Introduction to Probability and Statistics Eleventh Edition

Chapter 7

Sampling Distributions

Introduction

- **Parameters** are numerical descriptive measures calculated from the populations.
  - For the normal distribution, the location and shape are described by μ and σ.
  - For a binomial distribution consisting of n trials, the location and shape are determined by $\boldsymbol{p}$.
- Often the values of parameters that specify the exact form of a distribution are **unknown**.
- You must rely on the sample to learn about these parameters.

Sampling

- ✓ *A Sampling* is a process of taking a sample.

Why Sampling?

- An agronomist believes that the yield per acre of a variety of wheat is approximately normally distributed, but the mean μ and the standard deviation σ of the yields are unknown.
- A pollster is sure that the responses to his "agree/disagree" question will follow a binomial distribution, but p, the proportion of those who "agree" in the population, is unknown.

- ✓ *If you want the sample to provide reliable information about the population, you must select your sample in a certain way!*

Sampling Plans

- The **sampling plan** or **experimental design** is the way a sample is selected
- It determines the amount of information you can extract, and often allows you to measure the reliability of your inference.

- It involves
  1. **Non-randomization**
  2. **Randomization**

Sampling Plans

Non-randomized sampling plans should not be used for statistical inference!

1 **Convenience sampling:** A sample that can be taken easily without random selection.

- People walking by on the street

2 **Judgment sampling:** The sampler decides who will and won't be included in the sample.

Methods of (Randomized) Sampling

1. Simple random sampling
2. Stratified random sampling
3. Cluster sampling
4. 1-in-k systematic sampling

Simple Random Sampling

Simple random sampling is a method of sampling that allows each possible sample of size n an equal probability of being selected.

Example 1: There are 89 students in a statistics class. The instructor wants to choose 5 students to form a project group. How should he proceed?

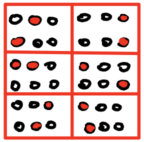
Example 1

1. Give each student a number from 01 to 89.
2. Choose 5 pairs of random digits from the random number table.
3. If a number between 90 and 00 is chosen, choose another number.
4. The five students with those numbers form the group.

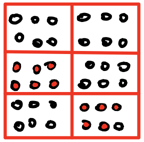
91977	05463	07972	18876	20922
14342	63661	10281	17453	18103
36857	53342	53988	53060	59533
69578	88231	33276	70997	79936
40961	48235	03427	49626	69445
93969	52636	92737	88974	33488
61129	87529	85689	48237	52267
97336	71048	08178	77233	13916

Methods of (Randomized) Sampling

1. **Simple random sampling:**
2. **Stratified random sampling**: Divide the population into subpopulations or **strata** and select a simple random sample from each stratum.



3. **Cluster sampling**: Divide the population into subgroups called **clusters**; select a simple random sample of clusters and take a census of every element in the cluster.



4. **1-in-k systematic sampling**: Randomly select one of the first k elements in an ordered population, and then select every *k*th element thereafter.

Examples 2-5

2. Divide Bangkok into districts and take a simple random sample within each district.

   Stratified

3. Divide Bangkok into districts and take a simple random sample of 10 districts.

   Cluster

4. Divide a city into city blocks, choose a simple random sample of 10 city blocks, and interview all who live there.

   Cluster

5. Choose an entry at random from the phone book, and select every 50^{th} number thereafter.

   1-in-50 Systematic

Sampling Distributions

- Numerical descriptive measures calculated from the sample are called **statistics**.
- Statistics vary from sample to sample and hence are random variables.
- The probability distributions for statistics are called **sampling distributions**.
- In repeated sampling, they tell us what values of the statistics can occur and how often each value occurs.

Sampling Distributions

Definition: The *sampling distribution of a statistic* is the probability distribution for the possible values of the statistic that results when random samples of size *n* are repeatedly drawn from the population.

Population: 2, 4, 5, 6

Population Mean, $\mu = 4.25$
Population Std, Dev., $\sigma = 1.479$

Draw all samples of size *n* = 3 without replacement.

Possible samples	$\bar{x}$	$p(\bar{x})$
2, 4, 5	11/3 = 3.67	1/4
2, 4, 6	12/3 = 4	1/4
2, 5, 6	13/3 = 4.33	1/4
4, 5, 6	15/3 = 5	1/4

Each value of x-bar is equally likely, with probability ¼.

A Sampling Distribution of $\bar{x}$

$\bar{x}$ is a statistic.

Its distribution is called the **sampling distribution of** $\bar{x}$ **.**

$\bar{x}$	$p(\bar{x})$
3.67	**1/4**
4	**1/4**
4.33	**1/4**
5	**1/4**

A Sampling Distribution of $\bar{x}$

$\bar{x}$	$p(\bar{x})$
3.67	**1/4**
4	**1/4**
4.33	**1/4**
5	**1/4**

One can find

$P(\bar{x} < 4.33)$,

the mean of $\bar{x}$,

the standard deviation of $\bar{x}$,

etc.

The standard deviation of *x*-bar is sometimes called the STANDARD ERROR (SE).

Sampling Distributions

How to find a **sampling distribution**?

Sampling distributions for statistics can be

- ✓Approximated with simulation techniques.
- ✓Derived using mathematical theorems
- ✓The Central Limit Theorem is one such theorem.

The Sampling Distribution of $\bar{x}$ (1)

Central Limit Theorem:

If random samples of n observations are drawn from any **population** (or any distribution) with finite mean μ and standard deviation σ, then, when n is large, the sampling distribution of the sample mean $\bar{x}$ is **approximately normally distributed**, with mean denoted by $\mu_{\bar{x}}$ and standard deviation denoted by $\sigma_{\bar{x}}$, where

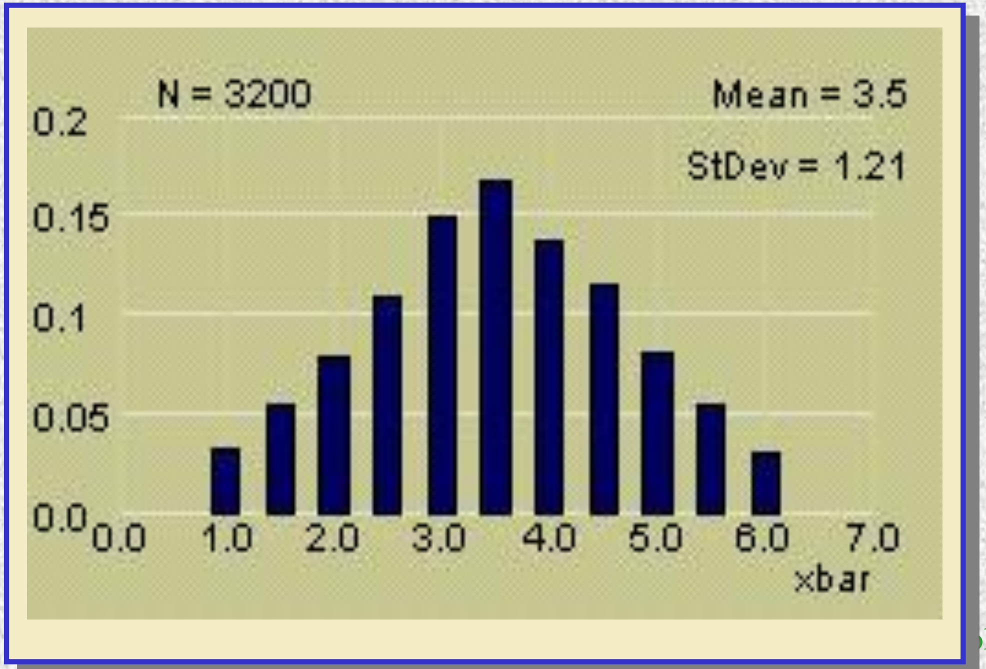
$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \; .$$

The approximation becomes more accurate as n becomes large.

Example 6.1

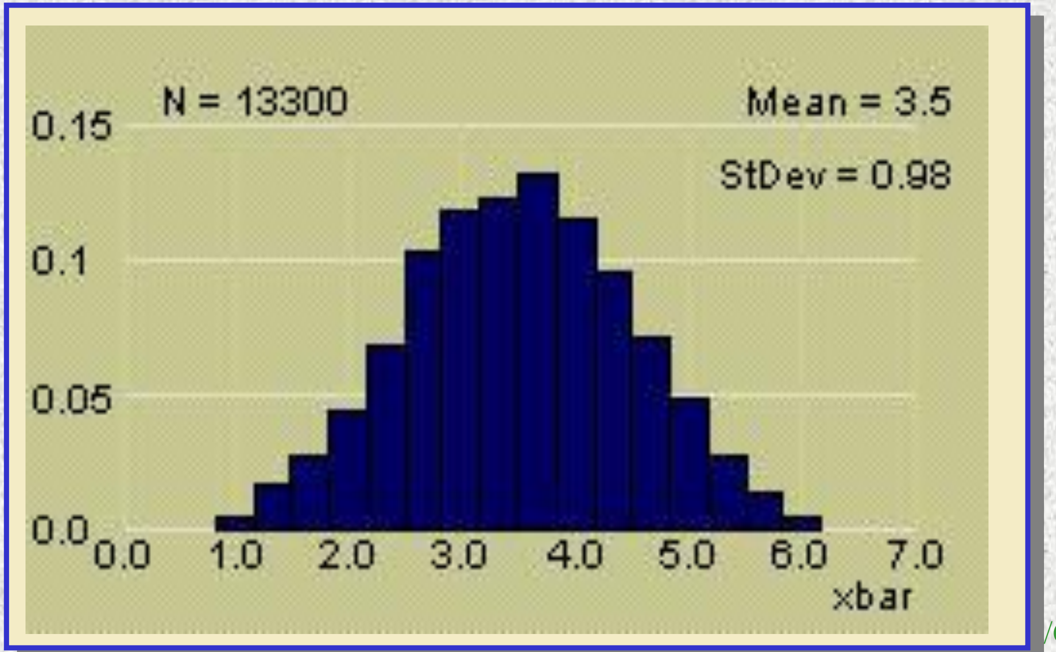
A pair of dice (***n* = 2**) is tossed N=3200 times.

The distribution of $\bar{x}$, the *average* number on the two upper faces is **mound-shaped.**



Example 6.2

Three dice (***n* = 3**) are tossed N=13,300 times. The distribution of $\bar{x}$, the *average* number on the three upper faces is **approximately normal.**



Notes

1. The standard deviation of a statistic is also called **the standard error.**
   - Thus, the standard deviation of $\bar{x}$, given by $\sigma/\sqrt{n}$, is called the **standard error of the mean,** and abbreviated as **SE.**

2. The spread of the distribution, $\dfrac{\sigma}{\sqrt{n}}$, *decreases* as the sample size *n* *increases*.

How Large is Large?

When the sample population is approximately **symmetric**, the distribution becomes approximately normal for relatively small values of n.

When the sample population is either **skewed or unknown,** the sample size should be **at least 30** before the sampling distribution of $\bar{x}$ becomes approximately normal.

Finding Probabilities for the Sample Mean

✓If the sampling distribution of $\bar{x}$ is normal or approximately normal, *standardize or rescale* the interval of interest in terms of

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

✓Find the appropriate area using Table 3.

Finding Probabilities for the Sample Mean

Example 7: A random sample of size $n = 36$ is taken from an unknown population (or distribution) with $\mu = 10$ and $\sigma = 8$. What is the probability that the sample mean will be more than 11?

$$P(\bar{x} > 11) = P\left(\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{11 - 10}{8/\sqrt{36}}\right)$$

$$= P(z > 0.75)$$

$$= 0.2266$$

Class Activity

1. A soda filling machine is supposed to fill cans of soda with 12 ounces. Suppose that the distribution of the fills is unknown but its mean is 12.05 oz. and standard deviation is 0.2 oz.
   a) What is the probability that the average fill for 49 soda cans is less than 12 oz.? Answer : 0.0401
   b) What is the probability that the average fill for 49 soda cans is between 12 and 12.1 oz.?

Answer : ~~0.0802~~

0.9198

1a) $P\left(Z < \frac{12-12.05}{\frac{0.2}{\sqrt{49}}}\right) = P(Z < -1.75)$

$= 0.0401$

b) $P\left(\frac{12-12.05}{0.2/\sqrt{49}} < Z < \frac{12.1-12.05}{0.2/\sqrt{49}}\right)$

Class Activity

2. A random sample of size n = 49 is selected from a population with a population mean of 53 and a population standard deviation of 21. μ

   σ

   a) What will be the approximate shape of the sampling distribution of the sample mean? Normal
   b) What will be the mean and standard deviation of the sampling distribution of the sample mean?
   c) What is the probability that the sample mean will be not more than 55?

b) $\mu_{\bar{x}} = \mu = 53 \, , \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{21}{\sqrt{49}}$

c) $P\left(z < \frac{55-53}{21/\sqrt{49}}\right)$