

# Basic Statistics

## Chapter 10: One-Way Analysis of Variance (ANOVA)

1

---

---

---

---

---

---

---

---



## Topical Outline

- Purpose of ANOVA
- Steps
- Assumptions
- Logic of ANOVA
- An example
- Group comparisons
- Effect size index (f)
- Presenting results (summary tables)
- Relationship between t and F

2

---

---

---


---

---

---

---

---



## Purpose of ANOVA

- more than 2 treatment levels
- these designs differ from those in t chapter by # of levels of the IV
- with a three level IV  
 $H_0: \mu_1 = \mu_2 = \mu_3$   
 $H_1: \text{not}(\mu_1 = \mu_2 = \mu_3)$

3

---

---

---

---

---

---

---

---



## Comparing Multiple Means

- o multiple t-tests
- o is that a good idea?
- o probability of making a type I error rises
- o sum the p values for all tests we run
- o 4 groups= 6 tests,  $p(\text{at least 1 type I error}) = 6 \times 0.05 = 0.30$
- o common in outcomes research
  - evaluate the relative efficacy of treatments
  - clinical trials, compare new drug with "old-standard" and a placebo control

4

---

---

---

---

---

---

---

---



## ANOVA ANALYSIS OF VARIANCE

- o One-way ANOVA allows comparisons among two or more sample means
- o Steps in hypothesis testing with ANOVA
- o 1. identify  $H_0$  and  $H_1$
- o 2. assume tentatively the equality hypothesis is true
  - deviations among the sample means within the bounds expected due to sampling error
- o 3. choose sampling distribution
  - t distribution worked for two sample means
  - will use the F distribution

5

---

---

---

---

---

---

---

---



## ANOVA hypothesis testing

- o 4. obtain data from the populations you are interested in and calculate an F statistic
- o 5. compare F observed to F critical
- o 6. reach a conclusion about the status of the null
  - if  $p \leq .05$  reject the null
  - if  $p > .05$  retain null and alternative as possibilities.
- o 7. tell the story in terms of the constructs being investigated
- o same process we followed with t test

6

---

---

---

---

---

---

---

---



## ANOVA assumptions

- o normal distribution of dependent variable
- o equal variances within groups - independent variable
- o random sample/random assignment
- o considered to be ROBUST, especially when equal numbers in groups and distributions are about the same shape

7

---

---

---

---

---

---

---

---



## ANOVA logic / DF

- o variance in a set of observations can be partitioned into within groups and between groups variance
- o F statistic - ratio of between groups variance and within groups variance
- o if no differences between group means, between and within groups variance will be about the same, resulting in an f ratio of around one
- o F statistic has two degrees of freedom
- o first is the between groups df which =  $n_{groups} - 1$
- o second is the within groups df which =  $n - n_{groups}$

8

---

---

---

---

---

---

---

---



## ANOVA example

- o exercise program example
- o enter data into SPSS
- o write paragraph summarizing findings
- o hired by a large corporation to implement a "wellness program."
- o 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
- o best way to present the program, in terms of scheduling?

9

---

---

---

---

---

---

---

---

## ANOVA example (cont.)

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

10

---

---

---

---

---

---

---

---

---

---

## 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

11

---

---

---

---

---

---

---

---

---

---

## Calculating F

	$X_{\text{days}}$	$X_{\text{days}}^2$	$X_{\text{nights}}$	$X_{\text{nights}}^2$	$X_{\text{saturday}}$	$X_{\text{saturday}}^2$
	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
$\Sigma X$	41	177	26	88	38	154
$\Sigma X^2$	1681		676		1444	
$s^2$	0.99		2.27		1.07	
$\bar{X}$	4.1		2.6		3.8	

	Calculate $MS_{\text{error}}$	$X$	$X^2$
		4.1	16.81
		2.6	6.76
Average variance		3.8	14.44
$MS_{\text{error}}$	1.44	$\Sigma X$	10.5
		$(\Sigma X)^2$	110.25
$F = MS_{\text{between}} / MS_{\text{error}}$	4.37	$\Sigma X^2$	38.01
$F_{\text{crit}}(2,27) p < .05 =$	3.35	$s_p^2$	0.63
$F_{\text{crit}}(2,27) p < .01 =$	5.49	$MS_{\text{between}}$	6.3
		n/group	$= MS_{\text{error}}$

12

---

---

---

---

---

---

---

---

---

---



## Components of F statistic

- two different estimates of the population variance ( $\sigma^2$ )
- the numerator of the F statistic represents a reasonable estimate of  $\sigma^2$  only when the null is true
- when the null is not true the estimate is too large

13

---

---

---

---

---

---

---

---



## Components of F statistic (cont.)

- If this variance, as estimated by the variance of the means is large relative to the average of the variances we reject the null hypothesis that there is no difference between treatment means
- expected value of F is 1.00 if null is true

14

---

---

---

---

---

---

---

---



## 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?

15

---

---

---

---

---

---

---

---

● ● ● | Multiple comparisons (HSD)

- Tukey's HSD (Honestly Significant Difference) test
- a priori or post hoc
- a priori have selected some small number of comparisons
- post hoc - comparing all of the group means
- lots of different test options

16

---

---

---

---

---

---

---

---

● ● ● | Tukey's HSD

$$HSD = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}}} \text{ where } \rightarrow s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}}$$

- Look up critical value in Table G, pg. 394, column is # of groups, row is error (within groups) df

17

---

---

---

---

---

---

---

---

● ● ● | Tukey's HSD (cont.)

- Night (2.6) and Saturday (3.8) means first (intermediate difference) gives  $1.2/\sqrt{1.44/10}$ ,  $1.2/.379=3.16$ ,  $HSD(.05)(3,27)=3.53$
- not statistically significant so smaller difference will not be significant either
- still need to test the Days - Nights
- $4.1-2.6/.379 = 3.96$

18

---

---

---

---

---

---

---

---



## Relation between t and f

- o t-test really just a special case of ANOVA
  - numerator (between groups)  $df = 1$
  - f statistic is equal to the t statistic squared
- o we can calculate the f and t statistics for the same data set if two groups
- o 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)

19

---

---

---

---

---

---

---

---

---

---

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	$\Sigma X$	2795	785925	2465	613325
Exercise	260	$(\Sigma X)^2$	7812025	6076225		
Exercise	255	$\Sigma X^2$	785925	613325		
Exercise	235	variance	524.72	633.61		
Exercise	270	M	279.5	246.5		
Exercise	230	$\Sigma$ variances	1158.33			
Exercise	285	$MS_w$	579.17			
Exercise	225					
Exercise	200					
Exercise	265	$f=MS_b/MS_w$	9.40	Calculate $MS_b$		
		$df1=n_{group}-1$	1	X		
		$df2=n_{group} * n_{pergroup}-1$	18	279.5	78120.25	
M1-M2	33	$f_{crit}(1,18), .05=$	4.41	246.5	60762.25	
S.E. <sub>-diff</sub>	10.76	$t^2=$	9.40	$\Sigma X$	526	
$t=M1-M2/S.E._{diff}$		square root of f=	3.067	$(\Sigma X)^2$	276676	
$t=$	3.07			$\Sigma X^2$	138882.5	
$t_{crit}(18), .05, 2tail=$	2.10	$t_{crit}^2=$	4.41	var	544.5	
				$MS_b$	5445	

---

---

---

---

---

---

---

---

---

---



## ANOVA problem

- o Example from study guide: Pg. 98 #2
- o Compute by hand
- o Run in SPSS

21

---

---

---

---

---

---

---

---

---

---



## Statistical vs. practical significance (revisit)

Sample Size	Sample Mean	Population Mean	$p$
4	110.0	100.0	0.05
25	104.0	100.0	0.05
64	102.5	100.0	0.05
400	101.0	100.0	0.05
2,500	100.4	100.0	0.05
10,000	100.2	100.0	0.05

22

---

---

---

---

---

---

---

---