**Unit 7: Investigation 5 (3 Days)**

**Interpreting Two-Way Frequency Tables**

**Common Core State Standards**

* ID-5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies) Recognize possible associations and trends in the data.
* CP-4. Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities. *For example, collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*

**Overview**

In the launch discussion, students are presented with results from two polls, a Gallup poll that asks about political preference and the distribution of money and wealth, and another poll that asks for participants’ gender and their beliefs about love at first sight. These two examples introduce students to two-way tables and conditional distributions. Students then organize data from the class on gender and (teacher assigned) political preference into a two-way table and calculate marginal, joint, and conditional relative frequencies/percentages. Next, students analyze data collected from a national survey of 12th-grade students. As part of the analysis they represent marginal and conditional percentages with bar graphs. From two-way tables of percentages and bar graphs, students interpret the percentages in the context of the data and identify associations and trends.

**Assessment Activities**

**Evidence of Success: What Will Students Be Able to Do?**

* Organize data from two survey questions into a two-way frequency table.
* Given data organized in a two-way frequency table, calculate the marginal relative frequencies/percentages.
* Given data organized in a two-way frequency table, calculate the joint relative frequencies/percentages.
* Given data organized in a two-way table, calculate conditional relative frequencies/percentages.
* Represent marginal and conditional percentages with bar graphs.
* Interpret marginal, joint, and conditional percentages in the context of the data.
* Describe associations and trends of data in two-way tables.

**Assessment Strategies: How Will They Show What They Know?**

* **Exit Slip 7.5.1** asks students to find marginal, joint, and conditional percentages from data organized in a two-way table.
* **Exit Slip 7.5.2** asks students to find conditional percentages, and represent them with bar graphs. Students then interpret these percentages in the context of the data and describe any trends/associations that they find.
* **Journal Entry** asks students to differentiate between a joint percentage and two conditional percentages. .

**Launch Notes**

In this unit students will analyze data sets that have been organized into two-way tables. One natural source for such data is public opinion polls. In preparation for Investigation 5, ask students to conduct an Internet search to find some results from recent public opinion polls. Two examples follow.

In a Google search, we entered the phrase *public opinion polls*. The search results listed some well-known polls such as the Gallup Poll, the Pew Research Center, and the Roper Center for Public Opinion Research. As a result of this search we clicked on the following website:

[www.pollingreport.com/life.htm](http://www.pollingreport.com/life.htm).

This site reported results from a Gallup Poll of 1015 U.S. adults. The poll was conducted over the period April 9 – 12, 2015. One of the questions on the survey is stated below.

“Do you feel that the distribution of money and wealth in this country today is fair, or do you feel that the money and wealth in this country should be more evenly distributed among a larger percentage of the people?”

Participants in the survey could choose among the following responses: Fair, Should be more even, or Unsure. Invite your class to respond to this survey question.

* Each student should put a tally mark under their response into a copy of Table 1. (Table 1 can be found in PowerPoint Presentation *Two-Way Tables*.)
* Convert the tally marks to frequencies for each response. Then convert the frequencies to relative frequencies by dividing them by the number of students in the class.
* Finally, to change frequencies to percentages, multiply the relative frequencies by 100.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Fair | Should be more even | Unsure |
| Tally |  |  |  |
| Frequency |  |  |  |
| Relative Frequency |  |  |  |
| Percent (%) |  |  |  |

Table 1. Student responses to the Gallup poll question.

The results from the actual Gallup poll appear in Table 2. Allow students a few minutes to compare their responses from Table 1 to the percentages given in Table 2.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Fair | Should be more even | Unsure |
| Percent | 31 | 63 | 6 |

Table 2. Percentage of participants’ responses in each category.

Because the responses to a question can vary from participant to participant, we can treat each survey question as a **variable**. A major problem with considering this one question (this one variable) is that participants’ political preferences might have a large impact on their responses to this question. Another survey question asked participants to check off their political preferences: Democrat, Independent, or Republican. Using the two questions together, we can see if there is an **association** between the responses to the two questions. The data from the two questions are organized into Table 3, which is called a **two-way table**. (Table 3 can be found in PowerPoint Presentation *Two-Way Tables*.) The numbers in this two-way table are percentages corresponding to each political affiliation.

|  |  |  |  |
| --- | --- | --- | --- |
| Percent | Fair | Should be more even | Unsure |
| Democrat | 12 | 86 | 2 |
| Independent | 32 | 61 | 7 |
| Republican | 57 | 34 | 9 |

Table 3. Two-way table records information on responses to two questions.

In Table 3, the rows are labeled with the possible outcomes for political preference. So, political preference is called the **row variable**. The columns are labeled with the possible outcomes for the question on wealth distribution. So, wealth distribution is the **column variable**.

Give students a few minutes to study Table 3 before discussing these results. Here are two questions that might help start a discussion:

* Are Democrats or Republicans more likely to feel that the distribution of money and wealth is fair? Support your answer with appropriate percentages.

Sample answer: Republicans; 57% of Republicans responded Fair compared to only 12% of Democrats.

* What makes the results presented in Table 3 more interesting/informative than presenting two tables, one for each question separately (similar to Table 1)?

Sample answer: Tables such as Table 1 show the results on one question at a time. Even if you were presented with a similar table showing the results of the question on political preference, the two tables together would not show how the responses to the two questions were related. In this case, that is important.

Don’t worry if students have trouble interpreting Table 3 at this point. This type of percentage (row percentages or conditional percentages of wealth distribution for each level of political preference) will be more formally introduced in **Activity 7.5.1**. However, without any further explanation, students probably will not have any trouble interpreting the results from the next survey.

Around Valentine’s Day each year, a number of polls ask men and women the following question: “Do you believe in love at first sight.” Table 4 presents the results from a recent Singles in America survey. (Again, the numbers in the table are percentages. A copy of this table can be found in the PowerPoint Presentation *Two-Way Tables*.) Give students a few minutes to discuss the results. Most students will note that a higher percentage of men (59%) believe in love at first site compared to women (49%).

|  |  |  |
| --- | --- | --- |
| Percent | Men | Women |
| Yes | 59 | 49 |
| No | 41 | 51 |

Table 4. Two-way table of men’s and women’s responses to “love at first sight” survey.

At this point ask students to conduct and Internet search to find recent surveys that they find interesting. If you have time, and students have access to computers at school, they can conduct their searches during class time. Otherwise, give this as a homework assignment. Of particular interest are surveys that report results on connections between two (or more) questions.

**Teaching Strategies**

After the launch discussion, students are ready to begin **Activity 7.5.1 Gender and Political Preference. Activity 7.5.1** is in two parts. In Part I, students use class data on gender and political preference (assigned by the teacher) to compute relative frequencies and percentages in which the “part” to the “whole” keeps changing. Students work their way through calculating marginal, joint, and conditional relative frequencies and percentages. In addition, they represent marginal and conditional distributions with bar charts. In Part II students analyze real data from a national survey of 12th-grade students on the same two variables, gender and political preference.

**Preparation for Activity 7.5.1**

First, you will need to prepare papers that will be used to assign students to a political preference. If possible use red paper to designate Republicans, blue paper to designate Democrats, and white paper to designate Independents. (In addition, you may want to write the words Democrat, Republican, and Independent on these papers.) In assigning political preference, you may want to stay close to the percentages from a General Social Survey (GSS) poll, which reported that 46% of the respondents were Democrats (or leaning Democratic), 20% were Independents, and 34% were Republicans (or leaning Republican). For example, in a class of 20 students, assign approximately 9 Democrats, 4 Independents, and 7 Republicans. Shuffle the political preference papers and then distribute to your students.

The next step in preparation for **Activity 7.5.1** requires students to organize the class data on gender and assigned political preference into a blank two-way table similar to Table 5. (A blank copy of the table below (see Table 1 Class Data) can be found in PowerPoint Presentation *Two-Way Tables*.) Each student should put a tally mark in the appropriate white box. Table 5 below shows the results after the first two students, a male Independent and a female Republican, entered their data. After all students have completed this task, convert the tally marks to frequencies and complete the table by entering the marginal totals (the last column holds the row totals and the last row holds the column totals).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Frequency | Democrat | Independent | Republican | Total |
| Female |  |  | **|** |  |
| Male |  | **|** |  |  |
| Total |  |  |  |  |

Table 5. Example of first two students classified by gender and political preference.

The number entered into the dark gray cell, which is the sum of the column totals or the sum of the row totals, is called the **grand total**, which in this case is the number of students in the class. The **row variable** is gender and the **column variable** is political preference.

**Activity 7.5.1** begins with three different types of percentages that were calculated from data collected in a General Social Survey (GSS) poll.

 (1) 58% of Democrats were female

 (2) 49% of females were Democrats

 (3) 20% of the respondents were female and Democrats.

(For information about GSS, see [www.norc.org/Research/Projects/Pages/general-social-survey.aspx](http://www.norc.org/Research/Projects/Pages/general-social-survey.aspx).)

Give students a few minutes to think about these three percentages and to consider why they are different. Do not clarify or try to explain the differences in these percentages. Students will revisit these percentages at the end of the activity, after they have had some experience calculating marginal, joint, and conditional percentages.

**Group Activity 7.5.1:** *Gender and Political Preference*

**Part I:** Is a whole class activity. Students use the class data from completed Table 5 to compute relative frequencies in which the “part” to the “whole” keeps changing. Starting with question 2, students should physically move into appropriate groups so that they can visualize both the “part” and the “whole” that make up the relative frequencies. For each new relative frequency requested, students compute the relative frequency and percentage appropriate for their group and then combine results from other groups to complete a table. (Blank tables corresponding to Tables 1–4 in the activity can be found in PowerPoint Presentation *Two-Way Tables*.) By the end of Part I, the class will have a complete set of marginal, joint, and conditional percentages for the class data.

**Part II:** Can be completed individually or in small groups. Students analyze data collected from a national survey of 12th-grade students on the same two questions used in the class data. If there is not sufficient time to complete Part II (questions 6–9) during class, it can be assigned for homework.

Here is a recap of the types of percentages covered in **Activity 7.5.1** along with the formulas for calculating each of them.

**Joint percentages** of two variables: [$\frac{Cell Entry}{Grand Total}$× 100]%.

Taken together the joint percentages form a distribution and sum to 100%.

**Marginal percentages** for one variable: [$\frac{Total Entry}{Grand Total}$× × 100]%.

The marginal percentages for the row variable form a distribution and sum to 100%.

The marginal percentages for the column variable also form a distribution and sum to 100%.

**Conditional percentages** provide the most insight into the association between two variables.

* To find the conditional percentages for the row variable for each level of the column variable, compute the **column percentages**: [$\frac{Cell Entry}{Column Total}$×100]%. The column percentages, for each level of the column variable, form a distribution and sum to 100%.
* To find conditional percentages for the column variable for each level of the row variable, compute the **row percentages**. [$\frac{Cell Entry}{Row Total}$×100]%. The row percentages, for each level of the row variable, form a distribution and sum to 100%.

After students have completed **Activity 7.5.1**, show the *Against All Odds* video from Unit 13 Two-Way Tables ([www.learner.org/courses/againstallodds](http://www.learner.org/courses/againstallodds)) or assign the video as homework. This video discusses results from a “Happiness Survey,” which was conducted in Somerville, Massachusetts. Town residents were asked to rate their happiness and the physical beauty of Somerville. The video provides a good review of marginal and conditional percentages. In addition, a bar graph is used to find trends between respondents’ happiness levels and how they rated the physical beauty of Somerville. This is helpful preparation for **Activity 7.5.2**. Even more importantly, this video sets up Unit 7’s performance task, in which students will conduct their own “Happiness Survey.”

**Exit Slip 7.5.1** should be assigned at this point to check whether students can calculate marginal, joint, and conditional percentages.

**Activity 7.5.2** **Calculations of Marginal, Joint, and Conditional Percentages**is designed for differentiated instruction (for learners needing more help). The activity uses (approximate) data from the Gallup poll that was part of the launch discussion. The focus is on calculating marginal, joint, and conditional relative frequencies and percentages. The formulas for each type of relative frequency/percentage are given prior to their use.

**Differentiated Instruction (For Learners Needing More Help)**

**Activity 7.5.2** shows the calculations for marginal, joint, and conditional relative frequencies and percentages. In each table that students complete, at least one of the entries has been calculated. This gives students a chance to check an initial calculation before completing the table.

In addition for use as differentiated instruction, feel free to use **Activity 7.5.2** for review either in class or as homework.

This would be a good point to assign the Journal Entry.

**Journal Entry** *Interpreting Three Percentages*

Students (12th graders) were asked the following question: How likely is it that you will graduate from college (four-year program)? Explain the differences between the following percentages based on their responses:

1. 69.5% of the female students responded definitely will.

2. 55.8% of the students who responded definitely will were female.

3. 34.94% of the participants identified themselves as female and responded definitely will.

Look for students to note that the first two percentages are conditional percentages. For the percentage 69.5%, only the responses from female students were considered. For the percentage 55.8%, only the responses from students who reported that they definitely will graduate from college were considered. The last percentage is a joint percentage.

In **Activity 7.5.3 Survey of 12th Grade** Students, students analyze data from 12th-grade students responses to six questions from a Monitoring the Future survey. After computing percentages and representing them with bar charts, students look for associations between variables (questions) and describe trends.

The data in **Activity 7.5.3** come from The Monitoring the Future (MTF) survey. This survey was administered to 15,198 students from high schools across the country. Because the sample size is large and the sample was selected to be representative of all 12th-grade students in the U.S., relative frequencies computed from MTF data can be used as estimates of probabilities. It should be noted that students participating in the survey have the option to not answer questions if they do not feel comfortable answering. Therefore, the grand total in each two-way table, which gives the number of students who answered both questions, will vary from table to table. To learn more about Monitoring the Future, visit [www.monitoringthefuture.org/](http://www.monitoringthefuture.org/).

After students have completed **Activity 7.5.3**, review their answers to questions 3 and 6. The appropriate conditional percentages for question 3 are row percentages while the appropriate conditional percentages for question 6 are column percentages. Make sure that students know how to draw a bar chart representing conditional percentages.

**Exit Slip 7.5.2** can be assigned after students have completed **Activity 7.5.3**.

**Activity 7.5.4** **Gender, Intelligence, and College** is designed for differentiated instruction (enrichment). Responses to three survey questions have been organized into a three-way table. Students calculate a variety of percentages. Then they collapse the table over one of the variables to form a two-way table. The last question asks students to analyze the data any way that they see fit and to write a brief report on their findings.

**Differentiated Instruction (Enrichment)**

**Activity 7.5.4** presents data organized into a three-way table. Adding the third variable makes analysis considerably more challenging. Students could simply answer questions 1 – 3. Question 4 is very open-ended and asks students to analyze the data any way that they see fit. A thorough analysis would be quite a challenge.

You should begin the **Unit 7 Performance Task** at this point. The **Performance Task: Happiness Survey** asks students to write questions to be used in a survey, administer the survey, organize the data into two-way tables, calculate appropriate percentages, create appropriate graphic displays and interpret the results. Students will need time to create questions for their questionnaire, collect the data, and enter the data into an Excel spreadsheet.

**Closure Notes**

In this investigation we have discussed three different types of relative frequencies and percentages: marginal, joint, and conditional. Ask students what information they learn from conditional distributions that they could not get from marginal percentages. Connect student responses to the differences between calculating *P*(*A*) and *P*(*B*) for two events *A* and *B*, and calculating *P*(*A*|*B*) or *P*(*B*|*A*). If events *A* and *B* are independent, *P*(*A*) = *P*(*A*|*B*) and *P*(*B*) = *P*(*B*|*A*) and the distinction does not matter. However, that is not the case when events *A* and *B* are dependent. In that case, understanding how the likelihoods of the two events depend on each other is what is interesting.

**Vocabulary**

**Column variable:** The variable whose possible outcomes are the labels for the columns in a two-way table.

**Conditional relative frequencies/percentages:** A relative frequency computed from a two-way frequency table by dividing a cell entry by a row or column total. To compute a conditional relative frequency to a percentage, multiply by 100 and add the percent symbol (%). Conditional percentages provide one way to explore the relationship between the row and column variables.

**Grand total:** The sum of all the entries in a two-way table.

**Joint relative frequencies/percentages:** A relative frequency computed from a two-way frequency table by dividing a cell entry by the grand total. To compute a joint relative frequency to a percentage, multiply by 100 and add the percent symbol (%).

**Marginal relative frequencies/percentages:** A relative frequency computed from a two-way frequency table by dividing the row or column totals by the grand total. To compute a marginal relative frequency to a percentage, multiply by 100 and add the percent symbol (%).

**Row variable:** The variable whose possible outcomes are the labels for the rows in a two-way table.

**Two-way table:** A table with *r* rows and *c* columns that organizes data on two categorical variables taken from the same respondent. The rows are labeled with the possible outcomes from the row variable and the columns are labeled with possible outcomes of the column variable.

**Variable:** Describes some characteristic, attribute, or response to a survey question that can vary in value.

**Resources and Materials**

Activity 7.5.1: Gender and Political Preference

Activity 7.5.2: Calculations of Marginal, Joint, and Conditional Percentages

Activity 7.5.3:

Activity 7.5.4: Gender, Intelligence, and College (Differentiated instruction for enrichment)

Exit Slip 7.5.1

Exit Slip 7.5.2

Journal Entry

Performance Task: Happiness Survey

PowerPoint presentation: *Two-Way Tables*

Web sites:

 For public opinion polls

 [www.pollingreport.com/life.htm](http://www.pollingreport.com/life.htm)

 [www.gallup.com/home.aspx](http://www.gallup.com/home.aspx)

 [www.rasmussenreports.com/](http://www.rasmussenreports.com/)

 [www.ropercenter.uconn.edu/](http://www.ropercenter.uconn.edu/)

 [www.pewresearch.org/](http://www.pewresearch.org/)

 For information about GSS, see [www.norc.org/Research/Projects/Pages/general-social-survey.aspx](http://www.norc.org/Research/Projects/Pages/general-social-survey.aspx).

 For video: [www.learner.org/courses/againstallodds](http://www.learner.org/courses/againstallodds), Two-Way Tables (Unit 13)

 For information on the Monitoring the Future study: [www.monitoringthefuture.org/](http://www.monitoringthefuture.org/)

Materials for **Activity 7.5.1**: Paper (preferably in three colors, red, white, and blue)