Chapter 12

Section 12-2

1. a. The chest deceleration measurements are categorized according to the one characteristic of size.
   b. The terminology of *analysis of variance* refers to the method used to test for equality of the three population means. That method is based on two different estimates of a common population variance.
3. The test statistic is $F = 3.288$, and the F distribution applies.
5. Test statistic: $F = 0.39$. P-value: 0.677. Fail to reject H_0: $\mu_1 = \mu_2 = \mu_3$. There is not sufficient evidence to warrant rejection of the claim that the three categories of blood lead level have the same mean verbal IQ score. Exposure to lead does not appear to have an effect on verbal IQ scores.
7. Test statistic: $F = 11.6102$. P-value: 0.000577. Reject H_0: $\mu_1 = \mu_2 = \mu_3$. There is sufficient evidence to warrant rejection of the claim that the three size categories have the same mean highway fuel consumption. The size of a car does appear to affect highway fuel consumption.
9. Test statistic: $F = 0.161$. P-value: 0.852. Fail to reject H_0: $\mu_1 = \mu_2 = \mu_3$. There is not sufficient evidence to warrant rejection of the claim that the three size categories have the same mean head injury measurement. The size of a car does not appear to affect head injuries.
11. Test statistic: $F = 27.2488$. P-value: 0.000. Reject H_0: $\mu_1 = \mu_2 = \mu_3$. There is sufficient evidence to warrant rejection of the claim that the three different miles have the same mean time. These data suggest that the third mile appears to take longer, and a reasonable explanation is that the third lap has a hill.
13. Test statistic: $F = 6.1413$. P-value: 0.0056. Reject H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$. There is sufficient evidence to warrant rejection of the claim that the four treatment categories yield poplar trees with the same mean weight. Although not justified by the results from analysis of variance, the treatment of fertilizer and irrigation appears to be most effective.
15. Test statistic: $F = 18.9931$. P-value: 0.000. Reject H_0: $\mu_1 = \mu_2 = \mu_3$. There is sufficient evidence to warrant rejection of the claim that the three different types of cigarettes have the same mean amount of nicotine. Given that the king-size cigarettes have the largest mean of 1.26 mg per cigarette, compared to the other means of 0.87 mg per cigarette and 0.92 mg per cigarette, it appears that the filters do make a difference (although this conclusion is not justified by the results from analysis of variance).
17. The Tukey test results show different P-values, but they are not dramatically different. The Tukey results suggest the same conclusions as the Bonferroni test.

Section 12-3

1. The load values are categorized using *two* different factors of (1) femur (left or right) and (2) size of car (small, midsize, large).
3. An interaction between two factors or variables occurs if the effect of one of the factors changes for different categories of the other factor. If there is an interaction effect, we should not proceed with individual tests for effects from the row factor and column factor. If there is an interaction, we should not consider the effects of one factor without considering the effects of the other factor.
5. For interaction, the test statistic is $F = 1.72$ and the P-value is 0.194, so there is not sufficient evidence to conclude that there is an interaction effect. For the row variable of femur (right, left), the test statistic is $F = 1.39$ and the P-value is 0.246, so there is not sufficient evidence to conclude that whether the femur is right or left has an effect on measured load. For the column variable of size of the car, the test statistic is $F = 2.23$ and the P-value is 0.122, so there is not sufficient evidence to conclude that the car size category has an effect on the measured load.
7. For interaction, the test statistic is $F = 1.05$ and the P-value is 0.365, so there is not sufficient evidence to conclude that there is an interaction effect. For the row variable of sex, the test statistic is $F = 4.58$ and the P-value is 0.043, so there is sufficient evidence to conclude that the sex of the subject has an effect on verbal IQ score. For the column variable of blood lead level (LEAD), the test statistic is $F = 0.14$ and the P-value is 0.871, so there is not sufficient evidence to conclude that blood lead level has an effect on verbal IQ score. It appears that only the sex of the subject has an effect on verbal IQ score.
9. For interaction, the test statistic is $F = 3.7332$ and the P-value is 0.0291, so there is sufficient evidence to conclude that there is an interaction effect. The measures of self-esteem appear to be affected by an interaction between the self-esteem of the subject and the self-esteem of the target. Because there appears to be an interaction effect, we should not proceed with individual tests of the row factor (target's self-esteem) and the column factor (subject's self-esteem).
11. a. Test statistics and P-values do not change.
    b. Test statistics and P-values do not change.
    c. Test statistics and P-values do not change.
    d. An outlier can dramatically affect and change test statistics and P-values.

Chapter 12: Quick Quiz

1. H_0: $\mu_1 = \mu_2 = \mu_3$. Because the displayed P-value of 0.000 is small, reject H_0.
2. No. Because we reject the null hypothesis of equal means, it appears that the three different power sources do not produce the same mean voltage level, so we cannot expect electrical appliances to behave the same way when run from the three different power sources.
3. Right-tailed.
4. Test statistic: $F = 183.01$. In general, larger test statistics result in smaller P-values.
5. The sample voltage measurements are categorized using only one factor: the source of the voltage.
6. Test a null hypothesis that three or more samples are from populations with equal means.
7. With one-way analysis of variance, the different samples are categorized using only one factor, but with two-way analysis of variance, the sample data are categorized into different cells determined by two different factors.
8. Fail to reject the null hypothesis of no interaction. There does not appear to be an effect due to an interaction between sex and major.