Module 3: Models for Data

3.3 Analysis of Variance

Analysis of Variance

- Analysis of Variance (ANOVA) is a general technique for splitting variability into its component parts to understand effects.
- It is an extension of two sample tests to situations where there are more than two treatments being studied
- In general, it is used to test questions like:
  - Are the mean contaminant concentrations different across several sites?
  - Do different treatments of contaminated materials give different results?

Module 3.3

Analysis of Variance

- Factors are variables that take on a finite number of levels.
- Examples:
  - Treatment type
  - Location
  - Species
- ANOVA can be used to study the effects of one factor at multiple levels (One-Way ANOVA), two factors at multiple levels (Two-Way ANOVA), and so forth.

Analysis of Variance

- Factors have fixed effects when they are set, or fixed, by the experimenter.
  - An example would be treatment levels of low, medium, and high
- Factors have random effects when the levels are a random selection from a set of possibilities.
  - Examples are randomly selected locations.

One-Way ANOVA

- The question to be answered in one-way ANOVA is:
  - Are the means of the different factor levels equal?
  - Said another way, are there significant effects associated with being at different factor levels?
- Notation:
  - X_{ij} = j^{th} observed data point at the i^{th} factor level
  - μ = overall mean

Module 3.3

One-Way ANOVA

- Notation
  - a_i = effect associated with the i^{th} level of factor A = deviation from the overall mean caused by the i^{th} factor level
  - ε_{ij} = random error associated with X_{ij}
- Model

$$X_{ij} = \mu + a_i + \varepsilon_{ij}$$

Module 3.3

One-Way ANOVA

- So, the observation X_{ij} can be explained by splitting it into component parts
  - an overall mean for the whole data set
  - an effect of being at the i^{th} factor level that makes the i^{th} level different from the overall mean
  - a random error

Module 3.3

One-Way ANOVA

Null hypothesis

$$H_o: a_1 = a_2 = \ldots = a_I = 0$$

(H_o implies that there is no effect in being at different factor levels of A)

Alternative hypothesis

$$H_A: a_1 = a_2 = \ldots = a_I \text{ not} = 0$$

(H_A implies that there is at least one factor level effect that is not zero)

Module 3.3

One-Way ANOVA

Another way to state this which is simpler and more intuitive is:

Null hypothesis:

$$H_o: \mu_1 = \mu_2 = \ldots = \mu_I$$

(H_o states that means are equal)

Alternative hypothesis:

$$H_A: \mu_1 \text{ not} = \mu_2 \text{ not} = \ldots \text{ not} = \mu_I$$

(H_A states that at least one mean is not equal to the others)

I would encourage you to use this simpler, more intuitive approach

Module 3.3

One-Way ANOVA

Running the ANOVA generates a Table such as Manly's Table 3.5 (page 76).

If the F statistic is significant ($p<0.05$), then Reject H_o.

Rejecting the null hypothesis leads to a conclusion that significant effects do exist or, stated another way, the factor means are not all equal.

Module 3.3

One-Way ANOVA Example

- Assume the data below came from three sites
- The research question is: are the mean contaminant concentrations at the sites different?

Location 1	Location 2	Location 3
10	14	5
12	18	9
8	25	15
15	15	11
11	18	8

Module 3.3

One-Way ANOVA Example

- The one-way ANOVA was performed using EXCEL.
- The output from the ANOVA is on the next slide
- Question: Does the test show that the means are different or equal?

Module 3.3

One-Way ANOVA Example

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Location 1	5	56	11.2	6.7		
Location 2	5	90	18	18.5		
Location 3	5	48	9.6	13.8		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	198.9	2	99.46667	7.6513	0.0072	3.8853
Within Groups	156	12	13			
Total	354.9	14				

One-Way ANOVA Example

- Answer:
  - Recall that the null hypothesis is that the means are equal which is equivalent to saying that the effect of moving from one location to another is zero
  - The p-value of the test is 0.0072
  - So, there is less than a 1% chance that these data could have occurred if the null hypothesis were true.

One-Way ANOVA Example

- Answer:
  - So, the null hypothesis is very unlikely to be true and is rejected
  - Conclusion: There is an effect of location
  - Said another way, the means are not all equal

Module 3.3 15

One-Way ANOVA

- The results of the ANOVA are that all of the means are not equal. BUT, this does not show which means are different.
- For example, all of the means EXCEPT ONE may be equal and only that one is different from the others.
- To determine this statistically, a Post Hoc (or multiple comparison) test must be done.

Module 3.3 16

One-Way ANOVA

- It may be tempting to do a series of t tests on pairs of factors levels. DON'T do this.
- The overall Type I error rate (alpha) of the set of pairwise comparisons is far higher than the fixed alpha level. This means that many of the means will appear significantly different when they are not.
- Post Hoc tests fix the alpha level for the entire set of comparisons.

One-Way ANOVA

- There are many good Post Hoc tests:
  - Least significant difference (LSD)
  - Duncan's multiple range test
  - Tukey's test
  - Newman-Keuls test
- We will look at the LSD test because it is easy although the others are better tests. A good statistical software package is a real help in carrying these out. Excel does not do any of these, you will need to tell Excel how to do the calculations (type the equations in in other words).

One-Way ANOVA

- Least significant difference test

$$LSD = t_{alpha,df}\sqrt{\frac{2s^2}{n}}$$

- Means different by more than this value are significantly different (regardless of sign), those different by less are not.

Module 3.3

19

One-Way ANOVA Example

- S^2_{pooled} = MS_{within} = 13.0
- n_i = size of each sample = 5
- df = $(n_1 -1)+(n_2 -1)+ (n_2 -1)$ = 4+4+4 = 12
- $t_{0.05, 12}$ = 2.179
- LSD = 2.179*sqrt((2*13)/5) = 2.179*2.280 = 4.97

Module 3.3

20

One-Way ANOVA Example

- $Mean_{Location\ 1}$ – $Mean_{Location\ 2}$ = 11.2 – 18 = -6.8
- $Mean_{Location\ 1}$ – $Mean_{Location\ 3}$ = 11.2 – 9.6 = 1.6
- $Mean_{Location\ 2}$ – $Mean_{Location\ 3}$ = 18 – 9.6 = 8.4
- Conclusion: Location 2 is significantly different from 1 and 3 but locations 1 and 3 are not different

Two-Way ANOVA

- ANOVA is easily extended to more than one factor
- Notation:
  - X_{ijk} = k^{th} observed data point at the i^{th} factor level of A and the j^{th} factor level of B
  - μ = overall mean

Two-Way ANOVA

- Notation
  - a_i = effect of the i^{th} level of A
  - b_j = effect of the j^{th} level of B
  - $(ab)_{ij}$ = interaction between A and B
  - ε_{ijk} = random error associated with X_{ijk}
- Model

$$X_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \varepsilon_{ijk}$$

Two-Way ANOVA

The ANOVA is carried out generating a table such as Manly's Table 3.6 (page 77)

Here the effect of factor A, B, and the interaction are separately tested for significance.

If they are significant ($p<0.05$) then the null hypothesis of no effect is rejected.