SUMMARY: The Five Steps of the Chi-Squared Test of Independence

1. **Assumptions:** Two categorical variables
   Randomization, such as random sampling or a randomized experiment
   Expected count ≥ 5 in all cells (otherwise, use small-sample test in Section 11.5)
2. **Hypotheses:**
   H_0: The two variables are independent.
   H_a: The two variables are dependent (associated).
3. **Test statistic:**

$$X^2 = \sum \frac{(\textit{observed count} - \textit{expected count})^2}{\textit{expected count}},$$

   where expected count = (row total × column total)/total sample size
4. **P-value:** Right-tail probability above observed X^2 value, for the chi-squared distribution with $\boldsymbol{df} = (\boldsymbol{r} - 1) \times (\boldsymbol{c} - 1)$
5. **Conclusion:** Report P-value and interpret in context. If a decision is needed, reject H_0 when P-value $\leq$ significance level (such as 0.05).

SUMMARY: Misuses of the Chi-Squared Test

The chi-squared test is often misused. Some common misuses are applying it

- When some of the expected frequencies are too small.
- When separate rows or columns are dependent samples,[2] such as when each row of the table has the same subjects.
- To data that do not result from a random sample or randomized experiment.
- To data by classifying quantitative variables into categories. This results in a loss of information. It is usually more appropriate to analyze the data with methods for quantitative variables, like those the next chapter presents.

SUMMARY: Fisher's Exact Test of Independence for 2 × 2 Tables

1. **Assumptions:**
   Two binary categorical variables
   Randomization, such as random sampling or a randomized experiment
2. **Hypotheses:**
   H_0: The two variables are independent (H_0: $p_1 = p_2$)
   H_a: The two variables are associated

   (Choose H_a: $p_1 \neq p_2$ or H_a: $p_1 > p_2$ or H_a: $p_1 < p_2$).
3. **Test statistic:** First cell count (this determines the others, given the margin totals).
4. **P-value:** Probability that the first cell count equals the observed value or a value even more extreme than observed in the direction predicted by H_a.
5. **Conclusion:** Report P-value and interpret in context. If a decision is needed, reject H_0 when P-value $\leq$ significance level (such as 0.05).