

Association between Gender and Sport Watching Preference?

Courtney Rubio, Barrow Arts and Sciences Academy

Published: November 2024

Overview of Lesson

Is there an association between gender and the sports students prefer to watch? This activity allows students to analyze data provided in a two-way table using one or more of the following methods: a chi-square test of independence, relative frequencies, segmented bar graph, mosaic plot, or bar chart. The lesson could be completed with or without technological tools, but this lesson plan specifically shows how to carry out the analysis using Stapplet.

Type of Data

- The explanatory variable is categorical (Gender) and the response variable is also categorical (Sport watching preference)
- A dataset is provided in this lesson plan, but it is also possible to collect data through a class survey.

Learning Objectives

- Students will read data presented in a two-way table.
- Students will use relative frequencies, bar graphs, mosaic plots, bar charts, or a chi-square test of independence to draw conclusions about the association between two categorical variables.
- After implementing this lesson fully, students will understand connections among solution methods.

Audience

- The audience for this lesson is a high school statistical reasoning course or an advanced placement high school statistics course.
- This lesson is designed to give students different starting points and extensions due to having multiple solution strategies.
- *Prerequisites:* Prior to this lesson, students should have experience with identifying variables as categorical or quantitative, analyzing data from a two-way table, calculating relative frequencies, creating segmented bar graphs, creating mosaic plots, creating bar charts, writing null/alternative hypotheses, calculating expected counts, calculating test statistics, calculating p-values, drawing conclusions, and performing chi-square test of independence.

Time Required

The time required to implement this lesson is roughly 1 class period (80 minutes)

Technology and Other Materials

- Students may use Chromebooks or other computers to access Stapplet (optional).
- Students may need a calculator to compute expected counts and the test statistic.
- Students will be provided with poster paper, where they will summarize the data and showcase their methods for determining if there is an association between gender and sport preference.

Lesson Plan

This lesson allows students to determine if there is an association between two categorical variables using various previously learned methods. These methods include calculating the relative frequencies of the data set, creating a segmented bar graph, creating a mosaic plot, creating a bar chart, and performing a chi-square test of independence. The students will be placed in assigned groups of five students in each group to analyze the following data:

A survey of 1000 students at a local high school was conducted. The survey asked each student the following two questions: “What gender do you identify with?” and “Out of these three sports, what sport do you prefer to watch?” The possible answer choices for the first question were Female or Male. The possible answer choices for the second question were Tennis, Football, or Soccer.

	Football	Tennis	Soccer	Total
Male	172	162	175	509
Female	110	191	190	491
Total	282	353	365	1000

Note: If you want students to generate their own data before implementing this lesson, one strategy is to have the students survey students in other classes using the prompting questions above and record their data into a two-way table. Then have the students combine their data collection to create a larger sample size. For this lesson, it is important that all groups of students use the same data set to be able to compare their strategies.

Introducing the Investigative Question and the Data

The first part of this lesson revolves around whole class discussion guided by the teacher. The whole class discussion will introduce the investigative question and analyze the two-way table. To motivate the students in engaging in the investigative question, the following video <https://www.youtube.com/watch?v=aR3htgTnzRs> will be played to showcase how gender and sports have been viewed in the past and how our beliefs may impact our assumption of what the conclusion for this investigative question may be. The teacher will follow the video asking the students to share their initial thoughts about the topic and the video. Then the teacher will introduce more guided questions to guide the conversation about the topic. A few questions the teacher may ask are “Do you think that if someone plays the sport, they are more likely to favor watching it?”, “Is football dominantly a female or male sport? How about Soccer? Tennis?”. This discussion encourages the students to explain their thoughts about the topic before analyzing the data and making conclusions about the investigative question.

This introductory conversation about the topic is mainly directed by the teacher with students engaging in the conversation. After this discussion, the teacher will present the description and the data set to the students. (See above)

The teacher will ask the students what they notice about the data set. Possible answers could be that there are 1000 students in the sample, there are an uneven amount of male and female students, most students in the sample answered that they prefer to watch soccer over tennis and football, etc.

Analyzing the Data

The teacher will provide the following instructions to the students:

Working with your group, create a poster following the given handout template to show your statistical process for answering the investigative question. You should have all the information included on the handout on your poster when you are finished.

Students will be working in groups with five students in each group. Students will be provided the handout seen below to guide them in creating a poster:

Names: _____

Investigative Question:

	Football	Tennis	Soccer	Total
Male	172	162	175	509
Female	110	191	190	491
Total	282	353	365	1000

Explanatory Variable:

-Categorical or Quantitative

Response Variable:

-Categorical or Quantitative

Show your statistical process to answer the investigative question:

Before students are able to choose a statistical process to answer the investigative question, the students need to determine the variables in the dataset. For students to determine this, they need to rely on their prerequisite knowledge of determining explanatory and response variables as categorical or quantitative. Students need to demonstrate that they know how to determine if there is an association between two categorical variables. This relies on their prerequisite knowledge of creating segmented bar graphs, bar charts, mosaic plots, calculating relative frequencies, and performing chi-square test of independence.

The teacher will be walking around during this stage and making sure that each group has a starting point and providing extensions as appropriate. If a group of students is stuck getting started, the teacher may encourage them to think about visualizing no association and ask them to describe how a dataset with no association should appear. The teacher may also ask what type

of test involves two categorical variables (assuming the students correctly identified the variables as categorical). The teacher will also ask students questions when walking around to better understand their thought process. This will help the teacher see each group's knowledge of the concepts and how well they can explain their statistical reasoning.

The teacher will also be thinking about how they want to sequence the student's solution methods based on the solutions they see when walking around the classroom and conversing with students.

Here are the possible solutions for the statistical processes:

#1: Relative frequencies

Students can compare the conditional probabilities.

$$P(\text{Football} \mid \text{Female}) = 110/491 = 0.224$$

$$P(\text{Football} \mid \text{Male}) = 172/509 = 0.338$$

$$P(\text{Tennis} \mid \text{Female}) = 191/491 = 0.389$$

$$P(\text{Tennis} \mid \text{Male}) = 162/509 = 0.318$$

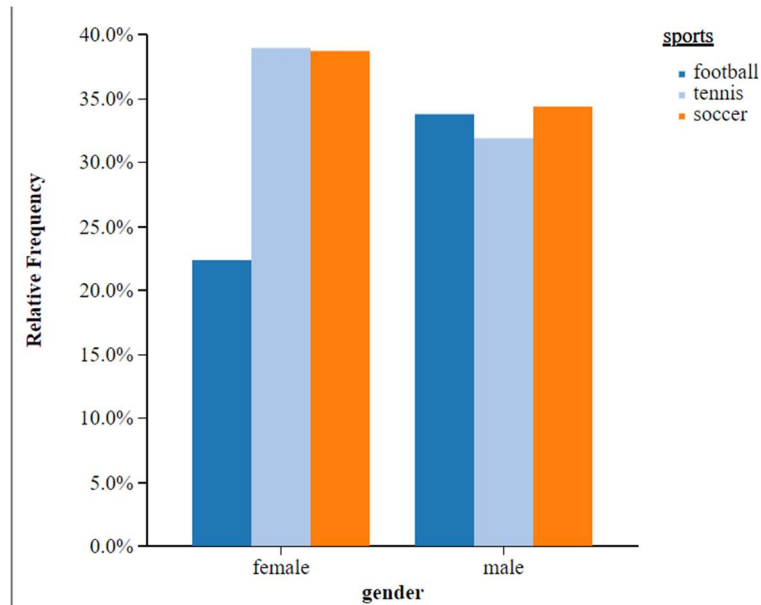
$$P(\text{Soccer} \mid \text{Female}) = 190/491 = 0.387$$

$$P(\text{Soccer} \mid \text{Male}) = 175/509 = 0.344$$

Students would need a statement making a conclusion to answer the investigative question. For example, the probability that a male student chooses football is not the same as the probability that a female student chooses football. This remains true for all sports. Therefore, there is an association between gender and sport watching preference.

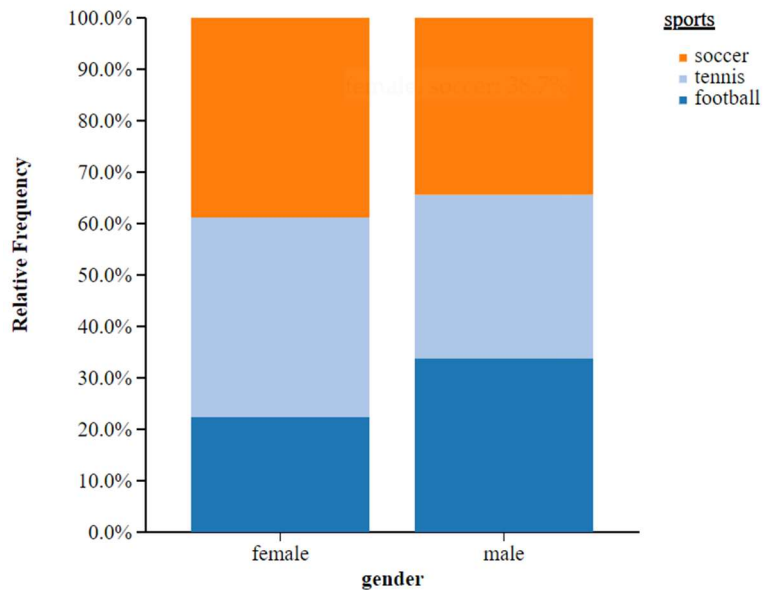
A common issue that may occur with this method is that students may be confused when trying to determine what probabilities they should be calculating to compare. A teacher directed question that may help students think through this issue, is asking the students "What does the probability you calculated tell you? Will this help you determine if there is a relationship between sport preference and gender?"

#2: Bar chart (Created using Stapplet)



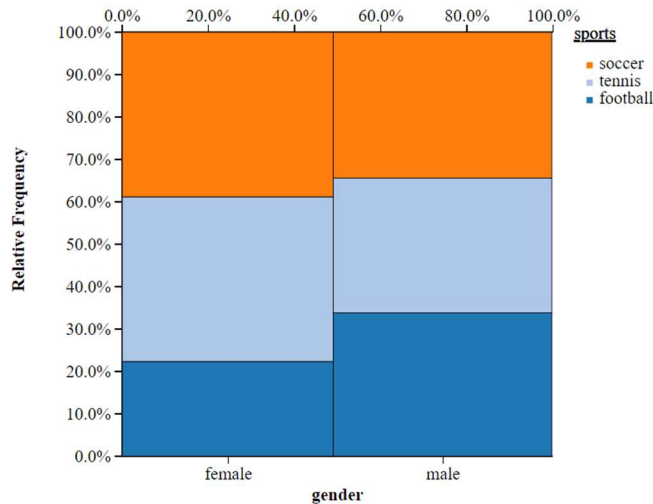
Students would need a statement making a conclusion to answer the investigative question. For example, the relative frequency for choosing football among female students is not equal to the relative frequency for choosing football among male students. This remains true for all sports. Therefore, there is an association between gender and sport watching preference.

#3: Segmented Bar Graph (Created using Stapplet)



Students would need a statement making a conclusion to answer the investigative question. For example, each colored area (sport preference) is not the same size (relative frequency) across genders. Therefore, there is an association between gender and sport watching preference.

#4: Mosaic Plot (Created using Stapplet)



Students would need a statement making a conclusion to answer the investigative question. For example, each colored area (sport preference) has different heights of the bars (relative frequency) across genders. Therefore, there is an association between gender and sport watching preference.

An issue that could occur for the visual representation methods could be that students are having a difficult time interpreting the visual representation and making a conclusion based on the representation. To help students through this issue, the teacher could ask “What do you think the visual representation would look like if there were no association?”, or in other words “What would the visual representation look like if there was no difference between the genders’ sport preferences?”

#5: Chi square test of independence

H₀: No association between gender and sport watching preference

H_a: There is an association between gender and sport watching preference

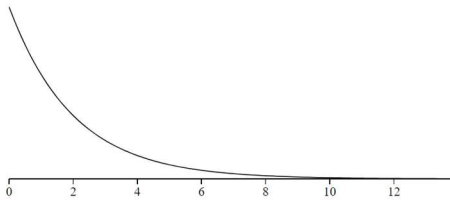
Expected counts:

	Football	Tennis	Soccer	Total
Male	172 (143.5)	162 (179.7)	175 (185.8)	509
Female	110 (138.5)	191 (173.3)	190 (179.2)	491
Total	282	353	365	1000

Conditions: All expected counts are >10

$X^2=16.36$ (Formula)

Df=2 (Formula)



Calculate the area Show labels on plot

Value:

Area = 0.0003

Students would need a statement making a conclusion to answer the investigative question. For example, the probability of getting our test statistic or more extreme assuming the null is true is 0.0003. (Calculated using Stapplet.) Therefore, there is strong evidence to suggest that gender is associated with sport watching preference.

An issue that may occur for this method is that students may have a difficult time in finding the test statistic or a difficult time finding the p-value to draw the conclusion. The teacher could help the students by providing them with the formulas or directing them to where they could find the formulas. The teacher could also remind them which applet would help them calculate the p-value.

In order to make the next stage of the lesson plan flow better, I would sequence the students answers in this order: Relative frequencies, Bar chart, Segmented Bar graph, Mosaic plot, and then Chi Square test of independence. If not all methods are created in the class, I would still use this as the sequence and I would fill in the missing methods by introducing them to the class.

Compare Strategies

To transition from the second stage of the lesson to the third stage of the lesson, the teacher will walk around to groups and collect their posters and let them know they are going to present their ideas to the rest of the class.

The teacher will then place the first poster so that it is visible for every student and call the first group up to present their ideas. The students will lead the discussion during this time to showcase their thoughts and answer any questions that the rest of the class may have for them about their process. If the students do not discuss their conclusion using their representation as evidence, the teacher will ask one of the group members to explain how they came to their conclusion using their representation. If not addressed naturally in the discussion among students, the teacher will be making the following connections between the representations as the methods are presented.

- What are we comparing in (method 1)?
 - The probability of a sport preference given male and given female
- Where could we see the probabilities calculated in (method 1), in (method 2)?
 - Each bar represents each calculated probability.
- What are we comparing in method 2?
 - Each colored bar for male and female

- Taking a first glance at (method 3), what do you notice is the same? What do you notice is different from (method 2)?
 - Same: displays the data in a visual bar representing the frequency of responses and distinguishes between genders
 - Different: displays data in one bar instead of three separate bars (proportions)
- What are we comparing in method 3?
 - Each colored segment of female and male
- Taking a first glance at (method 4), what do you notice is the same? What do you notice is different from (method 3)?
 - Same: displays data in one bar (proportions)
 - Different: displays the sample size (width of bar)
- What are we comparing in method 4?
 - Each colored segment of female and male
- How is (method 5) different? The same?
 - Same: uses the sample to conclude if there is an association between gender and sport watching preference
 - Different: This method goes beyond a description of the sample and is used to draw inferences about the population. This requires stating hypotheses, calculating expected counts and a test statistic, and using the distribution of chi square to estimate the p-value. This approach measures strength of evidence not just strength of association.

The discussion will continue as the groups are presenting. The group presentations will follow the order of the sequencing that the teacher planned in the previous stage. If any of the planned methods for discussion are missing from the student work, the teacher will introduce these methods to the students and have a whole class discussion about the method.

To close out the lesson, the students will be asked to write a five-sentence summary of takeaways from today's lesson. The goal is that students know how to carry out the methods for determining if there is an association between two categorical variables and understand the comparisons between strategies.

Attached Materials

- The template seen above will be provided to students along with a blank sheet of poster paper:

Names: _____

Investigative Question:

	Football	Tennis	Soccer	Total
Male	172	162	175	509
Female	110	191	190	491
Total	282	353	365	1000

Explanatory Variable:

-Categorical or Quantitative

Response Variable:

-Categorical or Quantitative

Show your statistical process to answer the investigative question:

- Slides:

https://www.canva.com/design/DAGD7-gAo2k/KOINU1rBUwT0TNfpdlZavA/edit?utm_content=DAGD7-gAo2k&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Reflections and Additional Recommendations

- If a group finishes their poster early, especially for method 1, you can challenge them to think of a different way to solve the problem. If they can think of a different way, have them try to make connections between the methods within their group.
- If you have more than one group finish early, you could have the groups discuss their methods with each other and see if they can make connections between their representations.
- If there is extra time remaining, you could challenge the students to think about the statistical process. For example, is it appropriate to generalize our conclusions to all students at this school? What about to all high school students in the United States?

References

Simpleshow Foundation. (2018, October 23). *Gender Equality and Sports*. [Video]. Youtube. <https://youtu.be/aR3htgTnzRs?si=WzidG17K13pZ-g5h>

Stapplet. *Two Categorical Variables*. [Applet]. <https://www.stapplet.com/cat2v.html>

Stapplet. *Chi-squared Distributions*. [Applet]. <https://www.stapplet.com/x2dist.html>