

The Basic Practice of Statistics Ninth Edition

David S. Moore

William I. Notz

Chapter 15
Sampling Distributions

Lecture Slides

In Chapter 15, we cover ...

- Parameters and statistics
- Statistical estimation and the law of large numbers
- Sampling distributions
- The sampling distribution of \bar{x}
- The central limit theorem
- Sampling distributions and statistical significance

Parameters and statistics

- As we begin to use sample data to draw conclusions about a wider population, we must be clear about whether a number describes a sample or a population.
- A **parameter** is a number that describes the population. In practice, the value of a parameter is not known because we can rarely examine the entire population.
- A **statistic** is a number that can be computed from the sample data without making use of any unknown parameters. In practice, we often use a statistic to estimate an unknown parameter.

Parameters and statistics

- Remember ***p*** and ***s***: ***p*** parameters come from **populations** and ***s*** statistics come from **samples**.
 - We write μ (the Greek letter mu) for the mean of the population and σ (the Greek letter sigma) for the **standard deviation of the population**. We write \bar{x} (“x-bar”) for the **mean of the sample** and ***s*** for the **standard deviation of the sample**.
-

Statistical estimation

- The process of **statistical inference** involves using information from a sample to draw conclusions about a wider population.
- Different random samples yield different statistics. We need to be able to describe the **sampling distribution** of possible statistic values in order to perform statistical inference.
- We can think of a statistic as a **random variable** because it takes numerical values that describe the outcomes of the random sampling process. Therefore, we can examine its probability distribution using concepts we learned in earlier chapters.



The law of large numbers

- If \bar{x} is rarely exactly right and varies from sample to sample, why is it nonetheless a reasonable estimate of the population mean μ ?
- Here is one answer: If we keep taking larger and larger samples, the statistic \bar{x} is guaranteed to get closer and closer to the parameter μ .

LAW OF LARGE NUMBERS

- Draw observations at random from any population with finite mean μ . As the number of observations drawn increases, the mean \bar{x} of the observed values tends to get closer and closer to the mean μ of the population.
-

Sampling distributions

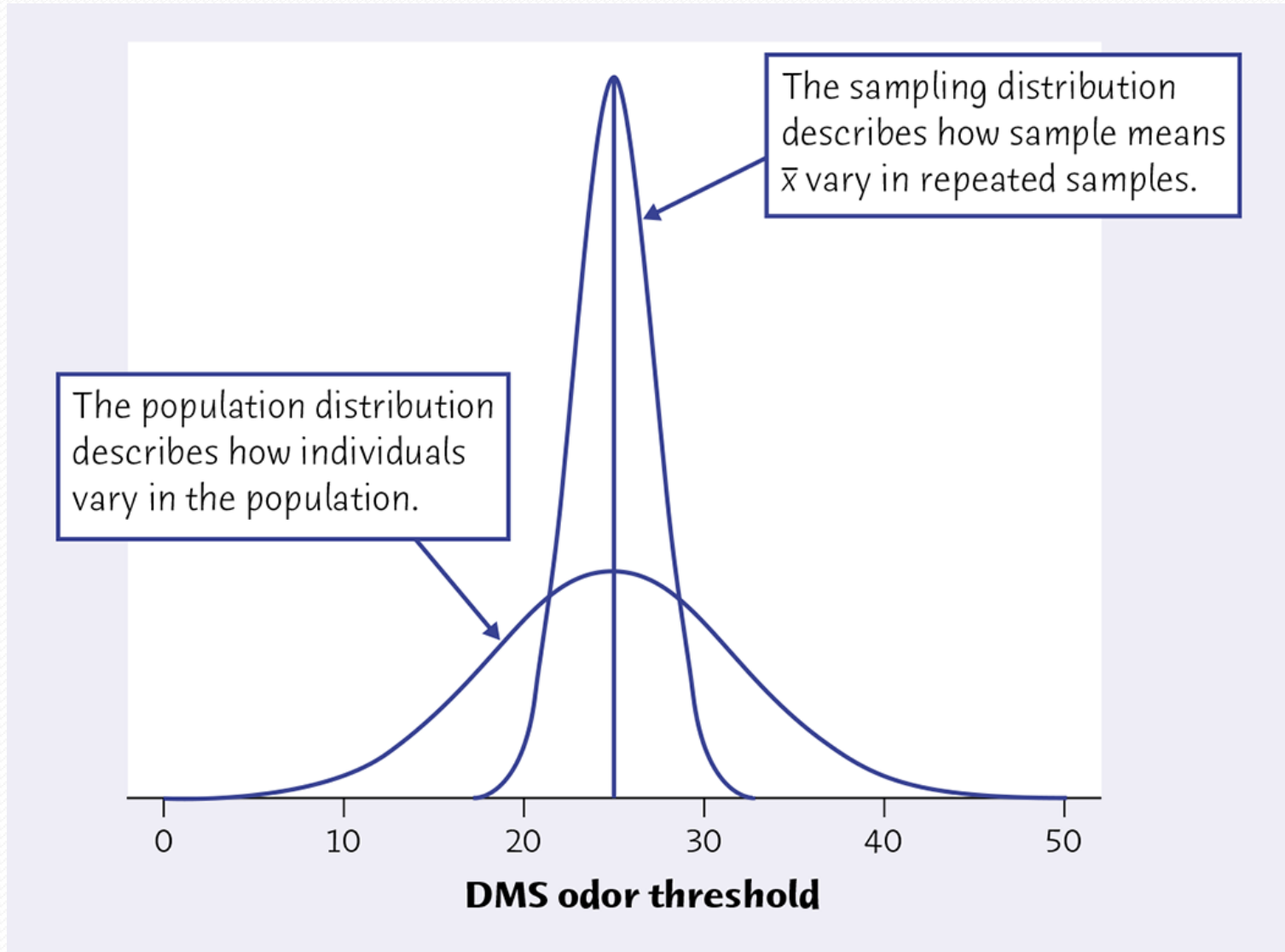
- The law of large numbers assures us that if we measure enough subjects, the statistic \bar{x} will eventually get very close to the unknown parameter μ .
- If we took every one of the possible samples of a certain size, calculated the sample mean for each, and graphed all of those values, we'd have a **sampling distribution**.
- When we use software to imitate chance behavior to carry out tasks such as exploring sampling distributions, this is called **simulation**.

Sampling distributions

- The **population distribution** of a variable is the distribution of values of the variable among all individuals in the population.
- The **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population.

Be careful: *The population distribution describes the individuals that make up the population. A sampling distribution describes how a statistic varies in many samples from the population.*

Population distributions vs. sampling distributions





Example using R - TBD

The sampling distribution of \bar{x} (part I)

- When we choose many SRSs from a population, the sampling distribution of the sample mean is centered at the population mean μ and is less spread out than the population distribution.

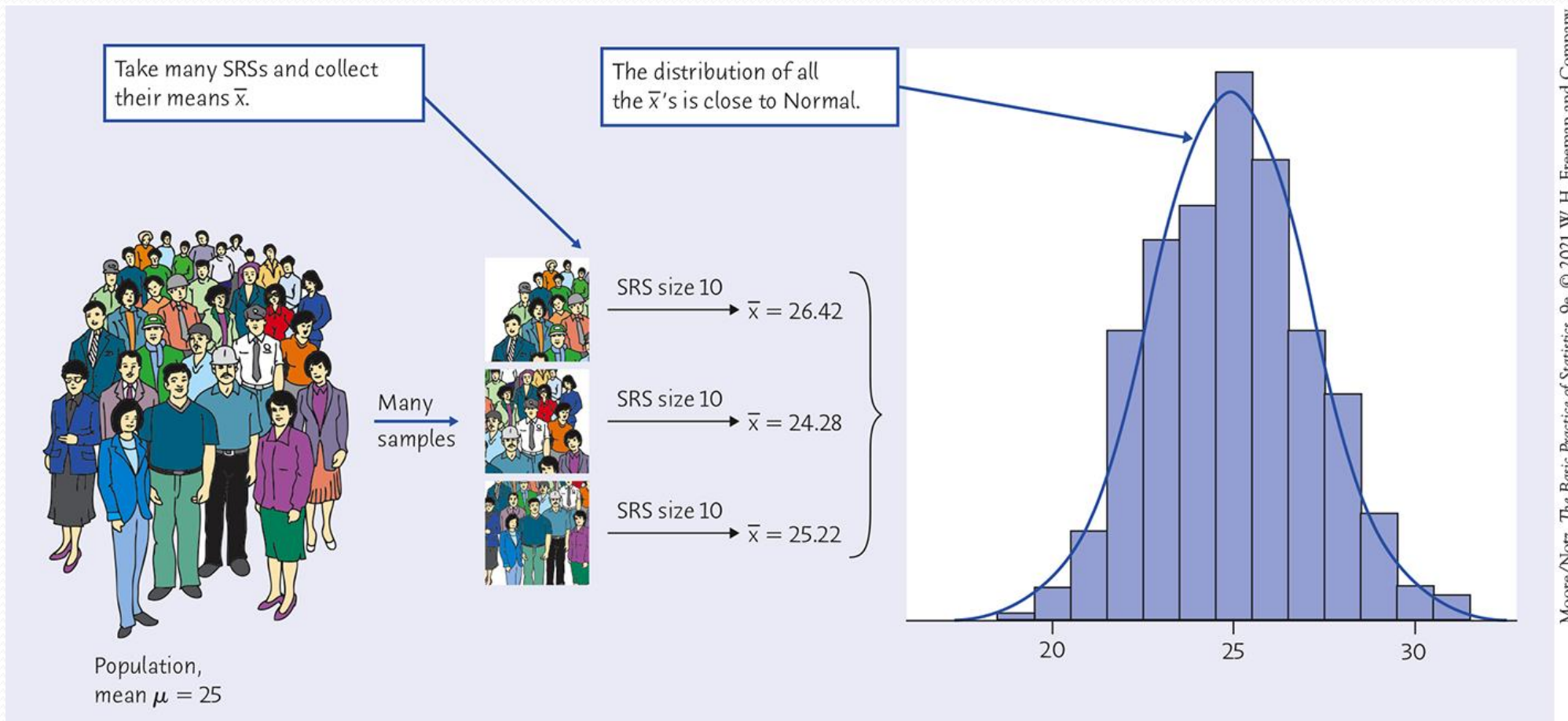
MEAN AND STANDARD DEVIATION OF A SAMPLE MEAN

- Suppose that \bar{x} is the mean of an SRS* of size n drawn from a large population with mean μ and standard deviation σ . Then the sampling distribution of \bar{x} has **mean μ** and **standard deviation σ/\sqrt{n}** .
- Because the mean of the statistic \bar{x} is always equal to the mean μ of the population (that is, the sampling distribution of \bar{x} is centered at μ), we say the statistic \bar{x} is an **unbiased estimator** of the parameter μ .

Note: On any particular sample, \bar{x} may fall above or below μ .

* SRS = Simple Random Sample

The sampling distribution of \bar{x} (illustrated)



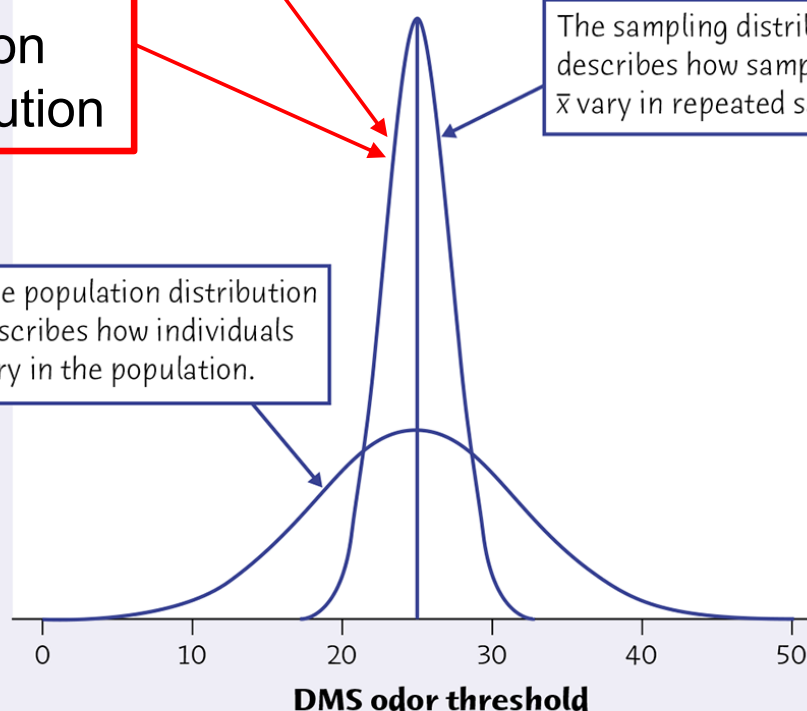
The sampling distribution of \bar{x} (part II)

- Because the standard deviation of the sampling distribution of \bar{x} is σ/\sqrt{n} , **the averages are less variable than individual observations.**

Sampling distribution has a smaller standard deviation than the population distribution

The population distribution describes how individuals vary in the population.

The sampling distribution describes how sample means \bar{x} vary in repeated samples.



The sampling distribution of \bar{x} (part II)

- Standard deviation of the sampling distribution of \bar{x} is σ/\sqrt{n} .
- Not only is the standard deviation of the distribution of \bar{x} smaller than the standard deviation of individual observations, **it gets smaller as we take larger samples**. The results of large samples are less variable than the results of small samples.

Note: *Although the standard deviation of the distribution of \bar{x} gets smaller, it does so at the rate of \sqrt{n} , not n . To cut the sampling distribution's standard deviation in half, for instance, you must take a sample four times as large, not just twice as large.*

The sampling distribution of \bar{x} (part III)

- We have described the center and variability of the sampling distribution of a sample mean \bar{x} , but not its shape. **The shape of the sampling distribution depends on the shape of the population distribution.**
- In one important case, there is a simple relationship between the two distributions: if the population distribution is Normal, then so is the sampling distribution of the sample mean.

The sampling distribution of \bar{x} (part III)

SAMPLING DISTRIBUTION OF A SAMPLE MEAN

- If individual observations have the $N(\mu, \sigma)$ distribution, then the sample mean \bar{x} of an SRS of size n has the $N(\mu, \sigma/\sqrt{n})$ distribution.
-

The central limit theorem

- Most population distributions are not Normal. What is the shape of the sampling distribution of sample means when the population distribution isn't Normal?
 - A remarkable fact is that as the sample size increases, the distribution of sample means changes its shape: It looks less like that of the population and more like a Normal distribution!
-

The central limit theorem

- Draw an SRS of size n from any population with mean μ and finite standard deviation σ . The **central limit theorem** says that when n is large, the sampling distribution of the sample mean \bar{x} is approximately Normal. That is,

$$\bar{x} \text{ is approximately } N\left(\mu, \sigma/\sqrt{n}\right)$$

- The central limit theorem allows us to use Normal probability calculations to answer questions about sample means from many observations, even when the population distribution is not Normal.
-

Central limit theorem (example, part I)

Based on service records from the past year, the time (in hours) that a technician requires to complete preventive maintenance on an air conditioner follows a distribution that is strongly right-skewed and whose most likely outcomes are close to 0 (see graph (a) in next slide). The mean time is $\mu = 1$ hour and the standard deviation is $\sigma = 1$.

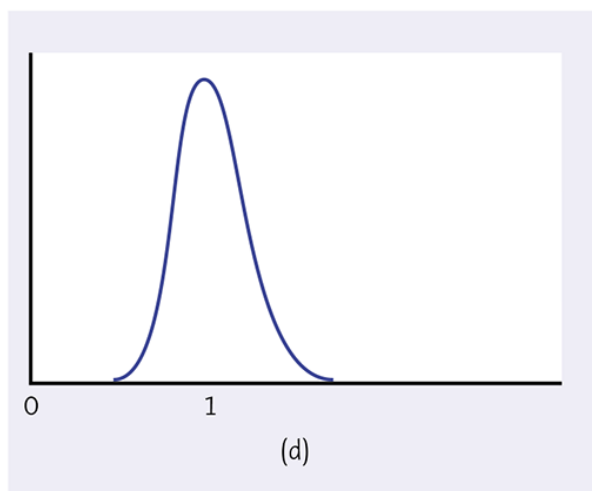
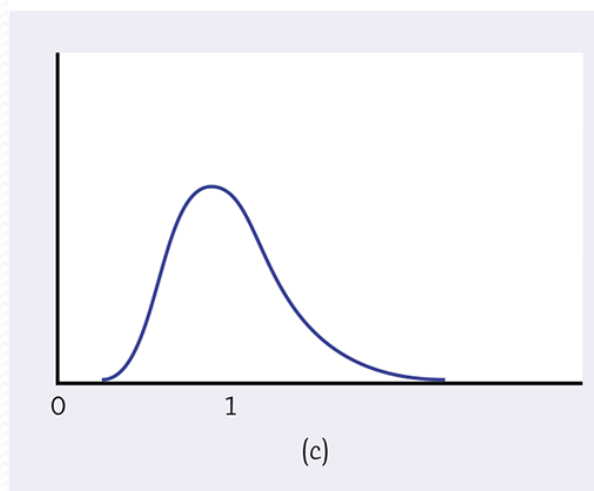
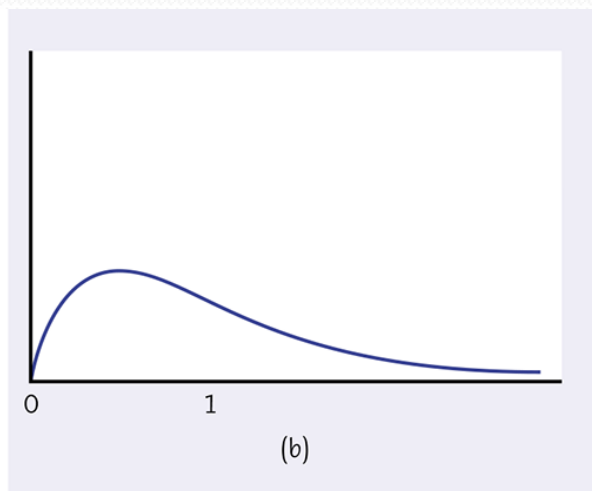
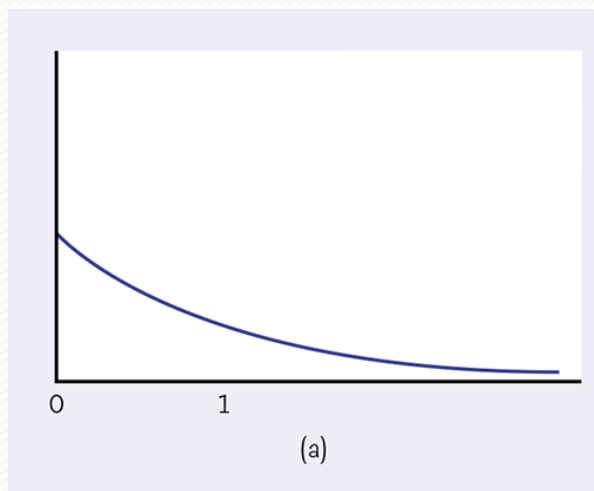
Your company will service an SRS of 70 air conditioners. You have budgeted 1.1 hours per unit. Will this be enough?

The central limit theorem states that the sampling distribution of the mean time spent working on the 70 units has

$$\mu_{\bar{x}} = \mu = 1,$$
$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{1}{\sqrt{70}} = 0.12,$$

and a Normal distribution shape.

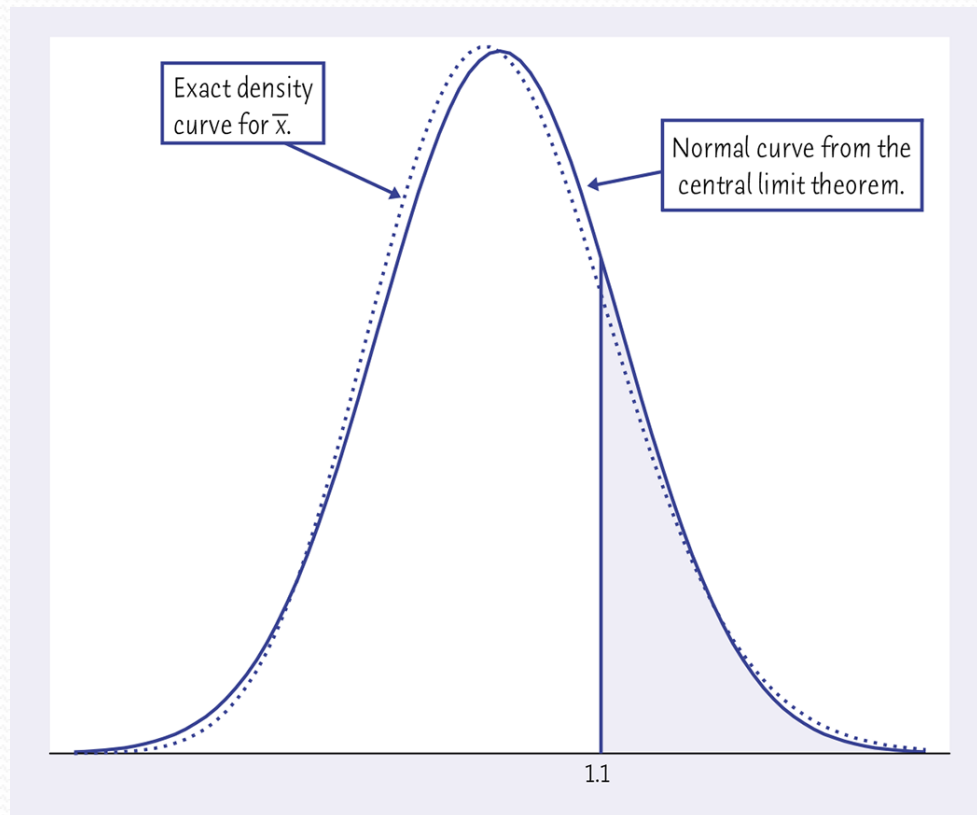
Central limit theorem (example, part II)



Central limit theorem (example, part III)

Your company will service an SRS of 70 air conditioners. You have budgeted 1.1 hours per unit. Will this be enough?

The sampling distribution of the mean time spent working is approximately $N(1, 0.12)$ because $n = 70 \geq 30$.



Moore/Notz, The Basic Practice of Statistics, 9e, © 2021 W. H. Freeman and Company

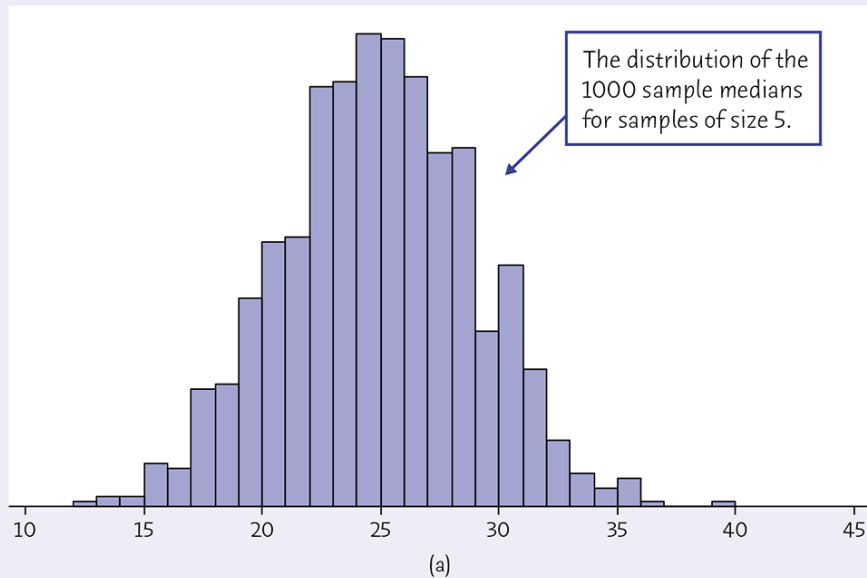
$$z = \frac{1.1 - 1}{0.12} = 0.83$$

$$P(\bar{x} > 1.1) = P(Z > 0.83) \\ = 1 - 0.7967 = 0.2033$$

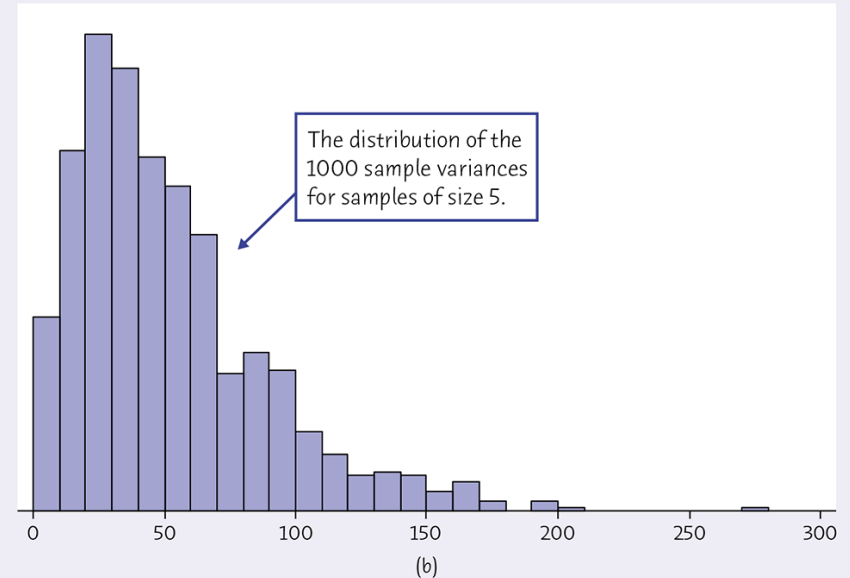
If you budget 1.1 hours per unit, there is a 20% chance that the technicians will not complete the work within the budgeted time.

Sampling distributions and statistical significance (part I)

- We have looked carefully at the sampling distribution of a sample mean.
- However, any statistic we can calculate from a sample will have a sampling distribution.



Moore/Notz, The Basic Practice of Statistics, 9e, © 2021 W. H. Freeman and Company



Moore/Notz, The Basic Practice of Statistics, 9e, © 2021 W. H. Freeman and Company

Sampling distributions and statistical significance (part II)

- The sampling distribution of a sample statistic is determined by the particular sample statistic we are interested in, the distribution of the population of individual values from which the sample statistic is computed, and the method by which samples are selected from the population.
- The sampling distribution allows us to determine the probability of observing any particular value of the sample statistic in another such sample from the population. An observed effect so large that it would rarely occur by chance is called **statistically significant**.
- Consider the second graph on the previous slide. We may decide, based on our observed set of 1000 samples, that because we see only 2 with variances above 200, this is a statistically significant event.

Example - statistically significant difference between means

TBD