Section 6-3

Sampling Distributions and Estimators

EXAMPLE

Because of rude sales personnel, a poor business plan, ineffective advertising, and a poor name, Polly Esther's Fashions was in business only three days. On the first day 1 dress was sold, 2 were sold on the second day, and only 5 were sold on the third day. Because 1, 2, and 5 are the entire population, the mean is $\mu = 2.7$ and the standard deviation is $\sigma = 1.7$.

Let's consider samples of size 2. There are only 9 different possible samples of size 2, assuming we sample *with replacement*.

WHY SAMPLE *WITH* REPLACEMENT?

1. When selecting a relatively small sample from a large population, it makes no significant difference whether we sample with replacement or without replacement.
2. Sampling with replacement results in independent events that are unaffected by previous outcomes, and independent events are easier to analyze and they result in simpler formulas.

SAMPLING DISTRIBUTION OF A STATISTIC

The **sampling distribution of a statistic** (such as a sample mean or sample proportion) is the distribution of all values of the statistic when all possible samples of the same size n are taken from the same population.

The sampling distribution of a statistic is typically represented as a probability distribution in the form of a table, probability histogram, or formula.

SAMPLING DISTRIBUTION OF THE SAMPLE MEAN

The **sampling distribution of the sample mean** is the distribution of all possible sample means (or the distribution of the variable $\bar{x}$), with all samples having the same size n taken from the same population.

The sampling distribution of the sample mean is typically represented as a probability distribution in the form of a table, probability histogram, or formula.

SAMPLING VARIABILITY

The value of a statistic, such as the sample mean $\bar{x}$, depends on the particular values included in the sample, and it generally varies from sample to sample. This variability of a statistic is called **sampling variability**.

SAMPLING DISTRIBUTION OF THE SAMPLE VARIANCE

The **sampling distribution of the sample variance** is the distribution of all possible sample variances (or the distribution of the variable s^2), with all samples having the same size n taken from the same population.

The sampling distribution of the sample variance is typically represented as a probability distribution in the form of a table, probability histogram, or formula.

SAMPLING DISTRIBUTION OF THE PROPORTION

The **sampling distribution of the sample proportion** is the distribution of sample proportions, with all samples having the same sample size n.

Notation for Proportions:

p = *population* proportion

$\hat{p}$ = *sample* proportion

The next slide shows the sampling distributions of several statistics for our example as well as the sampling distribution of the proportion of odd numbers for our example.

Table 6-7 Sampling Distributions of Statistics (for Samples of Size 2 Drawn with Replacement from the Population 1, 2, 5)

Sample	Mean $\bar{x}$	Median	Range	Variance s^2	Standard Deviation s	Proportion of Odd Numbers	Probability
1, 1	1.0	1.0	0	0.0	0.000	1	1/9
1, 2	1.5	1.5	1	0.5	0.707	0.5	1/9
1, 5	3.0	3.0	4	8.0	2.828	1	1/9
2, 1	1.5	1.5	1	0.5	0.707	0.5	1/9
2, 2	2.0	2.0	0	0.0	0.000	0	1/9
2, 5	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 1	3.0	3.0	4	8.0	2.828	1	1/9
5, 2	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 5	5.0	5.0	0	0.0	0.000	1	1/9
Mean of Statistic Values	8/3	8/3	16/9	26/9	1.3	2/3	
Population Parameter	8/3	2	4	26/9	1.7	2/3	
Does the sample statistic target the population parameter?	Yes	No	No	Yes	No	Yes	

PROPERTIES OF THE DISTRIBUTION OF SAMPLE MEANS

- The sample means target the value of the population mean. (That is, the mean of the sample means is the population mean. The expected value of the sample means is equal to the population mean.)
- The distribution of sample means tends to be a normal distribution. (This will be discussed further in the next section, but the distribution tends to become closer to a normal distribution as the sample size increase.)

PROPERTIES OF THE DISTRIBUTION OF SAMPLE VARIANCES

- The sample variances tend to target the value of the population variance. (That is, the mean of the sample variances is the population variance. The expected value of the sample variance is equal to the population variance.)
- The distribution of sample variances tends to be a distribution skewed to the right.

PROPERTIES OF THE DISTRIBUTION OF SAMPLE PROPORTIONS

- The sample proportions tend to target the value of the population proportion. (That is, the mean of the sample proportions is the population proportion. The expected value of the sample proportions is equal to the population proportion.)
- The distribution of sample proportions tends to be a normal distribution.

ESTIMATORS

An **estimator** is a statistic that is used to infer (or estimate) the value of a population parameter.

An **unbiased estimator** is a statistic that targets the value of the population parameter in the sense that the sampling distribution of the statistic has a mean that is equal to the mean of the corresponding parameter.

BIASED AND UNBIASED ESTIMATORS

- You will notice from Table 6-7 that some sample statistics (the mean, variance, and proportion) "target" the population parameters. These sample statistics are called **unbiased estimators**.
- Other sample statistics (the median, range, and standard deviation) either overestimate or underestimate the population parameter. These sample statistics are called **biased estimators**.

A COMMENT ON THE STANDARD DEVIATION

Even though the standard deviation is biased, the bias is relatively small in a ***large sample***. As a result s is often used to estimate σ.