Exploration 8.3: The Monday blues. Or are you more likely to have a heart attack on some days of the week than others?

Chi-Square Goodness-of-Fit Test

LEARNING GOALS

- Calculate expected counts based on the hypothesized model in a chi-square goodness-of-fit test.
- Conduct a simulation-based chi-square goodness-of-fit test using the MAD (mean absolute difference) statistic in the **Goodness of Fit** applet.
- Conduct a simulation-based chi-square goodness-of-fit test using the chi-square statistic in the **Goodness of Fit** applet.
- Identify whether or not validity conditions are met for a theory-based chi-square goodness-of-fit test.
- Conduct a theory-based chi-square goodness-of-fit test.

Step 1: Ask a research question.

Are heart attacks distributed equally across the seven days of the week or are you more likely to have a heart attack on some days than others?

Step 2: Design a study and collect data.

A study done in Augsburg, Germany (Willich et al., 1994) looked at this question. One subgroup of heart attack victims that they looked at were those who were employed. We'll start by analyzing that data. To do this, we first need to set up our hypotheses. The null hypothesis is:

H_0: Heart attacks are distributed equally across the days of the week for employed people in the population.

1. The alternative hypothesis is just the opposite of the null. Write out the alternative hypothesis.

The opposite of the null is that heart attacks are not distributed equally across the days of the week. Note that the alternative hypothesis is not saying that there is a difference just on one specific day or that every day is different than every other day. The alternative hypothesis is just that at least one day is different from the others in terms of the probability of having a heart attack.

Step 3: Explore the data.

The researchers found that the 884 heart attacks from employed people in Germany were distributed across the seven days of the week as shown in the following table.

Day	Sun	Mon	Tue	Wed	Thur	Fri	Sat
Number of heart attacks	106	160	123	115	141	107	132

We think of these 884 heart attacks as a sample. Although this isn't a random sample, this group probably is representative of some larger population. (We will discuss this in more detail later.)

2. What is the variable in this study? Is it categorical or quantitative? If it is categorical, how many categories are there?

These materials were developed by the STUB Network and supported by the National Science Foundation under Grant NSF-DBI 1730668. They are covered under the Creative Commons license BY-NC which allows users to distribute, adapt, and build upon the materials for noncommercial purposes only, and only so long as attribution is given to the STUB Network.

3. Let's compare the observed distribution to one that is distributed equally across the seven days.

    a. There are 884 total heart attacks, how many would we expect on each day if heart attacks were distributed equally across all seven days? (Round your answer to one decimal place.)

    b. In the observed data, what day of the week is farthest from what would be expected? What day is closest?

    c. If heart attacks are distributed equally across the seven days in the population, do the data collected here seem to be likely or surprising?

Step 4: Draw inferences.

To determine whether the variation in the observed data from what is expected is more than we would expect from sampling error alone, we need to conduct a goodness-of-fit test.

Definition
A ***goodness-of-fit test*** is used to compare observed counts for a categorical variable to hypothesized probabilities.

To conduct a goodness-of-fit test using the 3S strategy we first need to find a statistic that measures how far away the observed results are from the values we would expect if the days were equally likely in the population. The following table shows our original data for the 884 heart attacks, though we have renamed this row the **observed** number of heart attacks because this is the sample we observed. We have also added a row for the **expected** number of heart attacks for an equal distribution of heart attacks across the seven days.

	Day						
	Sun	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**
OBSERVED number of heart attacks	106	160	123	115	141	107	132
EXPECTED number of heart attacks	126.3	126.3	126.3	126.3	126.3	126.3	126.3

Simulation-based Goodness-of-fit Test: Using the MAD Statistic

We will first conduct a simulation analysis using the mean absolute difference (MAD) statistic.

Statistic

4. The MAD statistic is found by determining the mean of the absolute values of the differences between the observed and expected number of heart attacks.

    a. Complete the following table to find the seven differences between observed count and expected count.

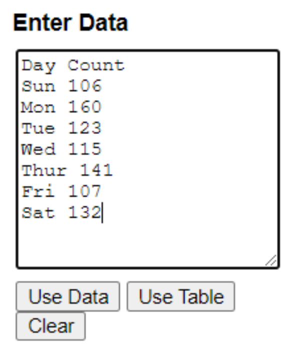
	Day						
	Sun	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**
OBSERVED number of heart attacks	106	160	123	115	141	107	132
EXPECTED number of heart attacks	126.3	126.3	126.3	126.3	126.3	126.3	126.3
Observed count – Expected count							

b. Now find the MAD statistic by finding the mean of the absolute values of your seven differences.

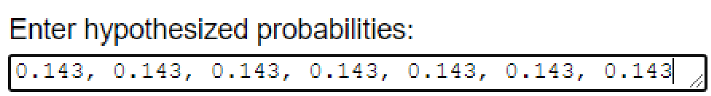
c. What kind of values of the MAD statistic will give strong evidence that heart attacks are not equally distributed across the seven days of the week? Large? Small? Positive? Negative?

Simulate

You could spin a spinner with seven equal sections to find out what values of the MAD statistic could happen by chance alone and repeat this 884 times, but we will use an applet.



- Open the **Goodness of Fit** applet.
- Press the **Clear** button, enter two column names (e.g., day count), and then specify each day of the week followed by a space and the count from your sample (e.g., Sun 106). Hit return and repeat for each day of the week. Then press **Use Table** (not Use Data). This should update the bar graph under "Sample Data."
- Near the bottom of the left panel, chose MAD for the **Statistic**. Confirm this is the value you calculated in #4(b).
- In the right panel, check the box next to **Show Sampling Options** and enter the seven hypothesized probabilities under the null to three decimal places (e.g., 1/7 = 0.143), separated by commas.



- Keep the **Number of Samples** set to 1 and push **Sample**.

5. What is the resulting "Most Recent Sample MAD"? Is this value more or less extreme than the one from the observed data? How are you deciding?

Now change the **Number of Samples** to be 999, for a total of 1000 samples, and push **Sample**. This will produce a simulated null distribution of 1,000 MAD statistics.

Strength of evidence

6. Does the observed MAD statistic appear unlikely to have happened by chance alone, that is, if the null hypothesis of all days being equally likely were to be true? How are you deciding?

7. Based on your answer to #4(c), in finding the p-value, will you count samples greater than or equal to the observed MAD from the data, less than or equal to the observed MAD, or more extreme (which will include positive and negative values) than the observed MAD? Why?

8. To find the p-value to quantify the strength of evidence against the null hypothesis, enter the observed value of the MAD statistic in the **Count Samples** box, choose the appropriate direction (Greater than, Less than, Beyond) and press **Count**.

9. Based on the p-value you obtained in #8 do you have strong evidence against the null hypothesis? How are you deciding?

Simulation-based Goodness-of-fit Test: Using the Chi-square Statistic

MAD statistics, while easy to understand, aren't standardized. A statistic that is standardized and will work in this situation is a chi-square statistic. Let's conduct a simulation analysis using the chi-square statistic to investigate whether the data provide evidence that heart attacks are not equally likely on all seven days of the week. The formula for the χ^2 statistic is:

$$\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}.$$

Note that the Σ in the formula is used to refer to summation across all the different terms.

Statistic

10. Now let's calculate the chi-square statistic.
    a. To find the chi-square statistic value for our data, first complete the following table:

	Day						
	Sun	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**
Observed count – expected count							
Square the difference							
Divide squared difference by expected count							

    b. The chi-square statistic is the sum of the values in the last row in the above table. What is the value of the chi-square statistic?

Simulation

We now need to determine typical values of the χ^2-statistics that would occur if our null hypothesis was true. Just as you ran the simulation using the MAD statistic, you can follow the same steps this time using the **Statistic** pull-down menu to select the chi-square statistic (**χ^2**). Create a null distribution for the chi-square statistic.

Strength of evidence

11. Where does the observed chi-square statistic lie on the simulated null distribution of the chi-square statistics? Does it appear surprising and unlikely to have happened by chance alone, that is, if the null hypothesis of all days being equally likely were to be true? How are you deciding?

12. Find the corresponding p-value. How does the p-value with the chi-square statistic compare to what you found with the MAD statistic?

13. Based on the p-value reported in #12 from the simulated null distribution for a chi-square statistic, do you have strong evidence against the null hypothesis? How are you deciding?

Theory-based Chi-square Goodness-of-fit Test

You just used simulation to find a p-value using the chi-square statistic. One advantage of using the chi-square statistic (instead of the MAD) is that the distribution of the chi-square statistic can be predicted by a theory-based distribution called the chi-square distribution.

14. Go back to the **Goodness of Fit** applet, and picking up where you left off, check the **Overlay chi-square distribution** box under the null distribution for the chi-square statistics. Report the theory-based p-value.

15. How does the theory-based p-value compare to the one from the simulation-based p-value using the chi-square test statistic? Do you still arrive at the same conclusion about whether there is convincing evidence that heart attacks are not distributed equally among the seven days of the week?

Remember that there are always certain validity conditions to be met before you can decide that it is valid to use a theory-based method to test hypotheses. In particular, the validity condition for a theory-based chi-square goodness-of-fit test is that each of the observed counts is at least 10.

16. Does the theory-based chi-square goodness-of-fit test appear to be valid in this context? Explain how you are deciding.

On the bottom of the left panel of the **Goodness of Fit** applet, there is a **Show χ^2 output** check box. When that box is checked you can see the chi-square statistic (reported as **Sum**) and the p-value for a theory-based test. You do not need to do a simulation to see these results.

STEP 5: Formulate conclusions.

In this study, the researchers collected data from 13 hospitals in and around Augsburg, Germany from 1985 to 1990. In particular, they looked at all the heart attacks that occurred in that period and found which day of the week they occurred. Furthermore, for this specific data set, they focused on heart attacks from people that were employed during that time.

17. State your complete conclusions. In particular, to what population are you willing to generalize your conclusions?

18. You should have found convincing evidence that heart attacks from this population are not equally distributed throughout the week. Which day of the week was the most "off"? That is, which day gave the largest component of the chi-square statistic when you calculated it back in the last row of the table in #10(a)?

STEP 6: Look back and ahead.

You should have found strong evidence that heart attacks are not distributed equally across the seven days of the week for the population of employed people represented by this sample. You should have also seen that the highest proportion of heart attacks occurred on Monday. In fact, there appears to be very strong evidence that the proportion of heart attacks on Mondays is more than 1/7. (What kind of test would you use to test this statement?) Does this same Monday effect occur with unemployed people?

19. The same researchers found that there were 1,191 heart attacks among those that were not employed. The heart attacks in this group were distributed as shown below. Again, we ask the same question as earlier: Do we have strong evidence that heart attacks are not distributed equally across the days of the week in this population?

Day	Sun	Mon	Tue	Wed	Thur	Fri	Sat
Number of heart attacks	168	180	157	169	180	169	168

a. Write the null and alternative hypotheses for this test.

b. Put the table of data into the **Goodness of Fit** applet (remember to press **Use Table**). Based on the distribution of heart attacks, do you think there will be strong evidence that heart attacks are not distributed equally across the seven days of the week? Why or why not?

c. Calculate the χ^2-statistic using the applet. How does this statistic compare with the 18.27 from the employed data? What does the difference between these two values mean in this context?

d. Calculate both a simulation-based p-value and theory-based p-value using the applet. Are these the types of numbers you would expect based on what the distribution looked like and the size of the χ^2-statistic?

e. Write out a conclusion in the context of this situation.

Reference

Willich SN, Löwel H, Lewis M, Hörmann A, Arntz HR, Keil U. Weekly variation of acute myocardial infarction. Increased Monday risk in the working population. Circulation. 1994 Jul;90(1):87-93. doi: 10.1161/01.cir.90.1.87. PMID: 8026056.