One-Way Between-Groups ANOVA

PSYC 300B - Lecture 1
Dr. J. Nicol

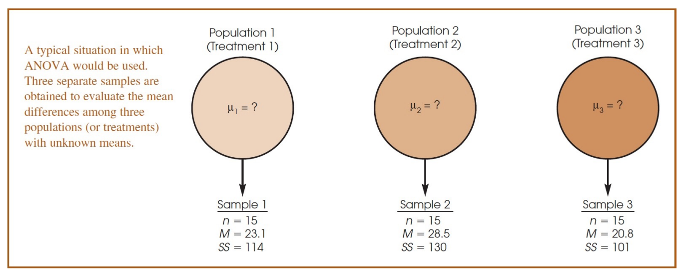
Lecture Introduction

- *In this lecture we will discuss the general logic and basic formulas for the hypothesis testing procedure known as **analysis of variance (ANOVA)***
- *The purpose of ANOVA is much the same as the t-tests—specifically the goal is to determine whether the mean differences that are obtained for sample data are sufficiently large to justify a conclusion that there are mean differences between the populations from which the samples were obtained*

One-Way Between-Groups ANOVA

- ANOVA allows us to draw inferences about means based two estimates of population variance
- The most used statistical technique in psychological research (Howell, 2017)
- Popularity and usefulness is attributable to:
  - Has no restrictions on the number of means that can be compared
  - Allows us to deal with two or more IVs simultaneously, asking not only about the individual effects of each variable separately but also about the interaction effect between those variables

The purpose of an ANOVA is to determine whether the mean differences in the sample data are large enough to justify a conclusion that there are mean differences between the populations from which the samples were obtained



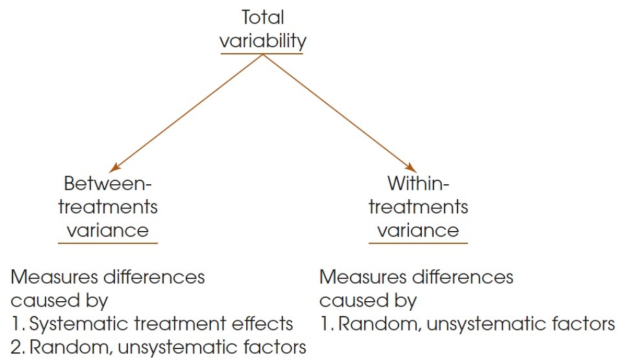
One-Way Between-Groups ANOVA

- Conducted to determine if there is a statistically significant difference between the means of an independent variable with more than two groups (and the groups are comprised of ***different*** participants)
- When significant it indicates the difference between ***at least one pair of group means*** is large enough that it is unlikely to be the result of sampling error
- Post-hoc pairwise comparisons of group means are performed to determine which pair(s) are different

Theory of ANOVA

- The *F*-statistic is a ratio of two sample variances, called mean squares (*MS*)
- The numerator reflects the variance in the group means (i.e., between groups) and the denominator reflects the average variance in groups (i.e., within groups)
- When we reject H_0 (i.e., there is a treatment effect), the variance between groups is significantly bigger than the variance within groups

The F-statistic in ANOVA is a measure of the ratio of systematic to unsystematic variation



The structure and sequence of calculating a one-way between groups ANOVA

The final goal for the ANOVA is an *F*-ratio	$F = \frac{\text{Variance between treatments}}{\text{Variance within treatments}}$	
Each variance in the *F*-ratio is computed as *SS/df*	Variance between treatments $= \frac{SS \text{ between}}{df \text{ between}}$	Variance within treatments $= \frac{SS \text{ within}}{df \text{ within}}$
To obtain each of the *SS* and *df* values, the total variability is analyzed into the two components	*SS* total → *SS* between, *SS* within	*df* total → *df* between, *df* within

Calculating Sum of Squares

$$SS_{TOTAL} = \sum (X - M_{GRAND})^2$$

$$SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{GRAND})^2$$

$$SS_{WITHIN} = SS_{TOTAL} - SS_{BETWEEN}$$

Calculating Degrees of Freedom

$$df_{TOTAL} = N - 1$$

$$df_{BETWEEN} = k - 1$$

$$df_{WITHIN} = df_{TOTAL} - df_{BETWEEN}$$

Calculating Mean Squares

$$MS_{BETWEEN} = \frac{SS_{BETWEEN}}{df_{BETWEEN}}$$

$$MS_{WITHIN} = \frac{SS_{WITHIN}}{df_{WITHIN}}$$

Calculating the *F*-Ratio

$$F = \frac{MS_{BETWEEN}}{MS_{WITHIN}}$$

$$F_{CRITICAL} = (df_{BETWEEN}, df_{WITHIN})$$

Assumptions of ANOVA

- Scores within groups are normally distributed
- Homogeneity of variance across groups
- ***NOTE: IF MET THEN THE POPULATIONS (GROUPS) HAVE THE SAME SHAPE AND DISPERSION—SO CAN ONLY DIFFER ON THEIR MEANS***
- Independence of errors/observations

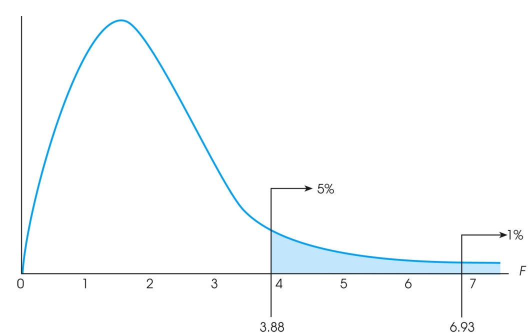
Assumptions of ANOVA

- ANOVA is robust with respect to violations of the assumptions of normality and homogeneity of variance (Howell, 2017)
- ANOVA is robust when the homogeneity of variance assumption is violated when sample sizes are equal, but when sample sizes are unequal the Type I error is not controlled (Wilcox, 2012)

Hypothesis Testing

- ANOVA is a one-tailed test of a non-directional H_0, it is an omnibus test that examines the overall difference between groups
  - H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_N$
  - H_1: μ's are not all equal
- A significant result (indicates that at least one pair of group means is significantly different, but it does not tell indicate which pair(s) of means are different

Variance cannot be less than zero so the F-statistic is always positive, and the distribution accumulates near 1.00 because when H_0 is true the variance between and within groups is about equal



Post-Hoc Testing

- ***Post-hoc tests*** are pairwise comparisons that compare different combinations of group means
- A problem with comparing group means is that unrestricted use of these comparisons leads to an excessively high probability of a Type I error
- Don't compare two means unless it is actually meaningful for what you are doing, and never run a comparison just because you can (Howell, 2017)

Post-Hoc Testing

- An ANOVA is based on a single hypothesis test so the Type I error is not inflated (i.e., it remains at α) regardless of how many means are being compared
- Each time a hypothesis test is conducted on a set of data there is a risk of committing a Type I error (α)
- For example, three group means would require three *t*-tests which would increase the ***family-wise error rate*** to $\alpha = 0.15$ (i.e., 3×0.05)

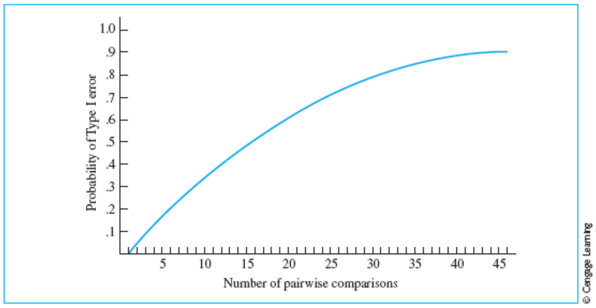


Figure 16.4
Probability of a Type I error as a function of the number of pairwise comparisons where $\alpha = .05$ for any one comparison.

Post-Hoc Testing

- In an attempt to control the likelihood of a Type I error, statisticians have developed a large number of procedures for comparing group means (see Howell, 2012)
- These techniques provide reasonable control of the probability of making a Type I error and are applicable to most multiple comparison problems
  - *Fisher's Least Significant Difference (LSD) test*
  - *Tukey's Honestly Significant Difference (HSD) test*
  - *Bonferroni correction*

Fisher's *LSD* Test

- One of the most ***liberal*** post-hoc tests
- Also referred to as a ***protected t-test*** because it requires the *F*-statistic be significant
- If the overall *F*-statistic is significant, you can proceed to make any (or all) pairwise comparisons between individual means using this modified *t*-test
- With three groups, Fisher's *LSD* test guarantees that the probability of a making at least one Type I error does not exceed $\alpha = .05$ (Howell, 2017)

The requirement of a significant overall F-statistic before running multiple comparisons (which is where the protection comes from) is an effective way of controlling family-wise error rates when there are only a few groups (Howell, 2017)

If the family-wise error rate goes up to $\alpha = .10$, it is not the end of the world. This advice is in line with several of the changes that have taken place in behavioural sciences over the last 30 years (Howell, 2014)

Tukey's *HSD* Test

- Computes a value (i.e., the *HSD*) that determines the minimum difference between two group means that is significant — a pair of means are significantly different when their difference exceeds the *HSD* value
- ***Requires that the number of participants in each group (n) is equal across all groups***
- The problem is that it involves the comparison of every mean with every other mean and adjusts the probability values accordingly, even though many comparisons aren't of interest (Howell, 2014)

Tukey's *HSD* Test

- $HSD = q\sqrt{(MS_{WITHIN}/n)}$
- q: look up in q-Table (based on k and df_{WITHIN})
- MS_{WITHIN}: taken from ANOVA summary table
- n = number of scores in each group

Post-Hoc Tests

$$Fisher's\ LSD:\ t = \frac{M_1 - M_2}{\sqrt{MS_{WITHIN}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Tukey's\ HSD = q\sqrt{\frac{MS_{WITHIN}}{n}}$$

$$Bonferroni\ correction = \frac{\alpha}{\#\ of\ comparisons}$$

Effect Size Measures

$$\eta^2 = \frac{SS_{BETWEEN}}{SS_{TOTAL}}$$

$$\omega^2 = \frac{SS_{BETWEEN} - (k-1)MS_{WITHIN}}{SS_{TOTAL} + MS_{WITHIN}}$$

$$d = \frac{M_1 - M_2}{\sqrt{MS_{WITHIN}}}$$

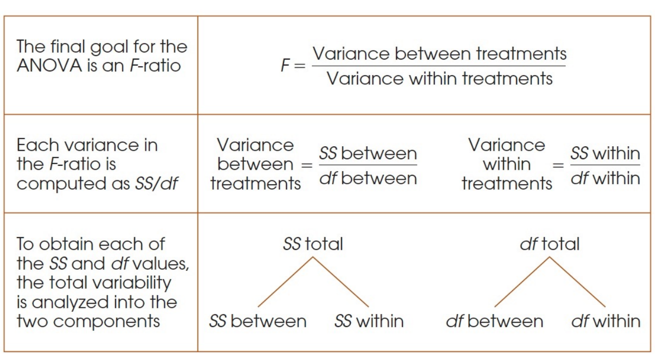
Effect Size Measures

- Eta-squared (η^2) reflects the proportion of variability in the DV that is related to variability in the IV (i.e., the group means), but it overestimates the true effect size in the population
- Omega-squared (ω^2) is a more accurate effect size measure, because it is basically an unbiased estimate of *r*
- For ω^2: 0.01, 0.06, and 0.14 represent small, medium, and large effect sizes, respectively (Kirk, 1996)

Conduct a hypothesis test (α = .05) to determine if there is an effect of perceived difficulty on cognitive performance. Scores indicate the number of questions solved. Perform post-hoc tests and report effect sizes

Easy	Medium	Difficult
9	4	1
12	6	3
4	8	4
8	2	5
7	10	2

The structure and sequence of calculating a one-way between groups ANOVA



Easy	Medium	Difficult	Total
9	4	1	
12	6	3	
4	8	4	
8	2	5	
7	10	2	$k = 3$
$n_E = 5$	$n_M = 5$	$n_D = 5$	$N = 15$
$M_E = 8$	$M_M = 6$	$M_D = 3$	$M_{GRAND} = 5.67$

- H_0: $\mu_E = \mu_M = \mu_D$
- H_1: μ's are not all equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 147.33$
- $SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{GRAND})^2$

  $= 5(8-5.67)^2 + 5(6-5.67)^2 + 5(3-5.67)^2 = 63.33$
- $SS_{WITHIN} = SS_{TOTAL} - SS_{BETWEEN} = 147.33 - 63.33 = 84$
- $df_{TOTAL} = N - 1 = 14$
- $df_{BETWEEN} = k - 1 = 2$
- $df_{WITHIN} = df_{TOTAL} - df_{BETWEEN} = 14 - 2 = 12$

- $MS_{BETWEEN} = SS_{BETWEEN} / df_{BETWEEN} = 63.33/2 = 31.67$
- $MS_{WITHIN} = SS_{WITHIN} / df_{WITHIN} = 84/12 = 7$
- $F = MS_{BETWEEN} / MS_{WITHIN} = 31.67/7 = 4.52$
- $F\,(2, 12)_{CRITICAL} = 3.88$
- $\omega^2 = 63.33\text{-}(2)(7) / (147.33 + 7) = 0.32$
- Reject H_0, perceived difficulty has a significant effect on cognitive performance, $F\,(2, 12) = 4.52$, $p < .05$, $\omega^2 = 0.32$

Source	*SS*	*df*	*MS*	*F*
Between	63.33	2	31.67	4.52
Within	84	12	7	
Total	147.33	14		

- $k = 3$
- $df_{WITHIN} = 12$
- $q = 3.77$
- $HSD = q\sqrt{(MS_{WITHIN}/n)}$
- $HSD = 3.77\sqrt{(7/5)} = 4.46$
- Pairwise comparisons:
  - $M_E - M_M = 8 - 6 = 2$ (*ns*)
  - $M_E - M_D = 8 - 3 = 5$ ($p < .05$); $d = 5/\sqrt{7} = 1.89$
  - $M_M - M_D = 6 - 3 = 3$ (*ns*)
- Participants in the perceived easy ($M = 8.0$) condition solved significantly more problems than participants in the perceived difficult ($M = 3.0$) condition ($p < .05$; $d = 1.89$). No other pairwise comparisons were significant

Conduct a hypothesis test (α = .01) to determine if there is an effect of strenuous physical exercise on the onset of puberty in girls. Scores indicate age at menarche. Perform post-hoc tests and report the effect sizes.

Control	Athlete	Musician
12	14	13
11	12	12
11	14	13
13	16	11
11	15	
12	13	
11		

Control	Athlete	Musician	
12	14	13	
11	12	12	
11	14	13	
13	16	11	
11	15		
12	13		
11			$k = 3$
$n_C = 7$	$n_A = 6$	$n_M = 4$	$N = 17$
$M_C = 11.57$	$M_A = 14.0$	$M_M = 12.25$	$M_G = 12.59$

- H_0: $\mu_{CONTROL} = \mu_{ATHELTE} = \mu_{MUSICIAN}$
- H_1: μ's are not all equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 36.12$
- $SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{GRAND})^2$

  $= 7(11.57\text{-}12.59)^2 + 6(14.0\text{-}12.59)^2 + 4(12.25\text{-}12.59)^2 = 19.65$
- $SS_{WITHIN} = SS_{TOTAL} - SS_{BETWEEN} = 36.12 - 19.65 = 16.46$
- $df_{TOTAL} = N - 1 = 16$
- $df_{BETWEEN} = k - 1 = 2$
- $df_{WITHIN} = df_{TOTAL} - df_{BETWEEN} = 16 - 2 = 14$

- $MS_{BETWEEN} = SS_{BETWEEN} / df_{BETWEEN} = 19.65/2 = 9.83$
- $MS_{WITHIN} = SS_{WITHIN} / df_{WITHIN} = 16.46/14 = 1.18$
- $F = MS_{BETWEEN} / MS_{WITHIN} = 9.83/1.18 = 8.36$
- $F\,(2, 14)_{CRITICAL} = 6.51$
- $\omega^2 = 19.65 - (2)(1.18) / (36.12 + 1.18) = 0.46$
- Reject H_0, strenuous physical exercise has a significant effect on the age of onset of puberty in girls, $F\,(2, 14) = 8.36$, $p < .01$, $\omega^2 = 0.46$

Source	*SS*	*df*	*MS*	*F*
Between	19.65	2	9.83	8.36
Within	16.46	14	1.18	
Total	36.12	16		

- t **(14)** critical = 2.977 **(NOTE: critical value from df_{WITHIN})**
- Athlete vs. Control:
  - $t = 14.0 - 11.57 / \surd((1.18(1/6+1/7)) = 4.02$, $p < .01$
  - $d = 14.0 - 11.57 / \surd 1.18 = 3.39$
- Athlete vs. Musician:
  - $t = 14.0 - 12.25 / \surd((1.18(1/6+1/4)) = 2.49$, *ns*
- Musician vs. Control:
  - $t = 12.25 - 11.57 / \surd((1.18(1/7+1/4)) = 1.0$, *ns*
- The onset of puberty was significantly later in girls in the athlete ($M = 14.0$) compared to girls in the control condition ($M = 11.57$) ($t\,(14) = 4.02$, $p < .01$; $d = 3.39$). No other pairwise comparisons were significant

---

Conduct a hypothesis test (α = .05) to determine if there is an effect of attitude on memory. Scores indicate the number of arguments recalled. Perform post-hoc tests and report the effect size.

Incidental-Agree	Incidental-Disagree	Intentional-Agree	Intentional-Disagree
8	2	6	7
7	3	8	9
7	2	9	8
9	4	5	5
4	4	8	7

Incidental-Agree	Incidental-Disagree	Intentional-Agree	Intentional-Disagree	
8	2	6	7	
7	3	8	9	
7	2	9	8	
9	4	5	5	
4	4	8	7	$k = 4$
$n = 5$	$n = 5$	$n = 5$	$n = 5$	$N = 20$
$M = 7.0$	$M = 3.0$	$M = 7.2$	$M = 7.2$	$M_G = 6.1$

- H_0: $\mu_{INCID-AGREE} = \mu_{INCID-DISAGREE} = \mu_{INTEN-AGREE} = \mu_{INTEN-DISAGREE}$
- H_1: μ's are not all equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 101.8$
- $SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{GRAND})^2$

  $= 5(7.0-6.1)^2 + 5(3.0-6.1)^2 + 5(7.2-6.1)^2 +$

  $5(7.2-6.1)^2 = 64.2$
- $SS_{WITHIN} = SS_{TOTAL} - SS_{BETWEEN} = 101.8 - 64.2 = 37.6$
- $df_{TOTAL} = N - 1 = 19$
- $df_{BETWEEN} = k - 1 = 3$
- $df_{WITHIN} = df_{TOTAL} - df_{BETWEEN} = 19 - 3 = 16$

- $MS_{BETWEEN} = SS_{BETWEEN} / df_{BETWEEN} = 64.2/3 = 21.4$
- $MS_{WITHIN} = SS_{WITHIN} / df_{WITHIN} = 37.6/16 = 2.35$
- $F = MS_{BETWEEN} / MS_{WITHIN} = 21.4/2.35 = 9.11$
- $F(3, 16)_{CRITICAL} = 3.24$
- $\omega^2 = 64.2\text{-}(3)(2.35) / (101.8 + 2.35) = 0.55$
- Reject H_0, attitude had a significant effect on memory, $F(3, 16) = 9.11$, $p < .05$, $\omega^2 = 0.55$

Source	*SS*	*df*	*MS*	*F*
Between	64.2	3	21.4	9.11
Within	37.6	16	2.35	
Total	101.8	19		

- $k = 4$ and $df_{WITHIN} = 16$, so $q = 4.05$
- $HSD = q\sqrt{(MS_{WITHIN}/n)} = 4.05\sqrt{(2.35/5)} = 2.78$
- Pairwise comparisons:
- $M_{INCID-AG} - M_{INCID-DISAG} = 7 - 3 = 4.0$ ($p < .05$); $d = 4/\sqrt{2.35} = 2.61$
- $M_{INCID-AG} - M_{INTEN-AG} = 7 - 7.2 = -0.2$ (*ns*)
- $M_{INCID-AG} - M_{INTEN-DISAG} = 7 - 7.2 = -0.2$ (*ns*)
- $M_{INCID-DISAG} - M_{INTEN-AG} = 3 - 7.2 = -4.2$ ($p < .05$); $d = 4.2/\sqrt{2.35} = 2.75$
- $M_{INCID-DISAG} - M_{INTEN-DISAG} = 3\text{-}7.2 = -4.2$ ($p < .05$); $d = 4.2/\sqrt{2.35}=2.75$
- $M_{INTEN-AG} - M_{INTEN-DISAG} = 7.2 - 7.2 = 0.0$ (*ns*)

- Significantly fewer statements were recalled in the incidental-disagree condition ($M = 3.0$) than the incidental-agree condition ($M = 7.0$) ($p < .05$; $d = 2.61$), the intentional-agree condition ($M = 7.2$)($p < .05$, $d = 2.75$), and the intentional-disagree ($M = 7.2$) condition than the incidental-disagree condition ($p < .05$, $d = 2.75$)
- No other pairwise comparisons were significant

- Marijuana consumption can increase or decrease locomotor activity (low doses of THC are known to increase locomotor activity and high doses of THC decrease locomotor activity)
- To determine if the effects of THC are acting on the nucleus accumbens (NA) researchers injected either: a placebo, or 0.1, 0.5, 1, or 2 micrograms of THC into the NA of rats, and then investigated the change in the animals' activity level
- Conduct a hypothesis test ($\alpha = .05$) to determine if there is an effect of marijuana consumption on locomotor activity

Placebo	.1μg	.5μg	1μg	2μg
30	60	71	33	36
27	42	50	78	27
52	48	38	71	60
38	52	59	58	51
20	28	65	35	29
26	93	58	35	34
8	32	74	46	24
41	46	67	32	17
49	63	61		50
49	44			53

Placebo	0.1μg	0.5μg	1.0μg	2.0μg	Total
30	60	71	33	36	
27	42	50	78	27	
52	48	38	71	60	
38	52	59	58	51	
20	28	65	35	29	
26	93	58	35	34	
8	32	74	46	24	
41	46	67	32	17	
49	63	61		50	
49	44			53	$k = 5$
$n_P = 10$	$n_{0.1} = 10$	$n_{0.5} = 9$	$n_{1.0} = 8$	$n_{2.0} = 10$	$N = 47$
$M_P = 34.0$	$M_{0.1} = 50.8$	$M_{0.5} = 60.33$	$M_{1.0} = 48.5$	$M_{2.0} = 38.1$	$M_G = 45.96$

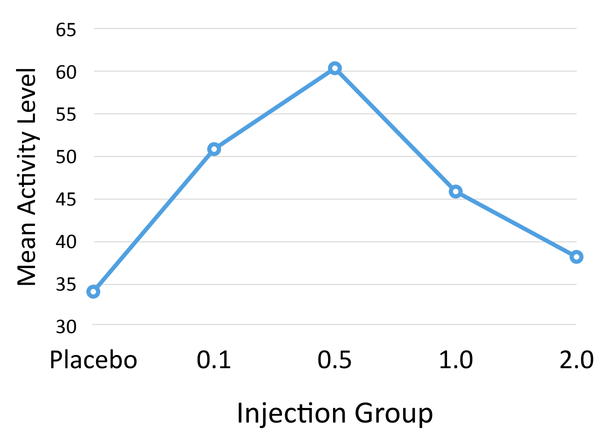
- H_0: $\mu_{PLACEBO} = \mu_{0.1} = \mu_{0.5} = \mu_{1.0} = \mu_{2.0}$
- H_1: μ's are not all equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 14,287.91$
- $SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{GRAND})^2$

  $= 10(34-45.96)^2 + 10(50.8-45.96)^2 + 9(60.33-45.96)^2$

  $+ 8(48.5-45.96)^2 + 10(38.1-45.96)^2 = 4,193.41$
- $SS_{WITHIN} = SS_{TOTAL} - SS_{BETWEEN} = 14287.91-4193.41 = 10094.5$
- $df_{TOTAL} = N - 1 = 46$
- $df_{BETWEEN} = k - 1 = 4$
- $df_{WITHIN} = df_{TOTAL} - df_{WITHIN} = 46 - 4 = 42$

- $MS_{BETWEEN} = SS_{BETWEEN} / df_{BETWEEN} = 4,193.41/4 = 1,048.35$
- $MS_{WITHIN} = SS_{WITHIN} / df_{WITHIN} = 10,094.5/42 = 240.35$
- $F = MS_{BETWEEN} / MS_{WITHIN} = 1,048.35/240.35 = 4.36$
- $F(4, 42)_{CRITICAL} = 2.61$
- $\omega^2 = 4193.41 - (4)(240.35) / (14,287.91+240.35) = 0.22$
- Reject H_0, marijuana consumption had a significant effect on locomotor activity, $F(4, 42) = 4.36$, $p < .05$, $\omega^2 = 0.22$

Source	*SS*	*df*	*MS*	*F*
Between	4,193.41	4	1,048.35	4.36
Within	10,094.50	42	240.35	
Total	14,287.91	46		



- t (42) critical = 2.02
- Placebo group vs. 0.5 group:
  - $t = (34.0-60.33)/\sqrt{(240.35(1/9+1/10))} = 3.70$, $p < .05$
  - $d = (34.0-60.33)/\sqrt{240.35} = 1.70$
- Placebo group vs. 2.0 group:
  - $t = (34.0-38.1)/\sqrt{(240.35(1/10 + 1/10))} = 0.59$, *ns*
- 0.5 group vs. 2.0 group:
  - $t = (60.33-38.1)/\sqrt{(240.35(1/9+1/10))} = 3.12$, $p < .05$
  - $d = (60.33-38.1)/\sqrt{240.35} = 1.43$
- Rats in the 0.5 group ($M = 60.33$) were significantly more active than rats in the placebo group ($M = 13.0$)($p < .05$, $d = 1.70$) and the 2.0 group ($M = 38.1$)($p < .05$, $d = 1.43$). No other pairwise comparisons were significant.

- Conduct a hypothesis test ($\alpha = .05$) to determine if there is an effect of smoking on performance in a reaction time test
- Perform post-hoc tests and report the effect sizes

Non-Smoker	Delay Smoker	Active Smokler
9	12	8
8	7	8
12	14	9
10	4	1
7	8	9
10	11	7
9	16	16
11	17	19
8	5	1
10	6	1
8	9	22
10	6	12
8	6	18
11	7	8
10	16	10
$M = 9.40$	$M = 9.60$	$M = 9.93$
$n = 15$	$n = 15$	$n = 15$

$GM = 9.64$
$N = 45$
$k = 5$

- H_0: $\mu_{NS} = \mu_{DS} = \mu_{AS}$
- H_1: μ's are not all equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 896.31$
- $SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{GRAND}) = 15(9.40-9.64)^2 + 15(9.60-9.64)^2 + 15(9.93-9.64)^2 = 2.18$
- $SS_{WITHIN} = SS_{TOTAL} - SS_{BETWEEN} = 896.31-2.18 = 894.13$
- $df_{TOTAL} = N - 1 = 45-1= 44$
- $df_{BETWEEN} = k - 1 = 5-1 = 4$
- $df_{WITHIN} = df_{TOTAL} - df_{WITHIN} = 44- 2 = 42$

- $MS_{BETWEEN} = SS_{BETWEEN} / df_{BETWEEN} = 2.18/2 = 1.09$
- $MS_{WITHIN} = SS_{WITHIN} / df_{WITHIN} = 894.13/42 = 21.29$
- $F = MS_{BETWEEN} / MS_{WITHIN} = 1.09/21.29 = 0.51$
- $F\ (2, 42)_{CRITICAL} = 3.21$
- $\omega^2 = 2.18 - (2)(21.29) / (896.31+21.29) = -0.044$
- Fail to reject H_0, there is not a significant effect of smoking on reaction time performance; $F\ (2, 42) = 0.51$, $p > .05$, $\omega^2 = -0.044$

Source	*SS*	*df*	*MS*	*F*
Between	2.18	2	1.09	0.51
Within	894.13	42	21.29	
Total	896.31	44		

Conduct a hypothesis test (α = .05) to determine if satisfaction differs across treatments. Data are satisfaction ratings, higher ratings reflect greater satisfaction with the treatment program. Perform post-hoc tests and report the effect sizes

Treatment A	Treatment B	Treatment C
8	7	6
13	3	4
10	8	2
9		

Treatment A	Treatment B	Treatment C	Total
8	7	6	
13	3	4	
10	8	2	
9			k =3
$n_A = 4$	$n_B = 3$	$n_C = 3$	$N = 10$
$M_A = 10$	$M_B = 6$	$M_C = 4$	$M_{GRAND} = 7$

- H_0: $\mu_A = \mu_B = \mu_C$
- H_1: μ's are not all equal
- $SS_{TOTAL} = \sum(X - M_{GRAND})^2 = 102.0$
- $SS_{BETWEEN} = \sum n_{GROUP}(M_{GROUP} - M_{GRAND})^2$
  $= 4(10-7)^2 + 3(6-7)^2 + 3(4-7)^2 = 66.0$
- $SS_{WITHIN} = SS_{TOTAL} - SS_{BETWEEN} = 102.0 - 66.0 = 36.0$
- $df_{TOTAL} = N - 1 = 9$
- $df_{BETWEEN} = k - 1 = 2$
- $df_{WITHIN} = df_{TOTAL} - df_{BETWEEN} = 9 - 2 = 7$

- $MS_{BETWEEN} = SS_{BETWEEN} / df_{BETWEEN} = 66/2 = 33.0$
- $MS_{WITHIN} = SS_{WITHIN} / df_{WITHIN} = 36/7 = 5.14$
- $F = MS_{BETWEEN} / MS_{WITHIN} = 33/5.14 = 6.42$
- $F\,(2, 7)_{CRITICAL} = 4.74$
- $\omega^2 = 66-(2)(5.14) / (102 + 5.14) = 0.52$
- Reject H_0, satisfaction ratings are significantly different across the treatment programs, $F\,(2, 7) = 6.42$, $p < .05$, $\omega^2 = 0.52$

Source	*SS*	*df*	*MS*	*F*
Between	66	2	33	6.42
Within	36	7	5.14	
Total	102	9		

- t (7) critical = 2.365 (from df_{WITHIN})
- Treatment A vs. Treatment B:
  - $t = 10 - 6 / \sqrt{(5.14(1/4 + 1/3))} = 2.31$, *ns*
- Treatment A vs. Treatment C:
  - $t = 10 - 4 / \sqrt{(5.14(1/4 + 1/3))} = 3.47$, $p < .05$, $d = 2.65$
- Treatment B vs. Treatment C:
  - $t = 6 - 4 / \sqrt{(5.14(1/3 + 1/3))} = 1.08$, *ns*
- Participants that received Treatment A ($M = 10.0$) gave significantly higher ratings of satisfaction than those in Treatment C ($M = 4.0$) ($p < .05$, $d = 2.65$). No other pairwise comparisons were significant.