**Chapter 12 FRAPPY!** Sample #1

*Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.*

Two statistics students wanted to know if including additional information in a survey question would change the distribution of responses. To find out, they randomly selected 30 teenagers and asked them one of the following two questions. Fifteen of the teenagers were randomly assigned to answer Question A, and the other 15 students were assigned to answer Question B.

> **A:** When choosing a college, how important is a good athletic program: very important, important, somewhat important, not that important, or not important at all?
>
> **B:** It's sad that some people choose a college based on its athletic program. When choosing a college, how important is a good athletic program: very important, important, somewhat important, not that important, or not important at all?

The table below summarizes the responses to both questions. For these data, the chi-square test statistic is $\chi^2 = 6.12$.

| | **Question A** | **Question B** | **Total** |
|---|---|---|---|
| Very important | 7 | 2 | **9** |
| Important | 4 | 3 | **7** |
| Somewhat important | 2 | 3 | **5** |
| Not that important | 1 | 2 | **3** |
| Not important at all | 1 | 5 | **6** |
| **Total** | **15** | **15** | **30** |

(a) State the hypotheses that the students are interested in testing.

Ho: The true distributions of answers to questions A and B are the same

Ha: They are not the same

(b) Describe a Type I error and a Type II error in the context of the hypotheses stated in part (a).

Type I: Not finding convincing evidence that the distributions of answers to questions A and B are different when they really are

Type II: Finding convincing evidence that the distributions of answers to questions A and B are different when they really are not different

(c) For these data, explain why it would *not* be appropriate to use a chi-square distribution to calculate the *P*-value.

Expected counts

| | |
|---|---|
| 4.5 | 4.5 |
| 3.5 | 3.5 |
| 2.5 | 2.5 |
| 1.5 | 1.5 |
| 3 | 3 |

All expected counts are <5

(d) To estimate the *P*-value, 100 trials of a simulation were conducted, assuming that the additional information didn't have an effect on the response to the question. In each trial of the simulation, the value of the chi-square statistic was calculated. These simulated chi-square statistics are displayed in the dotplot below.

Based on the results of the simulation, what conclusion would you make about the hypotheses stated in part (a)?

Because the value 6.12 could happen by chance alone, we do not have convincing evidence that the distributions of answers to questions A and B are different.

**Chapter 12 FRAPPY!** Sample #2

*Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.*

Two statistics students wanted to know if including additional information in a survey question would change the distribution of responses. To find out, they randomly selected 30 teenagers and asked them one of the following two questions. Fifteen of the teenagers were randomly assigned to answer Question A, and the other 15 students were assigned to answer Question B.

> **A:** When choosing a college, how important is a good athletic program: very important, important, somewhat important, not that important, or not important at all?
>
> **B:** It's sad that some people choose a college based on its athletic program. When choosing a college, how important is a good athletic program: very important, important, somewhat important, not that important, or not important at all?

The table below summarizes the responses to both questions. For these data, the chi-square test statistic is $\chi^2 = 6.12$.

| | **Question A** | **Question B** | **Total** |
|---|---|---|---|
| Very important | 7 | 2 | **9** |
| Important | 4 | 3 | **7** |
| Somewhat important | 2 | 3 | **5** |
| Not that important | 1 | 2 | **3** |
| Not important at all | 1 | 5 | **6** |
| **Total** | **15** | **15** | **30** |

(a) State the hypotheses that the students are interested in testing.

$H_0$: The distributions of responses were the same.

$H_a$: The distributions of responses were not the same.

(b) Describe a Type I error and a Type II error in the context of the hypotheses stated in part (a).

Type I Error - rejecting $H_0$ when $H_0$ is true

Type II Error - failing to reject $H_0$ when $H_a$ is true

(c) For these data, explain why it would *not* be appropriate to use a chi-square distribution to calculate the *P*-value.

Expected counts (4.5, 4.5, 3.5, 3.5, 2.5, 2.5, 1.5, 1.5, 3, 3) are all too small.

(d) To estimate the *P*-value, 100 trials of a simulation were conducted, assuming that the additional information didn't have an effect on the response to the question. In each trial of the simulation, the value of the chi-square statistic was calculated. These simulated chi-square statistics are displayed in the dotplot below.

Based on the results of the simulation, what conclusion would you make about the hypotheses stated in part (a)?

$\chi^2 = 6.12$

$df = 4$

p-value = 0.1904

Because the p-value > .05, we fail to reject $H_0$. There is not evidence that the distributions of responses were different.