Chapter 5. Sampling Distributions

Introduction

• A statistic such as a sample proportion \( \hat{p} \) or a sample mean \( \bar{X} \) from a random sample or randomized experiment is a random variable.

• The probability distribution of a statistic is its sampling distribution.

• The population distribution of a variable is the distribution of its values for all the members of the population.

• The population distribution is also the probability distribution of the variable for one individual chosen at random from the population.

Sampling Distributions for Counts and Proportions

• Let a random variable \( X \) be a count of the occurrences of some outcome in a fixed number of observations \( n \).

• The sample proportion is \( \hat{p} = \frac{X}{n} \).

What is in common?

• Observe the next 20 babies born at a local hospital. Let \( X \) be the number of girls.

• 60% of people in a population oppose the death penalty. Choose 100 adults at random from the population. Let \( X \) be the number of adults in the sample who oppose the death penalty.

• Draw two balls at random from an urn with 2 black balls and 3 white balls with replacement. Let \( X \) be the number of black balls in the two chosen balls.

• A telephone polling firm reports that the probability of a call reaching a live person is 0.2. Let \( X \) be the number of calls reaching a live person when the firm places 10 calls at randomly chosen residential phone numbers.

The Binomial Setting

• The number of observations \( n \) is fixed.

• The \( n \) observations are all independent.

• Each observation falls into one of two categories, either “success” or “failure”.

• The probability of a success \( p \) is the same for each observation.
Binomial Distribution

• The distribution of the count $X$ of successes in the binomial setting is called the *binomial* distribution.

• It is denoted by $B(n, p)$, where $n$ is the number of observations, and $p$ is the probability of a success on each observation.

• The possible values of $X$ are $0, 1, \ldots, n$.

(Exercise) Identify $n$ and $p$ in each of the four examples.

Binomial Probabilities

Suppose that $X$ has the binomial distribution $B(n, p)$.

• What is the probability that we have exactly $k$ successes?

For $k = 0, 1, \ldots, n$,

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the number of ways of arranging $k$ successes among $n$ observations.

• You may use Table C or Minitab for convenience. In Minitab, use the menu Calc; Probability Distributions; Binomial....

(EXAMPLE: Multiple Choice Exam)
A multiple choice exam has 10 questions, each with five possible answers. A student just guesses at all the answers. Let $X$ be the number of correct answers.

(a) What is the distribution of $X$?

(b) What is the probability that the student answers all wrong?

(c) What is the probability that the student answers all right?

(d) What is the probability that the student answers exactly two questions correctly?

2
Mean and Variance of $B(1, p)$

- Let $S_i$ be a random variable that indicates whether the $i$th observation is a success or failure, taking 1 for a success and 0 for a failure.
- The probability distribution of $S_i$ is

<table>
<thead>
<tr>
<th>Value</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>$1 - p$</td>
<td>$p$</td>
</tr>
</tbody>
</table>

- The mean of $S_i$: $\mu_{S_i} = 0 \times (1 - p) + 1 \times p = p$.
- The variance of $S_i$: $\sigma^2_{S_i} = (0 - p)^2 \times (1 - p) + (1 - p)^2 \times p = p(1 - p)$.

Mean and Standard Deviation of $B(n, p)$

- Decompose $X = S_1 + S_2 + \cdots + S_n$.
- Apply the addition rules for means and variances to the sum $X = \sum_{i=1}^{n} S_i$.
- $\mu_X = \mu_{S_1} + \mu_{S_2} + \cdots + \mu_{S_n} = np$.
- $\sigma^2_X = \sigma^2_{S_1} + \sigma^2_{S_2} + \cdots + \sigma^2_{S_n} = np(1 - p)$.
- $\sigma_X = \sqrt{np(1 - p)}$.

(EXAMPLE: Death Penalty)

60% of people in a population oppose the death penalty. Choose 100 adults at random from the population. Let $X$ be the number of adults in the sample who oppose the death penalty.

(a) Find the mean and standard deviation of $X$.

(b) What is $P(X \geq 50)$?
Normal Approximation to Binomial Distribution

- If $X$ is a count having the binomial distribution $B(n, p)$, then when $n$ is large, $X$ is approximately $N(np, \sqrt{np(1-p)})$.

- A rule of thumb is that we use this approximation when $np \geq 10$ and $n(1-p) \geq 10$.

Sample Proportions

- The sample proportion is $\hat{p} = X/n$.

- The mean of $\hat{p}$ is given by

$$\mu_{\hat{p}} = \mu_{\frac{X}{n}} = \frac{1}{n} \mu_X = \frac{1}{n}(np) = p.$$ 

- The sample proportion $\hat{p}$ in an SRS is an unbiased estimator of the population proportion $p$.

- The variance and standard deviation of $\hat{p}$

$$\sigma_{\hat{p}}^2 = \sigma_{\frac{X}{n}}^2 = \left(\frac{1}{n}\right)^2 \sigma_X^2 = \left(\frac{1}{n}\right)^2 np(1-p) = \frac{p(1-p)}{n},$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$ 

- The variability of $\hat{p}$ decreases as the sample size increases.

Normal Approximation for Proportions

- When $n$ is large, the sample proportion $\hat{p}$ in an SRS is approximately $N(p, \sqrt{\frac{p(1-p)}{n}})$.

- Use the same rule of thumb for this approximation ($np \geq 10$ and $n(1-p) \geq 10$).

(Example: Viral Infection)
A viral infection is spread by contact with an infected person. Let the probability that a healthy person gets the infection in one contact, be $p = 0.4$. An infected person has contact with 100 healthy persons independently. Let $\hat{p}$ be the proportion of persons who contract the infection.

(a) Find the mean and standard deviation of $\hat{p}$.

(b) What is $P(0.35 \leq \hat{p} \leq 0.45)$?