One-way Between Groups Analysis of Variance

Ψ320

Ainsworth

Major Points

- Problem with t-tests and multiple groups
- The logic behind ANOVA
- Calculations
- Multiple comparisons
- Assumptions of analysis of variance
- Effect Size for ANOVA

T-test

- So far, we have made comparisons between a single group and population, 2-related samples and 2 independent samples
- What if we want to compare more than 2 groups?
- One solution: multiple t-tests

T-test

- With 3 groups, you would perform 3 t-tests
- Not so bad, but what if you had 10 groups?
- You would need 45 comparisons to analyze all pairs
- That's right ***<u>45!!!</u>***

The Danger of Multiple t-Tests

- Each time you conduct a t-test on a single set of data, what is the probability of rejecting a true null hypothesis?
- Assume that H_0 **is true**. You are conducting 45 tests on the same set of data. How many rejections will you have?
- Roughly 2 or 3 false rejections!
- So, multiple t-tests on the same set of data artificially inflate α

Summary: The Problems With Multiple t-Tests

- **Inefficient** - too many comparisons when we have even modest numbers of groups.
- **Imprecise** - cannot discern patterns or trends of differences in subsets of groups.
- **Inaccurate** - multiple tests on the same set of data artificially inflate α
- What is needed: a single test for the overall difference among all means

  e.g. ANOVA

LOGIC OF THE ANALYSIS OF VARIANCE

Logic of the Analysis of Variance

- Null hypothesis h_0: Population means equal
  $\mu_1 = \mu_2 = \mu_3 = \mu_4$
- Alternative hypothesis: h_1
  – Not all population means equal.

Logic

- Create a measure of variability among group means
  – $MS_{BetweenGroups}$ AKA $s^2_{BetweenGroups}$
- Create a measure of variability within groups
  – $MS_{WithinGroups}$ AKA $s^2_{WithinGroups}$

Logic

- $MS_{BetweenGroups} / MS_{WithinGroups}$
  - Ratio approximately 1 if null true
  - Ratio significantly larger than 1 if null false
  - "approximately 1" can actually be as high as 2 or 3, but not much higher

"So, why is it called analysis of variance anyway?"

- Aren't we interested in mean differences?
- Variance revisited
  - Basic variance formula

$$s^2 = \frac{\sum \left(X_i - \bar{X} \right)^2}{n-1} = \frac{SS}{df}$$

"Why is it called analysis of variance anyway?"

- What if data comes from groups?
  - We can have different sums of squares

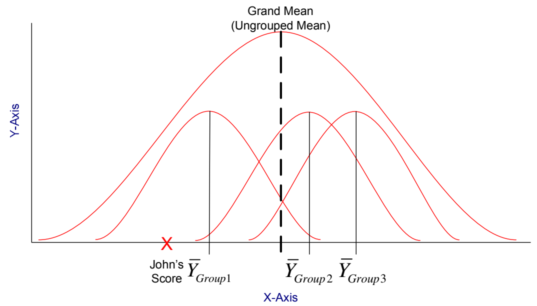
$$SS_1 = \sum \left(Y_i - \bar{Y}_{GM} \right)^2$$

$$SS_2 = \sum \left(Y_i - \bar{Y}_j \right)^2$$

$$SS_3 = \sum n_j \left(\bar{Y}_j - \bar{Y}_{GM} \right)^2$$

Where i represents the individual,
j represents the groups and
GM represent the ungrouped (grand) mean

Logic of ANOVA



Psy 320 - Cal State Northridge 13

CALCULATIONS

Psy 320 - Cal State Northridge 14

Sums of Squares

- The total variability can be partitioned into between groups variability and within groups variability.

$$\sum\left(Y_i-\bar{Y}_{GM}\right)^2=\sum n_j\left(\bar{Y}_j-\bar{Y}_{GM}\right)^2+\sum\left(Y_i-\bar{Y}_j\right)^2$$

$$SS_{Total}=SS_{BetweenGroups}+SS_{WithinGroups}$$

$$SS_T=SS_{BG}+SS_{WG}$$

$$SS_T=SS_{Effect}+SS_{Error}$$

Psy 320 - Cal State Northridge 15

Degrees of Freedom (*df*)

- Number of "observations" free to vary
  - $df_T = N - 1$
    - Variability of N observations
  - $df_{BG} = g - 1$
    - Variability of g means
  - $df_{WG} = g(n - 1)$ or $N - g$
    - n observations in each group = $n - 1$ *df* times g groups
  - $df_T = df_{BG} + df_{WG}$

Mean Square (i.e. Variance)

$$MS_T = s_T^2 = \frac{\sum\left(Y_i - \bar{Y}_{GM}\right)^2}{N - 1}$$

$$MS_{BG} = s_{BG}^2 = \frac{\sum n_j\left(\bar{Y}_j - \bar{Y}_{GM}\right)^2}{\# groups - 1}$$

$$MS_{WG} = s_{WG}^2 = \frac{\sum\left(Y_i - \bar{Y}_j\right)^2}{\# groups * (n - 1)}$$

F-test

- MS_{WG} contains **random sampling variation** among the participants
- MS_{BG} also contains random sampling variation but it can also contain **systematic (real) variation** between the groups (either naturally occurring or manipulated)

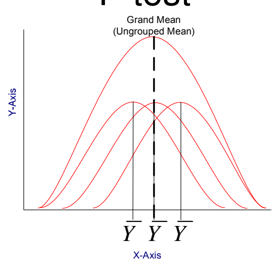
F-test

$$F_{Ratio} = \frac{Systematic\ BG\ Variance + Random\ BG\ Variance}{Random\ WS\ Variance}$$

- And if no "real" difference exists between groups

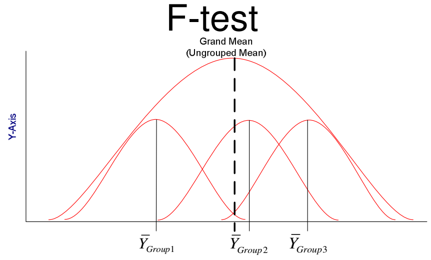
$$F_{Ratio} = \frac{Random\ BG\ Variance}{Random\ WS\ Variance} \approx 1$$

F-test



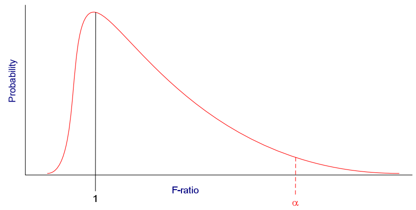
- The F-test is a ratio of the MS_{BG}/MS_{WG} and if the group differences are just random the ratio will equal 1 (e.g. random/random)

F-test



- If there are real differences between the groups the difference will be larger than 1 and we can calculate the probability and hypothesis test

F distribution



- There is a separate F distribution for every df like t but we need both df_{bg} and df_{wg} to calculate the F_{CV} from the F table D.3 for alpha = .05 and D.4 for alpha = .01

1-WAY BETWEEN GROUPS ANOVA EXAMPLE

Example

- A researcher is interested in knowing which brand of baby food babies prefer: Beechnut, Del Monte or Gerber.
- He randomly selects 15 babies and assigns each to try strained peas from one of the three brands
- Liking is measured by the number of spoonfuls the baby takes before getting “upset” (e.g. crying, screaming, throwing the food, etc.)

Hypothesis Testing

1. H_o: $\mu_{Beechnut} = \mu_{Del\ Monte} = \mu_{Gerber}$
2. At least 2 μs are different
3. $\alpha = .05$
4. More than 2 groups $\rightarrow$ ANOVA $\rightarrow$ F
5. For F_{cv} you need both $df_{BG} = 3 - 1 = 2$ and $df_{WG} = g\,(n - 1) = 3(5 - 1) = 12$
   Table D.3 $F_{cv}(2,12) = 3.89$, if $F_o > 3.89$ reject the null hypothesis

Step 6 – Calculate F-test

- Start with Sum of Squares (SS)
  - We need:
    - SS_T
    - SS_{BG}
    - SS_{WG}
- Then, use the SS and *df* to compute mean squares and *F*

Step 6 – Calculate F-test

`Brand	Baby	Spoonfuls (Y)	Group Means	$(Y_{ij} - \bar{Y}_{..})^2$	$(Y_{ij} - \bar{Y}_{.j})^2$	$n_j(\bar{Y}_{.j} - \bar{Y}_{..})^2$
Beechnut	1	3	4.6			
	2	4				
	3	4				
	4	4				
	5	8				
Del Monte	6	7	6	0.445	1	[5 * (6 - 6.333)²] = 0.555
	7	4		5.443	4	
	8	8		2.779	4	
	9	6		0.111	0	
	10	5		1.777	1	
Gerber	11	9	8.4	7.113	0.36	[5 * (8.4 - 6.333)²] = 21.36
	12	6		0.111	5.76	
	13	10		13.447	2.56	
	14	8		2.779	0.16	
	15	9		7.113	0.36	
Mean		6.333				
Sum				71.335	34.4	36.93

ANOVA summary table and Step 7

Source	SS	df	MS	F
BG	36.93			
WG	34.4			
Total	71.335			

- Remember
  - MS = SS/df
  - F = MS_{BG}/MS_{WG}
- Step 7 – Since _______ > 3.89, reject the null hypothesis

Conclusions

- The *F* for groups is significant.
  - We would obtain an *F* of this size, when H_0 true, less than 5% of the time.
  - The difference in group means cannot be explained by random error.
  - The baby food brands were rated differently by the sample of babies.

ALTERNATIVE COMPUTATIONAL APPROACH

Alternative Analysis – computational approach to SS

- Equations

$$SS_T = \sum Y^2 - \frac{\left(\sum Y\right)^2}{N} = \sum Y^2 - \frac{T^2}{N}$$

$$SS_{BG} = \frac{\sum\left(\sum a_j\right)^2}{n} - \frac{T^2}{N}$$

$$SS_{WG} = \sum Y^2 - \frac{\sum\left(\sum a_j\right)^2}{n}$$

  – Under each part of the equations, you divide by the number of scores it took to get the number in the numerator

Computational Approach Example

Brand	Baby	Spoonfuls (Y)
Beechnut	1	3
	2	4
	3	4
	4	4
	5	8
	Sum	23
Del Monte	6	7
	7	4
	8	8
	9	6
	10	5
	Sum	30
Gerber	11	9
	12	6
	13	10
	14	8
	15	9
	Sum	42
Total		95
Sum Y Squared		673

$$SS_T = \sum Y^2 - \frac{T^2}{N} = \underline{\qquad} - \frac{\underline{\quad}^2}{15} = 71.33$$

$$SS_{BG} = \frac{\sum\left(\sum a_j\right)^2}{n} - \frac{T^2}{N} = \frac{\underline{\quad}^2 + \underline{\quad}^2 + \underline{\quad}^2}{5} - \frac{\underline{\quad}^2}{15} =$$

$$= \underline{\qquad} - \underline{\qquad} = 36.93$$

$$SS_{WG} = \sum Y^2 - \frac{\sum\left(\sum a_j\right)^2}{n} =$$

$$= \underline{\qquad} - \frac{\underline{\quad}^2 + \underline{\quad}^2 + \underline{\quad}^2}{5} = \underline{\qquad} - \underline{\qquad} = 34.4$$

Note: You get the same SS using this method

Unequal Sample Sizes

- With one-way, no particular problem
  – Multiply mean deviations by appropriate n_i as you go
  – The problem is more complex with more complex designs, as shown in next chapter.
  – Equal samples only simplify the equation because when $n_1 = n_2 = \ldots = n_g$

$$\sum n_j\left(\bar{Y}_j - \bar{Y}_{GM}\right)^2 = n\sum\left(\bar{Y}_j - \bar{Y}_{GM}\right)^2$$

MULTIPLE COMPARISONS

Multiple Comparisons

- Significant *F* only shows that not all groups are equal
  - We want to know what groups are different.
- Such procedures are designed to control familywise error rate.
  - Familywise error rate defined
  - Contrast with per comparison error rate

More on Error Rates

- Most tests reduce significance level (α) for each *t* test.
- The more tests we run the more likely we are to make Type I error.
  - Good reason to hold down number of tests

Tukey
Honestly Significant Difference

- The honestly significant difference (HSD) controls for all possible pairwise comparisons
- The Critical Difference (CD) computed using the HSD approach

Tukey
Honestly Significant Difference

$$CD = q\sqrt{\frac{MS_{error}}{n_A}}$$

- where q is the studentized range statistic (table), MS_{error} is from the ANOVA and n_A is equal n for both groups

$$CD = q\sqrt{\left[MS_{error}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)\right]/2}$$

- for the unequal n case

Tukey

- Comparing Beechnut and Gerber
  - To compute the CD value we need to first find the value for q
  - q depends on alpha, the total number of groups and the DF for error.
  - We have 3 total groups, alpha = .05 and the DF for error is 12
  - q = 3.77

Tukey

- With a q of 3.77 just plug it in to the formula

$$CD = q\sqrt{\frac{MS_{error}}{n_A}} = 3.77\sqrt{\frac{2.867}{5}} = 2.86$$

- This give us the minimum mean difference
- The difference between gerber and beechnut is 3.8, the difference is significant

Fisher's LSD Procedure

- Requires significant overall *F,* or no tests
- Run standard *t* tests between pairs of groups.
  - Often we replace s^2_{pooled} with MS_{error} from overall analysis
    - It is really just a pooled error term, but with more degrees of freedom (pooled across all treatment groups)

Fisher's LSD Procedure

- Comparing Beechnut and Gerber

$$s^*_{\bar{X}_{Beechnut}-\bar{X}_{Gerber}} = \sqrt{\frac{MS_{WG}}{n_{Beechnut}} + \frac{MS_{WG}}{n_{Gerber}}} = \sqrt{\frac{2.867}{5} + \frac{2.867}{5}}$$

$$t = \frac{\bar{X}_{Beechnut} - \bar{X}_{Gerber}}{s^*_{\bar{X}_{Beechnut}-\bar{X}_{Gerber}}} = \frac{8.4 - 4.6}{1.071} = 3.55$$

- $t_{cv}(5+5-2=8)_{\alpha = .05} = 1.860$
- Since 3.55 > 1.860, the 2 groups are significantly different.

Bonferroni *t* Test

- Run *t* tests between pairs of groups, as usual
  - Hold down number of *t* tests
  - Reject if *t* exceeds critical value in Bonferroni table
- Works by using a more strict value of α for each comparison

Bonferroni *t*

- Critical value of α for each test set at $.05/c$, where c = number of tests run
  - Assuming familywise $\alpha = .05$
  - e. g. with 3 tests, each *t* must be significant at $.05/3 = .0167$ level.
- With computer printout, just make sure calculated probability $< .05/c$

Assumptions for Analysis of Variance

- Assume:
  - Observations normally distributed within each *population*
  - Population variances are equal
    - Homogeneity of variance or homoscedasticity
  - Observations are independent

ASSUMPTIONS

Assumptions

- Analysis of variance is generally robust to first two
  - A robust test is one that is not greatly affected by violations of assumptions.

EFFECT SIZE

Magnitude of Effect

- Eta squared (η^2)
  - Easy to calculate
  - Somewhat biased on the high side
  - Formula

$$\eta^2 = \frac{SS_{BG}}{SS_{Total}}$$

  - Percent of variation in the data that can be attributed to treatment differences

Magnitude of Effect

- Omega squared (ω^2)
  - Much less biased than η^2
  - Not as intuitive
  - We adjust both numerator and denominator with MS_{error}
  - Formula

$$\omega^2 = \frac{SS_{BG} - (k-1)MS_{WG}}{SS_T + MS_{WG}}$$

η^2 and ω^2 for Baby Food

$$\eta^2 = \frac{SS_{BG}}{SS_T} = \frac{36.93}{71.335} = .518$$

$$\omega^2 = \frac{SS_{BG} - (k-1)MS_{WG}}{SS_T + MS_{WG}} = \frac{36.93 - 2(2.867)}{71.335 + 2.867} = .420$$

- η^2 = .52: 52% of variability in preference can be accounted for by brand of baby food
- ω^2 = .42: This is a less biased estimate, and note that it is 20% smaller.

Other Measures of Effect Size

- We can use the same kinds of measures we talked about with *t* tests (e.g. *d* and *d-hat*)
- Usually makes most sense to talk about 2 groups at a time or effect size between the largest and smallest groups, etc.
- And there are methods for converting η^2 to *d* and vice versa