What is a statistic?

- Function of observable random variables
- Use random sampling to observe these random variables
- The point is to use the statistic to estimate a population parameter
- Also called *point estimates*

Examples include:

- $\bar{X}$ for μ
- s^2 for σ^2

What is random sampling?

What is a *simple random sample* (assume sample size n and population size N)?

- Each subset of size n from the population is equally likely to be chosen as the sample
- There are $\binom{N}{n}$ possible samples, since they are chosen without replacement
- This is one of the most common methods of sampling (although it is not always possible)
- When we talk about "random sampling" in this course, we will generally be referring to "simple random samples"

Sampling distributions

- We're interested in the probability distributions for statistics that we use $\rightarrow$ sampling distributions
- Understanding the distribution for the statistic is critical to understanding how good your estimate (obtained using the statistic) is
- Remember that statistics are random variables, because we're dealing with random samples (and the sample determines the value of the statistic)

What kinds of information would we like to have about the distribution for a statistic?

- What is the shape of the distribution? (normal, symmetric, etc.)
- What is the expected value of the statistic?
- What is the variance of the statistic?

Mean and variance of $\bar{X}$

Imagine a normally distributed population with mean μ and variance σ^2. We have a random sample of size n. Show that the expected value of $\bar{X}$ is μ and that the variance of $\bar{X}$ is $\frac{\sigma^2}{n}$.

Sampling dist. for $\bar{X}$, normal pop., known σ

Now that we know the mean and variance of $\bar{X}$ for this normally distributed population, we'd like to know about other aspects of the distribution. Does $\bar{X}$ have a distribution that we're already familiar with (like the normal)? If not, what can we say about it?

- Any linear combination of normal random variables is normally distributed (Ref thm. 6.3, pp. 274-5)
- Since the sample mean $\bar{X}$ is a linear combination of normal random variables, $\bar{X}$ must be normally distributed
- Conclude that $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Example: gasoline prices

Assume the prices for gasoline are normally distributed with standard deviation \$0.10. Suppose that a random sample of 25 gasoline stations is selected.

What is the probability that the simple random sample will provide a sample mean within 3 cents, \$0.03, of the population mean?

Sampling dist. for $\bar{X}$, normal pop., unknown σ

Previously, we assumed that we knew the population standard deviation σ. However, the more common case is that σ is unknown, so we estimate it using the sample standard deviation, which is given by

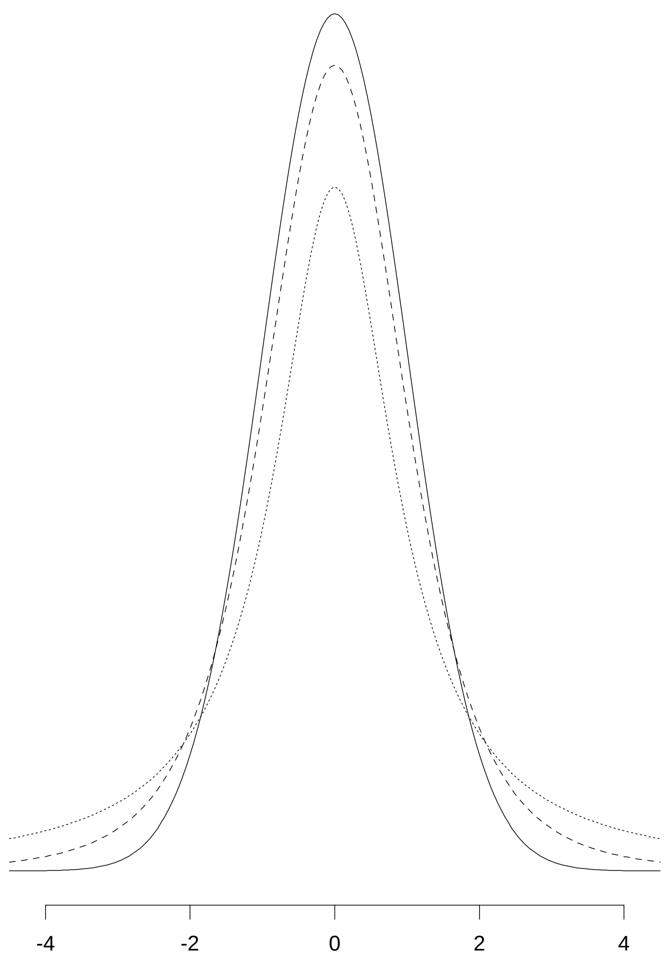
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- When we use $\frac{s^2}{n}$ as an estimate for $\frac{\sigma^2}{n}$, we can no longer assume that the sampling distribution for $\bar{X}$ is normal.
- When s is substituted for σ, $\bar{X}$ has a *Student's t* distribution with $n-1$ degrees of freedom (d.f.)

How are the t and normal distributions different?

- Both are symmetric and mound-shaped, but the t distribution has more area in the tails and shorter at the center
- The t distribution has a degrees of freedom (d.f.) ν; as $\nu \to \infty$, the t distribution converges to the normal
- Standardized table for t includes d.f. as rows and gives values for certain key quantiles in columns
- Standard t distribution has mean 0 and variance $\frac{\nu}{\nu-2} > 1$

How does the t dist. look?



Examples using the t table

- $P(1.476 < T^5 < 2.015)$

- $P(T^8 < -2.5)$

- What value marks the 99th percentile of the t distribution with 10 d.f.?

Example: gasoline prices (cont.)

Now assume the prices for gasoline are normally distributed *unknown* standard deviation. Suppose that a random sample of 25 gasoline stations is selected; the sample standard deviation is \$0.10.

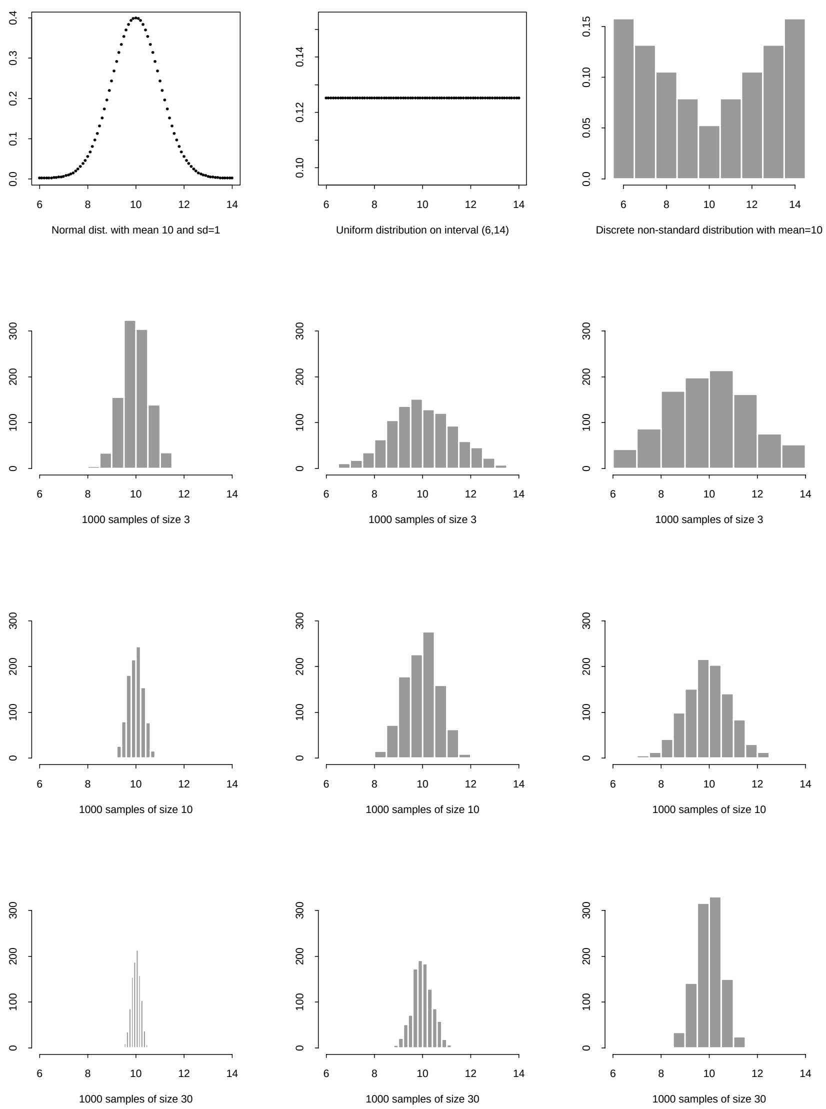
What is the probability that the simple random sample will provide a sample mean within 3 cents, \$0.03, of the population mean?

Central Limit Theorem

What if the population from which we are drawing the sample doesn't have a normal distribution?

- Often, we have no idea how the population is distributed
- Central Limit Theorem (CLT) tells us the as $n \to \infty$, the distribution of $\bar{X}$ converges to N(μ, $\frac{\sigma^2}{n}$)
- The closer to normally distributed the population is, the smaller the size that n needs to be before we say that $\bar{X}$ has a normal distribution
- Generally speaking, in most cases sample size $n > 30$ is big enough

Visual: Central Limit Theorem



Example using the CLT

Standardized test scores for a certain test are assumed to have variance 100 and mean 100. What is the probability that a sample of 25 will yield a sample mean of 105 or more?

Another example using the CLT

A freight elevator lifts 520 pounds. The average package is 10 pounds; standard deviation is 5 pounds. What is the probability the elevator will be overloaded if 49 packages are placed in the elevator?