

## EMILY ROBINSON 1

Chi-Square Goodness of Fit

December 7, 2023

### Anonymous

I am a college professor teaching applied statistics for life sciences. We are teaching chi-square goodness of fit. Help me write out the scenario introduction for this homework looking at mastitis in cows/heifers: Mastitis is xxx The proportion of cows of each breed in the entire dairy farmer Selberg's herd of cows is shown below. Research Question: Is there evidence to suggest that the breed pattern of cows with mastitis deviates from breed patterns in dairy farmer Selberg's herd? Identify the variable of interest (and categories). State the parameters and appropriate symbols (Hint: there should be 4). Write your null and alternative hypotheses. Consider the following table. The first row of this table contains the Observed Outcomes for the sample of cows from dairy farmer Selberg's herd of cows. The second row contains the Expected Outcome (under the null hypothesis) for each of the breeds. Brown Swiss Dutch Belted Holstein Jersey Total Observed 39 6 62 31 138 Expected What does the Total value for the Observed row represent? The value in the first row and second column is XXX (i.e. Observed count for the Brown Swiss breed). Explain where this number came from. What does this value represent? Complete the empty cells in the second row to contain the expected outcome for each of the breeds. Sketch a stacked bar plot of the Observed counts of mastitis versus Expected counts of mastitis for each breed. Sketch a stacked bar plot of the Observed proportions of mastitis versus Expected proportions of mastitis for each breed. Why is the Expected count for the Holstein's so much higher than the other breeds? Suppose your friend computes the following percentages: Brown Swiss: X/XXX X%; Dutch Belted: X/XXX X%; Holstein: X/XXX X%; and Jersey: X/XXX X%. Your friend then makes the following statement: "There is enough evidence for the research question because these percentages are different from the breed percentages for the entire herd (i.e., Brown Swiss = 17%, Dutch Belted = 10%, Holstein = 40%, and Jersey = 33%)." Why is this statement statistically incorrect? Explain. Calculate the Chi-square test statistic for our sample of cows (Hint: it may be helpful to add additional rows to the table above). Conduct 1000 replications via the Online Simulation Applets > `` to create a simulated distribution under the assumption that cows with mastitis do not deviate from the overall breed patterns in dairy farmer Selberg's herd. (Hint: you will need to change the expected probabilities in the applet). Paste/sketch the simulated null sampling distribution below. Is our observed Chi-square test statistic consistent with results we would expect to see if cows with mastitis do not deviate from the overall breed patterns in dairy farmer Selberg's herd? Explain. From your simulated distribution, compute the p-value, make a decision, and write a final conclusion for the original research question. Recall that in practice, statisticians conduct a Goodness of Fit test using the Chi-square distribution (and code) rather than simulating the distribution. Check the conditions for using the Chi-square distribution. How many degrees of freedom are used to determine the shape of the Chi-square distribution for our scenario? Select the correct code/output below to test the hypotheses.

### ChatGPT

### ChatGPT

Mastitis is a common and economically significant disease in dairy cows and heifers. It is an inflammation of the mammary gland tissue that can lead to reduced milk production and quality. Identifying whether there is a relationship between the breed of cows and the occurrence of mastitis is crucial for dairy farmer Selberg's herd.

The proportion of cows of each breed in the entire dairy farmer Selberg's herd of cows is shown below:

- Brown Swiss: 17%
- Dutch Belted: 10%
- Holstein: 40%
- Jersey: 33%

**Research Question:** Is there evidence to suggest that the breed pattern of cows with mastitis deviates from breed patterns in dairy farmer Selberg's herd?

**Identify the variable of interest (and categories):** Variable of Interest: Breed of cows Categories: Brown Swiss, Dutch Belted, Holstein, Jersey

**State the parameters and appropriate symbols (Hint: there should be 4):**

1.  $p_1$ : Proportion of Brown Swiss cows with mastitis
2.  $p_2$ : Proportion of Dutch Belted cows with mastitis
3.  $p_3$ : Proportion of Holstein cows with mastitis
4.  $p_4$ : Proportion of Jersey cows with mastitis

**Write your null and alternative hypotheses:** Null Hypothesis ( $H_0$ ): There is no significant difference in the breed pattern of cows with mastitis compared to the breed pattern in dairy farmer Selberg's herd. Alternative Hypothesis ( $H_a$ ): There is a significant difference in the breed pattern of cows with mastitis compared to the breed pattern in dairy farmer Selberg's herd.

**Consider the following table:** The Total value for the Observed row represents the total number of cows with mastitis in the sample from dairy farmer Selberg's herd.

The value in the first row and second column (6) is the Observed count for the Dutch Belted breed. This number came from observing 6 Dutch Belted cows with mastitis in the sample.

**Complete the empty cells in the second row to contain the expected outcome for each of the breeds:** Under the null hypothesis, the expected count for each breed can be calculated by multiplying the total number of cows with mastitis by the proportion of each breed in Selberg's herd:

- Expected for Brown Swiss = Total Observed \* Proportion of Brown Swiss in the herd
- Expected for Dutch Belted = Total Observed \* Proportion of Dutch Belted in the herd
- Expected for Holstein = Total Observed \* Proportion of Holstein in the herd
- Expected for Jersey = Total Observed \* Proportion of Jersey in the herd

**Sketch a stacked bar plot of the Observed counts of mastitis versus Expected counts of mastitis for each breed.**

**Sketch a stacked bar plot of the Observed proportions of mastitis versus Expected proportions of mastitis for each breed.**

**Why is the Expected count for the Holstein's so much higher than the other breeds?** The Expected count for Holstein cows is higher because Holsteins make up the largest proportion of the

entire herd (40%). Therefore, under the null hypothesis, we expect more Holstein cows to have mastitis simply due to their higher representation in the herd.

**Suppose your friend computes the following percentages:**

- Brown Swiss: X/XXX X%
- Dutch Belted: X/XXX X%
- Holstein: X/XXX X%
- Jersey: X/XXX X%

Your friend then makes the following statement: "There is enough evidence for the research question because these percentages are different from the breed percentages for the entire herd (i.e., Brown Swiss = 17%, Dutch Belted = 10%, Holstein = 40%, and Jersey = 33%)."

**Why is this statement statistically incorrect? Explain.** This statement is statistically incorrect because it only compares observed percentages with the overall breed percentages in the herd, without considering the expected values or conducting a formal statistical test. To determine if there is significant evidence that the breed pattern of cows with mastitis deviates from the breed patterns in the herd, a chi-square goodness-of-fit test should be performed, which takes into account both observed and expected values.

**Calculate the Chi-square test statistic for our sample of cows:** The chi-square test statistic can be calculated using the formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where:

- $O$  is the observed count.
- $E$  is the expected count.

**Conduct 1000 replications via the Online Simulation Applets:** To create a simulated distribution under the assumption that cows with mastitis do not deviate from the overall breed patterns in dairy farmer Selberg's herd, change the expected probabilities in the applet to match the herd's breed proportions.

**Paste/sketch the simulated null sampling distribution below.**

**Is our observed Chi-square test statistic consistent with results we would expect to see if cows with mastitis do not deviate from the overall breed patterns in dairy farmer Selberg's herd?**

**Explain.**

To answer this question, compare the observed chi-square test statistic with the distribution of chi-square test statistics obtained from the simulations. If the observed statistic falls within the range of values from the simulations, it suggests that there is no significant deviation from the expected breed patterns.

**From your simulated distribution, compute the p-value, make a decision, and write a final conclusion for the original research question:** The p-value is the probability of obtaining a chi-square test statistic as extreme as or more extreme than the observed statistic under the null

hypothesis. If the p-value is small (typically less than 0.05), it suggests that there is evidence to reject the null hypothesis and conclude that the breed pattern of cows with mastitis deviates from the breed patterns in the herd. Otherwise, if the p-value is large, you fail to reject the null hypothesis.

**Recall that in practice, statisticians conduct a Goodness of Fit test using the Chi-square distribution (and code) rather than simulating the distribution.**

**Check the conditions for using the Chi-square distribution:**

1. Randomness: The sample of cows should be selected randomly from the herd.
2. Independence: Each cow in the sample should be independent of the others.
3. Expected Frequencies: The expected count for each category should be at least 5.

**How many degrees of freedom are used to determine the shape of the Chi-square distribution for our scenario?** The degrees of freedom (df) for the chi-square goodness-of-fit test in this scenario would be  $k-1$ , where  $k$  is the number of categories (breeds) minus 1. So, in this case,  $df = 4 - 1 = 3$ .

**Select the correct code/output below to test the hypotheses:** You would typically use statistical software or coding (e.g., R, Python) to conduct the chi-square goodness-of-fit test. The code/output would involve calculating the chi-square test statistic, determining the p-value, and making a decision based on the p-value. The specific code or output would depend on the software or programming language you are using.

**Anonymous**

is a heifer a dairy cow?

ChatGPT

ChatGPT

Yes, a heifer can be a type of dairy cow, but it's important to understand the terminology in the context of cattle farming:

1. **Heifer:** A heifer is a young female cow that has not yet given birth to a calf. Heifers are typically raised on dairy farms with the intention of eventually becoming dairy cows once they reach maturity and start producing milk.
2. **Dairy Cow:** A dairy cow is a female cow that is specifically raised for the purpose of milk production. Once a heifer matures and has her first calf, she is referred to as a dairy cow, and her milk can be harvested for dairy products such as milk, cheese, and yogurt.

So, while a heifer is not a dairy cow in the sense that it hasn't started producing milk yet, it has the potential to become a dairy cow once it reaches the appropriate age and begins calving. Dairy farmers raise heifers with the goal of adding them to their milk-producing herd in the future.

**Anonymous**

what type of cows get mastitis?

ChatGPT

ChatGPT

Mastitis can affect all types of cows, regardless of their breed or age. Mastitis is a common and costly disease in dairy cattle, and it can also affect beef cattle. It is an inflammation of the mammary gland tissue and udder, which can be caused by various factors, including bacterial infections, environmental conditions, and poor milking practices.

Here are some key points about mastitis in cows:

1. **All Breeds:** Mastitis is not specific to any particular breed of cow. It can occur in Holsteins, Jerseys, Guernseys, Ayrshires, Brown Swiss, and other dairy cattle breeds, as well as in beef cattle.
2. **Age:** Cows of all ages, including heifers (young female cows), mature cows, and older cows, can develop mastitis.
3. **Causes:** Mastitis can be caused by a variety of factors, including bacteria entering the udder through the teat canal, poor hygiene practices, inadequate milking techniques, and environmental conditions such as wet and dirty bedding.
4. **Symptoms:** Common symptoms of mastitis in cows include swelling, redness, heat, and pain in the udder, as well as changes in milk consistency (e.g., clots, flakes, or pus) and reduced milk production.
5. **Treatment:** Treatment for mastitis typically involves antibiotics to address the bacterial infection, along with measures to alleviate pain and inflammation. Proper milking hygiene and management practices can also help prevent and control mastitis.
6. **Prevention:** Preventing mastitis is crucial in dairy farming. Practices such as maintaining clean and dry bedding, ensuring proper milking procedures, and regular monitoring of cow health can help reduce the risk of mastitis.

In summary, mastitis is a condition that can affect cows of all breeds and ages. It is essential for dairy farmers to implement good management practices to prevent and manage mastitis effectively, as it can have a significant impact on milk production and cow welfare.

## EMILY ROBINSON 2

Canvas Quiz on Mastitis

December 7, 2023

### Anonymous

Turn this homework into a canvas quiz with multiple choice questions. Mastitis is a common and costly condition in dairy cows and heifers. It is an inflammation of the mammary gland tissue and udder, which can be caused by various factors, including bacterial infections, environmental conditions, and poor milking practices. Mastitis can lead to reduced milk production and quality. Identifying whether the breed pattern of cows with mastitis consistent with the breed pattern of cows in the herd and is crucial for dairy farmer Selberg. The proportion of cows of each breed in the dairy farmer Selberg's herd is shown below. Farmer Selberg has collected data (shown below) on the heifers and dairy cows with mastitis from his herd.

```
library(tidyverse)
2mastitis <- read_csv("data/mastitis.csv")
3head(mastitis)
1 Load the tidyverse package.
2 Read in the mastitis data set.
3 View the top 6 rows of the mastitis data set.
# A tibble: 6 × 5
cow_id condition breed age_months type <dbl> <chr> <chr> <dbl> <chr>
1 1 1 mastitis BrownSwiss 39 dairy cow
2 2 2 mastitis BrownSwiss 35 heifer
3 3 3 mastitis BrownSwiss 48 dairy cow
4 4 4 mastitis Holstein 32 heifer
5 5 5 mastitis Jersey 32 heifer
6 6 6 mastitis BrownSwiss 30 heifer
```

Research Question: Is there evidence to suggest that the breed pattern of cows with mastitis deviates from breed patterns in dairy farmer Selberg's herd? Identify the variable of interest (and categories). State the parameters and appropriate symbols (Hint: there should be 4). Write your null and alternative hypotheses. Consider the following table. The first row of this table contains the Observed number of cows with mastitis. The second row contains the Expected number of cows with mastitis (under the null hypothesis) for each of the breeds.

Breed	Observed	Expected
Brown Swiss	39	23.46
Dutch Belted	6	23.46
Holstein	62	23.46
Jersey	31	23.46
Total	138	23.46

What does the Total value for the Observed row represent? The value in the first row and second column is 23.46 (i.e. Observed count for the Brown Swiss breed). Explain where this number came from. What does this value represent? Complete the empty cells in the second row to contain the expected outcome for each of the breeds. Sketch a stacked bar plot of the Observed counts of mastitis versus Expected counts of mastitis for each breed. Sketch a stacked bar plot of the Observed proportions of mastitis versus Expected proportions of mastitis for each breed (Hint: you may want to create a second table containing the observed and expected proportions) Why is the Expected count for the Holstein's so much higher than the other breeds? Suppose your friend computes the following percentages: Brown Swiss: 39/138 28%; Dutch Belted: 6/138 4%; Holstein: 62/138 45%; and Jersey: 31/138 23%. Your friend then makes the following statement: "There is enough evidence for the research question because these percentages are different from the breed percentages for the entire herd (i.e., Brown Swiss = 17%, Dutch Belted = 10%, Holstein = 40%, and Jersey = 33%)." Why is this statement statistically incorrect? Explain. Calculate the Chi-square test statistic for our sample of cows (Hint: it may be helpful to add additional rows to the table above). Conduct 1000 replications via the Online Simulation Applets > Goodness of Fit to create a simulated distribution of the test statistic under the assumption that cows with mastitis do not deviate from the overall breed patterns in dairy farmer Selberg's herd. (Hint: you will need to change the hypothesized probabilities in the applet). The breed data can be found at [https://raw.githubusercontent.com/earobinson95/stat218-campoly-f2023/main/02-homeworks/data/mastitis\\_breeds.csv](https://raw.githubusercontent.com/earobinson95/stat218-campoly-f2023/main/02-homeworks/data/mastitis_breeds.csv) Warning Watch the ordering of the categories when setting up your hypothesized probabilities. Paste/sketch the simulated null sampling distribution below. Is our observed Chi-square test statistic consistent with results we would expect to see if cows with mastitis do not deviate from the overall breed patterns in dairy farmer Selberg's herd? Explain. From your simulated distribution, compute the p-value, make a decision, and write a final conclusion

for the original research question. Recall that in practice, statisticians conduct a Goodness of Fit test using the Chi-square distribution (and code) rather than simulating the distribution. Check the conditions for using the Chi-square distribution. How many degrees of freedom are used to determine the shape of the Chi-square distribution for our scenario? Select the correct code/output below to test the hypotheses. Explain your selection. a. `library(infer) chisq_test(mastitis, response = breed, p = c("BrownSwiss" = 0.28, "DutchBelted" = 0.04, "Holstein" = 0.45, "Jersey" = 0.23))` # A tibble: 1 × 3  
statistic chisq\_df p\_value <dbl> <dbl> <dbl> 1 0.0625 3 0.996 b. `library(infer) chisq_test(mastitis, response = breed, p = c("BrownSwiss" = 0.25, "DutchBelted" = 0.25, "Holstein" = 0.25, "Jersey" = 0.25))` # A tibble: 1 × 3  
statistic chisq\_df p\_value <dbl> <dbl> <dbl> 1 46.4 3 4.65e-10 c. `library(infer) chisq_test(mastitis, response = breed, p = c("BrownSwiss" = 0.17, "DutchBelted" = 0.10, "Holstein" = 0.40, "Jersey" = 0.33))` # A tibble: 1 × 3  
statistic chisq\_df p\_value <dbl> <dbl> <dbl> 1 20.2 3 0.000156 From the output above, find the p-value, make a decision, and write a final conclusion for the original research question.

## ChatGPT

### ChatGPT

I can help you create a canvas quiz with multiple-choice questions based on your homework. Here are the questions and answer choices:

#### Question 1:

What is the variable of interest in this study, and what are its categories?

- A. Variable: Breed; Categories: Brown Swiss, Dutch Belted, Holstein, Jersey
- B. Variable: Age; Categories: Months, Years
- C. Variable: Condition; Categories: Healthy, Sick
- D. Variable: Type; Categories: Dairy cow, Heifer

#### Question 2:

State the null and alternative hypotheses for the research question.

A. Null Hypothesis: There is evidence to suggest that the breed pattern of cows with mastitis is different from the breed pattern in the herd.

Alternative Hypothesis: There is no evidence to suggest that the breed pattern of cows with mastitis is different from the breed pattern in the herd.

B. Null Hypothesis: There is no evidence to suggest that the breed pattern of cows with mastitis is different from the breed pattern in the herd.

Alternative Hypothesis: There is evidence to suggest that the breed pattern of cows with mastitis is different from the breed pattern in the herd.

C. Null Hypothesis: The breed pattern of cows with mastitis is the same as the breed pattern in the herd.

Alternative Hypothesis: The breed pattern of cows with mastitis is different from the breed pattern in the herd.

D. Null Hypothesis: The breed pattern of cows with mastitis is different from the breed pattern in the herd.

Alternative Hypothesis: The breed pattern of cows with mastitis is also different from the breed pattern in the herd.

**Question 3:**

What does the Total value for the Observed row in the table represent?

- A. The total number of cows with mastitis in the herd.
- B. The total number of cows in the herd.
- C. The total number of cows of each breed in the herd.
- D. The total number of cows without mastitis in the herd.

**Question 4:**

Explain where the value 23.46 (Observed count for the Brown Swiss breed) came from. What does this value represent?

- A. It represents the number of healthy Brown Swiss cows in the herd.
- B. It represents the number of cows with mastitis in the Brown Swiss breed.
- C. It represents the expected number of cows with mastitis in the Brown Swiss breed under the null hypothesis.
- D. It represents the total number of cows in the Brown Swiss breed.

**Question 5:**

Complete the empty cells in the second row to contain the expected outcome for each of the breeds.

- A. Brown Swiss: 39, Dutch Belted: 6, Holstein: 62, Jersey: 31
- B. Brown Swiss: 23.46, Dutch Belted: 3.6, Holstein: 46.92, Jersey: 23.46
- C. Brown Swiss: 17%, Dutch Belted: 10%, Holstein: 40%, Jersey: 33%
- D. Brown Swiss: 28%, Dutch Belted: 4%, Holstein: 45%, Jersey: 23%

**Question 6:**

Why is the Expected count for the Holstein breed so much higher than the other breeds?

- A. Because Holstein cows are more prone to mastitis.
- B. Because Holstein cows are healthier.
- C. Because of random variation, there is no specific reason.
- D. Because Holstein cows are milked more often.

**Question 7:**

Why is the statement made by your friend statistically incorrect?

- A. Because the percentages for each breed are different from the observed counts.
- B. Because the percentages should be calculated for healthy cows, not cows with mastitis.
- C. Because the statement doesn't consider the null hypothesis and expected values.
- D. Because the percentages should be calculated for the entire herd, not just cows with mastitis.

**Question 8:**

Calculate the Chi-square test statistic for our sample of cows.

- A. 0.0625
- B. 46.4
- C. 20.2
- D. 3.841

**Question 9:**

Is our observed Chi-square test statistic consistent with results we would expect to see if cows with mastitis do not deviate from the overall breed patterns in dairy farmer Selberg's herd? Explain.

- A. Yes, because the observed statistic is very close to the expected value.
- B. No, because the observed statistic is much higher than expected.
- C. Yes, because the observed statistic is much lower than expected.
- D. No, because the observed statistic is not relevant to the question.

**Question 10:**

From your simulated distribution, compute the p-value, make a decision, and write a final conclusion for the original research question.

- A. p-value: 0.996, Conclusion: Reject the null hypothesis.
- B. p-value:  $4.65e-10$ , Conclusion: Fail to reject the null hypothesis.
- C. p-value: 0.000156, Conclusion: Reject the null hypothesis.
- D. p-value: 0.05, Conclusion: Inconclusive.

**Question 11:**

How many degrees of freedom are used to determine the shape of the Chi-square distribution for our scenario?

- A. 1
- B. 2
- C. 3
- D. 4

**Question 12:**

Select the correct code/output below to test the hypotheses. Explain your selection.

- A. Code (a) is correct because it uses the correct probabilities for each breed under the null hypothesis.
- B. Code (b) is correct because it assigns equal probabilities to each breed under the null hypothesis.
- C. Code (c) is correct because it uses different probabilities for each breed under the null hypothesis.
- D. None of the codes are correct.

Answers:

1. A
2. B
3. B
4. C
5. B
6. C
7. C
8. A
9. B
10. C
11. C
12. A

## EMILY ROBINSON 3

Canvas Quiz: Research Questions

December 7, 2023

### Anonymous

Turn this homework into a canvas quiz with multiple choice questions: Homer's Iliad is an epic poem, compiled around 800 BCE, that describes several weeks of the last year of the 10-year siege of Troy (Ilium) by the Achaeans. The story centers on the rage of the great warrior Achilles. But it includes many details of injuries and outcomes, and is thus the oldest record of Greek medicine. The data report 146 recorded injuries for which both injury site and outcome are provided in the Iliad (Hutchinson, 2013). The top 6 rows of the recorded injuries is shown below: Research Question: Is the location of an injury associated with whether the injury was fatal? Identify the explanatory and response variables in the context of this study. How would you describe the orientation of this bar plot? Filled Stacked Dodged Based on the plot does there appear to be an association between the variables? Explain your answer. We can summarize the recorded injuries in a contingency table:

InjurySite	Fatal	Nonfatal	Total
Body	61	6	67
Head/neck	44	1	45
Limb	13	21	34
Unknown	37	1	38
Total	155	29	184

Which injury location has the smallest number of observations? Were there more fatal injuries or nonfatal injuries overall? What proportion of body injuries were fatal? What proportion of limb injuries were fatal? Set up the null and alternative hypotheses to test the research question. In the table below, fill in the expected counts. Lethal Body Head/neck Limb Unknown Fatal Nonfatal Use the observed and expected counts to compute the Chi-square test statistic. After running 100 simulations under the assumption the location of an injury is not associated with whether the injury was fatal, I obtained the following distribution. Use the simulated distribution of the Chi-square test statistic under the assumption location of an injury is not associated with whether the injury was fatal to estimate your p-value and make a decision about your null hypothesis. Is it appropriate to conduct a Chi-square Test using the Chi-square distribution to test our research question? How many degrees of freedom would the Chi-square distribution for our Test Statistic follow? Select the correct code/output for the Chi-square Test to investigate this question.

```
chisq_test(x = iliad, response = Lethal, explanatory = InjurySite) # A tibble: 1 × 3 statistic chisq_df p_value <dbl> <int> <dbl> 1 67.7 3 1.34e-14
chisq_test(x = iliad, response = Lethal) # A tibble: 1 × 3 statistic chisq_df p_value <dbl> <dbl> <dbl> 1 86.3 1 1.56e-20
chisq_test(x = iliad, response = Lethal, p = c("Fatal" = 0.84, "Nonfatal" = 0.16)) # A tibble: 1 × 3 statistic chisq_df p_value <dbl> <dbl> <dbl> 1 0.00783 1 0.929
```

Write a solution and make sure to include the chi-square test statistic, degrees of freedom, the p-value, and a conclusion written in everyday language.

ChatGPT

### ChatGPT

Here's your homework turned into a Canvas quiz with multiple-choice questions:

#### Question 1:

Homer's Iliad is an epic poem, compiled around 800 BCE, that describes several weeks of the last year of the 10-year siege of Troy (Ilium) by the Achaeans. The story centers on the rage of the great warrior Achilles. But it includes many details of injuries and outcomes, and is thus the oldest record of Greek medicine. The data report 146 recorded injuries for which both injury site and outcome are provided in the Iliad (Hutchinson, 2013). The top 6 rows of the recorded injuries are shown below:

Research Question: Is the location of an injury associated with whether the injury was fatal?

Identify the explanatory and response variables in the context of this study.

- A) Explanatory Variable: Injury Site, Response Variable: Fatality
- B) Explanatory Variable: Fatality, Response Variable: Injury Site
- C) Explanatory Variable: Research Question, Response Variable: Injury Site
- D) Explanatory Variable: Injury Site, Response Variable: Research Question

**Question 2:**

How would you describe the orientation of this bar plot?

- A) Filled
- B) Stacked
- C) Dodged
- D) None of the above

**Question 3:**

Based on the plot, does there appear to be an association between the variables? Explain your answer.

- A) Yes, there is a clear association.
- B) No, there is no association.
- C) It's unclear from the plot.
- D) The plot is not provided.

**Question 4:**

Which injury location has the smallest number of observations?

- A) Body
- B) Head/neck
- C) Limb
- D) Unknown

**Question 5:**

Were there more fatal injuries or nonfatal injuries overall?

- A) Fatal injuries
- B) Nonfatal injuries
- C) An equal number of both
- D) Data not provided

**Question 6:**

What proportion of body injuries were fatal?

- A) 91%
- B) 9%
- C) 14%
- D) 6%

**Question 7:**

What proportion of limb injuries were fatal?

- A) 38%
- B) 62%
- C) 32%
- D) 68%

**Question 8:**

Set up the null and alternative hypotheses to test the research question.

- A) Null Hypothesis: There is no association between injury location and fatality. Alternative Hypothesis: There is an association between injury location and fatality.
- B) Null Hypothesis: There is an association between injury location and fatality. Alternative Hypothesis: There is no association between injury location and fatality.
- C) Null Hypothesis: The Iliad is an epic poem. Alternative Hypothesis: The Iliad is not an epic poem.
- D) Null Hypothesis: The Iliad was written in 800 BCE. Alternative Hypothesis: The Iliad was written in 1000 BCE.

**Question 9:**

In the table below, fill in the expected counts.

Lethal Body Head/neck Limb Unknown Fatal Nonfatal

- A) You should enter numbers in these cells.
- B) Cannot be determined from the given information.
- C) Fatal: 88, 47, 20, 33; Nonfatal: 12, 23, 14, 5
- D) Data not sufficient to fill in expected counts.

**Question 10:**

Use the observed and expected counts to compute the Chi-square test statistic.

- A) You should enter the formula to calculate the Chi-square test statistic.
- B) 2.53
- C) 67.7
- D) 86.3

**Question 11:**

After running 100 simulations under the assumption the location of an injury is not associated with whether the injury was fatal, I obtained the following distribution.

Use the simulated distribution of the Chi-square test statistic under the assumption location of an injury is not associated with whether the injury was fatal to estimate your p-value and make a decision about your null hypothesis.

- A) I don't have enough information to answer this question.
- B) The p-value is 0.05, and the null hypothesis is rejected.
- C) The p-value is 0.01, and the null hypothesis is accepted.
- D) The p-value is 0.0001, and the null hypothesis is rejected.

**Question 12:**

Is it appropriate to conduct a Chi-square Test using the Chi-square distribution to test our research question?

- A) Yes, it is appropriate.
- B) No, it is not appropriate.
- C) Data not sufficient to determine appropriateness.
- D) Maybe, it depends on the type of research question.

**Question 13:**

How many degrees of freedom would the Chi-square distribution for our Test Statistic follow?

- A) 1
- B) 3
- C) 4
- D) 5

**Question 14:**

Select the correct code/output for the Chi-square Test to investigate this question.

- A) `chisq_test(x = iliad, response = Lethal, explanatory = InjurySite)`
- B) `chisq_test(x = iliad, response = Lethal)`
- C) `chisq_test(x = iliad, response = Lethal, p = c("Fatal" = 0.84, "Nonfatal" = 0.16))`

**Question 15:**

Write a solution and make sure to include the chi-square test statistic, degrees of freedom, the p-value, and a conclusion written in everyday language.

- A) This information is not provided.
- B) The chi-square test statistic is 67.7 with 3 degrees of freedom, and the p-value is  $1.34e-14$ . We reject the null hypothesis, indicating an association between injury location and fatality.
- C) The chi-square test statistic is 86.3 with 1 degree of freedom, and the p-value is  $1.56e-20$ . We accept the null hypothesis.
- D) The chi-square test statistic is 0.00783 with 1 degree of freedom, and the p-value is 0.929. The results are inconclusive.

## EMILY ROBINSON 4

Halloween Candy Survey Confidence

December 7, 2023

### Anonymous

Write me a scenario like this for Halloween candy: A November 7–10, 2013, Gallup poll asked 1,039 U.S. adults how much they planned to personally spend on Christmas gifts this year. The report cited an average of \$704, and the distribution is not strongly skewed. Researchers found the 95% confidence interval to be xxxx. Interpret the 95% confidence interval. Then explain what it means when we say we are 95% confident.

ChatGPT

### ChatGPT

Certainly! Let's create a Halloween candy scenario with statistics similar to the one you provided about Christmas gifts.

In a study conducted from October 25 to October 31, 2023, a group of 1,200 households across the United States were asked about their planned spending on Halloween candy for the year. The survey revealed an average intended expenditure of \$75, with a relatively normally distributed pattern of responses. Upon analysis, the researchers calculated the 95% confidence interval to be \$72 to \$78.

Interpreting the 95% confidence interval in this context means that if we were to conduct the same survey numerous times and construct confidence intervals using the same methodology, approximately 95 out of 100 of these intervals (with the same sample size and technique) would contain the true average amount that households plan to spend on Halloween candy.

Now, what does it mean to say we are 95% confident? This statement is a statistical interpretation, reflecting the level of certainty or precision in the estimation derived from the data. In this case, the 95% confidence level suggests that there's a strong likelihood (but not an absolute guarantee) that the true average expenditure on Halloween candy for all households in the United States falls within the calculated range of \$72 to \$78. The remaining 5% accounts for the margin of error or the possibility that the true average might be outside this specific interval due to sampling variability or other factors.

Therefore, when stating we are 95% confident in the context of this survey, it means there's a high degree of trust that the actual average spending on Halloween candy per household is encapsulated within the range of \$72 to \$78, based on the collected data and methodology used.

### Anonymous

Shorten this to be an exit ticket for class.

ChatGPT

### ChatGPT

Sure! Here's a concise exit ticket:

"In a recent survey, 1,200 households planned to spend an average of \$75 on Halloween candy, with a 95% confidence interval of \$72 to \$78. What does the 95% confidence interval mean? Explain what it signifies to be 95% confident in this context."

## EMILY ROBINSON 5

Fruit Flies VitaLong Lifespan

December 7, 2023

### Anonymous

I am writing a statistics quiz for college freshman and sophomores in applied statistics for life sciences (intro stat). Write me a scenario following this style but with a context in the life sciences, more modern (not years old) and is doing a lower tailed test instead of an upper tailed test. It is a well-known joke that Millennials LOVE Avocado Toast. It's also a well-known joke that all Millennials live in their parent's basements. Clearly, they aren't buying home because they are buying too much Avocado Toast! Was the Avocadopocalypse of 2017 real? In the past, an avocado cost \\$.136. To test the avocadopocalypse, we collected a random sample of avocados from 2017. \*\*Research Question:\*\* Is the price of an avocado in 2017 higher than the price of an avocado in the past?

ChatGPT

### ChatGPT

In the realm of life sciences, let's delve into a scenario that involves the potential impact of a particular nutrient supplement on the lifespan of a particular species. It's been suggested that a new dietary supplement, "VitaLong," might extend the lifespan of fruit flies, a common subject in life sciences due to their short lifespans.

**Research Question:** Does the intake of "VitaLong" extend the average lifespan of fruit flies compared to the average lifespan of fruit flies not receiving the supplement?

Previously, the average lifespan of fruit flies was recorded as 30 days. To investigate this, a study was conducted in 2023, where a random sample of fruit flies was given the "VitaLong" supplement.

The hypothesis under examination posits that the average lifespan of the fruit flies given the "VitaLong" supplement is greater than 30 days.

The data collected for the experiment revealed that the average lifespan of the fruit flies taking "VitaLong" was 32 days with a standard deviation of 3 days.

Now, the statistical analysis aims to determine if the average lifespan of the fruit flies receiving the "VitaLong" supplement is significantly different from the known average lifespan of 30 days.

A one-tailed lower test is conducted:

**Null Hypothesis (H0):** The average lifespan of fruit flies with the "VitaLong" supplement is not greater than 30 days.

**Alternative Hypothesis (H1):** The average lifespan of fruit flies with the "VitaLong" supplement is greater than 30 days.

Utilizing the collected data, a t-test can be performed to determine if there is enough evidence to reject the null hypothesis and conclude that the "VitaLong" supplement significantly extends the average lifespan of the fruit flies.

In this context, the statistical analysis would focus on assessing whether there is substantial evidence to support the claim that "VitaLong" does indeed extend the lifespan of fruit flies, thus drawing parallels to the statistical inquiry regarding the Avocadopocalypse in the original scenario.

**Anonymous**

This is an upper tailed test. I want to do a lower tailed test (something is less than the past).

ChatGPT

**ChatGPT**

Certainly! Let's reframe the scenario to fit a lower-tailed test, implying a decrease or a negative impact on the lifespan of fruit flies due to an external factor.

**Research Question:** Does exposure to a new pesticide, "EcoShield," reduce the average lifespan of fruit flies compared to their traditional average lifespan?

In the past, the average lifespan of fruit flies was recorded as 30 days. To investigate this, a study was conducted in 2023, where a random sample of fruit flies was exposed to the "EcoShield" pesticide.

The hypothesis under examination posits that the average lifespan of the fruit flies exposed to "EcoShield" is lower than the known average lifespan of 30 days.

The data collected for the experiment revealed that the average lifespan of the fruit flies exposed to "EcoShield" was 28 days, with a standard deviation of 4 days.

A one-tailed lower test is conducted:

**Null Hypothesis (H0):** The average lifespan of fruit flies with the "EcoShield" exposure is not less than 30 days.

**Alternative Hypothesis (H1):** The average lifespan of fruit flies with the "EcoShield" exposure is less than 30 days.

Utilizing the collected data, a t-test can be performed to determine if there is enough evidence to reject the null hypothesis and conclude that the "EcoShield" exposure significantly reduces the average lifespan of the fruit flies.

This scenario represents a lower-tailed test as it investigates whether the exposure to "EcoShield" results in a significant decrease in the average lifespan of the fruit flies, parallel to the statistical investigation of the Avocado price change.

**Anonymous**

I like that, help me write more of an intro for it: In the past, the average lifespan of fruit flies was recorded as 30 days. To investigate this, a study was conducted in 2023, where a random sample of fruit flies was exposed to the "EcoShield" pesticide. **Research Question:** Does exposure to a new pesticide, "EcoShield," reduce the average lifespan of fruit flies compared to their traditional average lifespan?

ChatGPT

**ChatGPT**

Certainly, here's a more detailed introduction for the scenario:

"In the intricate field of biological research, fruit flies have long been a cornerstone in genetic and life science studies due to their brief life cycle and genetic similarity to larger organisms. In the bustling landscape of 2023, amidst advancements in agricultural practices, the emergence of a new pesticide, 'EcoShield,' became a focal point of inquiry within the scientific community.

Traditionally, the average lifespan of fruit flies has been observed and documented at a stable 30 days. However, the introduction of 'EcoShield' sparked curiosity and concern about its potential impact on the lifespan of these minute creatures.

To scrutinize the ramifications of this novel pesticide, an experimental study was meticulously designed and executed in 2023. A randomized sample of fruit flies was subjected to controlled exposure to 'EcoShield,' aiming to assess whether this exposure indeed altered the average lifespan of the flies.

**Research Question:** Does exposure to the new pesticide, 'EcoShield,' diminish the average lifespan of fruit flies compared to their traditional average lifespan of 30 days?

The hypothesis formulated for this investigation posits that the average lifespan of fruit flies exposed to 'EcoShield' is significantly lower than the established average lifespan of 30 days. This sets the stage for a meticulous statistical examination to discern whether the introduction of 'EcoShield' elicits a measurable and concerning reduction in the lifespan of these quintessential subjects within the life sciences."

This preamble lays the groundwork for the subsequent statistical analysis by providing context and setting up the research question for the investigation into the impact of the 'EcoShield' pesticide on the fruit flies' lifespan.

**Anonymous**

shorten the intro

ChatGPT

**ChatGPT**

"In 2023, a groundbreaking pesticide known as 'EcoShield' emerged within the realm of agricultural sciences. This new pesticide sparked an investigation into its potential effects on the average lifespan of fruit flies, a species extensively studied in life sciences for its short life cycle and genetic correlations.

The traditional average lifespan of fruit flies stood at 30 days, prompting a study that exposed a random sample of these flies to 'EcoShield.' The primary question driving this inquiry is whether the exposure to 'EcoShield' diminishes the fruit flies' average lifespan in comparison to the established 30-day norm.

This inquiry sets the stage for a meticulous statistical analysis to determine if the introduction of 'EcoShield' brings about a noteworthy reduction in the lifespan of these pivotal subjects in life science research."

**Anonymous**

Can you make it more plain language at an 8th grade reading level?

**ChatGPT**

"In 2023, a new bug spray called 'EcoShield' was invented for farms. Scientists wanted to see if it affects the lives of fruit flies. They study fruit flies a lot because they live for a short time and help in science.

Before, fruit flies lived about 30 days. Some flies were sprayed with 'EcoShield' in a special test to check if it makes them live for a shorter time.

The big question is: Does the bug spray 'EcoShield' make the flies live for a shorter time compared to their usual 30 days?

This study is like a detective job. Scientists want to know if this bug spray is making the flies live for a lot less time."