Chapter 10

Study Questions

1. What is the sampling distribution of the test statistic for a one-way frequency hypothesis test?

2. What is the sampling distribution of the test statistic for a two-way frequency hypothesis test?

3. How are contingency tables and fixed-row or fixed-column tables different?

4. How are contingency tables and fixed-row or fixed-column tables similar?

5. The expected values used in calculating the test statistic are based on which hypothesis, null or alternative?

Analysis of Categorical Data
(Two Outcomes)

A **binomial experiment** is one that has the following properties:

1. The experiment consists of n identical trials.
2. Each trial results in one of two outcomes. We will label one outcome a success and the other a failure.
3. The probability of success on a single trial is equal to π and π remains the same from trial to trial.
4. The trials are independent; that is, the outcome of one trial does not influence the outcome of any other trial.
5. The random variable y is the number of successes observed during the n trials.

Frequently, it is of interest to estimate the proportion (π) of a population that is associated with a particular outcome (success).

A point estimate for π is given by

$$\hat{\pi} = \frac{Total\ number\ of\ successes}{Total\ number\ of\ observations}$$

Large Sample (1-α)100% Confidence Interval for a Single Population Proportion

$$\hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Large Sample Hypothesis Tests for a Single Population Proportion

Hypotheses:

H_0: $\pi = \pi_0$	H_0: $\pi = \pi_0$	H_0: $\pi = \pi_0$
H_a: $\pi < \pi_0$	H_a: $\pi > \pi_0$	H_a: $\pi \neq \pi_0$

Test Statistic:

$$z_{obs} = \frac{\hat{\pi} - \pi_o}{\sqrt{\frac{\pi_o(1-\pi_o)}{n}}}$$

Distribution for the Rejection Region:

The rejection region is found using a standard normal distribution.

Large Sample (1-α)100% Confidence Interval for the Difference of Two Population Proportions

$$(\hat{\pi}_1 - \hat{\pi}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

Large Sample Hypothesis Tests for the Difference of Two Population Proportions

Hypotheses:

H_0: $\pi_1 - \pi_2 = \pi_0$	H_0: $\pi_1 - \pi_2 = \pi_0$	H_0: $\pi_1 - \pi_2 = \pi_0$
H_a: $\pi_1 - \pi_2 < \pi_0$	H_a: $\pi_1 - \pi_2 > \pi_0$	H_a: $\pi_1 - \pi_2 \neq \pi_0$

Test Statistic:

A. Two samples - testing H_0: $(\pi_1 - \pi_2) = \pi_0$ (something other than zero)

$$z_{obs} = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - \pi_0}{\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}}$$

B. Two samples - testing H_0: $(\pi_1 - \pi_2) = 0$

$$z_{obs} = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \quad where\ \hat{\pi} = \frac{total\ no.\ successes\ from 1\ and\ 2}{total\ obs.\ from\ 1\ and\ 2}$$

Distribution for the Rejection Region:

The rejection region is found using a standard normal distribution.

Aspirin and Heart Attacks

Twenty thousand doctors participated in a study to determine if taking an aspirin every other day would result in a reduction of heart attacks. The physicians were randomly divided into two groups. The first group (called the treatment group) received an aspirin every other day, while the other group (called the control group) received a placebo. At the end of the study, there had been 104 heart attacks in the treatment group and 189 heart attacks in the control group. Use a significance level of .01.

State the hypotheses.

Determine the rejection region.

Calculate the test statistic.

What is your conclusion?