example

$$LSD = t_{20,.975}\sqrt{MS_{error}\left(\frac{2}{n}\right)} = (2.086)\sqrt{100.9\left(\frac{2}{6}\right)} = 12.1$$

Declare a significant difference if $|\bar{Y}_i - \bar{Y}_j| \geq$LSD.

Order sample means from smallest to largest:

Oil 4	Oil 1	Oil 3	Oil 2
12	22	26	35

The GLM Procedure

t Tests (LSD) for y

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	100.9
Critical Value of t	2.08596
Least Significant Difference	12.097

Means with the same letter are not significantly different.

t Grouping		**Mean**	**N**	**Oil**
	A	35.000	6	2
	A			
B	A	26.000	6	3
B				
B	C	22.000	6	1
	C			
	C	12.000	6	4

Tukey-Kramer Honest Significant Difference (HSD)

- Used to compare all pairs of treatment means
- Experiment-wise error rate is α for the entire set of the $\binom{r}{2}$ possible comparisons
- An exact solution for all pairwise comparisons with equal sample sizes (Tukey)
- Conservative for unequal sample sizes (Kramer modification)

Tukey-Kramer Honest Significant Difference (HSD)

- Based on the distribution of studentized range

$$q_{(r,N-r)} = \left(\max_i \bar{Y}_i - \min_i \bar{Y}_i \right) / (S_p/\sqrt{n})$$

- Use $\frac{1}{\sqrt{2}} q_{(r,N-r,1-\alpha)}$ in CIs
- For tests, declare a significant difference if

$$|\bar{Y}_i - \bar{Y}_j| \geq \frac{1}{\sqrt{2}} q_{(r,N-r,1-\alpha)} \sqrt{MS_{error} \left(\frac{1}{n} + \frac{1}{n} \right)}$$

Tukey HSD: Donut Example

$$\begin{aligned} HSD &= \frac{1}{\sqrt{2}} q_{(4,20,.95)} \sqrt{MS_{error}\left(\frac{2}{n}\right)} \\ &= \frac{1}{\sqrt{2}}(3.958)\sqrt{100.9\left(\frac{2}{6}\right)} = 16.23 \end{aligned}$$

Declare a significant difference if $|\bar{Y}_i - \bar{Y}_j| \geq$ HSD

Order sample means from smallest to largest:

Oil 4	Oil 1	Oil 3	Oil 2
12	22	26	35

The GLM Procedure

Tukey's Studentized Range (HSD) Test for y

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	100.9
Critical Value of Studentized Range	3.95825
Minimum Significant Difference	16.232

Means with the same letter are not significantly different.

Tukey Grouping		**Mean**	**N**	**oil**
	A	35.000	6	2
	A			
B	A	26.000	6	3
B	A			
B	A	22.000	6	1
B				
B		12.000	6	4

Scheffe's Method:

- Works for any number of (actually all possible) linear contrasts
- Most conservative procedure, but relatively easy to apply
- use $\sqrt{(r-1)F_{r-1,N-r,1-\alpha}}$ in place of $t_{N-r,1-\alpha/2}$
- Declare a significant difference if

$$|\bar{Y}_i - \bar{Y}_j| \geq \sqrt{(r-1)F_{r-1,N-r,1-\alpha}}\sqrt{MS_{error}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Scheffe: Donut Example

$$
\begin{aligned}
Sch &= \sqrt{3F_{3,20,.95}}\sqrt{MS_{error}\left(\frac{1}{6}+\frac{1}{6}\right)} \\
&= \sqrt{(3)(3.098)}\sqrt{100.9\left(\frac{2}{6}\right)} = 17.68
\end{aligned}
$$

Declare a significant difference if $|\bar{Y}_i - \bar{Y}_j| \geq Sch$.

Order sample means from smallest to largest:

Oil 4	Oil 1	Oil 3	Oil 2
12	22	26	35

The GLM Procedure

Scheffe's Test for y

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	100.9
Critical Value of F	3.09839
Minimum Significant Difference	17.681

Means with the same letter are not significantly different.				
Scheffe Grouping		**Mean**	**N**	**oil**
	A	35.000	6	2
	A			
B	A	26.000	6	3
B	A			
B	A	22.000	6	1
B				
B		12.000	6	4

Bonferroni Method:

- If we have m tests (or confidence intervals), use α/m instead of α in each test (or confidence interval)
- Easy to implement
- Declare a significant difference if

$$|\bar{Y}_i - \bar{Y}_j| \geq t_{N-r,1-\frac{\alpha}{2m}} \sqrt{MS_{error}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

- Conservative, especially if m is large and tests are not independent (experiment-wide type I error rate $< \alpha$)
- Need to pre-specify the number of comparisons m

Bonferroni: Donut Example

There are $r = 4$ treatments and $m = 6$ pairs of means to compare.

$$\begin{aligned} Bonf &= t_{20,1-\frac{0.05}{12}} \sqrt{MS_{error}\left(\frac{1}{6}+\frac{1}{6}\right)} \\ &= (2.927)\sqrt{100.9\left(\frac{2}{6}\right)} = 16.975 \end{aligned}$$

Declare a significant difference if $|\bar{Y}_i - \bar{Y}_j| \geq Bonf$.

Order sample means from smallest to largest:

Oil 4	Oil 1	Oil 3	Oil 2
12	22	26	35

The GLM Procedure

Bonferroni (Dunn) t Tests for y

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	100.9
Critical Value of t	2.92712
Minimum Significant Difference	16.976

Means with the same letter are not significantly different.				
Bon Grouping		**Mean**	**N**	**oil**
	A	35.000	6	2
	A			
B	A	26.000	6	3
B	A			
B	A	22.000	6	1
B				
B		12.000	6	4

Many, many other multiple comparison techniques

- Dunnet's procedure to compare each of r-1 treatment means to the mean for a control group
- Step down procedures, like the Student-Newman-Kuels (SNK) procedure, increase power
- Decision theory inspired procedures like Duncan's multiple range procedure
- Methods to control false discovery rates in genomic experiments

The GLM Procedure

Student-Newman-Keuls Test for y

Note: This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	100.9

Number of Means	**2**	**3**	**4**
Critical Range	12.097171	14.672442	16.232038

Means with the same letter are not significantly different.				
SNK Grouping		**Mean**	**N**	**Oil**
	A	35.000	6	2
	A			
B	A	26.000	6	3
B	A			
B	A	22.000	6	1
B				
B		12.000	6	4

- Set n=10 observations per group. Consider:
  - r=3, 3 comparisons, error df=27
  - r=10, 45 comparisons, error df=90
- Compare critical values, α = 0.05

Method	r=3	r=10
LSD (unadjusted t)	2.05	1.99
Tukey-Kramer ($q/\sqrt{2}$)	2.48	3.25
Bonferroni	2.55	3.37
Scheffe'	2.59	4.23

- Power to detect δ = 1.4, $\sigma = 1$, n=10

Method	r=3	r=10
LSD (unadjusted t)	84%	84%
Bonferroni	68%	27%

- Many possible approaches, many different opinions
- My philosophy: Treatment comparisons should be pre-selected to answer specific questions
- When a study has a relatively small number of planned comparisons or contrasts
  - perform tests or construct confidence intervals with Bonferroni adjustments
  - Prefer but don't require orthogonal contrasts (simple to interpret)
  - Use SNK or HSD to compare all pairs of means

- When a large number of unplanned comparisons are examined
  - What is the appropriate family of comparisons (all pairs of means, all possible contrasts)
  - Use the most powerful appropriate multiple comparison procedure (SNK, Bonferroni, Sheffe)
- Confidence intervals
  - Do I need an interval for only one comparison?
  - Or simultaneous intervals for several comparisons?
  - Use multiple comparison adjustment
    - HSD for pairs of means
    - Bonferroni for a few contrasts
    - Scheffe for unlimited contrasts

Questions?

Contact me:

Email: dmommen@iastate.edu
Student Office Hours: Thursdays @ 10-11 am