Advanced Experimental Design

Topic 6
Chapters 11 & 12
Oneway Analysis of Variance
Multiple Comparisons

Agenda

- Analysis of Variance (with computational illustration)
- Comparison of F and t, connection
- Magnitude of effect (η^2, ω^2)
- Power
- Multiple Comparisons

Multiple groups (>2)

- What if we have a situation where we have more than two means that we wish to compare with one another?
- How would we approach assessing mean differences?
- Situations this might be an issue?

Type I error rates

- could use multiple t-tests, but why might that not be a good idea?
- inflating the probability of making a type I error
- Error rate per experiment
  - sum the p values for all of the tests
  - expected rate of type 1 errors = 6 x 0.05 = 0.30.

4

Three or more groups

- Common in outcomes research, where we may wish to evaluate the relative efficacy of a number of different treatments
- Few studies we might be interested in doing have only two groups
- Oneway ANOVA allows comparisons among two or more sample means
- Examples of multiple group studies

5

ANOVA Structural Model

$$X_{ij} = \mu + \tau_j + \varepsilon_{ij} \qquad \tau_j = (\mu_j - \mu)$$

- i=# of people, j=# of groups, X_{ij} is the ith person in the jth group
- μ = grand mean
- τ = adjustment reflecting mean of jth group
- ε = "uniqueness"

6

ANOVA: ANalysis Of VAriance

- Assumptions - with a large enough sample can relax the assumptions
  - normal distribution of dependent variable
  - equal variances of groups
  - Independence of observations
- Considered to be ROBUST
  - especially insensitive to violations of normality when equal numbers in each group and the distributions are about the same shape

7

ANOVA conceptualization

- Variance in a set of observations can be partitioned into within groups and between groups variance
- F statistic: basically the ratio of the between groups variance to within groups variance
- If no differences between the groups means, the between and within groups variance will be about the same, resulting in an f ratio of around one

8

ANOVA degrees of freedom

- F statistic has two degrees of freedom:
- between groups df which = n_{groups}-1
- within groups df which = n - n_{groups}
- $F(2, 45) = 6.16, p < .01$

ANOVA

CORRECT NUMBER OF STEERING CORRECTIONS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1016.667	2	508.333	6.164	.004
Within Groups	3711.250	45	82.472		
Total	4727.917	47			

9

Steps involved in hypothesis testing w/ANOVA

1. Identify H_0 and H_1
2. Tentatively assume H_0
3. Choose a sampling distribution
4. Obtain data, calculate sample statistic
5. Compare F_{obs} with F_{crit}
6. Reach conclusion re: status of H_0
7. Tell the story in terms of constructs

10

ANOVA example

- hired by a large corporation to implement a “wellness program.”
- encouraging good dietary habits and exercise will decrease sick days, productivity will increase and have reduced health insurance benefit costs due to lower utilization of medical services
- best way to present the program, in terms of scheduling?

11

ANOVA example (cont.)

- freedom to schedule any time you wish
- attendance will be mandatory
- want to find out what people will prefer
- more satisfied with the program = receptive and attentive to the information presented
- group of 30 employees randomly assigned to each of the three conditions
- at the conclusion of the program (identical for all 3 groups) assess satisfaction with the program

12

The data

Condition	Satisfaction	Cond.	Satis.	Cond.	Satis.
days	5	nights	5	Saturday	5
days	5	nights	4	Saturday	5
days	5	nights	4	Saturday	5
days	5	nights	3	Saturday	4
days	4	nights	3	Saturday	4
days	4	nights	2	Saturday	4
days	4	nights	2	Saturday	3
days	4	nights	2	Saturday	3
days	3	nights	1	Saturday	3
days	2	nights	0	Saturday	2

Calculating F

$$\frac{\frac{1681+676+1444}{10}-\frac{105^2}{30}}{2(df)}=6.3$$

	X days	X^2	X nights	X^2	X saturday	X^2
	5	25	5	25	5	25
	5	25	4	16	5	25
	5	25	4	16	5	25
	5	25	3	9	4	16
	4	16	3	9	4	16
	4	16	2	4	4	16
	4	16	2	4	3	9
	4	16	2	4	3	9
	3	9	1	1	3	9
	2	4	0	0	2	4
ΣX	41	177	26	88	38	154
$(\Sigma X)^2$	1681		676		1444	
Variance	0.99		2.27		1.07	
Mean	4.1		2.6		3.8	
			Calculate MSb	X 4.1	X^2 16.81	
				2.6	6.76	
Average variance				3.8	14.44	
MS_w	1.44		ΣX	10.5		
			$(\Sigma X)^2$	110.25		
$f=MS_b/MS_w=$	4.37		ΣX^2	38.01		
$f_{crit}(2,27)$ $p<.05=$	3.35		σ^2_{means}	0.63		
$f_{crit}(2,27)$ $p<.01=$	5.49		MS_b	6.3		

Components of F statistic

- MS_w is often refered to as MS_{error}
- MS_b often referred to as $MS_{treatment}$
- Variance of means x n/group = MS_b
- two different estimates of the population variance (σ^2)

$$SS_{treat}=\frac{\sum \bar{X}_j}{n}-\frac{\left(\sum X\right)^2}{N} \qquad SS_{Total}=\sum X^2-\frac{\left(\sum X\right)^2}{N}$$

Components of F statistic (cont.)

- If this variance, as estimated by the variance of the means is large relative to the pooled variances (the second estimate of the population variance) we reject the null hypothesis that there is no difference between treatment means

16

Conclusions from ANOVA

- we can conclude that the three schedules do not result in equal satisfaction
- unsure about specifically where these significant differences lie
- probably a safe bet that the days mean is higher than the nights mean
- is days significantly higher than Saturday?
- is Saturday significantly higher than nights?
- Multiple comparison test will tell us this

17

Relation between t and F

- t-test really just a special case of ANOVA
  - numerator (between groups) df = 1
  - F statistic is equal to the t statistic squared
- we can calculate the f statistic for the earlier cholesterol example
- advantage to using the t statistic over ANOVA
  - t-gives directionality, that is, it can be positive or negative
  - indicates which mean is higher (obviously this is a pretty minor point)

18

Group	Chol		No Exer.	x^2	Exercise	x^2	
No Exercise	260		260	67600	240	57600	
No Exercise	255		255	65025	260	67600	
No Exercise	270		270	72900	255	65025	
No Exercise	300		300	90000	235	55225	
No Exercise	315		315	99225	270	72900	
No Exercise	280		280	78400	230	52900	
No Exercise	265		265	70225	285	81225	
No Exercise	295		295	87025	225	50625	
No Exercise	250		250	62500	200	40000	
No Exercise	305		305	93025	265	70225	
Exercise	240	ΣX	2795	785925	2465	613325	
Exercise	260	$(\Sigma X)^2$	7812025		6076225		
Exercise	255	ΣX^2	785925		613325		
Exercise	235	variance	524.72		633.61		
Exercise	270	M	279.5		246.5		
Exercise	230	Σvariances	1158.33				
Exercise	285	MS_w	579.17				
Exercise	225						
Exercise	200					Calculate MS_b	
Exercise	265	$f=MS_b/MS_w$	9.40			X	X^2
		$df1=n_{groups}-1$	1			279.5	78120.25
		$df2=n_{groups}*n_{pergroup}-1$	18			246.5	60762.25
M1-M2	33	$f_{crit}(1,18)$, .05=	4.41		ΣX	526	
$S.E._{diff}$	10.76	t^2=	9.40		$(\Sigma X)^2$	276676	
$t=M1-M2/S.E._{diff}$		square root of f=	3.067		ΣX^2	138882.5	
t=	3.07				var	544.5	
$t_{crit}(18)$, .05, 2tail=	2.10	t_{crit}^2=	4.41		MS_b	5445	

Magnitude of experimental effect

- SPSS gives us eta squared (η^2)
- η^2 is analogous to a squared correlation coefficient

  $\eta^2 = \dfrac{SS_{treat}}{SS_{total}}$ or $\dfrac{SS_{tot} - SS_{error}}{SS_{total}}$
- estimate of proportion of variance attributable to treatment/group membership
- η^2 is generally biased upward

20

Magnitude of experimental effect

- Omega squared (ω^2) alternative effect size index
- not biased upward
- SPSS does not calculate ω^2 automatically
- $\omega^2 = \dfrac{SS_{treatment} - (k-1)MS_{error}}{SS_{total} + MS_{error}}$

21

Power Analysis

- SPSS gives power after the fact
- g-power or charts to estimate before hand
- Computations in text for Cohen's f, Howell refers to as phi prime Φ'
- Lets look at g-power

22

Multiple Comparison Procedures

- A priori vs. Post hoc
- Linear contrasts/coefficients
- SPSS allows contrast coding
- MC procedures adjust α (or critical value of the test statistic) to control FW error rate
- Family-wise error rate
  - $\alpha = 1 - (1 - \alpha')^c$

23

A-Priori Comparisons

- t-tests
  - Bonferroni adjustment
  - Appendix t'
- Dunn-Sidák test
- Holm / Larzelere & Mulaik tests
  - Very similar
  - Holm for ANOVA / Larzelere & Mulaik for correlation matrices
  - Holm uses adjusted t, L & M adjust p directly

24

Post Hoc Tests

- Exhaustively exploring which groups differ from which
- No strong theory

- Fisher's LSD
  - significant overall F required
  - Multiple t's after significant F
  - > three groups, FW error rate can increase

25

Post Hoc Tests

- Studentized range statistic - q tables
- Critical value adjusted based on number of steps between means
- Furthest means apart most likely to be type I error
- Forms the basis for several other tests

26

Post Hoc Tests

- Student Newman Keuls (SNK)
- Sorts means into subsets
  - do not differ among themselves
  - different from means in other subsets
- One of the less conservative tests

- Tukey's HSD (honestly significant difference)
  - Similar to SNK except, uses maximum N of steps for all comparisons
  - Controls the FW error rate
  - loss of power

27

Post Hoc Tests

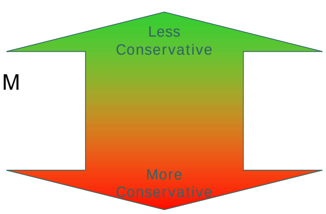
- Ryan Procedure (Ryan/Einot/Gabriel/Welsch Q: REGWQ)
  - Controls FW error at α
  - greater power than the Tukey HSD.
- Scheffe' test
  - Very conservative, least powerful
  - Controls FW for all possible pairwise and linear contrasts
- Dunnet's test
  - Comparing one treatment with several other treatments

28

Overall picture

- Ryan (REGWQ), SNK near middle on liberal-conservative continuum.

- Multiple t's
- LSD
- Holm / L & M
- SNK
- Ryan
- HSD
- Scheffe'



29