Is It Fair?

Overview and student objectives

In-Class Activity Length: 25 minutes

Overview

In this in-class activity, students will conduct a full chi-square test for goodness of fit, including checking conditions and drawing a conclusion. Students will perform this task with a public health scenario, which asks them to explore whether health resources are distributed fairly between counties in New York.

Objectives

Students will understand:

- For a chi-square test to be valid, certain conditions must be satisfied.
- A large chi-square test statistic value usually corresponds to a small P-value, meaning it's unlikely the observed values could differ more than expected by chance alone when the null hypothesis is true.
- A large chi-square test statistic value and small P-value indicate that we can reject the null hypothesis, supporting the conclusion that the population distribution differs from the claimed distribution.

Students will be able to:

- Conduct a full chi-square test for goodness of fit and interpret its conclusion.

Suggested resources and preparation

Materials and technology

- Computer, projector, document camera
- Student Pages for In-Class Activity
- Practice Assignment
- Access to the *DCMP Data Analysis Tools*

Prerequisite assumptions

Students should be able to:

- Calculate the value of a chi-square test statistic and explain what it means in the context of a claim.
- Write the null and alternative hypotheses of a chi-square test for goodness of fit.

Making connections

This activity:

- Connects back to the calculation of the chi-square test statistic.
- Connects forward to all future chi-square procedures (independence and homogeneity).

Background context

This activity focuses on mock data related to vaccine distribution during a pandemic. During public health crises, governments distribute treatments or aid to citizens through state and local channels. Often, imbalances of resource distribution between different areas or different demographics can become sources of controversy. This activity will explore mock data in such a context.

Suggested instructional plan

Frame the activity (3 minutes)

Resources and Structure	Instructor Suggestions
Think-Pair-Share	• Have students read and answer Question 1 independently. Then, have them discuss their answers in pairs before sharing with the whole class. • Students may be confused by the designations of the Bronx and Queens as "counties" since both are boroughs of a city (New York City). As it turns out, New York City is so large that each of its boroughs is its own county. Explain to students that we can think of each of these counties as their own areas in and around New York City (Queens and the Bronx in New York City and Westchester as a set of suburbs). Westchester is home to many wealthier suburban towns, which is why it has a higher median income than the city boroughs. • Transition to the in-class activity by briefly discussing the **Objectives for the activity**.

Activity flow (20 minutes)

Resources and Structure	Instructor Suggestions
Pairs	Questions 2 and 3 • Remind students of the steps in a hypothesis test—research question or claim, hypotheses, select a test, check assumptions, calculate, and conclusion in context. • Have students work in pairs, and then ask a few volunteers to share their answers with the whole class.

Small Groups	• Before Question 3, make sure to discuss the conditions with the whole class, along with the reasons why we check each one. This will be the first time students see these specific conditions; however, they've seen similar conditions for prior hypothesis tests. There is no need to go into too much detail unless necessary. Question 4 • Transition students to answer this question in small groups so they can help one another with the technology steps. Questions 5–7 • Students should be building intuition for the chi-square statistic and associated right-tail probability. Encourage them to refer to prior learning for help with questions of interpreting the value of the chi-square statistic or the P-value.

Wrap-up/transition (2 minutes)

Resources and Structure	Instructor Suggestions
Wrap-up	• Discuss Questions 5–7. • In discussing the test results in the context of the idea that vaccines disproportionately flow towards the wealthier area (Westchester County), note that the results are consistent with the idea that wealth is the causal mechanism behind why Westchester gets more vaccines than expected. However, this test does not prove that wealth is the true cause. There may be multiple confounding factors, other than wealth, that result in this disproportionate vaccine flow. • As a final formative assessment, ask students if the chi-square test statistic value would have been larger or smaller if the sample size was larger (and the proportions of vaccines actually given to each county were the same). • Have students refer back to the **Objectives for the activity** and check the ones they recognize. Alternatively, they may check the objectives throughout the activity.

Suggested assessment, assignments, and reflections

- Give Practice Assignment.

Is It Fair?

During public health crises, governments often aim to distribute resources and aid to citizens fairly. However, in practice, this may not always be the case. We will explore some mock data about vaccine distribution in New York during a pandemic.

Imagine a vaccine for a pandemic is produced, but there is a limited supply. Although the government promises that the vaccine will be distributed fairly, New Yorkers wonder whether the government is distributing more vaccines to wealthy individuals.

1) The State of New York claims that each county receives a number of vaccines that is proportional to its population size. An independent inspector is granted access to obtain a random sample of vaccine box shipping labels in a warehouse shipping to only three New York counties: the Bronx, Queens, and Westchester. The median income is highest in Westchester County, followed by Queens and finally the Bronx. Below are the populations of all three counties (as of 2019). Find the percentages of vaccine doses in the warehouse that should go to each county:

Source: U.S. Census Bureau

County	Population Size
Queens	2,287,000
Bronx	1,435,000
Westchester	968,000

Objectives for the activity

You will understand:

- ☐ For a chi-square test to be valid, certain conditions must be satisfied.
- ☐ A large chi-square test statistic value usually corresponds to a small P-value, meaning it's unlikely the observed values could differ more than expected by chance alone when the null hypothesis is true.
- ☐ A large chi-square test statistic value and small P-value indicate that we can reject the null hypothesis, supporting the conclusion that the population distribution differs from the claimed distribution.

You will be able to:

- ☐ Conduct a full chi-square test for goodness of fit and interpret its conclusion.

The independent inspector collects a random sample of 500 shipping labels for vaccine boxes being distributed by the warehouse (all boxes have an equal number of vaccine doses). She would like to test if there's convincing evidence against the claim that vaccine shipments are proportional to each county's population size using a **chi-square test for goodness of fit**.

2) Write the null and alternative hypotheses for this test.

Now, we have to check conditions for this test to make sure it's valid. Here are the conditions for the chi-square test for goodness of fit:

- **Random**: Observed counts must come from a random sample (to ensure our conclusions are free from sampling bias).
- **10%**: The sample size must be less than a tenth of the population size (to satisfy independence assumptions).
- **Large Sample**: The sample is large enough such that the expected counts are all five or greater (to ensure our sampling distribution resembles a chi-square distribution).

3) For the sample collected by the independent inspector, check each of the conditions described previously. Show your work in checking the large sample condition.

4) Below are the observed counts of vaccine shipments designated to each county.

County	Queens	Bronx	Westchester
Observed Count	204	132	164

Go to the *DCMP Chi-Squared Test* tool at https://dcmpdatatools.utdanacenter.org/chisquaredtest/. Select the **Goodness of Fit** tab at the top of the data analysis tool. Under "Enter Data," choose "Contingency Table." Change each of the counts to the observed counts from the previous table. Change each of the "props" to the expected proportions. Then press "Submit."

Report the degrees of freedom, chi-square test statistic value, and the *P*-value.

5) Was your chi-square value large or small? Explain. What does its value indicate about the claim that vaccine shipments are made proportionally to each county's population size?

6) Interpret the meaning of the *P*-value in the context of this problem.

7) Draw a conclusion about the hypotheses based on your test results. Use a significance level of $\alpha = 0.05$. In addition, comment on whether your test results are consistent with the idea that vaccine shipments flow disproportionately to wealthier areas.

Practice Assignment

1) A large corporation has offices in Chicago, New York, Boston, Los Angeles, and Houston. To boost employee morale, the corporation gives a gift basket to a randomly selected employee every week. The company claims that it randomly selects an office location each week and rewards the basket to an employee at that location. However, you've noticed that your location (Houston) has received fewer baskets than the other locations. You'd like to investigate the company's claims about their selection process. The actual number of basket giveaways per location are shared below:

Location	Chicago	New York	Boston	L.A.	Houston
Number of baskets	29	35	33	36	23

Perform a chi-square test for goodness of fit to state a conclusion about the company's claim. Make sure to include the hypotheses, expected counts, checked conditions, degrees of freedom, chi-square test statistic, P-value, and conclusion.

You may use the *DCMP Chi-Squared Test* tool at https://dcmpdatatools.utdanacenter.org/chisquaredtest/ to answer the question. Use a significance level of $\alpha = 0.05$.

2) Recently, a large music streaming service has taken steps to try to rid its "new music" recommendation algorithm of gender bias. Specifically, when suggesting a group of new artists to its users, the company's goal is for the gender distribution of the recommended artists to match the gender distribution of all artists on the streaming platform. Listed below are the proportions of all artists on the streaming service from each gender group (that the company collects information about):

Gender	Non-Binary	Female	Male
Proportion	0.08	0.43	0.49

The company collects a random sample of 120 recommended artists from its new algorithm. Here is the gender information for the 120 sampled recommended artists:

Gender	Non-Binary	Female	Male
Proportion	4	48	68

Perform a chi-square test for goodness of fit to determine if the algorithm is truly making recommendations proportional to the total number of artists of each gender category on the streaming platform. Make sure to include the hypotheses, expected counts, checked conditions, degrees of freedom, chi-square test statistic, P-value, and conclusion.

You may use the *DCMP Chi-Squared Test* tool at https://dcmpdatatools.utdanacenter.org/chisquaredtest/ to answer the question. Use a significance level of $\alpha = 0.05$.