# Assessment Questions

## Two Proportion Resampling Test App

### Homework question:

**Box Office ~** The Walt Disney Company, the production studio behind the Marvel Cinematic Universe, is trying to find the best release date for their new super hero movie. They are hesitating between the summer and winter holiday release, and they want to know if there is any difference between the earning potential during those seasons. The analytics division in the company examined box-office data for movies released during the past 5 years in the USA and Canada and found that 57 out of 1263 movies with summer release date earned over 400 million dollars. They also counted that out of 541 movies released in the winter, 6 earned over 400 million dollars. Round all answers to four decimal places.

Use this information to complete the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Over 400 million USD | Under 400 million USD | Total |
| Summer release | 57 |  | 1263 |
| Winter release | 6 |  | 541 |
| Total |  |  |  |

1. What are the hypotheses for testing whether there is a significant difference in the proportion of movies who earn over 400 million dollars for the two release seasons.
2. Calculate the difference in the proportions of movies that earned over 400 million dollars: =
3. The paragraph below describes the set up for a randomization test, if we were to conduct a hypothesis test without using software. Fill in the blanks with a number.

We write Summer on \_\_\_\_\_\_\_\_\_\_\_ cards and cards representing the movies with summer release date, and Winter on \_\_\_\_\_\_\_\_\_\_\_ cards. Then, we shuffle these cards and split them into two groups: one group of size \_\_\_\_\_\_\_\_\_\_\_ representing the movies with box-office over 400 million dollars, and another group of size \_\_\_\_\_\_\_\_\_\_\_ representing the rest of the movies. We calculate the difference in the proportions of movies that earned over 400 million dollars during the two release seasons to get . Finally, we build a histogram of these simulated differences. Use the Two Proportion Resampling Test App to do this.

4. Choose the **greater than/less than/beyond** for the counted samples and shuffling at least 3000 times. What is the approximate p-value for the hypothesis test rounded to four decimal places? p-value = \_\_\_\_\_\_\_\_\_\_\_

5. Based on this p-value, we should **reject/fail to reject** the null hypothesis.

### Exam question:

1. **Mass transit use between shift- and salaried employees ~** Kenpoolmore is a large manufacturing company with plants in several large metropolitan areas around the United States. In their ongoing effort to assess their environmental impact, they take a random sample of the salaried and shift-work employees at their plant in Kentucky. They find that in the sample of 200 salaried employees, 64 routinely take mass transit to work, while in the sample of 250 shift-work employees, 104 routinely take mass transit to work.
2. Consider the two variables in the scenario above, ‘**transit type**’ and **‘employee type**.’ The research question Kenpoolmore is interested in answering is…

*Is there a difference in the proportions of salaried and shift-employees who take mass transit?*

or, equivalently…

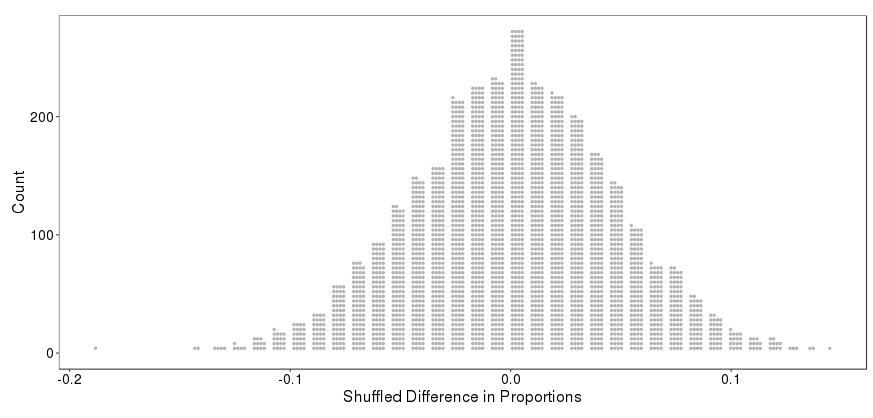
*Are the variables transit type and employee type independent of one another?*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Give a point estimate of the difference in proportions of salaried and shift employees who take mass transit to work. That is, compute .
2. The paragraph below describes the set up for a randomization technique, if we were to do it without using statistical software. Fill in the blanks with a number or phrase, whichever is appropriate. A number should be written in each blank in the paragraph below. Round non-integer answers to two decimal places.

We write *Salaried* on \_\_\_\_\_\_\_\_ cards and *Shift* on ­­\_\_\_\_\_\_\_ cards. Then, we shuffle these cards and split them into two groups: one group of size \_\_\_\_\_\_ representing those who take mass transit, and another group of size \_\_\_\_\_\_ representing those who did not. We calculate the difference between the proportion of *Salaried* respondents who take mass transit and the proportion of *Shift* respondents who take mass transit and record this value. We repeat this many times to build a distribution centered at \_\_\_\_\_\_\_\_\_\_. Lastly, we calculate the fraction of simulations where the simulated differences in proportions are (circle one) **less than greater than beyond** \_\_\_\_\_\_\_\_\_.



1. In 3000 repetitions, there were 44 simulations where and 64 simulations where . What is the approximate p-value for this hypothesis test?
2. At the significance level, it would be appropriate to …

**Circle one: *Reject Fail to Reject*** *…* the null hypothesis.