**Chapter 6 Exercises**

**Concepts**

1. List the assumptions of the analysis of variance and the impact on the decision-making process when the assumptions are not met. When we say that ANOVA is relative robust with respect to deviations from the assumptions, what do we mean?
2. What is a nonparametric test? Why might a nonparametric test not be as powerful as a test like the analysis of variance? When might a nonparametric test be more useful than a test like ANOVA?
3. Conceptually, what is the difference between a priori vs a posteriori contrasts? Which are more powerful in rejecting the null hypothesis that a contrast is equal to zero?

**Exercises**

1. Use the following data to answer the questions below.

University Graduation Rates by Major

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Computer science | History | Economics | Theater |
| 2008 | 0.63 | 0.75 | 0.74 | 0.47 |
| 2009 | 0.62 | 0.66 | 0.54 | 0.58 |
| 2010 | 0.65 | 0.86 | 0.76 | 0.61 |
| 2011 | 0.66 | 0.63 | 0.66 | 0.88 |
| 2012 | 0.63 | 0.74 | 0.62 | 0.78 |
| 2013 | 0.6 | 0.59 | 0.51 | 0.74 |

* 1. Find the mean and standard deviation for each major.
  2. Use the Levene’s test in SPSS to determine whether the assumption of homoscedasticity is justified.
  3. Carry out an analysis of variance to test the null hypothesis that graduation rate does not vary by major. Give the observed *F*-statistic and the critical *F*-value associated with 𝛂 = .05, and compare the two.

1. Use the following data to answer the questions below

Absences from Work by Division of a Company

|  |  |  |  |
| --- | --- | --- | --- |
| **Employee** | **Division I** | **Division II** | **Division III** |
| ***1*** | 10 | 6 | 6 |
| ***2*** | 8 | 4 | 6 |
| ***3*** | 7 | 3 | 5 |
| ***4*** | 6 | 5 | 7 |
| ***5*** | 4 | 6 | 6 |
| ***6*** | 5 | --- | 8 |
| ***7*** | 7 | --- | --- |
| ***8*** | 3 | --- | --- |

1. Carry out an analysis of variance to test the null hypothesis that the absentee rate does not vary by division. Give the observed *F*-statistic and the critical *F*-value associated with α = .05.
2. Calculate the Kruskal-Wallis statistics and compare the results to those from (a).
3. Assume that an analysis of variance is conducted for a study of N = 64 observations and   
   k = 4 categories. Fill in the blanks in the following ANOVA table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sums of squares | Degrees of freedom | Mean square | *F* |
| Between | \_\_\_\_\_\_ | \_\_\_\_\_\_ | 356.675 | \_\_\_\_\_\_ |
| Within | 703.655 | \_\_\_\_\_\_ | \_\_\_\_\_\_ |  |
| Total | \_\_\_\_\_\_ | \_\_\_\_\_\_ |  |  |

If 𝛂 = .05, assess the null hypothesis that the means of the different categories are equal.

1. Fill out the following analysis of variance table. Assuming that 𝛂 = .05, compare the F-value with the critical value:

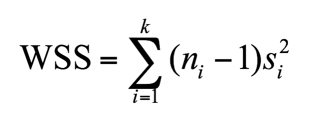
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sums of  squares | Degrees of freedom | Mean square | *F* |
| Between | 78.5 | 6 | \_\_\_\_\_\_ | \_\_\_\_\_\_ |
| Within | \_\_\_\_\_\_ | \_\_\_\_\_\_ | \_\_\_\_\_\_ |  |
| Total | 246.7 | 46 |  |  |

1. Use the following data to answer the questions:

Sightings per Year of Endangered Species Across Three Forests

|  |  |  |  |
| --- | --- | --- | --- |
| **Observation** | **Forest A** | **Forest B** | **Forest C** |
| **1** | 23 | 34 | 23 |
| **2** | 33 | 29 | 31 |
| **3** | 28 | 23 | 27 |
| **4** | 33 | 26 | 39 |
| **5** | 19 | 25 | 34 |
| **6** | 32 | 27 | 30 |
| ***Mean*** | ***28.0*** | ***27.3*** | ***30.7*** |
| ***Std dev*** | ***5.9*** | ***3.8*** | ***5.5*** |
| ***Overall mean*** | ***28.7*** | ***Overall std. dev.*** | ***5.1*** |

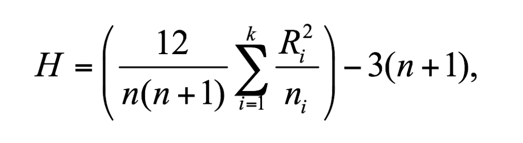
* 1. Find the within sum of squares for the data using the following definition:



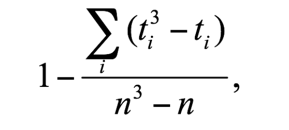
* 1. Find the value of the test statistic. Compare it with the critical value associated with

𝛂 = .05.

* 1. Rank the data, using 1 to indicate the lowest value and the average of the ranks for sets of tied observations. Find the Kruskall-Wallis statistic as follows:



and adjust the value of H by



where *ti* is the number of observations that are tied. Compare the test statistic with the critical chi-square value, which has k-1 degrees of freedom. How does this comparison relate to that generated in part (b)?

1. Find the variance and mean for each of the following categories of mineral deposit data. Then use ANOVA (with 𝛂 = .05) to test whether the population means differ, and present and test the null and alternative hypotheses.

|  |  |  |  |
| --- | --- | --- | --- |
| **Record** | **Deposit A** | **Deposit B** | **Deposit C** |
| **1** | 9 | 21 | 17 |
| **2** | 10 | 17 | 18 |
| **3** | 10 | 10 | 22 |
| **4** | 9 | 1 | 14 |
| **5** | 11 | 18 | 18 |
| **6** | 5 | 19 | 15 |
| **7** | 11 | 17 | 15 |
| **8** | 11 | 18 | 17 |
| **9** | 10 | 5 | 18 |
| **10** | 7 | 26 | 18 |
| **11** | 9 | 15 | 17 |
| **12** | 13 | 10 | 17 |
| **13** | 11 | 17 | 16 |
| **14** | 11 | 16 | 16 |

1. Use ANOVA (with 𝛂 = .10) to test the null hypothesis that the means of the following data are equal

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **Combined sample** |
| ***n*** | 15 | 11 | 14 | 40 |
| ***Mean*** | 79.6 | 74.5 | 78.9 | 78.0 |
| ***Std Dev*** | 3.1 | 8.8 | 3.9 | 5.7 |

1. Use the following data to:
   1. Present the post hoc contrasts and identify which (if any) of the paired differences are significant.
   2. Carry out an a priori test of the hypothesis that each of the categories differs from each of the other categories (A versus B, B versus C, A versus C).

**Height in Centimeters of Little League Baseball Team Players**

|  |  |  |
| --- | --- | --- |
| **A** | **B** | **C** |
| 140 | 140 | 143 |
| 132 | 131 | 156 |
| 125 | 117 | 151 |
| 145 | 123 | 157 |
| 152 | 113 | 127 |
| 118 | 112 | 121 |
| 135 | 127 | 128 |
| 147 | 106 | 149 |
| 140 | 112 | 129 |

1. Using SPSS, analyze the SPSS Housing dataset and determine whether the floor area varies with the number of bedrooms in a house. Report the results of Levene’s test and describe how much confidence a researcher might have in the results.
2. Divide the data in the Singapore Census file into areas with percent renter ≤ 0.1, 0.1 ≤ 0.2, 0.2 ≤ 0.3, 0.3 ≤. Using SPSS determine whether:
   1. the English speakers varies with the percent renters in a district
   2. the unemployment rate varies with the percent renters in a district.

Include a discussion of post hoc tests to identify which sub-sets appear to differ.

1. Using the SPSS Housing dataset and 𝛂 = .05, consider the following:
   1. Carry out an a priori test of the hypothesis that the average year in which houses were built in different regions varies. Specifically, test whether the mean year in Regions 1 and 3 differs.
   2. Consider the post hoc contrasts associated with exploring how the unemployment rate varies by region. Are there any significant differences between the mean unemployment rate in different regions?

**Chapter 6 Solutions**

|  |  |  |
| --- | --- | --- |
|  | Mean | Std. Deviation |
| Computer science | 0.632 | 0.021 |
| History | 0.705 | 0.098 |
| Economics | 0.638 | 0.102 |
| Theater | 0.677 | 0.150 |

* 1. The Levene statistic is 5.2, with a significance level of .008. This is lower than 0.05, which means we should consider the results cautiously.
  2. The *f*-statistic is .659, while the critical *f*-value is f3,20 = 3.10. The observed statistic is less than the critical value, so we fail to reject the null hypothesis that all the means are equal.
  3. *We can calculate the F-statistic as follows:*

*BSS = 8\*(6.25-5.9)2+5\*(4.8-5.9)2+6\*(6.3-5.9)2=8.2*

*WSS= (8-1)\*5.1+(5-1)\*1.7+(6-1)\*1.1 = 47.6*

*F = (BSS / (k-1)) / (WSS / (n-k)) = (8.2 / 2) / (47.6 / 16) = 1.37*

*The ANOVA test suggests none of the division absentee counts are significantly different from the others at the < 0.10 level. We therefore fail to reject the null hypothesis.*

* 1. *The KW statistic is given by summing the ranks of the different groups, so that we get*

*H = (12/(n(n+1)) sum{i=1->k} Ri2* / ni*) - 3(n+1)*

*= (12/(19(19+1)) sum{i=1->k} Ri2* / ni*) - 3(19+1)*

*= 0.032 (882/5 + 322/5 + 702/5) - 60*

*= 2.83*

*Comparing H with the chi-squared2 = 5.991, so that we fail to reject the null hypothesis. Both tests indicate that the divisions likely have similar absentee values.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sums of squares | Degrees of freedom | Mean square | *F* |
| Between | 1070.025 | 3 | ***356.675*** | 30.413 |
| Within | ***703.655*** | 60 | 11.728 |  |
| Total | 1773.680 | 63 |  |  |

If 𝛂 = .05, the critical value is F.05,3,60 = 8.57. The null hypothesis of no difference is therefore rejected.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sums of squares | Degrees of freedom | Mean square | *F* |
| Between | **78.5** | **6** | 13.1 | 3.1 |
| Within | 168.2 | 40 | 4.2 |  |
| Total | **246.7** | **46** |  |  |

If 𝛂 = .05, the critical value is F.05,6,40 = 3.77. We therefore fail to reject the null hypothesis of no difference.

* 1. WSS = 398.7
  2. F = 0.702, while F.05,2,15= 3.68, so that we fail to reject the null hypothesis.
  3. The adjusted value of H is 1.213, where the chi-squared value is 5.991. Thus we also fail to reject the null hypothesis, as we did in part (b)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Deposit A** | **Deposit B** | **Deposit C** |
| ***n*** | 14 | 14 | 14 |
| ***Mean*** | 9.8 | 15.0 | 17.0 |
| ***Std Dev*** | 2.0 | 6.5 | 1.9 |

From this, BSS = 416.2 and WSS = 698.2, so that F = (416.2 / 2) / (698.2 / 39) = 11.6, which is less than the F.05,2,15 value of approximately 3.68. Thus, we reject the null hypothesis and conclude that there are differences in the means.

1. BSS = 171.4, WSS = 1004.6, so that F = (171.4 / 2) / (1004.6 / 34) = 2.900, which is greater than the critical value F.10,2,34 of approximately 2.86. Thus we reject the null hypothesis and conclude that there are differences in the means.
   1. Group 1 has a significant contrast with Group 2, as does Group 3, at significance levels of .022 and .007 respectively. Groups 1 and 3 do not have significant contrasts.
   2. As we can assume equal variances based on the Levene test, we find statistically significant contrasts of 17 between A and B (p-value = .006) and -20 between B and C

(*p*-value = .002). The contrast between A and C is 3, with *p*-value = .601, which is not significant.

1. Levene’s test returns a statistic (10.7) with fairly extreme significance, supporting the null hypothesis that the variances of the different numbers of bedrooms could be equal. The

*F*-statistic of 64.455 is high as well, suggesting that we should reject the null hypothesis that the average floor areas for different numbers of bedroom are equal.

1. The ANOVA statistic returns values of 2.996 for English Speakers and 3.173 for the Unemployment Rate. Both values are significant at the 0.05 level. The post hoc test indicates that highest and lowest percent renter groups are different in English Speakers and Unemployment Rate. The other groups appear to be the same. However, the Levene’s test for English Speakers and Unemployment Rate return values of 3.719 and 4.932 respectively. Both values are significant at the 0.05 level indicating we likely fail to meet the assumption of equal variance. We should therefore interpret our results with caution.
   1. The mean does differ, at a statistically significant level. Giving Region 1 a coefficient of 1 and Region 3 a coefficient of -1, the value of the contrast is -20.82, with a *p*-value less than .001.

Region 1: differs from Regions 2, 3, and 6

Region 2: differs from Region 1

Region 3: differs from Regions 1 and 5

Region 5: differs from Region 3

Region 6: differs from Region 1