MATH 2300 – CHAPTER 7 – Sampling Distributions

- A **statistic** is a numerical summary of sample data such as a sample proportion or sample mean
- A **parameter** is a numerical summary of a population such as a population proportion or population mean.
- In practice, we seldom know the values of parameters.
- Parameters are estimated using sample data. We use statistics to estimate parameters.

The *sampling distribution* of a statistic is the *probability distribution* that specifies probabilities for the possible values the statistic can take. Sampling distributions describe the variability that occurs from study to study using statistics to estimate population parameters. Sampling distributions help to predict how close a statistic falls to the parameter it estimates.

Example 1: Let a population of size 5(consider 5 people A, B, C, D and E) and asked whether each individual likes Football or not. Their response to the Football and their age is given in the following table.

Name	Like football	Age
A	Yes	27
B	Yes	33
C	No	30
D	Yes	24
E	No	30

Consider the people who like football

Population proportion = $3/5$

Consider all possible sample of size 3 from this population and calculate sample proportion for each sample

Sample	Sample proportion($\hat{p}$)	Sample mean age($\bar{x}$)
S1 ={A,B,C}	2/3	30
S2 ={A,B,D}	3/3 = 1	28
S3 ={A,B,E}	2/3	30
S4 ={A,C,D}	2/3	27
S5 ={A,C,E}	1/3	29
S6 ={A,D,E}	2/3	27
S7 ={B,C,D}	2/3	29
S8 ={B,C,E}	1/3	31
S9 ={B,D,E}	2/3	29
S10 ={C,D,E}	1/3	28

Sampling Distribution for the sample proportion

Sample Proportion	Probability
1/3	3/10
2/3	6/10
1	1/10

Mean of the sampling distribution

Mean = $\sum x\, P(x) = (\frac{1}{3} \times \frac{3}{10}) + \left(\frac{2}{3} \times \frac{6}{10}\right) + (1 \times \frac{1}{10})$

$= \frac{3}{5}$

Mean and Standard Deviation of the Sampling Distribution of a Proportion

- For a random sample of size *n* from a population with proportion *p* of outcomes in a particular category, the sampling distribution of the proportion of the sample in that category has

$$\text{Mean} = p$$

$$\text{standard deviation} = \sqrt{\frac{p(1-p)}{n}}$$

The Standard Error

To distinguish the standard deviation of a *sampling distribution* from the standard deviation of an ordinary probability distribution, we refer to it as a *standard error.*

Example 2: Given that the exit poll had 2705 people and assuming 50% support the reelection of Schwarzenegger. Find the estimate of the population proportion and the standard error:

$$population\ \ proportion\ = p = .5$$

$$standard\ \ error\ = \sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.5(1-.5)}{2705}} = .0096$$

Three types of distributions

(1) **Population distribution**: This is the probability distribution from which we take the sample. Values of its parameters are usually unknown. They're what we'd like to learn about.

(2) **Data Distribution:** This is the distribution of the sample data. It's the distribution we actually see in practice. It's described by statistics. With random sampling, the larger the sample size *n*, the more closely the data distribution resembles the population distribution

(3) **Sampling Distribution:** This is the probability distribution of a sample statistic. With random sampling, the sampling distribution provides probabilities for all the possible values of the statistic. The sampling distribution provides the key for telling us how close a sample statistic falls to the corresponding unknown parameter. Its standard deviation is called the standard error.

Section 7.2 - How Close Are Sample Means to Population Means?

Example 3: Consider the distribution of sample mean ages in Example 1
Possible values for the random variable (sample mean age ($\bar{x}$)) are 27, 28, 29, 30, and 31

Sampling Distribution for the sample mean age

$\bar{x}$	27	28	29	30	31
$P(\bar{x})$	2/10	2/10	3/10	2/10	1/10

Population mean age= (27+33+30+24+30)/5 = 28.8

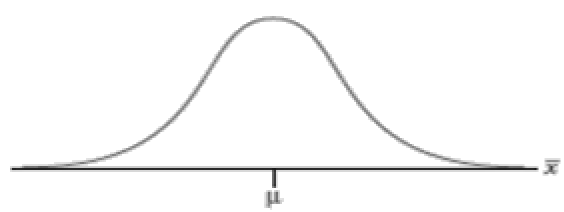
Mean of the sampling distribution

Mean = $\sum x\,P(x) = (27 \times \frac{2}{10}) + \left(28 \times \frac{2}{10}\right) + \left(29 \times \frac{3}{10}\right) + \left(30 \times \frac{2}{10}\right) + (31 \times \frac{1}{10})$

= 28.8

The Sampling Distribution of the Sample Mean

- The sample mean, $\bar{x}$, is a random variable.
- The sample mean varies from sample to sample.
- By contrast, the population mean, μ, is a single fixed number.



The Sampling Distribution of the Sample Mean

- For a random sample of size *n* from a population having mean μ and standard deviation σ, the sampling distribution of the sample mean has:
  - Center described by the **mean μ** (the same as the mean of the population).
  - Spread described by the standard error, which equals the population standard deviation divided by the square root of the sample size:

standard error of $\bar{x} = \sigma/\sqrt{n}$

Example 3: Daily sales at a pizza restaurant vary from day to day. The sales figures fluctuate around a mean μ = \$900 with a standard deviation σ = \$300. What are the center and spread of the sampling distribution?

$\mu = \$900$ $\quad$ $\text{standard error} = \frac{300}{\sqrt{7}} = 113$ $\quad$; n=7 (number of days per week)

Effect of the Standard Error *n* on

- Knowing how to find a standard error gives us a mechanism for understanding how much variability to expect in sample statistics "just by chance."
- The standard error of the sample mean = $\frac{\sigma}{\sqrt{n}}$
- As the sample size *n* increases, the denominator increases, so the standard error decreases.
- With larger samples, the sample mean is more likely to fall closer to the population mean.

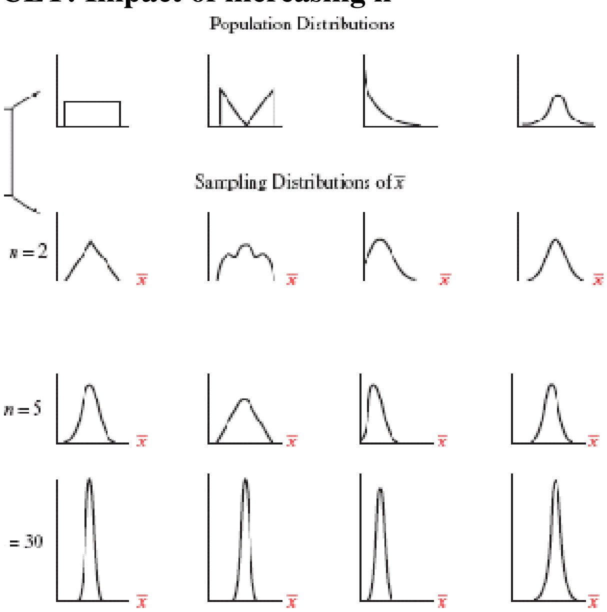
Central Limit Theorem (CLT)

- For random sampling with a large sample size *n*, the sampling distribution of the sample mean is approximately a normal distribution.
- This result applies *no matter what the shape* of the probability distribution from which the samples are taken.

How Large a Sample?CLT:

- The sampling distribution of the sample mean takes more of a bell shape as the random sample size *n* increases.
- The more skewed the population distribution, the larger *n* must be before the shape of the sampling distribution is close to normal.
- In practice, the sampling distribution is usually close to normal when the sample size *n* is at least about 30.
- If the population distribution is approximately normal, then the sampling distribution is approximately normal for *all* sample sizes.

CLT: Impact of increasing n



CLT: Making Inferences

- For large *n*, the sampling distribution is approximately normal even if the population distribution is not.
- This enables us to make inferences about population means regardless of the shape of the population distribution.

- For large *n*, the sampling distribution of $\bar{x}$ is approximately normal no matter what the shape of the underlying population distribution.

Example 4: The distribution of weights of milk bottles is normally distributed with a mean of 1.1 lbs and a standard deviation (σ)=0.20. What is the probability that the mean of a random sample of 5 bottles will be greater than 0.99 lbs?

First Calculate the mean and standard error for the sampling distribution of a random sample of 5 milk bottles

- By the CLT, $\bar{x}$ is approximately normal with mean=1.1 and
- standard error = $0.2/\sqrt{5}$ =0.0894
- P($\bar{x}$ >0.99)= 0.89
-

Example 5: Closing prices of stocks have a right skewed distribution with a mean (μ) of $25 and σ= $20. What is the probability that the mean of a random sample of 40 stocks will be less than $20?

Calculate the mean and standard error for the sampling distribution of a random sample of 40 stocks

- By the CLT, $\bar{x}$ is approximately normal with mean=25 and
- standard error = $20/\sqrt{40}$ =3.1623
- P($\bar{x}$ <20)=0.06

In practice, standard errors are estimated

- Standard errors have exact values depending on parameter values, e.g.,
  - $p(1-p)/n$ for a sample proportion
  - $\sigma/\sqrt{n}$ for a sample mean
- In practice, these parameter values are unknown. Inference methods use standard errors that substitute sample values for the parameters in the exact formulas above

These estimated standard errors are the numbers we use in practice

Sampling Distribution for a Proportion

- For a binomial random variable with *n* trials and probability *p* of success for each, the sampling distribution of the proportion of successes has
  - Mean = p
  - Standard error = $\sqrt{\frac{p(1-p)}{n}}$
- These values can be found by taking the mean *np* and the standard deviation $\sqrt{np(1-p)}$ for the binomial distribution of the number of successes and dividing by *n*

Example 6: An automobile insurer has found that repair claims have a mean of $920 and a standard deviation of $870. Suppose that the next 100 claims can be regarded as a random sample from the long-run claims process. What is the probability that the average of the 100 claims is larger than $900?