7.1-SJ: Executing the Chi-Square Test for One-Way Tables (Goodness-of-Fit)
Supreme Court Interruptions

INTRODUCTION

The Supreme Court of the United States of America is a strong and vital component of American democracy. As members of this highest court, the nine Supreme Court Justices are charged with the task of preserving the Constitution and the promise of equal justice for all. The composition of this court is, therefore, key in shaping laws that affect the lives of all Americans. In 1990, one woman served as a Supreme Court Justice. Two women served on the court in 2002, and 3 women served on the court in 2015.

Researchers at Northwestern University's Pritzker School of Law examined the behavior of United States Supreme Court Justices during oral arguments. "Oral arguments at the Supreme Court are important - they affect case outcomes and constitute the only opportunity for outsiders to directly witness the behavior of the justices of the highest court." To study justices' efforts to sway case outcomes during oral arguments, they examined how frequently justices interrupted each other during the presentation of arguments and identified the sex of the justices who were interrupted. The researchers focused on the 1990, 2002, and 2015 terms. Using the data from these terms, we will explore whether the interruptions were proportionally distributed among male and female justices when they were making oral arguments.

Recall that a chi-square goodness-of-fit test contains the following steps:

Step 1. State the null and alternative hypotheses.
Step 2. State the conditions (and examine if they are met).
Step 3. Compute the chi-square test statistic.
Step 4. Compute (or approximate) the *P*-value.
Step 5. Make a decision and interpret the decision in the context of the situation.

We will begin by examining the data from the **1990** Supreme Court term. In 1990, 8 men and 1 woman served on the court. A total of 38 interruptions were observed, where 10 occurred when female justices made arguments, and 28 occurred when male justices made arguments. Let's treat the results from the 1990 term as a *random sample* of results from all terms that had this same male / female composition of justices.

1 At the 1% significance level, perform a chi-square goodness-of-fit test to test the claim that interruptions were proportionally distributed among male and female justices in Supreme Court terms similar to the 1990 term.

A If interruptions were proportionally distributed among male and female justices, what percent of interruptions would be directed at male justices, and what percent of interruptions would be directed at female justices?

B State the null and alternative hypotheses for this test.

C If interruptions were proportionally distributed among male and female justices, how many interruptions would you expect would occur when male justices made arguments, and when female justices made arguments? These values are called the *expected frequencies*. Round the values to two decimal places.

D Complete the table below.

	Interruptions of Female Justices	**Interruptions of Male Justices**
Observed frequencies (O) (from 1990 data)		
Expected frequencies (E) (based on Null Hypothesis)		
$O - E$		
$(O - E)^2$		
$\frac{(O-E)^2}{E}$		

E A condition for the chi-square goodness-of-fit test is that each expected frequency is at least 5. Is this condition met?

F Compute the chi-square test statistic: $\chi^2 = \sum \frac{(O-E)^2}{E}$

G Compute the *P*-value for this test. You must use the χ^2 distribution to determine the *P*-value. Open this Desmos calculator: https://carnegiemathpathways.org/go/desmoschi

In this document, the symbol $X_{ChiSquare}$ refers to your χ^2 statistic. The document only needs the test statistic, because it correctly assumes you are conducting a right-tailed test.

H Make a decision about the hypotheses and state a conclusion in context.

Let's now focus on data from the **2002** Supreme Court term. In 2002, 7 men and 2 women served on the court. A total of 170 interruptions were observed, where 77 occurred when female justices made arguments, and 93 occurred when male justices made arguments. Let's treat the results from the 2002 term as a *random sample* of results from all terms that had this same male/female composition of justices.

2 At the 1% significance level, perform a chi-square goodness-of-fit test to test the claim that interruptions were proportionally distributed among male and female justices.

A If interruptions were proportionally distributed among male and female justices, what percent of interruptions would be directed at male justices, and what percent of interruptions would be directed at female justices?

B State the null and alternative hypotheses for this test.

C If interruptions were proportionally distributed among male and female justices, how many interruptions would you expect would occur when male justices made arguments, and when female justices made arguments? These values are called the *expected frequencies*. Round the values to two decimal places.

D Complete the table.

	Interruptions of Female Justices	**Interruptions of Male Justices**
Observed frequencies (O) (from 2002 data)		
Expected frequencies (E) (based on Null Hypothesis)		
$O - E$		
$(O - E)^2$		
$\frac{(O-E)^2}{E}$		

E A condition for the chi-square goodness-of-fit test is that each expected frequency is at least 5. Is this condition met?

F Compute the chi-square test statistic: $\chi^2 = \sum \frac{(O-E)^2}{E}$

G Compute the *P*-value for this test. You must use the χ^2 distribution to determine the *P*-value. Open this Desmos calculator: https://carnegiemathpathways.org/go/desmoschi

In this document, the symbol $X_{ChiSquare}$ refers to your χ^2 statistic. The document only needs the test statistic, because it correctly assumes you are conducting a right-tailed test.

H Make a decision about the hypotheses and state a conclusion in context.

Finally, let's focus on data from the **2015** Supreme Court term. In 2015, 6 men and 3 women served on the court. A total of 214 interruptions were observed, where 141 occurred when female justices made arguments, and 73 occurred when male justices made arguments. Let's treat the results from the 2015 term as a *random sample* of results from all terms that had this same male/female composition of justices.

3 At the 1% significance level, perform a chi-square goodness-of-fit test to test the claim that interruptions were proportionally distributed among male and female justices.

A If interruptions were proportionally distributed among male and female justices, what percent of interruptions would be directed at male justices, and what percent of interruptions would be directed at female justices?

B State the null and alternative hypotheses for this test.

C If interruptions were proportionally distributed among male and female justices, how many interruptions would you expect would occur when male justices made arguments, and when female justices made arguments? These values are called the *expected frequencies*. Round the values to two decimal places.

D Determine the observed and expected frequencies of interruptions for male and female justices. Fill in the table below.

	Interruptions of Female Justices	**Interruptions of Male Justices**
Observed frequencies (O) (from 2015 data)		
Expected frequencies (E) (based on Null Hypothesis)		

E A condition for the chi-square goodness-of-fit test is that each expected frequency is at least 5. Is this condition met?

F Compute the chi-square test statistic: $\chi^2 = \sum \frac{(O-E)^2}{E}$

G Compute the *P*-value for this test. You must use the χ^2 distribution to determine the *P*-value. Open this Desmos calculator: https://carnegiemathpathways.org/go/desmoschi

In this document, the symbol $X_{ChiSquare}$ refers to your χ^2 statistic. The document only needs the test statistic, because it correctly assumes you are conducting a right-tailed test.

H Make a decision about the hypotheses and state a conclusion in context.

4 Overall, what does this data suggest about the relationship between gender and interruptions in the United States Supreme Court? Has the distribution become more or less equal over time? Use your findings to justify your conclusion. What are some possible implications of these findings?

Exercise 7.1-SJ

(1) A fair jury selection process is critical to a functioning democracy. All adults in a population who are eligible to serve on a jury should have the opportunity to serve, and the composition of juries should reflect the racial, ethnic, and gender makeup of the communities where the trials occur. Historically in the United States, this has not been the case. Historians have documented numerous instances where people of color and women were systematically excluded for jury participation.

Suppose a researcher was interested in whether a current jury-selection system in a state showed evidence of a racial bias. The researcher randomly collected data on the racial makeup of adults who served on juries in the state. The table below summarizes the races of the adults who served. This data represents a random sample of all juries which served in the state during that year.

White	Black	Hispanic	Other
212	38	42	28

Census data shows the following racial composition for adults in the state: 60% white, 14% black, 18% Hispanic, and 8% other. Let's assume that eligible jurors in the state have the same racial composition as adults in the state. Conduct a chi-square goodness-of-fit test to examine the claim that the racial composition of people selected to be jurors differs from the racial composition of adults in the state. Use a 1% significance level for the test.

(a) State the null hypothesis.

(b) State the alternative hypothesis.

(c) What is the sample size?

(d) What are the expected frequencies of White, Black, Hispanic, and Other?

(e) Compute $\frac{(O-E)^2}{E}$ for White, Black, Hispanic, and Other.

(f) Compute the Chi-Square Test Statistic.

(g) Compute the *P*-value.

(h) Make a decision and a conclusion in context.

(2) Suppose that a cereal consumer feedback survey of a random sample of 558 consumers yielded the results that are presented in the table below:

Yes	No	Maybe
400	103	55

Conduct a chi-square goodness-of-fit test to examine the claim that the population distribution of responses differs from the desired proportions of 65% "Yes", 25% "No", and 10% "Maybe". Use a 5% significance level for the test.

Complete all of the steps for the chi-square goodness-of-fit test. Be sure to begin the test by stating the null and alternative hypotheses. When you state your final conclusion, remember to interpret your conclusion in the context of the cereal consumer feedback survey case.

(a) State the null hypothesis.

(b) State the alternative hypothesis.

(c) What is the sample size?

(d) What are the expected frequencies of Yes, No, and Maybe.

(e) Compute $\frac{(O-E)^2}{E}$ for Yes, No, and Maybe.

(f) Compute the Chi-Square Test Statistic.

(g) Compute the *P*-value.

(h) Make a decision and a conclusion in context.