Chapter 11

Sampling Distributions

Sampling Terminology

- **Parameter**
  - fixed, unknown number that describes the population
  - Example: population mean
- **Statistic**
  - known value calculated from a sample
  - a statistic is often used to estimate a parameter
  - Example: sample mean
- **Variability**
  - different samples from the same population may yield different values of the sample statistic

Parameter vs. Statistic

A properly chosen sample of 1600 people across the United States was asked if they regularly watch a certain television program, and 24% said *yes*. The *parameter* of interest here is the true proportion of all people in the U.S. who watch the program, while the *statistic* is the value 24% obtained from the sample of 1600 people.

Parameter vs. Statistic

- The mean of a population is denoted by μ – this is a parameter.
- The mean of a sample is denoted by $\bar{x}$ – this is a statistic. $\bar{x}$ is used to estimate μ.
- The true proportion of a population with a certain trait is denoted by p – this is a parameter.
- The proportion of a sample with a certain trait is denoted by $\hat{p}$ ("*p-hat*") – this is a statistic. $\hat{p}$ is used to estimate p.

The Law of Large Numbers

Consider sampling at random from a population with true mean μ. As the number of (independent) observations sampled increases, the mean of the sample gets closer and closer to the true mean of the population.

($\bar{x}$ gets closer to μ)

The Law of Large Numbers Gambling

- The "house" in a gambling operation is not gambling at all
  - the games are defined so that the gambler has a negative expected gain per play (the true mean gain is negative)
  - each play is independent of previous plays, so the *law of large numbers* guarantees that the average winnings of a large number of customers will be close the the (negative) true average

Sampling Distribution

- The **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same size (*n*) from the same population
  – to visualize a distribution we use a histogram
  – to describe a distribution we need to specify the shape, center, and spread
  – we will discuss the distribution of the sample mean (x-bar) in this chapter

BPS - 5th Ed. Chapter 11 7

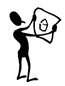
Case Study

Does This Wine Smell Bad?

Dimethyl sulfide (DMS) is sometimes present in wine, causing “off-odors”. Winemakers want to know the odor threshold – the lowest concentration of DMS that the human nose can detect. Different people have different thresholds, and of interest is the mean threshold in the population of all adults.

BPS - 5th Ed. Chapter 11 8

Case Study



Does This Wine Smell Bad?

Suppose the mean threshold of all adults is μ=25 micrograms of DMS per liter of wine, with a standard deviation of σ=7 micrograms per liter and the threshold values follow a bell-shaped (normal) curve.

BPS - 5th Ed. Chapter 11 9

Where should 95% of all <u>individual</u> threshold values fall?

- mean plus or minus two standard deviations
  25 - 2(7) = 11
  25 + 2(7) = 39
- 95% should fall between 11 & 39
- What about the ***mean*** (average) of a sample of *10* adults? What values would be expected?

BPS - 5th Ed. Chapter 11 10

Sampling Distribution

- What about the ***mean*** (average) of a sample of *10* adults? What values would be expected?
- Answer this by thinking: “What would happen if we took many samples of *10* subjects from this population?”
  – take a large number of samples of 10 subjects from the population
  – calculate the sample mean (x-bar) for each sample
  – make a histogram of the values of x-bar
  – examine the graphical display for shape, center, spread

BPS - 5th Ed. Chapter 11 11

Case Study

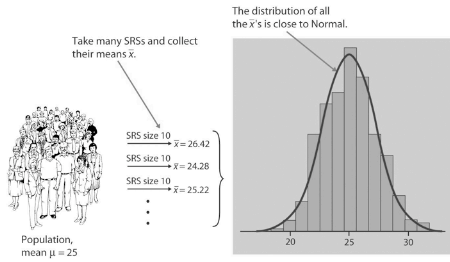
Does This Wine Smell Bad?

Mean threshold of all adults is μ=25 micrograms per liter, with a standard deviation of σ=7 micrograms per liter and the threshold values follow a bell-shaped (normal) curve.

Many (1000) samples of *n*=10 adults from the population were taken and the resulting histogram of the 1000 x-bar values is on the next slide.

BPS - 5th Ed. Chapter 11 12

Case Study
Does This Wine Smell Bad?



BPS - 5th Ed. Chapter 11 13

Mean and Standard Deviation of Sample Means

If numerous samples of size *n* are taken from a population with mean μ and standard deviation σ, then the *mean* of the sampling distribution of $\bar{X}$ is μ (the population mean) and the *standard deviation* is: $\sigma / \sqrt{n}$
(σ is the population s.d.)

BPS - 5th Ed. Chapter 11 14

Mean and Standard Deviation of Sample Means

- Individual observations have standard deviation σ, but sample means $\bar{X}$ from samples of size *n* have standard deviation $\sigma / \sqrt{n}$.
- **Averages are less variable than individual observations**.

BPS - 5th Ed. Chapter 11 15

Sampling Distribution of Sample Means

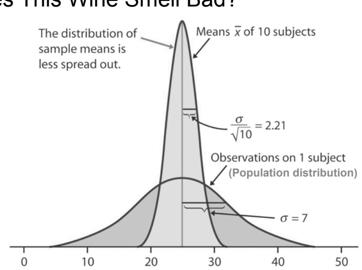
If individual observations have the N(μ, σ) distribution, then the sample mean $\bar{X}$ of *n* independent observations has the **N(μ, $\sigma / \sqrt{n}$)** distribution.

"If measurements in the population follow a Normal distribution, then so does the sample mean."

BPS - 5th Ed. Chapter 11 16

Case Study
Does This Wine Smell Bad?

Mean threshold of all adults is μ=25 with a standard deviation of σ=7, and the threshold values follow a bell-shaped (normal) curve.



BPS - 5th Ed. Chapter 11 17

Central Limit Theorem

If a random sample of size *n* is selected from ANY population with mean μ and standard deviation σ , then *when n is large* the sampling distribution of the sample mean $\bar{X}$ is approximately Normal:

$\bar{X}$ is approximately N(μ, $\sigma / \sqrt{n}$)

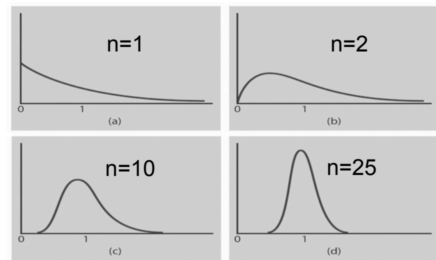
"No matter what distribution the population values follow, the sample mean will follow a Normal distribution if the sample size is large."

BPS - 5th Ed. Chapter 11 18

Central Limit Theorem: Sample Size

- How large must *n* be for the CLT to hold?
  - depends on how far the population distribution is from Normal
    - the further from Normal, the larger the sample size needed
    - a sample size of **25 or 30** is typically large enough for any population distribution encountered in practice
    - recall: if the population is Normal, any sample size will work ($n \geq 1$)

Central Limit Theorem: Sample Size and Distribution of *x-bar*



Statistical Process Control

- Goal is to make a process stable over time and keep it stable unless there are planned changes
- All processes have variation
- Statistical description of stability over time: the *pattern of variation* remains stable (does not say that there is *no* variation)

Statistical Process Control

- A variable described by the same distribution over time is said to be *in control*
- To see if a process has been disturbed and to signal when the process is *out of control*, **control charts** are used to monitor the process
  - distinguish natural variation in the process from additional variation that suggests a change
  - most common application: industrial processes

Example

- Testing a new drug
- Measure levels of certain analytes in blood
- Current practice:
- Measure "normal levels" of blood analytes in subject
- Administer drug and observe analytes levels
- A flag is raised when level reaches 40 (preset), or three times higher than normal levels (whichever is smaller)

- Does this make sense?

$\bar{x}$ Charts

- There is a true mean μ that describes the center or aim of the process
- Monitor the process by plotting the means (x-bars) of small samples taken from the process at regular intervals over time
- Process-monitoring conditions:
  - measure quantitative variable *x* that is Normal
  - process has been operating in control for a long period
  - know process mean μ and standard deviation σ that describe distribution of *x* when process is in control

$\overline{x}$ Control Charts

- Plot the means (x-bars) of regular samples of size n against time
- Draw a horizontal center line at μ
- Draw horizontal control limits at $\mu \pm 3\sigma/\sqrt{n}$
  - almost all (99.7%) of the values of x-bar should be within the mean plus or minus 3 standard deviations
- *Any x-bar that does not fall between the control limits is evidence that the process is out of control*

BPS - 5th Ed. Chapter 11 25

Case Study

Making Computer Monitors

Need to control the tension in millivolts (mV) on the mesh of fine wires behind the surface of the screen.

- Proper tension is 275 mV (target mean μ)
- When in control, the standard deviation of the tension readings is σ=43 mV

BPS - 5th Ed. Chapter 11 26

Case Study

Making Computer Monitors

Proper tension is 275 mV (target mean μ). When in control, the standard deviation of the tension readings is σ=43 mV.

Take samples of *n*=4 screens and calculate the means of these samples

- the control limits of the x-bar control chart would be

$$\mu \pm 3\frac{\sigma}{\sqrt{n}} = 275 \pm 3\left(\frac{43}{\sqrt{4}}\right) = 275 \pm 64.5$$

$$= \underline{210.5} \text{ and } \underline{339.5}$$

BPS - 5th Ed. Chapter 11 27

Case Study

Making Computer Monitors (data)

TABLE 10.1 Twenty control chart samples of mesh tension

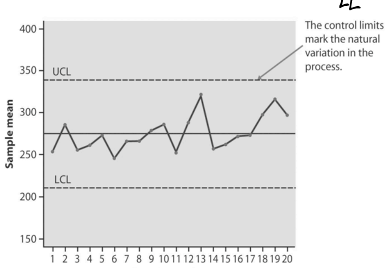
Sample	Tension measurements				$\overline{x}$
1	234.5	272.3	234.5	272.3	253.4
2	311.1	305.8	238.5	286.2	285.4
3	247.1	205.3	252.6	316.1	255.3
4	215.4	296.8	274.2	256.8	260.8
5	327.9	247.2	283.3	232.6	272.7
6	304.3	236.3	201.8	238.5	245.2
7	268.9	276.2	275.6	240.2	265.2
8	282.1	247.7	259.8	272.8	265.6
9	260.8	259.9	247.9	345.3	278.5
10	329.3	231.8	307.2	273.4	285.4
11	266.4	249.7	231.5	265.2	253.2
12	168.8	330.9	333.6	318.3	287.9
13	349.9	334.2	292.3	301.5	319.5
14	235.2	283.1	245.9	263.1	256.8
15	257.3	218.4	296.2	275.2	261.8
16	235.1	252.7	300.6	297.6	271.5
17	286.3	293.8	236.2	275.3	272.9
18	328.1	272.6	329.7	260.1	297.6
19	316.4	287.4	373.0	286.0	315.7
20	296.8	350.5	280.6	259.8	296.9

BPS - 5th Ed. Chapter 11 28

Case Study

Making Computer Monitors ($\overline{x}$ chart)

(In control)

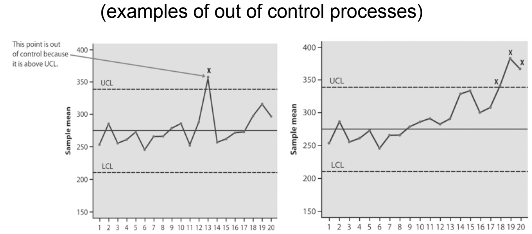


BPS - 5th Ed. Chapter 11 29

Case Study

Making Computer Monitors

(examples of out of control processes)



BPS - 5th Ed. Chapter 11 30