15.1 Exercises

Basic Concepts

1. Give two examples where you might be interested in comparing several population means.
2. **a.** What are experimental units? **b.** In the agricultural experiment in the beginning of 15.1, what were the experimental units?
3. **a.** What is a treatment? **b.** In the agricultural experiment in the beginning of 15.1, what was the treatment?
4. Explain how box plots can be useful in analyzing data when comparing population means.
5. How is the total variation broken down in an analysis of variance?
6. What does the total sum of squares describe? What are its degrees of freedom?
7. What is the mathematical expression for the sum of squares for treatments?
8. What is the mean square for treatments?
9. What is the relationship between the Total Sum of Squares, SST, and SSE? Explain why this relationship makes sense.
10. Why is it important to validate the assumptions upon which a hypothesis test is based?
11. What are the assumptions for an ANOVA F-test?
12. If you found that MST is much larger than MSE, would you tend to think that the population means were similar or different? Explain how this ratio brings you to this conclusion.
13. If the variability among the sample means is very similar to the variability among the sample observations, what value will F be close to? Explain why.
14. Is the null hypothesis generally rejected for large or small values of the F-statistic? Explain why this is the case.
15. What are the null and alternative hypotheses for the one-way ANOVA F-test?
16. Explain the completely randomized experimental design.
17. Why would data derived from an experimental design likely be better data than observational data?

Exercises

18. Consider the following table containing daily production data from a particular week for three different employee shifts.

Items Produced			
	First Shift (7 AM-3 PM)	Second Shift (3 PM-11 PM)	Third Shift (11 PM-7 AM)
Monday	140	168	77
Tuesday	181	224	123
Wednesday	127	162	77

Thursday	172	182	101
Friday	161	219	147
Saturday	152	171	145
Sunday	173	217	111

a. Identify the experimental units and the treatment in the context of this problem.

b. Compute the mean and median numbers of items produced for each shift.

c. Compute the values of the minimum, maximum, first, and third quartiles for each shift.

d. Construct side-by-side box plots for the three shifts.

e. Based on the box plots, do you think that there may be a significant difference in the average numbers of items produced during the first and second shifts? Explain.

f. Based on the box plots, do you think that there may be a significant difference in the average numbers of items produced during the second and third shifts? Explain.

g. Based on the box plots, do you think that there may be a significant difference in the average numbers of items produced during the first and third shifts? Explain.

h. Based on your analysis, which shift would you say is the most productive, on average? Explain your answer.

19. Consider the production data given in Exercise 18.

a. What is the value of the grand mean, $\bar{\bar{x}}$?

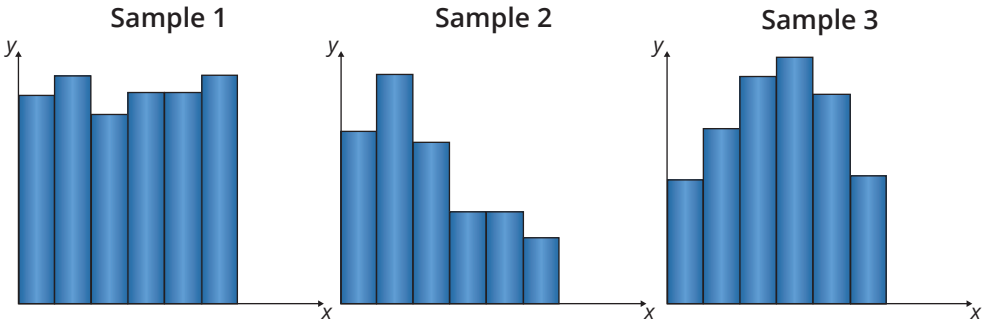
b. What is the value of n_2?

c. What is the value of k?

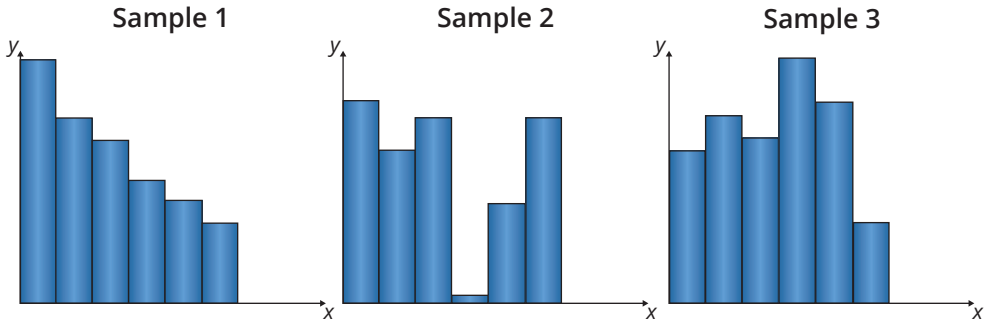
d. What is the value of N?

e. For this data, identify the degrees of freedom associated with the total sum of squares, the degrees of freedom associated with the sum of squares for treatments, and the degrees of freedom associated with the sum of squares for error. Verify that the relationship between the degrees of freedom (Total = Treatment + Error) holds.

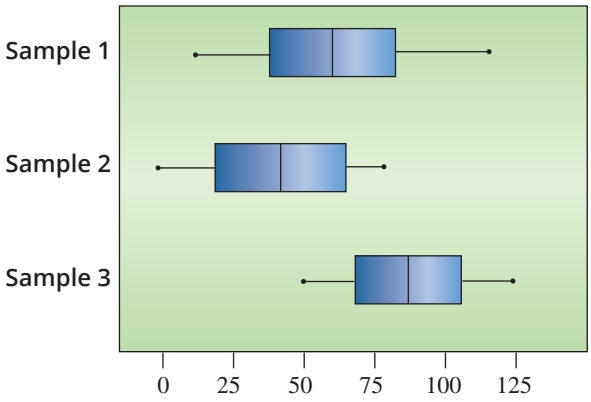
20. For each of the following histograms of sample data, decide whether or not you think it is reasonable to assume that the data was drawn from a population that has an approximately normal distribution.



21. For each of the following histograms of sample data, decide whether you think it is reasonable to assume that the data was drawn from a population that has an approximately normal distribution.

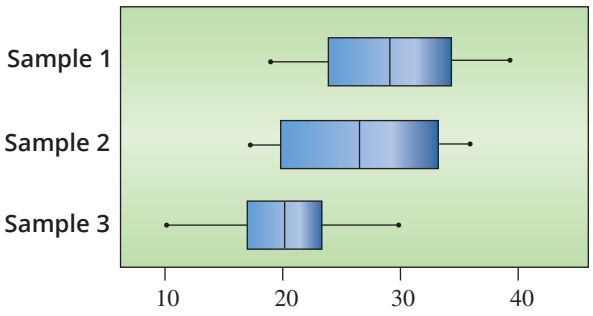


22. Consider the following box plots.



Do you think it is reasonable to assume that the three populations represented by the sample data in these box plots have equal variances? Explain.

23. Consider the following box plots.



Do you think it is reasonable to assume that the three populations represented by the sample data in these box plots have equal variances? Explain.

24. The results of a comparison of four popular minivans are reported in the following table. One of the features the researchers compared was the distance (in feet) required for the minivan to come to a complete stop when traveling at a speed of 60 miles per hour (braking distance). Suppose the braking distances were measured for five minivans of each type with the following results.

Braking Distances (Feet)			
Minivan A	Minivan B	Minivan C	Minivan D
150	153	155	167
152	150	150	164
151	156	157	169
149	151	158	162
153	155	155	173

a. Can the researchers conclude at $\alpha = 0.10$ that there is a difference among average braking distances for the four minivan models?

b. What assumptions did the researchers make in performing the test procedure in part **a.**? Does the data appear to satisfy these assumptions? Explain.

25. A steel company is considering the relocation of one of its manufacturing plants. The company's executives have selected four areas that they believe are suitable locations. However, they want to determine if the average wages are significantly different in any of the locations, since this could have a major impact on the cost of production. A survey of hourly wages of similar workers in each of the four areas is performed with the following results.

Hourly Wages ($)			
Area 1	Area 2	Area 3	Area 4
13	18	16	23
15	19	17	19
14	21	18	21
16	20	18	20
13	17	15	19

a. Does the data indicate a significant difference among the average hourly wages in the four areas at $\alpha = 0.05$?

b. What assumptions were made in performing the test in part **a.**? Does the data appear to satisfy these assumptions? Explain.

26. Consider the following table containing yields for mutual funds in different asset classes (small, mid, and large cap).

Fund Yield by Asset Class					
Small Cap		Mid Cap		Large Cap	
Fund	Yield (%)	Fund	Yield (%)	Fund	Yield (%)
Explorer Value	2.04	Capital Value	0.96	Equity Income	3.24
Small-Cap Value Index Admiral	2.46	Mid-Cap Value Index Admiral	1.57	High Dividend Yield Index	3.50
Small-Cap Index Admiral Shares	1.49	Extended Market Index Admiral Shares	1.22	500 Index Admiral Shares	1.57

Strategic Small-Cap Equity	0.38	Mid-Cap Index Admiral Shares	1.52	Diversified Equity	1.23
Explorer	0.17	Mid-Cap Growth	2.76	FTSE Social Index	1.42
Small-Cap Growth Index Admiral	0.21	Capital Value	0.32	Growth Equity	2.52
Explorer Value	2.55	Strategic Equity	1.54	U. S. Growth	0.37
Samll-Cap ETF	1.44	Capital Opportunity Admiral Shares	2.14	Windsor	1.64

Sum of squares for treatments ≈ 1.5986

Sum of squares for error ≈ 18.4205

a. What are the degrees of freedom associated with the sum of squares for treatments?

b. Find the mean square for treatments. Round your answer to two decimal places, if necessary.

27. A physical trainer has four workouts that he recommends for his clients. The workouts have been designed so that the average maximum heart rate achieved is the same for each workout. To test this design, he randomly selects 12 people and randomly assigns three of them to use each of the workouts. During each workout, he measures the maximum heart rate in beats per minute with the following results.

Maximum Heart Rates (Beats per Minute)			
Workout #1	**Workout #2**	**Workout #3**	**Workout #4**
180	160	175	185
185	170	180	190
170	175	170	180

a. Can the physical trainer conclude at $\alpha = 0.05$ that there is a difference among the average maximum heart rates which are achieved during the four workouts?

b. What assumptions did the physical trainer make in performing the test procedure in part **a.**? Does the data appear to satisfy these assumptions? Explain.

28. The results of a survey comparing the costs of staying one night in a full-service hotel (including food, beverages, and telephone calls, but not taxes or gratuities) for several major cities are given in the following table.

Hotel Costs per Night ($)				
New York	**Los Angeles**	**Atlanta**	**Houston**	**Phoenix**
300	240	190	195	238
320	250	198	190	240
325	230	185	200	236
350	245	195	192	248
275	235	182	198	228

a. Does the data suggest that there is a significant difference among the average costs of one night in a full-service hotel for the five major cities at $\alpha = 0.05$?

b. What assumptions were made in performing the test procedure in part **a.**? Does the data appear to satisfy these assumptions? Explain.

c. Based on the analysis you performed in part **b.**, which cities, if any, do you think have significantly different average costs for a one-night stay in a full-service hotel? Explain.

29. Consider the following information regarding the dividends paid per share by companies in the banking, transportation, and energy industries.

Dividends per Share ($)		
Banking	**Transportation**	**Energy**
1.52	1.00	2.08
3.12	1.20	2.68
1.32	0.20	0.70
0.60	0.40	2.00
1.20	1.09	1.91
1.00	0.61	1.60
1.19	0.35	1.28

a. Does the data provide sufficient evidence to conclude that there is a significant difference among the average dividends paid per share for the three different industries? Use $\alpha = 0.10$.

b. What assumptions were made in performing the test procedure in part **a.**? Does the data appear to satisfy these assumptions? Explain.

c. Based on the analysis you performed in part **b.**, which industries, if any, do you think pay significantly different average dividends per share? Explain.

30. The sales strategy data given below yields the following statistics for the sum of squares for treatments and the sum of squares for error.

SST = 24.875

SSE = 579.1

Sales by Strategy (Millions of Dollars)			
Strategy 1	**Strategy 2**	**Strategy 3**	**Strategy 4**
15	5	3	8
8	4	12	6
4	14	4	7
7	16	9	5
10	9	9	4
4	5	10	13
5	7	5	5
9	8	5	8
14	3	6	13
2	16	2	10

a. What are the degrees of freedom associated with the total sum of squares?
b. What are the degrees of freedom associated with the sum of squares for treatments, SST?
c. Find the mean square for treatments, MST. Round your answer to four decimal places.
d. Find the mean square for error, MSE. Round your answer to four decimal places.

15.2 Multiple Comparison Procedures

In the previous section, we used one-way ANOVA to test whether differences existed between population means. In the earlier examples, when we rejected the null hypothesis that all of the population means were equal, we were only testing if differences existed. However, the results of the one-way ANOVA test do not indicate which population means are different. To determine which population means are different, we need to perform more tests to determine if there are statistically significant differences between two population means such as $\mu_1 - \mu_2$, for example. Multiple comparison procedures present several options to the analyst when comparing means after finding significance when performing a one-way ANOVA. The ones we will consider are Fisher's Least Significance Difference (LSD) Method, Tukey's Honest Significant Difference (HSD), and performing *t*-tests to make pairwise comparisons between means.

Data were collected to study the use of media by three age groups–tweens (8–12 years old), teens (13–18 years old), and adults (over 18 years old). A sample of 50 participants in each age group was asked how frequently they engaged in activities such as time spent on cell phones, watching online videos, watching television, and playing mobile games. The table below presents some summary statistics from the data collected showing the sample size, the sample mean time spent on devices, and the sample standard deviation of the data by age group.

Table 15.2.1 – Screen Time Summary Statistics by Age Group

	Tweens 8–12 Years Old	Teens 13–18 Years Old	Adults Over 18 Years Old
n	50	50	50
Mean	128	137.48	100.12
Standard Deviation	20.90	25.57	20.32

Test to determine if there is a significant difference between average time spent on devices among the three age groups.

Let

μ_1 = population mean time spent on devices for participants 8–12 years old,

μ_2 = population mean time spent on devices for participants 13–18 years old, and

μ_3 = population mean time spent on devices for participants over 18 years old.

Data

This data set can be found on stat.hawkeslearning.com under **Discovering Statistics and Data, Fourth Edition > Data Sets > Screen Time by Age Group.**