

## CHAPTER 11

### Analysis of Variance: Independent Samples

#### Summary

You were introduced to a new NHST technique in this chapter called *analysis of variance (ANOVA)*. This technique will be the topic of the next two chapters, as well. Like  $t$  tests, ANOVA allows you to draw conclusions about differences between population means based on sample data. Unlike  $t$  tests, though, ANOVA lets you compare two or more populations rather than just two.

The *one-way independent ANOVA* described in Chapter 11 is similar to the independent samples  $t$  test you learned in Chapter 10. They both use one independent variable (hence the “one” in “one-way”) when the levels of the independent variable are independent of each other (rather than paired). The next chapter will review ANOVA for *paired samples* designs with one independent variable. Learning ANOVA may seem like a lot of trouble when you could just do multiple  $t$  tests for the same purpose, but ANOVA protects you from increased chances of a Type I error that all those  $t$  tests would bring.

The null hypothesis of a one-way independent ANOVA is that the means of the populations from which the samples are drawn are equal. The alternative is that one or more of the population means is different from the others.

The test statistic for ANOVA is  $F$ .  $F$  is a ratio of two estimates of the same thing; the *variability of the population* all the samples are assumed to have come from.  $F = 1$  when the null hypothesis is true, and  $F > 1$  when the null hypothesis is false. The rationale of ANOVA (i.e., how it works) is important. To understand the rationale, you must have a clear idea of how the two estimates of the population variance are found.

If you think back to Chapter 4, you will remember variance is a measure of variability (differences between scores) and *chance* variance is sometimes called “error” in statistics. Thus, the denominator of the  $F$  (an estimate of the chance variation in the population) is sometimes called the *error term* or the *within groups estimate*. It is called the within groups estimate because this estimate of the population variance is obtained by calculating the variance that occurs *within* each sample you have and averaging those variances together. By now, it should make sense to you that a good estimate of the population variance is the variance in a sample that is meant to represent the population, and averaging more than one estimate together is even better than using just one. This estimate works well as an estimate of the population variance whether the null is true or false because differences between population means (which occur when the null is false) have no effect on the within group variance. In formulas for  $F$ , this within groups estimate it is identified as  $MS_{error}$ .

The numerator of the  $F$  is sometimes called the *between groups estimate*. It is called the between groups estimate because you obtain this estimate of the population variance by determining the variance *between* the sample means that you have. From your study of sampling distributions in Chapter 7, you learned that even if all of the sample means came from populations with the same mean (i.e., the null hypothesis is true), you would still expect some variance in the sample means just by chance, and the amount of variance depends on the population variance. If the null hypothesis is true—that is, the means of all populations from which samples are drawn are equal—then this is a good estimate of the population variance. If the null is false—that is, one or more of the samples came from a population with a mean that differs from the rest—then the differences in the sample means are due to *two* factors not just one: 1) chance, just like before *and* 2) the independent or grouping variable. In other words, when the null hypothesis is false, some of the reason the sample means are different from each other is chance, but some of the reason the sample means are different from each other is due to exposure to different levels of the independent variable (i.e. the *treatment* participants are exposed to). In formulas for  $F$  it is identified as  $MS_{treat}$ .

In short, the denominator ( $MS_{error}$ ) of  $F$  is a good estimate of population variance. If the null is true, the numerator ( $MS_{treat}$ ) is also a good estimate of population variance. The  $F$  test then becomes one good estimate of the population variance, divided by another good estimate of that same thing, making  $F = 1$ . If the null is false, however,  $MS_{treat}$  will *not* be a good estimate of the population variance, because it will be too large (variance due to chance *plus* variance due to the independent variable). The  $F$  test then becomes one inflated estimate of the population variance divided by a more accurate estimate of that same thing, making  $F > 1$ . But, how much bigger than 1 does an  $F$  ratio

have to be to reject the null hypothesis and conclude that at least one sample comes from a population with a mean that differs from the rest? The ***F distribution*** (Table F) shows *F* values that occur 5% and 1% of the time when the null hypothesis is true. If your data produce a calculated *F* value that is larger than the tabled *F* value, reject the null hypothesis and conclude that the samples did not all come from populations with equal means.

For ANOVA, if the data allow you to reject the null hypothesis, you investigate with ***multiple comparison tests*** to determine *which* population means are different from the others. Remember that an ANOVA tells you *if* there are differences in the means, but it does not tell you *which* means differ from one another. ***Multiple comparisons tests*** allow you to examine whether particular pairs of means are significantly different from each other, much like a *t* test, and are either decided upon *a priori* or *post hoc*. ***A priori tests*** require choosing a *limited* number of comparisons on logical grounds *prior to* data analysis. There is no requirement that the ANOVA be significant to do these tests. ***Post hoc tests*** allow you to make many, often all, possible paired comparison to determine which means are different from each other, but usually only after the ANOVA led to a rejection of the null hypothesis. One post hoc test, the ***Tukey Honestly Significant Difference test*** (Tukey HSD) is described in this chapter.

Like other NHST techniques you have learned, you can also calculate effect size indexes (here *f* and *d*) to determine the size of the effects in ANOVA. Remember that an ANOVA and multiple comparisons tests tell you *whether* there are differences in means but do not tell you *how big* those differences are. In ANOVA, an effect size index, *f*, indicates the size of the overall effect of the independent variable on the dependent variable as small ( $f = 0.10$ ), medium ( $f = 0.25$ ), or large ( $f = 0.40$ ). Just as you have used *d* in past chapters, *d* can be used here to measure the size of the difference between any two means as small ( $d = 0.20$ ), medium ( $d = 0.50$ ), or large ( $d = 0.80$ ).

After conducting an ANOVA, appropriate multiple comparisons tests, and finding effect size estimates, your final task is to explain these results in a good interpretation. Be sure to use the terms of the experiment and to include the size and direction of any effects or differences.

The ANOVA technique described in this chapter is appropriate for analyzing quantitative data from independent samples if certain assumptions about the populations hold true. These assumptions are that the population variances are equal (***homogeneity of variance***) and the populations are normally distributed. In addition, if cause-and-effect conclusions between the independent and dependent variables are to be drawn, the participants must be randomly assigned to levels of the independent variable.

### Multiple-Choice Questions

1. An  $F$  distribution is always a \_\_\_\_\_.
  - a. normal distribution
  - b.  $t$  distribution
  - c. sampling distribution
  - d. none of the above
2. The null hypothesis tested by ANOVA is that \_\_\_\_\_.
  - a. all samples have the same mean
  - b. each sample is drawn from a different population
  - c. the populations from which the samples are drawn have the same mean
  - d. at least one of the populations from which the samples are drawn has a mean that is different from the others
3. The ANOVA technique described in this chapter can be used on \_\_\_\_\_.
  - a. paired-samples designs
  - b. independent-samples designs
  - c. both a. and b.
  - d. neither a. nor b.
4. If the null hypothesis is true, the \_\_\_\_\_ will be a good estimate of the population variance.
  - a.  $MS_{treat}$
  - b.  $MS_{error}$
  - c. both a. and b.
  - d. neither a. nor b.
5. If the null hypothesis is false, the \_\_\_\_\_ will be a good estimate of the population variance.
  - a. numerator of the  $F$  ratio
  - b. denominator of the  $F$  ratio
  - c. both a. and b.
  - d. neither a. nor b.
6. The larger the population variance is, the larger the \_\_\_\_\_ is (are).
  - a.  $F$
  - b.  $df_{error}$
  - c.  $MS_{error}$
  - d. all of the above
7. If a tabled value of  $F$  is 10.00 and the  $F$  obtained from the data is 9.00, you should \_\_\_\_\_.
  - a. retain the null hypothesis
  - b. reject the null hypothesis
  - c. calculate the  $F$  value again, such a number is not possible
  - d. cannot be determined from the information given

8. Suppose  $MS_{treat}$  is calculated for three samples that are drawn from a population with the same mean. Under which condition below would  $MS_{treat}$  become substantially larger?

- a. the addition of a sample from a population with the same mean
- b. the addition of a sample from a population with a mean one half as big as the original mean
- c. the removal of one of the three samples from the calculations
- d. none of the above, adding and removing samples for other populations never affects the size of  $MS_{treat}$

9. In a properly calculated one-way independent ANOVA problem \_\_\_\_\_.

- a.  $SS_{treat} + SS_{error} = SS_{tot}$
- b.  $MS_{treat} + MS_{error} = MS_{tot}$
- c. both a. and b.
- d. neither a. nor b.

10. For a one-way independent ANOVA, an effect size index ( $f$ ) that qualifies as large is

- a. 0.25
- b. 0.40
- c. both a. and b.
- d. neither a. nor b.

11. *A priori* and *post hoc* are terms that refer to \_\_\_\_\_.

- a. whether the null hypothesis should be rejected
- b. whether the assumptions of ANOVA have been met
- c. statistics that test the size of the effect found in an ANOVA
- d. none of the above

12. The Tukey Honestly Significant Difference test is a(n) \_\_\_\_\_ test.

- a. *a priori*
- b. *post hoc*
- c. both a. and b.
- d. neither a. nor b.

13. Suppose the following  $F$  values were calculated from different experiments. If  $\alpha = .01$ , which of them leads to *rejection* of the null hypothesis?

- a.  $F = 18.50, df = 2, 2$
- b.  $F = 2.55, df = 10, 17$
- c.  $F = 2.41, df = 20, 42$
- d. none of the above

14. Which of the following statements accurately compares the  $F$  test and the  $t$  test?

- a. The  $F$  test can be used with only two groups; the  $t$  test can be used with more than two groups.
- b. The  $t$  test can be used with only two groups; the  $F$  test can be used with more than two groups.
- c. The  $F$  test is used only one independent variable; the  $t$  test is used with more than one independent variable.
- d. None of the above.

15. Suppose you have designed a study and are using a one-way independent ANOVA to analyze the data. You have just calculated the  $F$  test. How many degrees of freedom are there for the numerator and denominator respectively if there are four groups with eight participants in each group?

- a. 4, 8
- b. 3, 8
- c. 3, 28
- d. 4, 21

16. Suppose you have designed a study and are using a one-way independent ANOVA to analyze the data. You have just calculated the  $F$  test. How many degrees of freedom are there for the numerator and denominator respectively if there are three groups with 10 participants in each group?

- a. 3, 10
- b. 2, 9
- c. 3, 30
- d. 2, 27

17. The  $F$  test in an ANOVA can be used to determine \_\_\_\_\_.

- a. that at least one population mean in a study is different from the rest
- b. how big the difference is between population means in a study
- c. *which* population means are different from the other means in a study
- d. all of the above

18. One difference between *post hoc* and *a priori* comparisons is that \_\_\_\_\_.

- a. *post hoc* tests are multiple-comparisons tests, *a priori* tests are not
- b. *post hoc* tests are appropriate after examination of the data, *a priori* tests must be planned before examination of the data
- c. *post hoc* tests give information about the size of the differences between groups in a study and *a priori* tests tell whether differences are significant
- d. *post hoc* tests tell whether differences between groups in a study are significant and *a priori* tests give information about the size of the differences.

### **Short Answer**

1. What is the difference between the function of an  $F$  test and Tukey HSD tests?

2. A wine maker wanted to market a new red wine blend of several varieties grown in his vineyards. He developed four blends, and he wanted to decide which was the best. He employed the services of eight wine tasters and had each taster rate each of the four wines on a 7-point scale, ranging from awful (1) to exceptional (7). Explain why the ANOVA method described in this chapter is inappropriate for the analysis of the data.

3. The wine maker from the last question goes another route. He is wondering if men and women have differing views on which of the four wines is best. This time, he asked potential customers (80 men and 80 women) to help him. Each of the 80 men was randomly given one of the four wines to taste and each of the 80 women was randomly given one of the four wines to taste. Each rated it on a 7-point scale, ranging from awful (1) to excellent (7). In this study, 40 different tasters (20 men and 20 women) tasted each wine. Explain why the ANOVA method described in this chapter is still inappropriate to determine if men and women have differing views on which of the wines is best.

4. Assume that five groups of rats each receive different dosages of a drug. You believe that the dosages are going to result in different running speeds among the rats. What type of analysis will allow you to determine if your belief is correct? Briefly explain why that analysis is the appropriate option.

5. Walking is valuable exercise and a researcher wanted to know if there is any relationship between the amount of walking people do and the size of the city where they live. A researcher collected the following data, conducted an ANOVA, and then conducted a series of follow-up Tukey HSD tests. The following summary data (in miles per day) are hypothetical. The place names are ordered based on size, from large to small.

- Identify the independent and dependent variables.
- Calculate the sample means, identify necessary critical values, and write a conclusion based on the data.
- Do the data support the hypothesis that city size is related to the amount of walking people do?

	Gotham	Middletown	Grover's Corners
$\Sigma X$	16	24	32
$\Sigma X^2$	70	60	140
$N$	4	12	8

Source	$SS$	$df$	$MS$	$F$	$p$
Treatment	24.00	2	12.00	8.39	< .05
Error	30.00	21	1.43		
TOTAL	54.00	23			

$$HSD_{Gotham v. Middletown} = 4.10$$

$$HSD_{Gotham v. Grover's} = 0.00$$

$$HSD_{Middletown v. Grover's} = 5.18$$

$$f = 1.11$$

$$d_{Gotham v. Middletown} = 1.67$$

$$d_{Gotham v. Grover's} = 0$$

$$d_{Middletown v. Grover's} = 1.67$$

### Problems

1. A group of 72 participants was divided equally into four groups. A Tukey HSD test produced a value that led to the conclusion that Mean 1 was significantly larger than Mean 2,  $p < .05$ . Could either of the following situations lead to such a conclusion?

- $\bar{X}_1 = 7, \bar{X}_2 = 0, MS_{error} = 80$
- $\bar{X}_1 = 24, \bar{X}_2 = 13, MS_{error} = 180$

2. A group of 36 participants was equally divided into three groups. A Tukey HSD produced a value that led to the conclusion that Mean 1 was significantly larger than Mean 2,  $p < .05$ . Could either of the following situations lead to such a conclusion?

- $\bar{X}_1 = 9, \bar{X}_2 = 2, MS_{error} = 50$
- $\bar{X}_1 = 24, \bar{X}_2 = 14, MS_{error} = 84$

3. The pineal gland is a gland in your body that does a number of things. Among them, it releases melatonin, which is closely tied to the timing and quality of sleep. Researchers in Seoul, South Korea (Park et al., 2018) were interested in whether lifetime coffee consumption was related to differences in pineal gland size in healthy older (60-94 years old) individuals. Using the number of years they had been drinking coffee and the number of cups per day, Park and colleagues divided their participants into three groups based on their lifetime coffee consumption: low, moderate, and high. They also measured the volume of the pineal gland using magnetic resonance imaging (MRI). The data that follow are hypothetical pineal gland volume created so the conclusions and means you reach will mimic those of Park et al. (2018).

Low Coffee Consumption	Moderate Coffee Consumption	High Coffee Consumption
5	5	3
4	5	4
4	5	4
4	4	4
5	4	4
5	4	3
5	6	4

Conduct an analysis using  $\alpha = .05$  to determine if lifetime coffee consumption is related to differences in pineal gland size. Perform the  $F$  test, create an ANOVA summary table, complete a Tukey HSD analysis (if appropriate), calculate the effect size index  $f$ , and the three  $d$ 's. Finally, write a conclusion about the relationship of lifetime coffee consumption and pineal gland size.

4. Researchers in Poland (Gasiorowska, Chaplin, Zaleskiewicz, Wygrab, & Vohs, 2016) were interested in whether children who were primed to think about money shared less than children primed to think about other things. These researchers had 3-6 year old Polish children sort either coins, buttons, or candy for 10 minutes (this technique is known as “priming”). They considered both the button- and candy-sorting groups as “neutral” conditions. Afterward, the children were shown six Disney character stickers (Note: If you do not know much about children, Disney character stickers are highly coveted items). They were allowed to take as many stickers as they wanted. Next, they were given the opportunity to give some, all, or none of their stickers away to a child who they were told did not get to participate in the study. The researchers used the percentage of the child’s stickers they shared with the other child as the dependent variable. The data below are hypothetical percentages for six participants in each group created so the conclusions and means that you reach will mimic those of Gasiorowska and colleagues (2016).

Money Prime	Candy Prime	Button Prime
26	45	27
25	43	28
17	35	69
33	47	64
35	43	67
0	53	33

Conduct an analysis using  $\alpha = .05$  to determine if what children are primed to think about effects how much they are willing to share. Perform the  $F$  test, create an ANOVA summary table, complete a Tukey HSD analysis (if appropriate), calculate the effect size index  $f$ , and the three  $d$ 's. Finally, write a conclusion about what you found.

5. A researcher wonders if the type of pain relief administered to a mother during delivery effects the well-being of the newborn. In most hospital delivery rooms, newborn infants are evaluated at one minute of age and again at five minutes of age. The evaluation is based on the Apgar Scale, which uses certain criteria to rate the infant's well-being using heart rate, respiratory effort, crying, muscle tone, and color. Scores from 0 to 10 are possible, with 10 indicating the highest general well-being. For this study, the newborn's Apgar Score is the dependent variable. The independent variable is the type of pain relief administered to the mother: twilight sleep induced by sedatives, spinal block, opiate, and the Lamaze method (no drugs). The researcher has 40 mothers consent to be randomly assigned to one of the four groups when it is time for her delivery. Below are summary data for the five-minute Apgar scores for the 40 newborns.

	Sedatives	Spinal Block	Opiate	Lamaze
$\Sigma X$	49	74	78	83
$\Sigma X^2$	269	574	636	711
$N$	10	10	10	10

Conduct an analysis using  $\alpha = .05$  to determine if the type of pain relief given to a mother during delivery effects the well-being of the newborn at five minutes of age. Perform the  $F$  test, create an ANOVA summary table, complete a Tukey HSD analysis (if appropriate), calculate the effect size index  $f$ , and the six  $d$ 's. Finally, write a conclusion about the effect of type of pain relief administered to a mother during delivery on the well-being of the newborn.

6. Hanna, Kee, and Robertson (2017) examined whether use of social media to connect with co-workers was associated with differences in job satisfaction. They used adult employees at two tech companies in Northern California. They assessed the amount of time participants connected with co-workers by asked each participating employee "In a typical work week, approximately how much time per day (including during breaks, after work, during the weekend, etc.) do you spend on Facebook interacting with people from work?" They divided their participants into three groups based on their answers to this question: less than 10 minutes, 10-30 minutes, and 31 minutes to 2 hours. Here we'll call this low, medium, and high connection. They also assessed their job satisfaction using a set of five questions, with higher scores meaning greater job satisfaction. The data below are hypothetical job satisfaction scores for five participants in each group created so the conclusions and means that you reach will mimic those of Hanna and colleagues (2017).

Low connection (Less than 10 minutes)	Medium connection (10-30 minutes)	High connection (31 minutes to 2 hours)
2	2	5
3	6	5
4	4	4
4	4	6
2	3	5

Conduct an analysis using  $\alpha = .05$  to determine if time spent connecting with co-workers through Facebook is related to differences in job satisfaction. Perform the  $F$  test, create an ANOVA summary table, complete a Tukey HSD analysis (if appropriate), calculate the effect size index  $f$ , and the three  $d$ 's. Finally, write a conclusion about the relationship of time spent connecting with co-workers through Facebook and job satisfaction.



7. Farmer Marc A., who also serves as his community's Shakespearean promoter, delivered a plea at the county fair, asking farmers from the county's three groups to lend him unshelled corn. (After examining the groups, you might be able to figure out how he would phrase his plea.) The number of bushels offered by the farmers is shown below.

Friends	Romans	Countrymen
7	8	10
9	7	12
3	4	16
5	7	14

Conduct an analysis using  $\alpha = .05$  to determine whether there are differences in the county's three groups' willingness to lend a bushel of corn. Perform the  $F$  test, create an ANOVA summary table, complete a Tukey HSD analysis (if appropriate), calculate the effect size index  $f$ , and the three  $d$ 's. Finally, write a conclusion about the relationship between county group membership and willingness to lend corn.

## ANSWERS

### Multiple-Choice Questions

1. c

**Explanation:** Unlike the normal distribution, which is always, of course, normally distributed,  $F$  is a family of distributions which are most often positively skewed. The  $F$  distribution is only normally distributed when  $df_{treat} = 1$  and  $df_{error} = \infty$ .

2. c

3. b

**Explanation:** The ANOVA technique described in this chapter is used for independent-samples designs with one independent variable. The ANOVA technique you will learn in the next chapter is used for paired-samples designs with one independent variable.

4. c

**Explanation:** If the null hypothesis is true, that means that all of our samples were drawn from populations with means equal to one another. Another way to think of this is that all of our samples came from the *same* population. If that's the case, any variability between groups (represented by  $MS_{treat}$ ) should just be due to sampling error, making it roughly equal to variance within groups ( $MS_{error}$ ). Thus, *both*  $MS_{treat}$  and  $MS_{error}$  will be good estimates of population variance. If this idea is still confusing, you are encouraged to re-read the summary and return to the section in your textbook labeled, 'Rationale of ANOVA.'

5. b

**Explanation:** If the null hypothesis is false, that means that *not* all of our samples were drawn from populations with means equal to one another. Another way to think of this is that at least one of our samples came from a *different* population. If that's the case, variability between groups (represented by  $MS_{treat}$ ) represents variability due to sampling error *and* variability due to the independent variable, making it *larger than* variance within groups ( $MS_{error}$ ). Now, only  $MS_{error}$  will be a good estimate of population variance. If this idea is still confusing, you are encouraged to re-read the summary and return to the section in your textbook labeled, 'Rationale of ANOVA.'

6. c

**Explanation:** In ANOVA,  $MS_{error}$  is an estimate of population variance. Having a larger population variance can make  $F$  smaller not larger. *Degrees of freedom* are not affected by the amount of population variance, most often they are determined by  $N$  and/or the number of levels of an independent variable.

7. a

8. b

**Explanation:** The  $MS_{treat}$  is the estimate of the population variance obtained by examining differences *between* the sample means. If all of the samples are drawn from a single population, as in the question stem, the sample means will not be very different from each other. Taking a fourth sample from that same population, as in choice a., almost certainly yields a fourth mean that would still be very close to the others. This would not change the variance between the sample means very much.

Adding in a sample that was drawn from a population with a mean that is half as big as the original mean, as in choice b., is likely to yield a sample mean close to half as big as the original three means, and that would substantially increase the variability between the means,  $MS_{treat}$ . (If you have three numbers that are not very different from each other and you add in another number that is half as big, it will greatly increase the variability).

For choice c., eliminating one of the samples might reduce the variance between the means by a small amount, but it would definitely not increase it. If you have three numbers that are not very different from each other, eliminating one of them will not make the two numbers that are left *more* different from each other.

As for option d., the last paragraphs should convince you that adding and subtracting samples from other populations would likely change  $MS_{treat}$ .

9. a

10. b

11. d

**Explanation:** *A priori* and *post hoc* are terms that refer to multiple comparisons tests that examine *which* population means are different from the others. Specifically, these terms indicate when decisions about multiple comparisons tests were made: *prior* to data collection or *after* collecting and inspecting the data.

12. b

**Explanation:** The Tukey HSD test is considered a *post hoc* test, because it examines all possible two-group comparisons.

13. c

**Explanation:** First, look up a critical  $F$  ( $F_{.01}$ ) for each of the experiments. Then compare each obtained  $F$  with the appropriate critical  $F$  and

make a decision about the null hypothesis. Reject the null when the observed  $F$  is greater than the critical  $F$ . The correct critical  $F$ s follow:

- a.  $F_{.01} = 99.00$  ( $18.50 < 99.00$ ); retain the null
- b.  $F_{.01} = 3.59$  ( $2.55 < 3.59$ ); retain the null
- c.  $F_{.01} = 2.35$  ( $2.41 > 2.35$ ); reject the null

14. b

15. c

**Explanation:** To find  $N_{tot}$ , multiply the number of groups by the number of participants in each group; here  $N_{tot} = (4)(8) = 32$ .

For a one-way independent ANOVA, the  $df$  for the numerator is  $df_{treat} = K - 1$ , where  $K$  represents the number of groups or levels of the independent variable; here that is  $df_{treat} = 4 - 1 = 3$ .

The  $df$  for the denominator is

$$df_{error} = N_{tot} - K; \text{ here that is}$$

$$df_{error} = 32 - 4 = 28.$$

16. d

**Explanation:** To find  $N_{tot}$ , multiply the number of groups by the number of participants in each

group; here  $N_{tot} = (3)(10) = 30$ .

For a one-way independent ANOVA, the  $df$  for the numerator is  $df_{treat} = K - 1$ , where  $K$  represents the number of groups or levels of the independent variable; here that is  $df_{treat} = 3 - 1 = 2$ .

The  $df$  for the denominator is

$$df_{error} = N_{tot} - K; \text{ here that is}$$

$$df_{error} = 30 - 3 = 27.$$

17. a

**Explanation:** The  $F$  test in an ANOVA can only be used to determine *if* there are any population mean differences. The effect size index,  $f$ , is used to determine *how big* the difference is between population means in a study. Tukey HSD, or some other multiple comparisons test, is used to determine *which* population means are different from the other means in a study.

18. b

### Short Answer

1.  $F$  test tells you *if* there are any differences between the means. The Tukey HSD test tells you *where* those differences are.
2. The ANOVA described in this chapter requires the levels of the independent variable (here, blend) to be independent of each other. Here the wine tasters gave a score to each of the four blends (a paired design because it is repeated-measures). Those scores are not independent.  
**Explanation:** If you had trouble with this problem, start by identifying the independent variable in the question. Next, identify the levels of the independent variable. Then decide if the levels are independent or paired. If you are having trouble determining if the levels of an independent variable are independent or paired, review that section in your book in Chapter 10 labeled 'Paired-Samples Designs and Independent-Samples Designs'. You will learn how to analyze the data for a paired-samples design with more than two groups appropriately with a one-factor repeated measures ANOVA in the next chapter.
3. Even though the levels of the independent variables are independent in this version of the study, the ANOVA described in this chapter only works with one independent variable at a time. This study has two independent variables (grape blend with four levels) and sex of taster (men and women).  
**Explanation:** You will learn how to analyze the data from this study in Chapter 13.
4. A one-way independent ANOVA, which is described in this chapter, will allow you to determine if there are significant differences between the drug dosages. One-way independent ANOVA is most appropriate, because there is only one independent variable (dosage amount) with more than two levels. If the  $F$  test is significant, then use a series of Tukey HSD tests to determine which dosages are significantly different from each other. Follow this by calculating an  $f$  and a series of  $d$ 's to determine the size of the effects you are examining.
5. a. The independent variable is the city (it has three levels) and the dependent variable is the number of miles walked per day.  
b.  $\bar{X}_{Gotham} = 4.00$ ;  $\bar{X}_{Middletown} = 2.00$ ;  $\bar{X}_{Groveris} = 4.00$   
Critical value:  $F_{.05}(2, 21) = 3.47$   
Critical value: Tukey HSD $_{.05}(3, 21) = 3.58$

According to the data, we reject the null hypothesis. City size is significantly related to the amount of walking people do,  $F(2, 21) = 8.39, p < .05$ . The size of the effect is large ( $f = 1.11$ ). Follow-up Tukey HSD tests showed that people who live in cities that are large ( $\bar{X}_{Gotham} = 4.00$ ) and small ( $\bar{X}_{Groveris} = 4.00$ ) walk significantly more miles per day than people who live in middle size cities ( $\bar{X}_{Middletown} = 2.00$ ),  $HSD_{Gotham v. Middletown} = 4.10$ ,  $HSD_{Middletown v. Groveris} = 5.18$ . These differences are both large (both  $ds = 1.67$ ). There is not a significant difference in the amount of walking people do who live in large and small cities ( $HSD_{Gotham v. Groveris} = 0.00; d = 0.00$ ).

APA format: There is a large, significant relationship between city size and the amount of walking people do,  $F(2, 21) = 8.39, p < .05, f = 1.11$ . People who live in large and small cities walk significantly more miles per day ( $M = 4.00$  and  $M = 4.00$ , respectively) than people who live in medium-sized cities ( $M = 2.00$ ),  $HSD(3, 21) = 4.10$  and  $5.18$ , respectively (both  $ps < .05$ ). These differences were large (both  $ds = 1.67$ ). There is not a significant difference between large and small cities in the amount of walking people do,  $HSD(3, 21) = 0.00, p > .05, d = 0.00$ .

c. Yes, size does matter, sometimes. There is no difference between people who live in large and small cities, but people who live in medium cities walk significantly less.

### Problems

1. No, neither a. nor b. can lead to this conclusion.

**Explanation:** For this problem, calculate the Tukey HSD value for both choices a. and b. and then compare those calculated Tukey HSD values with the critical value from the table to determine if the null could be rejected (i.e., the difference was significant).

Finding the Tukey HSD critical value:

Go to Table G. Find the page labeled  $\alpha = .05$  because the  $p$  reported in the problem is less than .05.

Find the column labeled 4, because the problem states there are four groups in this study (in other words, four levels of the independent variable).

Calculate the  $df_{error}$  for this problem.

$$df_{error} = N_{tot} - K$$

$$df_{error} = 72 - 4$$

$$df_{error} = 68$$

Find the intersection of the column marked 4 and row marked 68; there isn't one with 68 so choose the next smaller number (here 60). This  $HSD_{.05} = 3.74$ .

For choice a.

Start by calculating the standard error. Don't forget that  $N_i$  is the number of scores in each group, here 72 divided equally into 4 groups, 18.

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{80}{18}} = \sqrt{4.444} = 2.12$$

Now Tukey HSD

$$HSD = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}}} = \frac{7 - 0}{2.11} = \frac{7}{2.11} = 3.32$$

Compare Tukey HSD = 3.32 to  $HSD_{.05} = 3.74$ ; 3.32 is smaller than the critical value so you must retain the null and conclude this is not a significant difference.

Now repeat these steps for choice b.

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{180}{18}} = \sqrt{10} = 3.16$$

$$HSD = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}}} = \frac{24 - 13}{3.16} = \frac{11}{3.16} = 3.48$$

Compare Tukey HSD = 3.48 to  $HSD_{.05} = 3.74$ ; 3.48 is smaller than the critical value so you must retain the null and conclude this is not a significant difference.

2. Yes, b.  $\bar{X}_1 = 24$ ,  $\bar{X}_2 = 14$ ,  $MS_{error} = 84$  could lead to such a conclusion.

**Explanation:** For this problem, calculate the Tukey HSD value for both choices a. and b. and then compare those calculated Tukey HSD values with the critical value from the table to determine if the null could be rejected (i.e., the difference was significant).

Finding the Tukey HSD critical value:

Go to Table G. Find the page labeled  $\alpha = .05$  because the  $p$  reported in the problem is less than .05.

Find the column labeled 3, because the problem states there are 3 groups in this study (in other words, three levels of the independent variable).

Calculate the  $df_{error}$  for this problem.

$$df_{error} = N_{tot} - K$$

$$df_{error} = 36 - 3$$

$$df_{error} = 33$$

Find the intersection of the column marked 3 and row marked 33; there isn't one with 33 so choose the next smaller number (here 30). This  $HSD_{.05} = 3.49$ .

For choice a.

Start by calculating the standard error. Don't forget that  $N_i$  is the number of scores in each group, here 36 divided equally into 3 groups, 12.

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{50}{12}} = \sqrt{4.167} = 2.04$$

Now Tukey HSD

$$HSD = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}}} = \frac{9-2}{2.04} = \frac{7}{2.04} = 3.43$$

Compare Tukey HSD = 3.43 to  $HSD_{.05} = 3.49$ ; 3.43 is smaller than the critical value so you must retain the null and conclude this is not a significant difference.

Now repeat these steps for choice b.

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{84}{12}} = \sqrt{7} = 2.65$$

$$HSD = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}}} = \frac{24-14}{2.65} = \frac{10}{2.65} = 3.77$$

Compare Tukey HSD = 3.77 to  $HSD_{.05} = 3.49$ ; 3.77 is larger than the critical value so you reject the null and conclude this is a significant difference.

Source	SS	df	MS	F	p
Treatment	4.10	2	2.05	5.54	< .05
Error	6.57	18	0.37		
TOTAL	10.67	20			

Critical value:  $F_{.05}(2, 18) = 3.55$

According to the data, we reject the null hypothesis. Lifetime coffee consumption was related to significant differences in pineal gland size,  $F(2, 18) = 5.54$ ,  $p < .05$ . The effect size index,  $f = 0.66$ , was large. Follow-up Tukey HSD tests showed that high coffee consumers had significantly smaller pineal glands ( $\bar{X}_{High} = 3.71$ ) compared to moderate ( $\bar{X}_{Moderate} = 4.71$ ) (HSD = 4.35) and low ( $\bar{X}_{Low} = 4.57$ ) (HSD = 3.74) coffee consumers. There was no significant difference in pineal gland size between moderate and low coffee consumers (HSD = -0.61). The differences in size of pineal glands between the high coffee consumers and the moderate and low coffee consumers were both large ( $d = 1.64$ ,  $d = 1.41$  respectively). The difference between the moderate and low coffee consumers was small ( $d = 0.23$ ).

APA format: Lifetime coffee consumption was related to large significant differences in pineal gland size,  $F(2, 18) = 5.54$ ,  $p < .05$ ,  $f = 0.66$ . High coffee consumers had significantly smaller pineal glands ( $M = 3.71$ ) compared to moderate ( $M = 4.71$ ),  $HSD(3, 18) = 4.35$ ,  $p < .05$ ,  $d = 1.64$ , or low ( $M = 4.57$ ),  $HSD(3, 18) = 3.74$ ,  $p < .05$ ,  $d = 1.41$  coffee consumers. Each of these differences was large. There was a small, but not significant, difference in pineal gland size between low and moderate consumers,  $HSD(3, 18) = -0.61$ ,  $p > .05$ ,  $d = 0.23$ .

**Explanation:**

Low Coffee Consumption	$X^2_{Low}$	Moderate Coffee Consumption	$X^2_{Mod}$	High Coffee Consumption	$X^2_{High}$	
5	25	5	25	3	9	
4	16	5	25	4	16	
4	16	5	25	4	16	
4	16	4	16	4	16	
5	25	4	16	4	16	
5	25	4	16	3	9	
5	25	6	36	4	16	
$\Sigma X_{Low} = 32$	$\Sigma X^2_{Low} = 148$	$\Sigma X_{Mod} = 33$	$\Sigma X^2_{Mod} = 159$	$\Sigma X_{High} = 26$	$\Sigma X^2_{High} = 98$	$\Sigma X_{tot} = 91$
$\bar{X}_{Low} = 4.57$		$\bar{X}_{Mod} = 4.71$		$\bar{X}_{High} = 3.71$		$\Sigma X^2_{tot} = 405$

$$SS_{tot} = \sum X_{tot}^2 - \frac{(\Sigma X_{tot})^2}{N_{tot}} = 405 - \frac{91^2}{21} = 405 - \frac{8281}{21} = 405 - 394.333 = 10.67$$

$$SS_{treat} = \sum \left[ \frac{(\Sigma X_t)^2}{N_t} \right] - \frac{(\Sigma X_{tot})^2}{N_{tot}} = \left[ \frac{(32)^2}{7} \right] + \left[ \frac{(33)^2}{7} \right] + \left[ \frac{(26)^2}{7} \right] - 394.333$$

$$SS_{treat} = \left[ \frac{1024}{7} \right] + \left[ \frac{1089}{7} \right] + \left[ \frac{676}{7} \right] - 394.333$$

$$SS_{treat} = 146.286 + 155.571 + 96.571 - 394.333$$

$$SS_{treat} = 398.428 - 394.333 = 4.10$$

$$SS_{error} = \sum \left[ \sum X_t^2 - \frac{(\Sigma X_t)^2}{N_t} \right]$$

$$SS_{error} = \left( 148 - \frac{(32)^2}{7} \right) + \left( 159 - \frac{(33)^2}{7} \right) + \left( 98 - \frac{(26)^2}{7} \right)$$

$$SS_{error} = (148 - 146.286) + (159 - 155.571) + (98 - 96.571)$$

$$SS_{error} = 6.57$$

$$df_{tot} = N_{tot} - 1 = 21 - 1 = 20$$

$$df_{treat} = K - 1 = 3 - 1 = 2$$

$$df_{error} = N_{tot} - K = 21 - 3 = 18$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{4.10}{2} = 2.05$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{6.57}{18} = 0.37$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{2.05}{0.37} = 5.54$$

For the Tukey tests, the critical value is  $HSD_{.05}(3, 18) = 3.61$

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{0.37}{7}} = \sqrt{0.053} = 0.23$$

$$HSD_{low \ v. \ mod} = \frac{\bar{X}_{low} - \bar{X}_{mod}}{s_{\bar{x}}} = \frac{4.57 - 4.71}{0.23} = \frac{-0.14}{0.23} = -0.61 \ ns$$

$$\text{HSD}_{\text{low v. high}} = \frac{\bar{X}_{\text{low}} - \bar{X}_{\text{high}}}{s_{\bar{x}}} = \frac{4.57 - 3.71}{0.23} = \frac{0.86}{0.23} = 3.74 *$$

$$\text{HSD}_{\text{mod v. high}} = \frac{\bar{X}_{\text{mod}} - \bar{X}_{\text{high}}}{s_{\bar{x}}} = \frac{4.71 - 3.71}{0.23} = \frac{1.00}{0.23} = 4.35 *$$

$$f = \frac{\sqrt{\frac{K-1}{N_{\text{tot}}} (MS_{\text{treat}} - MS_{\text{error}})}}{\sqrt{MS_{\text{error}}}} = \frac{\sqrt{\frac{3-1}{21} (2.05 - 0.37)}}{\sqrt{0.37}}$$

$$f = \frac{\sqrt{0.095(1.68)}}{.608} = \frac{0.400}{.608} = .66 \text{ large}$$

$$d_{\text{low v. mod}} = \frac{\bar{X}_{\text{low}} - \bar{X}_{\text{mod}}}{\sqrt{MS_{\text{error}}}} = \frac{4.57 - 4.71}{\sqrt{0.37}} = \frac{-0.140}{0.608} = -0.23 \text{ small}$$

$$d_{\text{low v. high}} = \frac{\bar{X}_{\text{low}} - \bar{X}_{\text{high}}}{\sqrt{MS_{\text{error}}}} = \frac{4.57 - 3.71}{\sqrt{0.37}} = \frac{0.860}{0.608} = 1.41 \text{ large}$$

$$d_{\text{mod v. high}} = \frac{\bar{X}_{\text{mod}} - \bar{X}_{\text{high}}}{\sqrt{MS_{\text{error}}}} = \frac{4.71 - 3.71}{\sqrt{0.37}} = \frac{1.000}{0.608} = 1.64 \text{ large}$$

Note: If you are interested, Park et al. (2018) also determined that smaller pineal glands were associated with poorer sleep in older individuals.

4.

Source	SS	df	MS	F	p
Treatment	2249.33	2	1124.67	5.41	< .05
Error	3118.67	15	207.91		
TOTAL	5368.00	17			

Critical value:  $F_{.05}(2, 15) = 3.68$

According to the data, we reject the null hypothesis. What children are primed to think about has a significant effect on how much they were willing to share,  $F(2, 15) = 5.41, p < .05$ . The effect size index,  $f = 0.70$ , was large. Follow-up Tukey HSD tests showed that children primed to think about money shared a significantly smaller percentage of their stickers ( $\bar{X}_{\text{money}} = 22.67\%$ ) compared to children primed to think about candy ( $\bar{X}_{\text{candy}} = 44.33\%$ ) (HSD = -3.68) and buttons ( $\bar{X}_{\text{button}} = 48.00\%$ ) (HSD = -4.30). There was no significant difference in sharing between children in the two neutral conditions (candy and buttons) (HSD = -0.62). The differences in sharing between the children primed to think about money and the two neutral conditions were both large ( $d_{\text{money v. candy}} = 1.50, d_{\text{money v. button}} = 1.76$ ). The difference between the two neutral conditions was small ( $d = 0.25$ ).

APA format: What children were primed to think about had a large significant effect on how much they were willing to share,  $F(2, 15) = 5.41, p < .05, f = 0.70$ . Children primed to think about money shared a significantly smaller percentage of their stickers ( $M = 22.67\%$ ) compared to children primed to think about candy ( $M = 44.33\%$ ),  $\text{HSD}(3, 15) = -3.68, p < .05, d = 1.50$ , or those primed to think about buttons ( $M = 48.00\%$ ),  $\text{HSD}(3, 15) = -4.30, p < .05, d = 1.76$ . Each of these differences was large. There was a small, but not significant, difference in how much children were likely to share between the two neutral conditions of candy and buttons,  $\text{HSD}(3, 15) = -0.62, p > .05, d = 0.25$ .

**Explanation:**

Money Prime	$X^2_{Money}$	Candy Prime	$X^2_{Candy}$	Button Prime	$X^2_{Button}$	
26	676	45	2025	27	729	
25	625	43	1849	28	784	
17	289	35	1225	69	4761	
33	1089	47	2209	64	4096	
35	1225	43	1849	67	4489	
0	0	53	2809	33	1089	
$\Sigma X_{Money} = 136$	$\Sigma X^2_{Money} = 3904$	$\Sigma X_{Candy} = 266$	$\Sigma X^2_{Candy} = 11966$	$\Sigma X_{Button} = 288$	$\Sigma X^2_{Button} = 15948$	$\Sigma X_{tot} = 690$
$\bar{X}_{Money} = 22.67$		$\bar{X}_{Candy} = 44.33$		$\bar{X}_{Button} = 48.00$		$\Sigma X^2_{tot} = 31818$

$$SS_{tot} = \sum X_{tot}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 31818 - \frac{690^2}{18} = 31818 - \frac{476100}{18} = 31818 - 26450 = 5368$$

$$SS_{treat} = \sum \left[ \frac{(\sum X_t)^2}{N_t} - \frac{(\sum X_{tot})^2}{N_{tot}} \right] = \left[ \frac{(136)^2}{6} \right] + \left[ \frac{(266)^2}{6} \right] + \left[ \frac{(288)^2}{6} \right] - 26450$$

$$SS_{treat} = \left[ \frac{18496}{6} \right] + \left[ \frac{70756}{6} \right] + \left[ \frac{82944}{6} \right] - 26450$$

$$SS_{treat} = 3082.667 + 11792.667 + 13824 - 26450$$

$$SS_{treat} = 28699.334 - 26450 = 2249.33$$

$$SS_{error} = \sum \left[ \sum X_t^2 - \frac{(\sum X_t)^2}{N_t} \right]$$

$$SS_{error} = \left( 3904 - \frac{(136)^2}{6} \right) + \left( 11966 - \frac{(266)^2}{6} \right) + \left( 15948 - \frac{(288)^2}{6} \right)$$

$$SS_{error} = (3904 - 3082.667) + (11966 - 11792.667) + (15948 - 13824)$$

$$SS_{error} = 3118.67$$

$$df_{tot} = N_{tot} - 1 = 18 - 1 = 17$$

$$df_{treat} = K - 1 = 3 - 1 = 2$$

$$df_{error} = N_{tot} - K = 18 - 3 = 15$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{2249.33}{2} = 1124.67$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{3118.67}{15} = 207.91$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{1124.67}{207.91} = 5.41$$

For the Tukey tests, the critical value is  $HSD_{.05}(3, 15) = 3.67$

$$s_{\bar{X}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{207.91}{6}} = \sqrt{34.652} = 5.89$$

$$HSD_{money \ v. \ candy} = \frac{\bar{X}_{money} - \bar{X}_{candy}}{s_{\bar{X}}} = \frac{22.67 - 44.33}{5.89} = \frac{-21.66}{5.89} = -3.68 *$$

$$HSD_{money \ v. \ button} = \frac{\bar{X}_{money} - \bar{X}_{button}}{s_{\bar{X}}} = \frac{22.67 - 48.00}{5.89} = \frac{-25.33}{5.89} = -4.30 *$$

$$HSD_{candy \ v. \ button} = \frac{\bar{X}_{candy} - \bar{X}_{button}}{s_{\bar{X}}} = \frac{44.33 - 48.00}{5.89} = \frac{-3.67}{5.89} = -0.62 \ ns$$



$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{3-1}{18}(1124.67 - 207.91)}}{\sqrt{207.91}}$$

$$f = \frac{\sqrt{0.111(916.76)}}{14.419} = \frac{10.088}{14.419} = .70 \text{ large}$$

$$d_{money \text{ v. } candy} = \frac{\bar{X}_{money} - \bar{X}_{candy}}{\sqrt{MS_{error}}} = \frac{22.67 - 44.33}{\sqrt{207.91}} = \frac{-21.66}{14.419} = -1.50 \text{ large}$$

$$d_{money \text{ v. } button} = \frac{\bar{X}_{money} - \bar{X}_{button}}{\sqrt{MS_{error}}} = \frac{22.67 - 48.00}{\sqrt{207.91}} = \frac{-25.33}{14.419} = -1.76 \text{ large}$$

$$d_{candy \text{ v. } button} = \frac{\bar{X}_{candy} - \bar{X}_{button}}{\sqrt{MS_{error}}} = \frac{44.33 - 48.00}{\sqrt{207.91}} = \frac{-3.67}{14.419} = -0.25 \text{ small}$$

Note: If you are interested, Gasiorowska et al. (2016) thought interacting with buttons might not lead to the same enhancement of mood as interacting with money. Therefore, they used a second neutral condition, candy, that they thought would produce positive mood like money, but was not money. Analyses suggested they were correct: Both money and candy enhanced mood to the same degree and buttons did not. This suggests it wasn't an enhancement of mood that explained the differences you saw in sharing behavior. They also determined that children who sorted money took more stickers in the first place than those who sorted candy or buttons.

5.

Source	SS	df	MS	F	p
Treatment	68.60	3	22.87	7.83	< .05
Error	105.00	36	2.92		
TOTAL	173.60	39			

Critical value:  $F_{.05}(3, 36) = 2.86$

According to the data, we reject the null hypothesis. Type of pain relief administered to a mother during delivery significantly effects the well-being of the newborn,  $F(3, 36) = 7.84, p < .05$ . The effect size index,  $f = 0.71$ , was very large. Follow-up Tukey HSD tests showed that newborns whose mothers were administered sedatives had significantly lower Apgar scores five minutes after birth ( $\bar{X}_{Sedatives} = 4.90$ ) compared to all three other groups ( $\bar{X}_{Spinal \text{ block}} = 7.40$ , HSD = -4.63;  $\bar{X}_{Opiate} = 7.80$ , HSD = -5.37;  $\bar{X}_{Lamaze} = 8.30$ , HSD = -6.30). There were no significant differences between the Spinal Block, Opiate, and Lamaze groups. The differences between Apgar scores of newborns whose mothers were administered sedative and those given spinal block ( $d = 1.46$ ), opiate ( $d = 1.70$ ), and Lamaze ( $d = 1.99$ ) were large. The difference between spinal block and Lamaze was medium ( $d = .53$ ). The differences between spinal block and opiates ( $d = 0.23$ ) and opiate and Lamaze ( $d = .29$ ) were both small.

APA format: Type of pain relief administered to a mother during delivery has a large significant effect on the well-being of the newborn,  $F(3, 36) = 7.84, p < .05, f = 0.71$ . There are large significant reductions in Apgar scores of newborns five minutes after birth whose mothers were administered sedatives ( $M = 4.90$ ) compared to all three other groups ( $M_{Spinal \text{ block}} = 7.40$ , HSD (3, 9) = -4.63,  $p < .05, d = 1.46$ ;  $M_{Opiate} = 7.80$ , HSD (3, 9) = -5.37,  $p < .05, d = 1.70$ ;  $M_{Lamaze} = 8.30$ , HSD (3, 9) = -6.30,  $p < .05, d = 1.99$ ). The difference between spinal block and Lamaze was medium,  $d = .53$ . The differences between spinal block and opiates,  $d = 0.23$ , and opiate and Lamaze,  $d = .29$ , were both small, but not significant (all  $ps > .05$ ).

	Sedatives	Spinal Block	Opiate	Lamaze
$\Sigma X$	49	74	78	83
$\Sigma X^2$	269	574	636	711
$N$	10	10	10	10
$\bar{X}$	49/10 = 4.9	74/10 = 7.4	78/10 = 7.8	83/10 = 8.3

$$\Sigma X_{tot} = 49 + 74 + 78 + 83 = 284$$

$$\Sigma X^2_{tot} = 269 + 574 + 636 + 711 = 2190$$

$$SS_{tot} = \sum X_{tot}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 2190 - \frac{284^2}{40} = 2190 - \frac{80656}{40} = 2190 - 2016.40 = 173.60$$

$$SS_{treat} = \sum \left[ \frac{(\sum X_t)^2}{N_t} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[ \frac{(49)^2}{10} \right] + \left[ \frac{(74)^2}{10} \right] + \left[ \frac{(78)^2}{10} \right] + \left[ \frac{(83)^2}{10} \right] - 2016.40$$

$$SS_{treat} = \left[ \frac{2401}{10} \right] + \left[ \frac{5476}{10} \right] + \left[ \frac{6084}{10} \right] + \left[ \frac{6889}{10} \right] - 2016.40$$

$$SS_{treat} = 240.10 + 547.60 + 608.40 + 688.90 - 2016.40$$

$$SS_{treat} = 2085 - 2016.40 = 68.60$$

$$SS_{error} = \sum \left[ \sum X_t^2 - \frac{(\sum X_t)^2}{N_t} \right]$$

$$SS_{error} = \left( 269 - \frac{(49)^2}{10} \right) + \left( 574 - \frac{(74)^2}{10} \right) + \left( 636 - \frac{(78)^2}{10} \right) + \left( 711 - \frac{(83)^2}{10} \right)$$

$$SS_{error} = (269 - 240.10) + (574 - 547.60) + (636 - 608.40) + (711 - 688.90)$$

$$SS_{error} = 105.00$$

$$df_{tot} = N_{tot} - 1 = 40 - 1 = 39$$

$$df_{treat} = K - 1 = 4 - 1 = 3$$

$$df_{error} = N_{tot} - K = 40 - 4 = 36$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{68.60}{3} = 22.87$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{105.00}{36} = 2.92$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{22.87}{2.92} = 7.83$$

For the Tukey tests, the critical value is  $HSD_{.05}(3, 36) = 3.84$ . Because there is no row labeled 36, use the next smallest  $df$  on the table (here, 30).

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{2.92}{10}} = \sqrt{0.292} = 0.54$$

$$HSD_{Sedatives\ v.\ Spinal} = \frac{\bar{X}_{Sedatives} - \bar{X}_{Spinal}}{s_{\bar{x}}} = \frac{4.90 - 7.40}{0.54} = \frac{-2.50}{0.54} = -4.63 *$$

$$HSD_{Sedatives\ v.\ Opiate} = \frac{\bar{X}_{Sedatives} - \bar{X}_{Opiate}}{s_{\bar{x}}} = \frac{4.90 - 7.80}{0.54} = \frac{-2.90}{0.54} = -5.37 *$$

$$HSD_{Sedatives\ v.\ Lamaze} = \frac{\bar{X}_{Sedatives} - \bar{X}_{Lamaze}}{s_{\bar{x}}} = \frac{4.90 - 8.30}{0.54} = \frac{-3.40}{0.54} = -6.30 *$$

$$HSD_{Spinal\ v.\ Opiate} = \frac{\bar{X}_{Spinal} - \bar{X}_{Opiate}}{s_{\bar{x}}} = \frac{7.40 - 7.80}{0.54} = \frac{-0.40}{0.54} = -0.74\ ns$$

$$HSD_{Spinal\ v.\ Lamaze} = \frac{\bar{X}_{Spinal} - \bar{X}_{Lamaze}}{s_{\bar{x}}} = \frac{7.40 - 8.30}{0.54} = \frac{-0.90}{0.54} = -1.67\ ns$$

$$HSD_{Opiate\ v.\ Lamaze} = \frac{\bar{X}_{Opiate} - \bar{X}_{Lamaze}}{s_{\bar{x}}} = \frac{7.80 - 8.30}{0.54} = \frac{-0.50}{0.54} = -0.93\ ns$$

$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{4-1}{40}(22.87 - 2.92)}}{\sqrt{2.92}}$$

$$f = \frac{\sqrt{0.075(19.95)}}{1.709} = \frac{1.22}{1.709} = 0.71 \text{ large}$$

$$d_{Sedatives v.Spinal} = \frac{\bar{X}_{Sedatives} - \bar{X}_{Spinal}}{\sqrt{MS_{error}}} = \frac{4.90 - 7.40}{\sqrt{2.92}} = \frac{-2.50}{1.709} = -1.46 \text{ large}$$

$$d_{Sedatives v.Opiate} = \frac{\bar{X}_{Sedatives} - \bar{X}_{Opiate}}{\sqrt{MS_{error}}} = \frac{4.90 - 7.80}{\sqrt{2.92}} = \frac{-2.90}{1.709} = -1.70 \text{ large}$$

$$d_{Sedatives v.Lamaze} = \frac{\bar{X}_{Sedatives} - \bar{X}_{Lamaze}}{\sqrt{MS_{error}}} = \frac{4.90 - 8.30}{\sqrt{2.92}} = \frac{-3.40}{1.709} = -1.99 \text{ large}$$

$$d_{Spinal v.Opiate} = \frac{\bar{X}_{Spinal} - \bar{X}_{Opiate}}{\sqrt{MS_{error}}} = \frac{7.40 - 7.80}{\sqrt{2.92}} = \frac{-0.40}{1.709} = -0.23 \text{ small}$$

$$d_{Spinal v.Lamaze} = \frac{\bar{X}_{Spinal} - \bar{X}_{Lamaze}}{\sqrt{MS_{error}}} = \frac{7.40 - 8.30}{\sqrt{2.92}} = \frac{-0.90}{1.709} = -0.53 \text{ medium}$$

$$d_{Opiate v.Lamaze} = \frac{\bar{X}_{Opiate} - \bar{X}_{Lamaze}}{\sqrt{MS_{error}}} = \frac{7.80 - 8.30}{\sqrt{2.92}} = \frac{-0.50}{1.709} = -0.29 \text{ small}$$

6.

Source	SS	df	MS	F	p
Treatment	10.13	2	5.07	4.12	< .05
Error	14.80	12	1.23		
TOTAL	24.93	14			

Critical value:  $F_{.05}(2, 12) = 3.88$

According to the data, we reject the null hypothesis. Time spent connecting with co-workers through Facebook was significantly related to differences in job satisfaction,  $F(2, 12) = 4.12, p < .05$ . The effect size index,  $f = 0.64$ , was large. Follow-up Tukey HSD tests showed that people spending the lowest amount of time connecting with co-workers on Facebook (less than 10 minutes per day) were significantly less satisfied with their jobs ( $\bar{X}_{Low} = 3.00$ ) than people spending the highest amounts of time (31 minutes to 2 hours per day) ( $\bar{X}_{High} = 5.00$ ),  $HSD = -4.00$ . People spending medium amounts of time connecting with co-workers through Facebook (10 to 30 minutes per day) ( $\bar{X}_{Medium} = 3.80$ ) were not significantly different in job satisfaction from either the low ( $HSD = -1.60$ ) or high ( $HSD = -2.40$ ) groups. The difference in job satisfaction between the low and high users was large ( $d = 1.80$ ), as was the difference between medium and high users ( $d = 1.08$ ), even though this difference was not significant. The difference between the medium and low users was medium to large ( $d = 0.72$ ). These researchers might consider repeating their study with a larger sample size to increase power.

APA format: Time spent connecting with co-workers through Facebook had a large significant association with job satisfaction,  $F(2, 12) = 4.12, p < .05, f = 0.64$ . Employees who spend the least amount of time connecting to co-workers through Facebook (less than 10 minutes per day) had significantly lower job satisfaction ( $M = 3.00$ ) than employees who spend the most time (31 minutes to 2 hours) connecting in this way ( $M = 5.00$ ),  $HSD(3, 12) = -4.00, p < .05, d = 1.80$ . This difference was very large. Medium users (10 to 30 minutes per day) fell between these two groups on job satisfaction ( $M = 3.80$ ). However the medium-use group did not significantly differ in job satisfaction from either the low-use group,  $HSD(3, 12) = -1.60, d = -0.72$ , or the high-use group,  $HSD(3, 12) = -2.40, d = 1.08$  (both  $ps > .05$ ), though the magnitudes of these differences were medium to large.

**Explanation:**

Low Connection	$X^2_{Low}$	Medium Connection	$X^2_{Medium}$	High Connection	$X^2_{High}$	
2	4	2	4	5	25	
3	9	6	36	5	25	
4	16	4	16	4	16	
4	16	4	16	6	36	
2	4	3	9	5	25	
$\Sigma X_{Low} = 15$	$\Sigma X^2_{Low} = 49$	$\Sigma X_{Med} = 19$	$\Sigma X^2_{Medium} = 81$	$\Sigma X_{High} = 25$	$\Sigma X^2_{High} = 127$	$\Sigma X_{tot} = 59$
$\bar{X}_{Low} = 3.00$		$\bar{X}_{Med} = 3.80$		$\bar{X}_{High} = 5.00$		$\Sigma X^2_{tot} = 257$

$$SS_{tot} = \sum X_{tot}^2 - \frac{(\Sigma X_{tot})^2}{N_{tot}} = 257 - \frac{59^2}{15} = 257 - \frac{3481}{15} = 257 - 232.067 = 24.93$$

$$SS_{treat} = \sum \left[ \frac{(\Sigma X_t)^2}{N_t} \right] - \frac{(\Sigma X_{tot})^2}{N_{tot}} = \left[ \frac{(15)^2}{5} \right] + \left[ \frac{(19)^2}{5} \right] + \left[ \frac{(25)^2}{5} \right] - 232.067$$

$$SS_{treat} = \left[ \frac{225}{5} \right] + \left[ \frac{361}{5} \right] + \left[ \frac{625}{5} \right] - 232.067$$

$$SS_{treat} = 45 + 72.20 + 125 - 232.067$$

$$SS_{treat} = 242.20 - 232.067 = 10.13$$

$$SS_{error} = \sum \left[ \sum X_t^2 - \frac{(\Sigma X_t)^2}{N_t} \right]$$

$$SS_{error} = \left( 49 - \frac{(15)^2}{5} \right) + \left( 81 - \frac{(19)^2}{5} \right) + \left( 127 - \frac{(25)^2}{5} \right)$$

$$SS_{error} = (49 - 45) + (81 - 72.20) + (127 - 125)$$

$$SS_{error} = 14.80$$

$$df_{tot} = N_{tot} - 1 = 15 - 1 = 14$$

$$df_{treat} = K - 1 = 3 - 1 = 2$$

$$df_{error} = N_{tot} - K = 15 - 3 = 12$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{10.13}{2} = 5.07$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{14.80}{12} = 1.23$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{5.07}{1.23} = 4.12$$

For the Tukey tests, the critical value is  $HSD_{.05}(3, 12) = 3.77$

$$s_{\bar{X}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{1.23}{5}} = \sqrt{0.246} = 0.50$$

$$HSD_{low \ v. \ medium} = \frac{\bar{X}_{low} - \bar{X}_{medium}}{s_{\bar{X}}} = \frac{3.00 - 3.80}{0.50} = \frac{-0.80}{0.50} = -1.60 \ ns$$

$$HSD_{low \ v. \ high} = \frac{\bar{X}_{low} - \bar{X}_{high}}{s_{\bar{X}}} = \frac{3.00 - 5.00}{0.50} = \frac{-2.00}{0.50} = -4.00 \ *$$

$$HSD_{medium \ v. \ high} = \frac{\bar{X}_{medium} - \bar{X}_{high}}{s_{\bar{X}}} = \frac{3.80 - 5.00}{0.50} = \frac{-1.20}{0.50} = -2.40 \ ns$$

$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{3-1}{15}(5.07 - 1.23)}}{\sqrt{1.23}}$$

$$f = \frac{\sqrt{0.133(3.84)}}{1.109} = \frac{0.715}{1.109} = .64 \text{ large}$$

$$d_{low \text{ v. } medium} = \frac{\bar{X}_{low} - \bar{X}_{medium}}{\sqrt{MS_{error}}} = \frac{3.00 - 3.80}{\sqrt{1.23}} = \frac{-0.80}{1.109} = -0.72 \text{ medium to large}$$

$$d_{low \text{ v. } high} = \frac{\bar{X}_{low} - \bar{X}_{high}}{\sqrt{MS_{error}}} = \frac{3.00 - 5.00}{\sqrt{1.23}} = \frac{-2.00}{1.109} = -1.80 \text{ large}$$

$$d_{medium \text{ v. } high} = \frac{\bar{X}_{medium} - \bar{X}_{high}}{\sqrt{MS_{error}}} = \frac{3.80 - 5.00}{\sqrt{1.23}} = \frac{-1.20}{1.109} = -1.08 \text{ large}$$

Source	SS	df	MS	F	p
Treatment	122.00	2	61.00	11.21	< .05
Error	49.00	9	5.44		
TOTAL	171.00	11			

Critical value:  $F_{.05}(2, 9) = 4.26$

According to the data, we reject the null hypothesis. County group membership is significantly associated with willingness to lend corn,  $F(2, 9) = 11.21, p < .05$ . The effect size index,  $f = 1.31$ , was very large. Follow-up Tukey HSD tests showed that Countrymen lent significantly more bushels of corn ( $\bar{X}_{CM} = 13.00$ ) compared to Friends ( $\bar{X}_{Friends} = 6.00$ ) (HSD = -5.98) and Romans ( $\bar{X}_{Romans} = 6.50$ ) (HSD = -5.56). There was no significant difference in how much corn Friends and Romans were willing to give (HSD = -0.43). The differences between the number of bushels lent by Countrymen and both Friends ( $d = 3.00$ ) and Romans ( $d = 2.79$ ) were huge. The difference between the Friends and Romans was small ( $d = 0.21$ ).

APA format: County group membership is largely and significantly associated with willingness to lend corn,  $F(2, 9) = 11.21, p < .05, f = 1.31$ . Countrymen lent significantly more corn ( $M = 13.00$ ) than either Friends ( $M = 6.00$ ), HSD (3, 9) = -5.98,  $p < .05, d = 3.00$ , or Romans ( $M = 6.50$ ), HSD (3, 9) = -5.56,  $p < .05, d = 2.79$ . Both of these differences were huge. There was a small, but not significant, difference in willingness to lend corn between Friends and Romans, HSD (3, 9) = -0.43,  $p > .05, d = 0.21$ .

#### Explanation:

Friends	$X^2_{Friends}$	Romans	$X^2_{Romans}$	Countrymen	$X^2_{CM}$	
7	49	8	64	10	100	
9	81	7	49	12	144	
3	9	4	16	16	256	
5	25	7	49	14	196	
$\Sigma X_{Friends} = 24$	$\Sigma X^2_{Friends} = 164$	$\Sigma X_{Romans} = 26$	$\Sigma X^2_{Romans} = 178$	$\Sigma X_{CM} = 52$	$\Sigma X^2_{CM} = 696$	$\Sigma X_{tot} = 102$
$\bar{X}_{Friends} = 6.00$		$\bar{X}_{Romans} = 6.50$		$\bar{X}_{CM} = 13.00$		$\Sigma X^2_{tot} = 1038$

$$SS_{tot} = \sum X_{tot}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 1038 - \frac{102^2}{12} = 1038 - \frac{10404}{12} = 1038 - 867 = 171$$

$$SS_{treat} = \sum \left[ \frac{(\sum X_t)^2}{N_t} - \frac{(\sum X_{tot})^2}{N_{tot}} \right] = \left[ \frac{(24)^2}{4} \right] + \left[ \frac{(26)^2}{4} \right] + \left[ \frac{(52)^2}{4} \right] - 867$$

$$SS_{treat} = \left[ \frac{576}{4} \right] + \left[ \frac{676}{4} \right] + \left[ \frac{2704}{4} \right] - 867$$

$$SS_{treat} = 144 + 169 + 676 - 867$$

$$SS_{treat} = 989 - 867 = 122$$

$$SS_{error} = \sum \left[ \sum X_t^2 - \frac{(\sum X_t)^2}{N_t} \right]$$

$$SS_{error} = \left(164 - \frac{(24)^2}{4}\right) + \left(178 - \frac{(26)^2}{4}\right) + \left(696 - \frac{(52)^2}{4}\right)$$

$$SS_{error} = (164 - 144) + (178 - 169) + (696 - 676)$$

$$SS_{error} = 49$$

$$df_{tot} = N_{tot} - 1 = 12 - 1 = 11$$

$$df_{treat} = K - 1 = 3 - 1 = 2$$

$$df_{error} = N_{tot} - K = 12 - 3 = 9$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{122}{2} = 61$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{49}{9} = 5.44$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{61}{5.44} = 11.21$$

For the Tukey tests, the critical value is  $HSD_{.05}(3, 12) = 3.77$

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{5.44}{4}} = \sqrt{1.36} = 1.17$$

$$HSD_{Friends \ v. \ Romans} = \frac{\bar{X}_{Friends} - \bar{X}_{Romans}}{s_{\bar{x}}} = \frac{6.00 - 6.50}{1.17} = \frac{-0.50}{1.17} = -0.43 \text{ ns}$$

$$HSD_{Friends \ v. \ CM} = \frac{\bar{X}_{Friends} - \bar{X}_{CM}}{s_{\bar{x}}} = \frac{6.00 - 13.00}{1.17} = \frac{-7.00}{1.17} = -5.98 *$$

$$HSD_{Romans \ v. \ CM} = \frac{\bar{X}_{Romans} - \bar{X}_{CM}}{s_{\bar{x}}} = \frac{6.50 - 13.00}{1.17} = \frac{-6.50}{1.17} = -5.56 *$$

$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{3-1}{12}(61.00 - 5.44)}}{\sqrt{5.44}}$$

$$f = \frac{\sqrt{0.167(55.56)}}{2.33} = \frac{3.046}{2.332} = 1.31 \text{ large}$$

$$d_{Friends \ v. \ Romans} = \frac{\bar{X}_{Friends} - \bar{X}_{Romans}}{\sqrt{MS_{error}}} = \frac{6.00 - 6.50}{\sqrt{5.44}} = \frac{-0.50}{2.33} = -0.21 \text{ small}$$

$$d_{Friends \ v. \ CM} = \frac{\bar{X}_{Friends} - \bar{X}_{CM}}{\sqrt{MS_{error}}} = \frac{6.00 - 13.00}{\sqrt{5.44}} = \frac{-7.00}{2.33} = -3.00 \text{ huge!}$$

$$d_{Romans \ v. \ CM} = \frac{\bar{X}_{Romans} - \bar{X}_{CM}}{\sqrt{MS_{error}}} = \frac{6.50 - 13.00}{\sqrt{5.44}} = \frac{-6.50}{2.33} = -2.79 \text{ huge!}$$

## **References**

Gasiorowska, A., Chaplin, L. N., Zaleskiewicz, T., Wygrab, S., & Vohs, K. (2016). Money cues increase agency and decrease prosociality among children: Early signs of market-mode behaviors. *Psychological Science*, 27(3), 331-344. doi: 10.1177/0956797615620378

Hanna, B., Kee, K., & Robertson, B. W. (2017). Positive impacts of social media at work: Job satisfaction, job calling, and Facebook use among co-workers. *SHS Web of Conferences*, 33. doi: 10.1051/shsconf/20173300012

Park, J., Han, J. W., Lee, J. R., Byun, S., Suh, S. W., Kim, T., . . . Kim, K. W. (2018). Lifetime coffee consumption, pineal gland volume, and sleep quality in late life. *Sleep*, Advance online publication. doi: 10.1093/sleep/zsy127