P8130: Biostatistical Methods I

Lecture 8: One-Way Analysis of Variance (ANOVA)

Cody Chiuzan, PhD
Department of Biostatistics
Mailman School of Public Health (MSPH)

Lectures 6&7: Recap

- Inferences for one- and two-samples means
  - Confidence intervals
  - Hypothesis testing
  - Power and sample size calculation

Lectures 8: Outline

- Analysis of Variance (ANOVA)
    - One-Way ANOVA (one treatment group with more than two levels)
    - Hypothesis testing in One-Way ANOVA (fixed effects model)
    - Comparisons of specific groups
        - Multiple comparisons – methods of adjusting for multiple comparisons

What is Analysis of Variance (ANOVA)?

- ANOVA is used for comparing the means of more than two distributions
  - Multi-sample inferences – imply more than two independent groups
  - Compare the response (continuous variable) with respect to the levels of a factor (categorical variable)

- In One-Way ANOVA we have <u>one</u> variable with more than two levels
  - E.g., Treatment group: placebo, high dose, low dose.

- We want to test for any differences in mean response among those levels
  - A direct generalization of the t-test methodology

What is Analysis of Variance (ANOVA)?

- One-Way ANOVA: test for any differences in mean response among different levels of a factor (>2 levels)

- Can we simply subtract the sample mean values? No.

- Instead, we analyze the variance in the data by partitioning it
  - Is the variance **within** each group small compared to the variance **between** the groups (specifically, between the group means)?

One-Way ANOVA: General Hypotheses

- Suppose that in an experiment we have *k* treatment groups (or number of levels).
- The multi-sample hypotheses to be tested are:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k, \text{ i.e., } ALL\ treatment\ means\ are\ equal$$

vs.

H_1: at least two of the treatment population means differ,

i.e., not *ALL treatment means are equal*

One-Way ANOVA: Model

- Suppose that there are k groups with n_i observations in the i^{th} group.
- Let y_{ij} denote the j^{th} observation from the i^{th} group.
- The One-Way ANOVA model can be specified as:

$$y_{ij} = \mu + \alpha_i + e_{ij}\,, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i.$$

Where:

- μ is a constant representing the underlying mean of all groups taken together (the 'grand mean')
- α_i is a constant representing the difference between the mean of the i^{th} group and the 'grand mean'
- e_{ij} represent the random error terms, $e_{ij} \sim N(0, \sigma^2)$
  - An observation from the i^{th} group is normally distributed with mean $\mu + \alpha_i$ and variance σ^2.

One-Way ANOVA: Assumptions

- There are *k* populations of interest (*k* > 2)
- The samples are drawn independently from the underlying populations
- Homoscedasticity: the variances of the *k* populations are equal
  - Variance of the outcome does not depend on the sample
- Normality: the outcomes are normally distributed
- Any assumption violation jeopardizes the method validity!

One-Way ANOVA: Hypothesis Testing

- The One-Way ANOVA model is specified as:

$$y_{ij} = \mu + \alpha_i + e_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i \ (*)$$

- Let *n* be the total number of observations: $\sum_{i=1}^{n} n_i = n$.
- Let $\overline{y_i}$ denote the mean from the i^{th} group.
- Let $\bar{\bar{y}}$ denote the 'grand mean', $\bar{\bar{y}} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}}{n}$.
- We can write:

$$y_{ij} - \bar{\bar{y}} = (y_{ij} - \overline{y_i}) \quad + \quad (\overline{y_i} - \bar{\bar{y}})$$

Within group variability + Between group variability

One-Way ANOVA: Hypothesis Testing

- We can partition the total variability:

$$y_{ij} - \bar{\bar{y}} = (y_{ij} - \bar{y}_i) \quad + \quad (\bar{y}_i - \bar{\bar{y}})$$

Within group variability + Between group variability

**Squaring both sides of the equation results in:

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{\bar{y}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_i - \bar{\bar{y}})^2$$

Total Sum of Squares (Total SS) = Within Sum of Square (SS) + Between Sum of Squares (SS)

- Thus, we have partitioned the total variability into two components.

** The cross-product is equal to zero.

One-Way ANOVA: Hypothesis Testing

- Short computational forms:

$$\text{Total SS} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} {y_{ij}}^2 - \frac{(\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij})^2}{n}$$

$$\text{Between SS} = \sum_{i=1}^{k} n_i{\overline{y_i}}^2 - \frac{(\sum_{i=1}^{k} n_i\overline{y_i})^2}{n} = \sum_{i=1}^{k} n_i{\overline{y_i}}^2 - \frac{{y_{..}}^2}{n}$$

$$\text{Within SS} = \sum_{i=1}^{k}(n_i - 1)s_i^2$$

Where:

$y_{..}$ represents the sum of all observations across all groups, the ‘grand total’.

n represents the total number of observations over all groups.

s_i^2 represents the sample variances for the i^{th} group.

One-Way ANOVA: Hypothesis Testing

The Mean Squares can be calculated from the Sum of Squares as:

Between Mean Square (MS) = Between SS / (k-1)

Within Mean Square (MS) = Within SS / (n-k)

- Note that the Within Mean SS ‘pools the sample variances’ of the k groups:

$$\text{Within Mean SS} = \frac{\sum_{i=1}^{k}(n_i-1)s_i^2}{\sum_{i=1}^{k}(n_i-1)} = \frac{\sum_{i=1}^{k}(n_i-1)s_i^2}{n-k}.$$

- Note that the sum of Between and Within MS is <u>not equal</u> to the Total MS.

One-Way ANOVA: Summary Table

ANOVA results are usually summarized in the following table:

Source	**Sum of Square (SS)**	**Degrees of freedom (df)**	**Mean Sum of Square**	**F-Statistics**
Between	Between SS	k-1	Between SS/(k-1)	$F = \frac{\text{Between SS}/(k-1)}{\text{Within SS}/(n-k)}$
Within	Within SS	n-k	Within SS/(n-k)	
Total	Between SS + Within SS	n-1		

One-Way ANOVA: Overall F-test

Testing the hypotheses:

H_0: $\mu_1 = \mu_2 = \ ... = \mu_k$

H_1: at least two means are not equal

is equivalent to:

H_0: $\alpha_1 = \alpha_2 = \ ... = \alpha_k = 0$ (from *)

H_1: at least two α's are not equal

- Compute the test statistic:

$$F = \frac{\text{Between SS}/(k-1)}{\text{Within SS}/(n-k)} \sim F_{k-1,n-k} \text{ distribution under } H_0$$

- Reject H_0: if $F > F_{k-1,n-k,1-\alpha}$
- Fail to reject H_0: if $F \leq F_{k-1,n-k,1-\alpha}$
- P-value: area to the right: $P(F_{k-1,n-k} > F)$.

One-Way ANOVA: Example

A study is examining the effect of glucose on insulin release. Specimens of pancreatic tissue from experimental animals were treated with five different stimulants. Later, the amounts of insulin released were recorded.

Let y_{ij} denote the amount of insulin released from the j^{th} mouse in the i^{th} experimental group ($i = 1, 2, \dots, 5;\ j = 1, 2, \dots, n_i$).

Question: are there any significant differences among the population means for the five stimulants?

One-Way ANOVA: Example

	Responses (y_{ij}) by Stimulant Group				
Mouse	**1**	**2**	**3**	**4**	**5**
1	1.53	3.15	3.89	8.18	5.86
2	1.61	3.96	3.68	5.64	5.46
3	3.75	3.59	5.70	7.36	5.69
4	2.89	1.89	5.62	5.33	6.49
5	3.26	1.45	5.79	8.82	7.81
6		1.56	5.33	5.26	9.03
7				7.10	7.49
8					8.98
$\sum_{j=1}^{n_i} y_{ij}$	**13.04**	**15.60**	**30.01**	**47.69**	**56.81**
$\sum_{j=1}^{n_i} {y_{ij}}^2$	**37.98**	**46.60**	**154.68**	**337.17**	**417.92**
$\bar{y}_i$	**2.61**	**2.60**	**5.01**	**6.81**	**7.10**
s_i^2	**0.99**	**1.21**	**0.92**	**2.04**	**2.02**

One-Way ANOVA: Example

Calculate the Sum of Squares:

Total SS $=\sum_{i=1}^{k}\sum_{j=1}^{n_i} {y_{ij}}^2 - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}\right)^2}{n} = 994.35 - \frac{163.15^2}{32} = 162.54$

Between SS = $\sum_{i=1}^{k} n_i {\overline{y}_i}^2 - \frac{{y_{..}}^2}{n} = \frac{13.04^2}{5} + \frac{15.60^2}{6} + \frac{30.01^2}{6} + \frac{47.69^2}{7} + \frac{56.81^2}{8} - \frac{163.15^2}{32} = 121.19$

Within SS = Total SS - Between SS = 162.54 – 121.19 = 41.35

F = $\frac{\text{Between SS}/(k-1)}{\text{Within SS}/(n-k)} = \frac{121.19/4}{41.35/27} = 19.78 > F_{4,27,0.95} = 2.73.$

At 0.05 significance level, we reject the null hypothesis and conclude that at least two of mean insulins from the five stimulant groups are different.

Next question: Which two means are different?

One-Way ANOVA: Multiple Comparisons

- The overall F-test can only indicate a significant difference between the group means (if the case), but does not tell which means are different.

- Problem: multiple looks/comparisons can increase the type I error, i.e., a higher probability of rejecting the null hypothesis in error
  - In our insulin example, there are 10 possible pairwise comparisons. If you make 10 comparisons (t-tests), each at 0.05 significance level, then the probability of rejecting at least 1 out of 10 is approximately 0.40 (why?)

- We need to control and preserve the overall (family-wise) error rate at the pre-specified alpha level:

$$\text{FWER} \leq 1 - (1 - \alpha)^{total\ number\ of\ comparisons}$$

In our example: FWER $\leq 1 - (1 - 0.05)^{10} = 0.401$

Adjusting for Multiple Comparisons

- We need to modify the critical regions so that we can control the probability of rejecting when there are no real differences.

- Bonferroni adjustment: $\alpha^* = \frac{\alpha}{\binom{k}{2}}$

  Reject H_0: if $|t| > t_{n-k,1-\alpha^*/2}$

  Fail to reject H_0: if $|t| \leq t_{n-k,1-\alpha^*/2}$

- Bonferroni is the most conservative method, i.e., the most stringent in declaring significance (thus, less powerful).

Bonferroni adjustment for Insulin Example

- In our insulin example: α^*=0.05/10=0.005 and we would reject if: $|t| > t_{27,0.9975} = 3.06$.

T-test statistics for all 10 pairwise comparisons

Means	(1)	(2)	(3)	(4)
(2)	-0.011	-	-	-
(3)	3.194*	3.361*	-	-
(4)	5.802*	6.118*	2.630	-
(5)	6.368*	6.734*	3.140*	0.450
* Indicates statistical significance after Bonferroni adjustment for an overall type I error set at 0.05.				

Other Multiple Comparisons Adjustments

- Tukey’s method – controls for all pairwise comparisons and it is less conservative than Bonferroni.
- Scheffe’s method – allows analysts to perform any comparison they might think of, not just pairwise comparisons.
- Dunnett’s method – mainly focuses on comparisons with a pre-defined control arm.
- Sequential procedures: Benjamini-Hochberg’s method that controls for the false-discovery rate (FDR).
  - FWER guards against any ‘false positives’
  - FDR is designed to control the proportions of ’false-positives’ among the set of rejected hypotheses (allows for ‘some’ false positives)

Readings

Rosner, *Fundamentals of Biostatistics*, Chapter 12

- Sections: 12.1 – 12.4
- Read section about Linear Contrasts (also in Recitation 4)
- Two-Way (Multi-Way) ANOVA and ANCOVA will be covered as part of linear regression