

Quiz 10.

Part 1.

Textbook problems to be completed on another paper and turned in as “page 0” file of this quiz: Put them in the order listed here. (I prefer that you start each section on a separate piece of paper.) (You may discuss these Part 1 problems on the Disc Bd)

Sec. 7.1 7.13, 7.19

Sec. 7.2 7.45, 7.53

Sec. 8.1 8.29, 8.37

Part 2.

1. Refer to exercise 7.26 for background about Benford’s Law. Professor Rick Cleary of Bentley University has given several public lectures about Benford’s law. As part of his presentation, he rips out pages of a telephone book and asks audience members to select entries at random and record the first digit of the street address. Counts for the leading digits of 1188 such addresses are shown below.

Do not type this data in. It is already entered (both the expected probabilities and the data) in the drop-down menu in StatKey for the appropriate menu item for this test. It is called **Benford (Address digits)**.

Perform a chi-square goodness of fit test to address the question of whether these counts are inconsistent with the probabilities given by Benford’s law.

Table 7.18 *Leading digits for 1188 addresses*

Leading digit	1	2	3	4	5	6	7	8	9
Observed count	345	197	170	126	101	72	69	51	57

- a. State hypotheses.

Use Statkey to calculate the χ^2 test statistic for a Test of Goodness of Fit.

- i. Include a screenshot of the Original Sample pane with this calculation and write the value here.

$$\chi^2 = \underline{\hspace{2cm}}$$

- b. What are the correct degrees of freedom for this test?
- c. Obtain an exact p-value from the appropriate theoretical distribution in StatKey and make a quick sketch of the graph, with the horizontal axis labeled, and shade the appropriate tail area

exact p-value from StatKey:

- d. State your conclusion, in context, for the 10% significance level.

2. Our textbook includes the dataset **NutritionStudy** which contains, among other things, information about vitamin use and gender of the participants. Is there a significant association between these two variables? Use Statkey with the variables **Vitamin** (vitamin use: 1=Regularly, 2=Occasionally, 3=No) and **Gender** to conduct a chi-square test of association.
- State hypotheses.
 - Are conditions met for using a chi-square distribution? Clearly indicate what you checked. (List all the conditions and discuss each one separately.)
 - Use Statkey to calculate the χ^2 test statistic for a χ^2 Test of Association.
 - Include a screenshot of the Original Sample pane with this calculation and write the value here. _____
 - What are the correct degrees of freedom for this test? (Show work) _____
 - Generate a χ^2 distribution (*not* a randomization dotplot) with the correct degrees of freedom in Statkey and get a p-value.
 - Include a screenshot of the distribution curve showing the p-value.
p-value from Statkey: _____
 - State your conclusion in context using “strength of evidence” language.
 - Which cell contributes most to the chi-squared statistic?
 - For that cell, is the observed data greater than or less than the expected value?
 - What does that cell indicate about the differences between males and females and their vitamin use?

3. Consider the same question as in the previous question (Vitamins and Gender), but now use the randomization method to test the same hypotheses you used in that question.
- a. Test the hypotheses using the randomization method and get a p-value.
Include a screenshot of the randomization distribution showing the p-value.
p-value from Statkey: _____
- b. How much different are the p-values for the same hypothesis test on the same data from the two methods (theoretical dist'n method and randomization dist'n method) ? (Show how you computed the difference as well as the difference itself.)
- c. Does these p-values seem very different to you? Explain. (Say "Yes" or "No" and explain.)
- d. Generally it is acceptable to find the p-value using either the theoretical distribution or the randomization distribution. Do you have an opinion about which you'd rather use? Why?

4. A university student decided to set up a formal experiment to see if ants are more attracted to some common sandwich fillings. He chose three types of fillings to compare: vegemite, peanut butter, and ham & pickles. After several minutes he placed a jar over the sandwich bit and counted the number of ants. He repeated this process eight times to get eight samples. The data are stored in the file **SandwichAnts**.

- a. Open the data set in Descriptive Statistics to generate comparative histograms and summary statistics. Check conditions for using the F-distribution. Explain how each condition is verified, violated, or not possible to verify.

i. **Include a screenshot of the descriptive statistics page you used.**

- b. Open the data set in ANOVA for Difference in Means. Fill in the blanks below, and make a quick sketch of the appropriate F distribution graph and shade in the p-value.

Include a screen shot of the information under Original Sample.

F = _____

Degrees of freedom for groups: _____

Degrees of freedom for error _____

p-value estimate = _____

- c. Write a conclusion to the hypothesis test, in context, using the “strength of evidence” language.

- d. If there is evidence of a difference, which means appear to be the most different? Write the conclusion in context (about what, on average, ants like to eat!)