Section 6.4

Definition: A **statistic** is a numerical descriptive measure of a *sample*.

Definition: A **parameter** is a numerical descriptive measure of a *population*.

Definition: A **sampling distribution** is a probability distribution of a sample statistic based on all possible simple random samples of the *same* size from the same population.

Exercise 1. (a) Suppose the population consists of the following three numbers: $X = \{2, 5, 8\}$. Form the sampling distribution of sample means for samples of size $n = 2$. (In this exercise, the statistic is the mean.)

Remark: The samples must be chosen with replacement or else the independence* criteria won't be satisfied. (If we are sampling **without** replacement and the first number taken is 2, then it is not possible for the next number to be 2.)

There are 9 possible samples:

For each of the 9 samples, we calculate the mean:

Based on the above table, we make a frequency table:

The last table gives the sampling distribution for the sample means of size $n = 2$.

(b) Draw a probability histogram for the sampling distribution.

P(sample mean = 2) =

P(sample mean = 3.5) =

P(sample mean = 5) =

P(sample mean = 6.5) =

P(sample mean = 8) =

Exercise 2. Repeat the last exercise for a population of $X = \{1, 2, 3, 4, 5\}$.

(a) There are 25 possible samples:

For each of the 25 samples, we calculate the mean:

Sample	**Mean**	**Sample**	**Mean**	**Sample**	**Mean**
(1,1)	1	(2,5)	3.5	(4,4)	4
(1,2)	1.5	(3,1)	2	(4,5)	4.5
(1,3)	2	(3,2)	2.5	(5,1)	3
(1,4)	2.5	(3,3)	3	(5,2)	3.5
(1,5)	3	(3,4)	3.5	(5,3)	4
(2,1)	1.5	(3,5)	4	(5,4)	4.5
(2,2)	2	(4,1)	2.5	(5,5)	5
(2,3)	2.5	(4,2)	3	.	.
(2,4)	3	(4,3)	3.5	.	.

Based on the above table, we make a frequency table:

(b)

Exercise 3. Let's suppose we have a population of 10,000,000. How many possible samples of size 30 (with replacement) can we take from the population?

There are   possibilities for the first member of the sample. Likewise, there are   possibilities for the second member of the sample. There are   possibilities for each of the 30 members of the sample. Therefore, there are a total of   possible samples.

Remark: Each of the probability histograms resemble the normal curve. It turns out that if we took a larger sample (say $n = 10$), the histogram would bare a lot closer resemblance to the normal curve.

Homework

C problems

1. Let's say we have a population of 10 people. How many possible samples of size 2 (without replacement) are there? How many possible sample of size 2 (with replacement) are there? **Answers: 90, 100**

Remark: In typical sampling situations, we don't sample with replacement. However, for large enough populations, the sampling distribution of the means with replacement is similar to sampling distribution of the means without replacement. That's why we can still use the Central Limit Theorem.

2. Consider the population of $\{1, 3, 5\}$.
(a) Give the sampling distribution of the sum of the values for all samples of size $n = 2$.
(b) Construct a histogram of the sampling distribution in part(a). Comment on the shape of the histogram. (6.1.4)

3. Consider the population of $\{3, 4, 5, 7, 9, 12, 20\}$.
(a) How many samples of size $n = 2$ are there possible if the sampling is done with replacement? **Answer: 49**
(b) If the sampling is done without replacement, how many samples of size $n = 2$ are possible? **Answer: 42**
(c) How many samples of size $n = 3$ are there possible if the sampling is done with replacement? **Answer: 343**

4. Let's suppose we have a population of 9,000,000. How many possible samples of size 25 (with replacement) can we take from the population?

B problems

Section 6.4: 1, 3, 5

A problems

Section 6.4: 7, 9

Section 6.5

Central Limit Theorem: Draw an SRS of size n from any population with mean μ and finite standard deviation σ. When n is large, the sampling distribution of the sample mean $\bar{x}$ is approximately normal:

$$\bar{x} \text{ is approximately } N(\mu, \tfrac{\sigma}{\sqrt{n}}).$$

Practical Rules:

1. If the original population is *not normally distributed*, here is a common guideline: for $n > 30$, the distribution of the sample means can be approximated reasonably well by a normal distribution. (There are exceptions, such as populations with very non-normal distributions requiring sample sizes larger than 30, but such exceptions are relatively rare.) The distribution of sample means gets closer to a normal distribution as the sample size n becomes larger.

2. If the original population is *normally distributed*, then for *any* sample size n, the sample means will be normally distributed.

Notation: The mean of the sample means is denoted by $\mu_{\bar{x}}$. (So, $\mu_{\bar{x}} = \mu$.)

Notation: The standard deviation of the sample means is denoted by $\sigma_{\bar{x}}$. (So, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.)

Exercise 4. What is the distribution of the sample means (samples of size 36) from a population with $\mu = 5$ and $\sigma = 4$?

By the Central Limit Theorem, the distribution of the sample means is

Class Exercise 1. What is the distribution of sample means (samples of size 48) from a population with $\mu = 5$ and $\sigma = 4$?

In the next exercise, we find the probability of an event occurring using the Central Limit Theorem.

Exercise 5. For a sample of size $n = 36$, find the probability that the sample mean is less than 12.2 if $\mu = 12$ and $\sigma = 0.95$.

Let $\bar{X} =$ By the CLT, $\bar{X} \sim$ N(12, 0.95/6).

We are trying to find

Class Exercise 2. For a sample of size $n = 75$, find the probability of a sample mean being greater than 221 if $\mu = 220$ and $\sigma = 3.9$. **Answer: 0.0132**

The Central Limit Theorem is used again in the next exercise.

Exercise 6. A bottling company uses a filling machine to fill plastic bottles with a popular cola. The bottles are suppose to contain 300 milliliters (ml). In fact, the contents vary according to an unknown distribution with mean $\mu = 298$ ml and standard deviation $\sigma = 30$ ml. What is the probability that the mean contents of the bottles in a thirty six pack is less than 295 ml?

Remark: There are millions of bottles of cola. That means that there are an ENORMOUS number of samples of size 36. The question is asking what is the probability that one of these samples contains less than 295 ml.

Let X =

By the CLT, We are looking for

Class Exercise 3. The population mean annual salary for plumbers is \$44,700. A random sample of 42 plumbers is drawn from this population. What is the probability that the mean salary of the sample is less than \$44,000? Assume σ = \$1700. **Answer: 0.00381**

We use the Central Limit Theorem yet again in the next exercise.

Exercise 7. The time that a technician requires to perform preventive maintenance on an air conditioning unit is governed by an unknown distribution. The mean time is $\mu = 1$ hour and the standard deviation is $\sigma = 1$ hour. Your company operates 70 of these units. What is the probability that their average maintenance exceeds 50 minutes?

Let X =

By the CLT,

Remark: Note how in none of the previous exercises did we assume that the population was normally distributed. (Before this handout, we had always assumed that the population followed a normal distribution). This is why the Central Limit Theorem is so powerful.

Exercise 8. A bottling company uses a filling machine to fill plastic bottles with a popular cola. The bottles are suppose to contain 300 milliliters (ml). In fact, the contents vary according to a normal distribution with mean $\mu = 298$ ml and the standard deviation $\sigma = 30$ ml. What is the probability that the mean contents of the bottles in a six pack is less than 295 ml?

Remark: This differs from the previous bottling exercise in two ways:
(1) The contents
(2) The number of bottles

Let $X =$

By the 2nd practical rule,

Class Exercise 4. A manufacturer claims that the mean weight of its ice cream cartons is 10 ounces with a standard deviation of 0.5 ounce. Assume the weights are normally distributed. You test 12 cartons and find their mean weight is 10.21 ounces. Assuming that the manufacturer's claim is correct, what is the probability the mean of the sample is 10.21 ounces or more? **Answer: 0.0728**

Class Exercise 5. Your lumber company has bought a machine that automatically cuts lumber. The seller of the machine claims that the machine cuts lumber to a mean length of 8 feet (96 inches) with a standard deviation of 0.5 inch. Assume the lengths are normally distributed. You randomly select 12 boards. What is the probability that the mean length is between 95 inches and 98 inches? **Answer: 1**

Class Exercise 6. The average sales price of a single-family home in the United States is 306,258 dollars. You randomly select 12 single-family houses. What is the probability that the mean sales price is more than 280,000 dollars? Assume that the sales prices are normally distributed with a standard deviation of 44,000 dollars. **Answer: 0.981**

Homework

C problems

Section 6.5: 5(a), 5(b), 5(c), 13(a), 13(b), 15(a), 15(b), 15(c), 15(d), 17(a), 17(b), 19(a), 19(b), 19(c), 19(d)

B problems

Section 6.5: 1, 3, 5(d), 13(c), 15(e), 17(c), 17(d), 19(e)

A problems

Section 6.5: 9, 11, 21