Chi-Square testing and ANOVA

Overview and Rationale

In this assignment, you will use your knowledge of chi-square and ANOVA testing to solve various types of problems.

Summary

Complete the following problems using R. Be sure to show your work and include the hypothesis tests, the critical values, the computed test values, and the resulting decisions where applicable.

Section 11-1

Perform these steps:

a. *State the hypotheses and identify the claim.*
b. *Find the critical value.*
c. *Compute the test value.*
d. *Make the decision.*
e. *Summarize the results.*

Use the traditional method of hypothesis testing unless otherwise specified. Assume all assumptions are met.

6. Blood Types A medical researcher wishes to see if hospital patients in a large hospital have the same blood type distribution as those in the general population. The distribution for the general population is as follows: type A, 20%; type B, 28%; type O, 36%; and type AB = 16%. He selects a random sample of 50 patients and finds the following: 12 have type A blood, 8 have type B, 24 have type O, and 6 have type AB blood.

At $\alpha = 0.10$, can it be concluded that the distribution is the same as that of the general population?

8. On-Time Performance by Airlines According to the Bureau of Transportation Statistics, on-time performance by the airlines is described as follows:

Action	% of Time
On time	70.8
National Aviation System delay	8.2
Aircraft arriving late	9.0

Other (because of weather and other conditions)	12.0

Records of 200 randomly selected flights for a major airline company showed that 125 planes were on time; 40 were delayed because of weather, 10 because of a National Aviation System delay, and the rest because of arriving late. At $\alpha = 0.05$, do these results differ from the government's statistics?

Source: www.transtats.bts.gov

Section 11-2

Perform the following steps.

a. *State the hypotheses and identify the claim.*
b. *Find the critical value.*
c. *Compute the test value.*
d. *Make the decision.*
e. *Summarize the results.*

Use the traditional method of hypothesis testing unless otherwise specified. Assume all assumptions are valid.

8. Ethnicity and Movie Admissions Are movie admissions related to ethnicity? A 2014 study indicated the following numbers of admissions (in thousands) for two different years. At the 0.05 level of significance, can it be concluded that movie attendance by year was dependent upon ethnicity?

	Caucasian	**Hispanic**	**African American**	**Other**
2013	724	335	174	107
2014	370	292	152	140

Source: MPAA Study.

10. Women in the Military This table lists the numbers of officers and enlisted personnel for women in the military. At $\alpha = 0.05$, is there sufficient evidence to conclude that a relationship exists between rank and branch of the Armed Forces?

	Officers	**Enlisted**
Army	10,791	62,491
Navy	7,816	42,750
Marine Corps	932	9,525
Air Force	11,819	54,344

Source: New York Times Almanac.

Section 12-1

Assume that all variables are normally distributed, that the samples are independent, that the population variances are equal, and that the samples are simple random samples, one from each of the populations. Also, for each exercise, perform the following steps.

a. *State the hypotheses and identify the claim.*
b. *Find the critical value.*
c. *Compute the test value.*
d. *Make the decision.*

e. Summarize the results, and explain where the differences in the means are.

Use the traditional method of hypothesis testing unless otherwise specified.

8. Sodium Contents of Foods The amount of sodium (in milligrams) in one serving for a random sample of three different kinds of foods is listed. At the 0.05 level of significance, is there sufficient evidence to conclude that a difference in mean sodium amounts exists among condiments, cereals, and desserts?

Condiments	Cereals	Desserts
270	260	100
130	220	180
230	290	250
180	290	250
80	200	300
70	320	360
200	140	300
		160

Source: The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.

Section 12-2

Perform a complete one-way ANOVA. If the null hypothesis is rejected, use either the Scheffé or Tukey test to see if there is a significant difference in the pairs of means. Assume all assumptions are met.

10. Sales for Leading Companies The sales in millions of dollars for a year of a sample of leading companies are shown. At $\alpha = 0.01$, is there a significant difference in the means?

Cereal	Chocolate Candy	Coffee
578	311	261
320	106	185
264	109	302
249	125	689
237	173	

Source: Information Resources, Inc.

12. Per-Pupil Expenditures The expenditures (in dollars) per pupil for states in three sections of the country are listed. Using $\alpha = 0.05$, can you conclude that there is a difference in means?

Eastern third	Middle third	Western third
4946	6149	5282
5953	7451	8605
6202	6000	6528
7243	6479	6911

6113

Source: New York Times Almanac.

Section 12-3

Assume that all variables are normally or approximately normally distributed, that the samples are independent, and that the population variances are equal.

a. *State the hypotheses.*
b. *Find the critical value for each* F *test.*
c. *Complete the summary table and find the test value.*
d. *Make the decision.*
e. *Summarize the results. (Draw a graph of the cell means if necessary.)*

10. Increasing Plant Growth A gardening company is testing new ways to improve plant growth. Twelve plants are randomly selected and exposed to a combination of two factors, a “Grow-light” in two different strengths and a plant food supplement with different mineral supplements. After a number of days, the plants are measured for growth, and the results (in inches) are put into the appropriate boxes.

	Grow-light 1	**Grow-light 2**
Plant food A	9.2, 9.4, 8.9	8.5, 9.2, 8.9
Plant food B	7.1, 7.2, 8.5	5.5, 5.8, 7.6

Can an interaction between the two factors be concluded? Is there a difference in mean growth with respect to light? With respect to plant food? Use $\alpha = 0.05$.

Use R to complete the following steps. Be sure to include all code in an appendix at the end of your submission. Assume the expected frequencies are equal and 〈 = 0.05.

1. Download the file ‘baseball.csv’ from the course resources and import the file into R.
2. Perform EDA on the imported data set. Write a paragraph or two to describe the data set using descriptive statistics and plots. Are there any trends or anything of interest to discuss?
3. Assuming the expected frequencies are equal, perform a Chi-Square Goodness-of-Fit test to determine if there is a difference in the number of wins by decade. Be sure to include the following:
   a. *State the hypotheses and identify the claim.*
   b. *Find the critical value (〈 = 0.05) (From table in the book).*
   c. *Compute the test value.*
   d. *Make the decision. Clearly state if the null hypothesis should or should not be rejected and why.*
   e. *Does comparing the critical value with the test value provide the same result as comparing the p-value from R with the significance level?*

Here is some code to get you started. Be sure to import the dplyr and tidyverse packages.

```
# Extract decade from year
bb$Decade <- bb$Year - (bb$Year %% 10)

# Create a wins table by summing the wins by decade
wins <- bb %>%
    group_by(Decade) %>%
```

```
                                                    summarize(wins = sum(W))
%>%
    as.tibble()
```

4. Download the file ‘crop_data.csv’ from the course resources and import the file into R.
5. Perform a Two-way ANOVA test using *yield* as the dependent variable and *fertilizer* and *density* as the independent variables. Explain the results of the test. Is there reason to believe that fertilizer and density have an impact on yield?

   ** Be sure to convert the variables density, fertilizer and block to R factors.
   **Include a null and alternate hypothesis for both factors and the interaction.