CHAPTER 12

Analysis of Variance: Repeated-measures

Summary

The design described in this chapter is the ANOVA version of the paired-samples *t* test. The only difference is that a **one-factor repeated-measures ANOVA** can analyze more than two levels of the independent variable, while a *t* test can only analyze two levels.

Like a one-way ANOVA, a one-factor repeated-measures ANOVA has one *independent variable* (factor) that has two or more levels and one *dependent variable*. However, when using a one-factor repeated-measures ANOVA, there is a reason to match up the scores in the different treatments.

In a display of the data for a one-factor repeated-measures ANOVA, columns represent the treatments and each row represents a matched group of scores. A one-factor repeated-measures ANOVA partitions the total variance in the dependent variable scores into *three* identifiable sources. The first source of variability is due to the independent variable (the treatment; MS_{treat}). The second source of variance is due to participants (called between subject variance). The third source of variance is the error variance (MS_{error}). Just like in the one-way ANOVA from the last chapter, the *F* test for the one-factor repeated-measures, which determines if there are any significant differences among the treatment means, is a ratio of the mean square for the treatments divided by the mean square for error.

As a comparison, a one-way ANOVA partitions the total variance into only two sources, treatment variance and error variance. A one-factor repeated-measures ANOVA, however, removes the effect of participants from the total variance, which makes it a more powerful test than a one-way ANOVA. This occurs because participants are either matched on some specific, important criteria (for example, they all come from the same family), or because the same participants experience all of the treatments.

One advantage of one-factor repeated-measures ANOVA is that it is more efficient; that is, it requires less time and effort to obtain a given number of scores when using a repeated-measures pairing. For example, you might only need to recruit 10 participants that you test three times rather than 30 separate participants you test once each to get the same data in a one-way ANOVA. A second advantage is that it is more powerful because the error variance (the denominator of the *F* test) is smaller. The error variance is smaller because the variance due to the participants has been removed. A disadvantage of a one-factor repeated-measures ANOVA is that it may contain unwanted *carryover* effects, which occur when one level of the treatment continues to affect the participant's during subsequent treatments.

Like the one-way ANOVA, it is expected that the population scores are normally distributed when conducting a one-factor repeated-measures ANOVA. Pairwise comparisons among the several treatments can still be made with Tukey HSD tests and effect size indices (*f* and *d*) can be used to determine the size of differences.

Multiple-Choice Questions

Data Set 12-1

Participants	Treatment A (X_A)	Treatment B (X_B)	Treatment C (X_C)	ΣX_k
1	2	3	7	12
2	3	5	6	14
ΣX_t	5	8	13	26

1. In Data set 12-1, SS_{tot} is equal to _____.

  a. 19.33
  b. 106.00
  c. 127.67;

d.   none of the above

2. In Data Set 12-1, SS_{treat} is equal to _____.

   a.   0.67
   b.   16.33
   c.   124.67
   d.   none of the above

3. In Data Set 12-1, MS_{error} is equal to _____.

   a.   2.33
   b.   1.50
   c.   1.17
   d.   none of the above

4. In Data Set 12-1, the F value for the repeated-measures ANOVA is _____.

   a.   3.50
   b.   4.44
   c.   10.20
   d.   none of the above

5. A repeated-measures ANOVA removes _____ variance from further consideration in the analysis.

   a.   between treatments
   b.   between subjects
   c.   error
   d.   all of the above

6. The F ratio in a repeated-measures ANOVA consists of the _____.

   a.   between subjects variance divided by the between treatments variance
   b.   between treatments variance divided by the between subjects variance
   c.   between subjects variance divided by the error variance
   d.   between treatments variance divided by the error variance

7. A one-factor repeated-measures ANOVA is like a one-way ANOVA with respect to _____.

   a.   the number of pieces that the total variance is partitioned into
   b.   its efficiency
   c.   its power
   d.   none of the above

8. The number of independent variables that can be handled by a one-factor repeated-ANOVA is _____.

   a.   1
   b.   2
   c.   3
   d.   cannot be determined from the information given

9. A one-factor repeated-measures ANOVA partitions the total variance into _____ component(s).

   a.   1
   b.   2
   c.   3

d. 4

10. A Type I error occurs when you _____. A Type II error occurs when you _____.

    a. retain a false null hypothesis; reject a true null hypothesis
    b. retain a false null hypothesis; retain a true null hypothesis
    c. reject a false null hypothesis; retain a true null hypothesis
    d. reject a true null hypothesis; retain a false null hypothesis

11. An F value of 3.50 with $df_{treat} = 4$ and $df_{error} = 10$ is obtained from the data. Using an $\alpha = .05$, you should _____ the null hypothesis, even though you might be making a _____ error.

    a. reject, Type I
    b. reject, Type II
    c. retain, Type I
    d. retain, Type II

12. For which of the descriptions that follow would a repeated-measures ANOVA be appropriate? A study in which _____.

    a. participants are matched along a variable of importance
    b. there are no unwanted carryover effects from one administration of one level of the independent variable to the next
    c. both a. and b.
    d. neither a. nor b.

13. What is the critical value for a repeated-measures ANOVA that has three subjects who provide data for four conditions? Let $\alpha = .05$.

    a. 3.84
    b. 4.76
    c. 5.14
    d. none of the above

14. Which of the following is an advantage of a repeated-measures ANOVA over one-way ANOVA?

    a. repeated-measures ANOVA is more efficient
    b. repeated-measures ANOVA eliminates variance due to participants
    c. repeated-measures ANOVA is more powerful
    d. all of the above

15. A psychologist finds a significant difference between two conditions. She is very happy, and decides to repeat the study. This time, she finds no significant difference. Suppose she runs the study ten additional times, and fails to find a significant difference. There is a good chance that the significant difference she found the first time was due to _____.

    a. Type I error
    b. Type II error
    c. both a. and b.
    d. cannot be determined from the information given

16. Type II errors are made when the researcher _____.

    a. rejects the null hypothesis when it shouldn't be rejected
    b. retains the null hypothesis when it shouldn't be rejected
    c. rejects the null hypothesis when it should be rejected

d.   retains the null hypothesis when it should be rejected

17. A one-factor repeated-measures ANOVA is like a one-way ANOVA with respect to _____.

   a.   the number of factors
   b.   the number of dependent variables
   c.   use of the F distribution
   d.   all of the above

18. An F value obtained from data is 3.45 based on $df_{treat} = 3$ and $df_{error} = 12$. You should _____ the null hypothesis, even though you might be making a _____ error.

   a.   reject, Type I
   b.   reject, Type II
   c.   retain, Type I
   d.   retain, Type II

19. The repeated-measures designs that are common with repeated-measures ANOVA are _____.

   a.   matched pairs
   b.   pre-test, post-test, follow-up
   c.   both a. and b.
   d.   neither a. nor b.

Short Answer Questions

1. List the advantages of using a repeated-measures design over an independent-samples design. If you were to design a study using a repeated-measure design, what would be your main concern?

2. Design a repeated-measures experiment you could analyze with a one-factor repeated-measures ANOVA. In doing so, identify a research hypothesis, the independent and dependent variables, and explain how your design is repeated measures.

Problems

1. Harvie et al. (2015) studied the influence of sensory information on the experience of neck pain. The researchers wanted to know if visual cues are an early signal about the potential experience of pain. In this study, participants were 24 adults who experienced moderate neck pain. The participants viewed scenes wearing virtual reality glasses and the scenes either understated how rotated their heads were, gave accurate visual feedback, or overstated how rotated their heads were. Participants were asked to rotate their head to the extent that they could without feeling pain while wearing the virtual reality glasses. Once they indicated they were uncomfortable, measures were taken of the extent of their head rotation. Below are summary data from Harvie et al. (2015; retrieved from http://thedata.harvard.edu/dvn/dv/PainData on July 17, 2018) that will allow you to replicate their findings. Conduct a repeated-measures ANOVA using $\alpha = .05$ to determine if the accuracy of sensory information influenced how far people could move before experiencing neck pain. Perform the F test, create an ANOVA summary table, calculate the effect size index f, and write a conclusion about the influence of sensory input on the experience of pain. To focus on the results the researchers were most interested in, conduct a Tukey HSD test to compare the difference between understated and overstated visual feedback,

Understated Visual Feedback (X_A)	Accurate Visual Feedback (X_B)	Overstated Visual Feedback (X_C)
$\Sigma X_A = 25.57$	$\Sigma X_B = 24$	$\Sigma X_C = 22.35$
$\Sigma X_A^2 = 27.49$	$\Sigma X_B^2 = 24$	$\Sigma X_C^2 = 21.03$

$\bar{X}_A = 1.07$	$\bar{X}_B = 1.00$	$\bar{X}_C = 0.93$

For this problem, because of the large number of participants (which would require you to sum 24 fractions), use this number as the first term when calculating $SS_{subjects}$:

$$\sum \left[\frac{(\sum X_k)^2}{N_k} \right] = 71.94$$

2. T. S. Spatz (1991) conducted workshops teaching women to conduct breast self-examinations. She gave participants a 27-item true/false test of their knowledge before they were trained (pre-test), after they were trained (post-test), and again three months later (follow-up). The scores below mimic her results. Perform a repeated-measures ANOVA on the data below with $\alpha = .01$, create an ANOVA summary table, calculate the effect size index f, conduct a Tukey HSD test for each pair of means, and write a conclusion about the influence of training on knowledge about breast self-examination.

Participant	Pre-test (X_A)	Post-test (X_B)	Follow-up (X_C)
1	7	22	16
2	10	20	22
3	10	23	24
4	13	21	16
5	15	26	22
6	17	26	26

3. In this study, children were tested three times, at ages 10, 11, and 12. At each age, they worked a set of 25 logic problems. The scores below represent the percentage of problems they got correct. Perform a repeated-measures ANOVA on the data below with $\alpha = .05$, create an ANOVA summary table, calculate the effect size index f, conduct a Tukey HSD test for each pair of means, and write a conclusion about the influence of age on performance on this logic test.

Participant	Age 10 (X_A)	Age 11 (X_B)	Age 12 (X_C)
1	0	8	40
2	8	12	36
3	12	24	36
4	16	20	48

4. Testing new drugs often involves giving participants each of several dosages of the drug, as well as a placebo (an inert substance, which serves as a control condition). In this study, the dependent variable was reaction time in hundredths of a second after taking the new drug. Perform a repeated-measures ANOVA on the data below with $\alpha = .05$, create an ANOVA summary table, calculate the effect size index f, conduct one Tukey HSD test of your choice, and write a conclusion about the influence of dosage on reaction time.

Participant	Placebo (0 mg) (X_A)	15 mg (X_B)	30 mg (X_C)	60 mg (X_D)
1	36	32	20	18
2	35	40	25	27
3	40	42	30	24

ANSWERS

Multiple-Choice Questions

1. a. 19.33
   Explanation: $\sum X_{tot} = 26$ $\sum X_{tot}^2 = 132$
   $$SS_{tot} = \sum X_{tot}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 132 - \frac{26^2}{6} = 132 - 112.67 = 19.33$$
   Remember that in a one-factor repeated measures ANOVA, N_{tot} is the total number of scores (here 6), not the total number of participants (here 2).

2. b. 16.33
   Explanation:
   $$SS_{treat} = \sum\left[\frac{(\sum X_t)^2}{N_t}\right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{5^2}{2}\right] + \left[\frac{8^2}{2}\right] + \left[\frac{13^2}{2}\right] - \frac{26^2}{6}$$

   $$SS_{treat} = 12.5 + 32 + 84.5 - 112.67 = 16.33$$

   Just like one-way ANOVA, in a one-factor repeated measures ANOVA, N_t is the total number of scores per treatment (here 2 for all groups).

3. c. 1.17
   Explanation:
   $$SS_{subjects} = \sum\left[\frac{(\sum X_k)^2}{N_k}\right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{12^2}{3}\right] + \left[\frac{14^2}{3}\right] - \frac{26^2}{6}$$

   $$SS_{subjects} = 48 + 65.33 - 112.67 = 0.66$$

   $$SS_{error} = SS_{tot} - SS_{subjects} - SS_{treat} = 19.33 - 0.66 - 16.33 = 2.34$$

   $$df_{error} = (N_t - 1)(N_k - 1) = (2 - 1)(3 - 1) = 2$$

   Remember that in a one-factor repeated measures ANOVA N_t is the number of scores in one group, N_k is the number of groups (levels of the independent variable).

   $$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{2.34}{2} = 1.17$$

4. d. none of the above
   Explanation: $df_{treat} = N_k - 1 = 3 - 1 = 2$ $MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{16.33}{2} = 8.17$

   $$F = \frac{MS_{treat}}{MS_{error}} = \frac{8.17}{1.17} = 6.98$$

5. b. between subjects
6. d. between treatments variance divided by the error variance
7. d. none of the above
   Explanation: Repeated-measures ANOVA is more powerful and efficient than one-way ANOVA. It also divides total variability into three parts (subjects, treatment, and error), whereas one-way ANOVA divides total variability into two parts (treatment and error). They are similar in that they both tell whether there are any differences between two or more group means for a study that has one independent variable and one dependent variable.
8. a. 1
9. c. 3

10. d. reject a true null hypothesis, retain a false null hypothesis
11. a. reject, Type I
   Explanation: Look up a critical value in the F distribution table. With $df_{treat} = 4$ and $df_{error} = 10$, you get $F_{.05} = 3.48$. Because the F you obtained here (3.50) is larger than the critical value, you can reject the null hypothesis.
   A Type I error is rejecting a true null hypothesis, and a Type II error is retaining a false null hypothesis. Because you can't make a Type II error when rejecting, only a Type I error is possible.
12. c. both a. and b.
13. b. 4.76
   Explanation: $df_{treat} = N_k - 1 = 4 - 1 = 3 \quad df_{error} = (N_t - 1)(N_k - 1) = 3 - 1 = 6 \quad F_{.05}(3, 6) = 4.76$
14. d. all of the above
15. a. Type I error
16. d. retains the null hypothesis when it should be rejected
17. d. all of the above
18. d. retain, Type II
   Explanation: The critical value would be $F_{.05}(3, 12) = 3.49$. Your value is smaller than the critical value, so you should retain the H_0. Because you can't make a Type I error when retaining, only a Type II error is possible.
19. c. both a. and b.

Short Answer Questions

1. In a repeated-measures design, you have more statistical power and less error because there are fewer differences between participants. One additional benefit if using a repeated-measures pairing is that you will not need as many participants, making the study design more efficient. The main concern would be the possibility of carryover effects if participants receive multiple levels of the treatment. But, it is also worth noting that not all studies can be conducted with pairing and this research design would not work for some research questions.

2. Whatever you come up with, the experiment must involve participants who contribute data to every level of the independent variable or participants who are matched into groups with the same number of participants in each grouping as there are levels of the independent variable. You will also need to identify the thing that is being decided by the researcher (the independent variable) and the thing that is being measured (the dependent variable).

   Here's an example: Jordan wanted to know if college student loneliness is influenced by day of the week, because she believed students might be less likely to feel lonely on the weekend. She had participants complete a measure of loneliness, which was used to create a loneliness score, three different times in one week: on Tuesday, Thursday, and Saturday. The independent variable is the day of the week and the dependent variable is the loneliness score. Because the same participants complete the measure on three different days, this is a repeated measures design.

Problems

1.

Source	SS	df	MS	F	p
Subjects	0.10	23			
Treatment	0.22	2	0.11	11.00	< .05
Error	0.36	46	0.01		
TOTAL	0.68	71			

Critical value: $F_{.05}(2, 46) = 3.20$

According to the data, we reject the null hypothesis. Participants rotated their necks further when viewing a scene that underestimated the rotation of their heads ($\bar{X}_A = 1.07$) than when viewing

accurate visual information ($\bar{X}_B = 1.00$). In contrast, participants rotated their necks less far when viewing a scene that overestimated the rotation of their head movement ($\bar{X}_C = 0.93$). The effect size index, $f = 0.53$, was large. There was a significant difference between understated and overstated visual feedback on neck rotation (HSD = 7.00). Thus, sensory input does affect the experience of neck pain to a large degree.

APA format: Sensory input has a significant effect on neck pain. Participants who viewed a scene that underestimated the rotation of their neck rotated the most before feeling pain ($M = 1.07$) and participants who viewed a scene that overestimated the rotation of their neck moved their head the least ($M = 0.93$), both relative to neck movement after receiving accurate visual information ($M = 1.00$), $F(2, 46) = 11.00$, $p < .05$. The size of the difference was large, $f = 0.53$. There was a significant difference between understated and overstated visual feedback on neck rotation, $HSD(3, 46) = 7.00$, $p < .01$.

$$SS_{tot} = \sum X_{tot}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 72.52 - \frac{71.92^2}{72} = 72.52 - 71.84 = 0.68$$

$$SS_{subjects} = \sum \left[\frac{(\sum X_k)^2}{N_k} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = 71.94 - 71.84 = 0.10$$

$$SS_{treat} = \sum \left[\frac{(\sum X_t)^2}{N_t} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{(25.57)^2}{24} \right] + \left[\frac{(24)^2}{24} \right] + \left[\frac{(22.35)^2}{24} \right] - 71.84$$

$$SS_{treat} = 27.243 + 24 + 20.813 - 71.84 = 72.056 - 71.84 = 0.22$$

$$SS_{error} = SS_{tot} - SS_{subjects} - SS_{treat} = 0.68 - 0.10 - 0.22 = 0.36$$

$$df_{tot} = N_{tot} - 1 = 72 - 1 = 71$$

$$df_{subjects} = N_t - 1 = 24 - 1 = 23$$

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

$$df_{error} = (N_t - 1)(N_k - 1) = (24 - 1)(3 - 1) = 46$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{0.22}{2} = 0.11$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{0.36}{46} = 0.01$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{0.11}{0.01} = 11.00$$

$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{3-1}{72}(0.11 - 0.01)}}{\sqrt{0.01}}$$

$$f = \frac{\sqrt{(.028)(0.10)}}{\sqrt{.010}} = \frac{0.053}{.10} = .53$$

For the Tukey tests, the critical value is $HSD_{.01}$ (3, 40) = 3.44
(Note: you use df_{error} = 40, because there is no df_{error} = 46 listed in the table and you use the next smallest value)

$$s_{\bar{x}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{0.01}{24}} = 0.02$$

$$HSD_{understated\ v.overstated} = \frac{\bar{X}_A - \bar{X}_C}{s_{\bar{x}}} = \frac{1.07 - 0.93}{0.02} = 7.00$$

2.

Source	SS	df	MS	F	p
Subjects	130.67	5			
Treatment	412	2	206	37.25	< .01
Error	55.33	10	5.53		
TOTAL	512	17			

Critical value: $F_{.01}(2, 10) = 7.56$

According to the data, we reject the null hypothesis. Training significantly affected participant knowledge about breast self-examination. Participants had significantly less knowledge before training ($\bar{X}_A$ = 12.0) than during the post-test ($\bar{X}_B$ = 23.0) or at the follow-up three months later ($\bar{X}_C$ = 21.0). The effect size index, f = 2.01, was large. Thus, training increased the amount of information participants had. There was a significant difference between knowledge pre-test and post-test (HSD = -11.46) and between knowledge pre-test and at the follow-up test three months later (HSD = -9.38). However, there was no significant difference between knowledge between the post-test and at the follow-up test three months later (HSD = 2.08).

APA format: Training significantly affected participant knowledge about breast self-examination. Participants had less knowledge on the pre-test (M = 12.0) than on the post-test (M = 23.0) or the follow-up test three months later (M = 21.0), F (2, 10) = 37.25, $p < .01$. The size of the effect was large, f = 2.01. There was a significant difference between knowledge on the pre-test and post-test, HSD(3, 10) = -11.46, $p < .01$. There was a significant difference between knowledge on the pre-test and the follow-up test, HSD(3, 10) = -9.38, $p < .01$. But, there was not a significant difference between the knowledge post-test or at the follow-up, HSD (3, 10) = 2.08, $p > .01$.

Participant	Pre-test (X_A)	Post-test (X_B)	Follow-up (X_C)	
1	7	22	16	$\sum X_1 = 45$
2	10	20	22	$\sum X_2 = 52$
3	10	23	24	$\sum X_3 = 57$
4	13	21	16	$\sum X_4 = 50$
5	15	26	22	$\sum X_5 = 63$
6	17	26	26	$\sum X_6 = 69$
	$\sum X_A = 72$	$\sum X_B = 138$	$\sum X_C = 126$	$\sum X_{tot} = 336$
	$\bar{X}_A = 12$	$\bar{X}_B = 23$	$\bar{X}_C = 21$	

$$\sum X_{tot}{}^2 = 6870.00$$

$$SS_{tot} = \sum X_{tot}{}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 6870 - \frac{336^2}{18} = 6870 - 6272 = 598$$

$$SS_{subjects} = \sum \left[\frac{(\sum X_k)^2}{N_k} \right] - \frac{(\sum X_{tot})^2}{N_{tot}}$$

$$SS_{subjects} = \left[\frac{(45)^2}{3} \right] + \left[\frac{(52)^2}{3} \right] + \left[\frac{(57)^2}{3} \right] + \left[\frac{(50)^2}{3} \right] + \left[\frac{(63)^2}{3} \right] + \left[\frac{(69)^2}{3} \right] - 6272$$

$$SS_{subjects} = [675] + [901.333] + [1083] + [833.333] + [1323] + [1587] - 6272$$

$$SS_{subjects} = 6402.666 - 6272 = 130.67$$

$$SS_{treat} = \sum \left[\frac{(\sum X_t)^2}{N_t} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{(72)^2}{6} \right] + \left[\frac{(138)^2}{6} \right] + \left[\frac{(126)^2}{6} \right] - 6272$$

$$SS_{treat} = 864 + 3174 + 2646 - 6272 = 6684 - 6272 = 412$$

$$SS_{error} = SS_{tot} - SS_{subjects} - SS_{treat} = 598 - 130.67 - 412 = 55.33$$

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

$$df_{subjects} = N_t - 1 = 6 - 1 = 5$$

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

$$df_{error} = (N_t - 1)(N_k - 1) = (6 - 1)(3 - 1) = 10$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{412}{2} = 206$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{55.33}{10} = 5.53$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{206}{5.53} = 37.25$$

$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{3-1}{18}(206 - 5.53)}}{\sqrt{5.53}}$$

$$f = \frac{\sqrt{(.111)(200.47)}}{\sqrt{5.53}} = \frac{\sqrt{22.252}}{2.352} = \frac{4.717}{2.352} = 2.01$$

For the Tukey tests, the critical value is HSD$_{.01}$ (3, 10) = 5.27

$$s_{\bar{X}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{5.53}{6}} = \sqrt{.922} = .96$$

$$HSD_{pre\ v.post} = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}}} = \frac{12 - 23}{.96} = -11.46$$

$$HSD_{pre\ v.follow\ up} = \frac{\bar{X}_A - \bar{X}_C}{s_{\bar{X}}} = \frac{12 - 21}{.96} = -9.38$$

$$HSD_{post\ v.follow\ up} = \frac{\bar{X}_B - \bar{X}_C}{s_{\bar{X}}} = \frac{23 - 21}{.96} = 2.08$$

3.

Source	SS	df	MS	F	p
Subjects	260.00	3			
Treatment	2114.67	2	1057.34	46.64	< .05
Error	136.00	6	22.67		
TOTAL	2510.67	11			

Critical value: $F_{.05}(2, 6) = 5.14$

According to the data, we reject the null hypothesis. Age had an effect on test performance. Participants had significantly different performance solving logic problems at age 10 ($\bar{X}_A = 9.0$), age 11 ($\bar{X}_B = 16.0$), and age 12 ($\bar{X}_C = 40.0$), $F (2, 6) = 46.64$, $p < .05$. The effect size index, $f = 2.76$, was large. There was not a significant difference between performance at age 10 and 11 (HSD = -2.94), but there were significant differences between performance at ages 10 and 12 (HSD = -13.03) and 11 and 12 (HSD = -10.08). The 12-year-olds solved more problems than either the 10- or 11-year-olds.

APA format: Age had a significant effect on test performance. Children did better on this logic test as they got older, with 12-year-olds solving a higher percentage of problems ($M = 40.0$) than 10- ($M = 9.0$) and 11-year-olds ($M = 16.0$), $F (2, 6) = 46.64$, $p < .05$. The effect size was large, $f = 2.76$. There was a significant difference between the performance of 10- and 12-year-olds, HSD(3, 6) = -13.03, $p < .01$, and the performance of 11- and 12-year-olds, HSD(3, 6) = -10.08, $p < .01$. But, there was not a significant difference between the performance of 10- and 11-year-olds, HSD(3, 6) = -2.94, $p > .01$.

Participant	Age 10 (X_A)	Age 11 (X_B)	Age 12 (X_C)	
1	0	8	40	$\sum X_1 = 48$
2	8	12	36	$\sum X_2 = 56$
3	12	24	36	$\sum X_3 = 72$
4	16	20	48	$\sum X_4 = 84$
	$\sum X_A = 36$	$\sum X_B = 64$	$\sum X_C = 160$	$\sum X_{tot} = 260$
	$\bar{X}_A = 9$	$\bar{X}_B = 16$	$\bar{X}_C = 40$	

$$\sum X_{tot}{}^2 = 8144.00$$

$$SS_{tot} = \sum X_{tot}{}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 8144 - \frac{260^2}{12} = 8144 - 5633.333 = 2510.67$$

$$SS_{subjects} = \sum \left[\frac{(\sum X_k)^2}{N_k} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{(48)^2}{3} \right] + \left[\frac{(56)^2}{3} \right] + \left[\frac{(72)^2}{3} \right] + \left[\frac{(84)^2}{3} \right] - 5633.333$$

$$SS_{subjects} = [768] + [1045.333] + [1728] + [2352] - 5633.333$$

$$SS_{subjects} = 5898.333 - 5633.333 = 260$$

$$SS_{treat} = \sum \left[\frac{(\sum X_t)^2}{N_t} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{(36)^2}{4} \right] + \left[\frac{(64)^2}{4} \right] + \left[\frac{(160)^2}{4} \right] - 5633.333$$

$$SS_{treat} = 324 + 1024 + 6400 - 5633.333 = 7748 - 5633.333 = 2114.67$$

$$SS_{error} = SS_{tot} - SS_{subjects} - SS_{treat} = 2510.67 - 260 - 2114.67 = 136$$

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

$$df_{subjects} = N_t - 1 = 4 - 1 = 3$$

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

$$df_{error} = (N_t - 1)(N_k - 1) = (4 - 1)(3 - 1) = 6$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{2114.67}{2} = 1057.34$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{136}{6} = 22.67$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{1057.34}{22.67} = 46.64$$

$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{3-1}{12}(1057.34 - 22.67)}}{\sqrt{22.67}}$$

$$f = \frac{\sqrt{(.167)(1034.67)}}{\sqrt{22.67}} = \frac{\sqrt{172.90}}{4.761} = \frac{13.145}{4.761} = 2.76$$

For the Tukey tests, the critical value is $HSD_{.05}(3, 6) = 4.34$

$$s_{\bar{X}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{22.67}{4}} = \sqrt{5.668} = 2.38$$

$$HSD_{10\,v.11} = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}}} = \frac{9 - 16}{2.38} = -2.94$$

$$HSD_{10\,v.12} = \frac{\bar{X}_A - \bar{X}_C}{s_{\bar{X}}} = \frac{9 - 40}{2.38} = -13.03$$

$$HSD_{11\,v.12} = \frac{\bar{X}_B - \bar{X}_C}{s_{\bar{X}}} = \frac{16 - 40}{2.38} = -10.08$$

4.

Source	SS	df	MS	F	p
Subjects	118.50	2			
Treatment	554.25	3	184.75	25.48	< .05
Error	43.50	6	7.25		
TOTAL	716.25	11			

Critical value: $F_{.05}(3, 6) = 4.76$

[NOTE: these conclusions describe all possible Tukey tests, but you were only asked to complete one in solving the problem.]

According to the data, we reject the null hypothesis. Drug dosage had a significant effect on reaction times. The reaction times associated with the placebo ($\bar{X}_A = 37.0$), 15 mg ($\bar{X}_B = 38.0$), 30 mg ($\bar{X}_C = 25.0$) and 60 mg ($\bar{X}_D = 23.0$) were significantly different, $F(3, 6) = 25.48$, $p < .05$. The effect size index, $f = 2.47$, was large. There were not significant differences between the placebo and 15 mg (HSD = -.65) or between 30 and 60 mg (HSD = 1.29) conditions. There were significant differences between the placebo and 30 mg and the placebo and 60 mg (HSD = 7.74 and HSD = 9.03, respectively). There were significant differences between 15 mg and 30 mg and 15 mg and 60 mg (HSD = 8.39 and HSD = 9.68, respectively). Faster reaction times were associated with higher doses of the drug.

APA format: Drug dosage had a significant effect on reaction times. Participants' reaction times were significantly different, depending on whether they received the placebo ($M = 37.0$), 15 mg ($M = 38.0$), 30 mg ($M = 25.0$), or 60 mg ($M = 23.0$) of the drug, $F(3, 6) = 25.48$, $p < .05$. The effect size was large, $f = 2.47$. Tukey tests revealed that there was no significant difference between the placebo and 15 mg condition, HSD(4, 6) = -.65, $p > .05$, and there was no significant difference between the 30 mg and 60 mg conditions, HSD(4, 6) = 1.29, $p > .05$. However, there were significant differences between the placebo and the 30 mg, HSD (4, 6) = 7.74, $p < .05$, and 60 mg, HSD (4, 6) = 9.03, $p < .05$, conditions and between the 15 mg and the 30 mg, HSD(4, 6) = 8.39, $p < .05$. and 60 mg, HSD (4, 6) = 9.68, $p < .05$, conditions. These findings suggest that the placebo and 15 mg had the same effect on reaction times, which were longer than the reaction times produced by 30 and 60 mg.

Participant	Placebo (0 mg) (X_A)	15 mg (X_B)	30 mg (X_C)	60 mg (X_D)	
1	36	32	20	18	$\sum X_1 = 106$
2	35	40	25	27	$\sum X_2 = 127$
3	40	42	30	24	$\sum X_3 = 136$
	$\sum X_A = 111$	$\sum X_B = 114$	$\sum X_C = 75$	$\sum X_D = 69$	$\sum X_{tot} = 369$
	$\bar{X}_A = 37$	$\bar{X}_B = 38$	$\bar{X}_C = 25$	$\bar{X}_D = 23$	

$$\sum X_{tot}{}^2 = 12063$$

$$SS_{tot} = \sum X_{tot}{}^2 - \frac{(\sum X_{tot})^2}{N_{tot}} = 12063 - \frac{369^2}{12} = 12063 - 11346.75 = 716.25$$

$$SS_{subjects} = \sum \left[\frac{(\sum X_k)^2}{N_k} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{(106)^2}{4} \right] + \left[\frac{(127)^2}{4} \right] + \left[\frac{(136)^2}{4} \right] - 11346.75$$

$$SS_{subjects} = [2809] + [4032.25] + [4624] - 11346.75$$

$$SS_{subjects} = 11465.25 - 11346.75 = 118.50$$

$$SS_{treat} = \sum \left[\frac{(\sum X_t)^2}{N_t} \right] - \frac{(\sum X_{tot})^2}{N_{tot}} = \left[\frac{(111)^2}{3} \right] + \left[\frac{(114)^2}{3} \right] + \left[\frac{(75)^2}{3} \right] + \left[\frac{(69)^2}{3} \right] - 11346.75$$

$$SS_{treat} = 4107 + 4332 + 1875 + 1587 - 11346.75 = 11901 - 11346.75 = 554.25$$

$$SS_{error} = SS_{tot} - SS_{subjects} - SS_{treat} = 716.25 - 118.50 - 554.25 = 43.50$$

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

$$df_{subjects} = N_t - 1 = 3 - 1 = 2$$

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

$$df_{error} = (N_t - 1)(N_k - 1) = (3 - 1)(4 - 1) = 6$$

$$MS_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{554.25}{3} = 184.75$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{43.5}{6} = 7.25$$

$$F = \frac{MS_{treat}}{MS_{error}} = \frac{184.75}{7.25} = 25.48$$

$$f = \frac{\sqrt{\frac{K-1}{N_{tot}}(MS_{treat} - MS_{error})}}{\sqrt{MS_{error}}} = \frac{\sqrt{\frac{4-1}{12}(184.75 - 7.25)}}{\sqrt{7.25}}$$

$$f = \frac{\sqrt{(.25)(177.5)}}{\sqrt{7.25}} = \frac{\sqrt{44.375}}{2.693} = \frac{6.661}{2.693} = 2.47$$

For the Tukey tests, the critical value is HSD$_{.05}$ (4, 6) = 4.90

$$s_{\bar{X}} = \sqrt{\frac{MS_{error}}{N_t}} = \sqrt{\frac{7.25}{3}} = \sqrt{2.417} = 1.55$$

$$\text{HSD}_{placebo\ v.15} = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}}} = \frac{37 - 38}{1.55} = |{-.65}|$$

$$\text{HSD}_{placebo\ v.30} = \frac{\bar{X}_A - \bar{X}_C}{s_{\bar{X}}} = \frac{37 - 25}{1.55} = 7.74$$

$$\text{HSD}_{placebo\ v.60} = \frac{\bar{X}_A - \bar{X}_D}{s_{\bar{X}}} = \frac{37 - 23}{1.55} = 9.03$$

$$\text{HSD}_{15\ v.30} = \frac{\bar{X}_B - \bar{X}_C}{s_{\bar{X}}} = \frac{38 - 25}{1.55} = 8.39$$

$$\text{HSD}_{15\ v.60} = \frac{\bar{X}_B - \bar{X}_D}{s_{\bar{X}}} = \frac{38 - 23}{1.55} = 9.68$$

$$\text{HSD}_{30\ v.60} = \frac{\bar{X}_C - \bar{X}_D}{s_{\bar{X}}} = \frac{25 - 23}{1.55} = 1.29$$

References

Harvie, D. S., Broecker, M., Smith, R. T., Meulders, A., Madden, V. J., & Moseley, G. L. (2015). Bogus visual feedback alters onset of movement-evoked pain in people with neck pain. *Psychological Science*, *26*(4), 385-392. doi: 10.1177/0956797614563339

Spatz, T.S. (1991). Improving breast self-examination training by using the 4MAT instructional model. *Journal of Cancer Education*, *6*, 179-183.