Review of Central Limit Theorem and Confidence Intervals

Central Limit Theorem (CLT):

1. Consider all $M = \binom{N}{n}$ sample means, $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_M$, of size n from a population of size N. The mean of the population is equal to the mean of the sample means ($\mu_{\bar{x}} = \mu$) and the standard deviation of the sample means (called the standard error of the means), $\sigma_{\bar{x}}$, can be found by using the following formula: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-1}{N-n}}$. $\sqrt{\frac{N-1}{N-n}}$ is called the finite population correction factor.
2. If $n \geq 30$ and N is "large" then the distribution of sample means, $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_M$, is approximately normal, and $\sqrt{\frac{N-1}{N-n}}$, the finite population correction factor is approximately 1. Thus, we may assume $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Consider the population of size N=5. X: 19, 25, 40, 52, 63

1. Find μ
2. Find the population standard deviation, σ
3. Find $M = \binom{N}{n}$, where N=5 and n=3. M is the number of samples of size n one can obtain from a population of size N.
4. List all M samples of size n from our population.

{ , , }	{ , , }	{ , , }	{ , , }	{ , , }
{ , , }	{ , , }	{ , , }	{ , , }	{ , , }

5. Find the mean of each of the M samples; that is, find $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M$.

---

6. Find $\mu_{\bar{x}}$. Is $\mu = \mu_{\bar{x}}$?
7. Find the standard error of the mean, $\sigma_{\bar{x}}$.

   Find the finite population correction factor, $c = \sqrt{\frac{N-n}{N-1}}$.
8. Refer to problem #2 then find $c\frac{\sqrt{\sigma}}{n}$. Does this product equal to the standard error of the mean found in problem #7?
9. If N were 10,000 and n where 100 then what would be the value of c?

Suppose we roll *one* die 1,000 times and record the outcome of each roll, which can be the number 1, 2, 3, 4, 5, or 6.

Figure 1 shows a histogram of outcomes. All six outcomes have roughly the same relative frequency, because the die is equally likely to land in each of the six possible ways. That is, the histogram shows a (nearly) *uniform distribution*

It turns out that the distribution in Figure 1 has a mean of 3.41 and a standard deviation of 1.73.

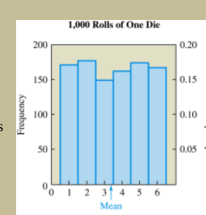


Figure 1 Frequency and relative frequency distribution of outcomes from rolling one die 1,000 times.

Now suppose we roll *two* dice 1,000 times and record the *mean* of the two numbers that appear on each roll. To find the mean for a single roll, we add the two numbers and divide by 2.

Figure 2 shows a typical result. The most common values in this distribution are the central values 3.0, 3.5, and 4.0. These values are common because they can occur in several ways.

The mean and standard deviation for this distribution are 3.43 and 1.21, respectively.

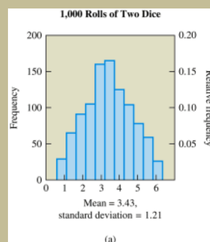


Figure 2 Frequency and relative frequency distribution of sample means from rolling two dice 1,000 times.

Suppose we roll *five* dice 1,000 times and record the mean of the five numbers on each roll. A histogram for this experiment is shown in Figure 3.

Once again we see that the central values around 3.5 occur most frequently, but the spread of the distribution is narrower than in the two previous cases.

The mean and standard deviation are 3.46 and 0.74, respectively.

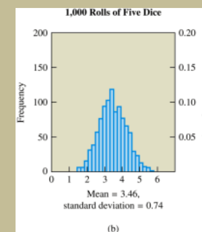


Figure 3 Frequency and relative frequency distribution of sample means from rolling five dice 1,000 times.

If we further increase the number of dice to *ten* on each of 1,000 rolls, we find the histogram in Figure 4, which is even narrower.

In this case, the mean is 3.49 and standard deviation is 0.56.

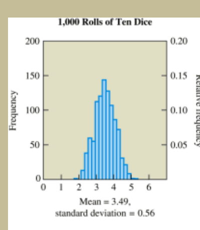


Figure 4 Frequency and relative frequency distribution of sample means from rolling ten dice 1,000 times.

Table 5.2 shows that as the sample size increases, the mean of the distribution of means approaches the value 3.5 and the standard deviation becomes smaller (making the distribution narrower).

Table 5.2 Summary of Dice Rolling Experiments

Number of dice rolled each time	Mean of the distribution of means	Standard deviation of the distribution of means
1	3.41	1.73
2	3.43	1.21
5	3.46	0.74
10	3.49	0.56

More important, the distribution looks more and more like a normal distribution as the sample size increases.

Confidence Intervals

A **point estimate** is the value of a statistic that estimates the value of a parameter.

For example, the sample mean, $\bar{x}$, is a point estimate of the population mean μ.

Parallel Example 1: *Computing a Point Estimate*

Pennies minted after 1982 are made from 97.5% zinc and 2.5% copper. The following data represent the weights (in grams) of 17 randomly selected pennies minted after 1982. Assume we know that σ=0.02 grams.

2.46 2.47 2.49 2.48 2.50 2.44 2.46 2.45 2.49

2.47 2.45 2.46 2.45 2.46 2.47 2.44 2.45

Treat the data as a simple random sample. **The distribution of the weights of pennies is normal; thus, the sample size does not have to be at least 30 for the sample means to be normally distributed.** Estimate the population mean weight of pennies minted after 1982.

The sample mean is

$$\bar{x} = \frac{2.46 + 2.47 + \cdots + 2.45}{17} \approx 2.464$$

The point estimate of μ is 2.464 grams.

The shape of the distribution of all possible sample means will be normal, provided the population is normal **or** approximately normal, if the sample size is large (n≥30), with

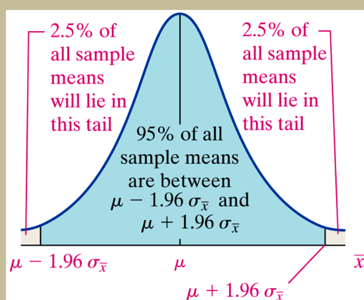
- mean $\mu_{\bar{x}} = \mu$
- and standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.02}{\sqrt{17}} = .0048507$$

95% Confidence Interval

1.96 = invNorm(.975)

-1.96=invNorm(.025)



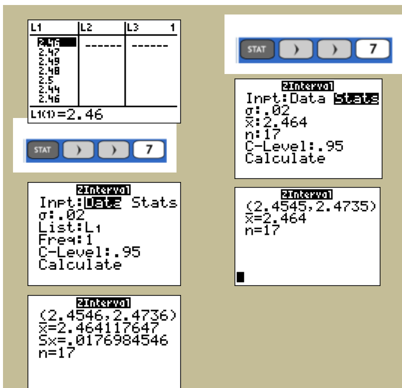
Constructing a (1- α)·100% Confidence Interval for μ, σ Known

Suppose that a simple random sample of size n is taken from a population with unknown mean, μ, and known standard deviation σ. A (1-α)·100% confidence interval for μ is given by

Lower Bound: $\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	Upper Bound: $\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

where $z_{\alpha/2}$ is the critical Z-value.

Note: The sample size must be large (n≥30) or the population must be normally distributed.



9-15

Confidence Intervals about a Population Mean When the Population Standard Deviation is Unknown

Comparing the Standard Normal Distribution to the t-Distribution Using Simulation

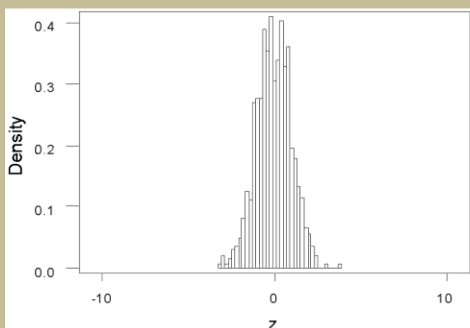
a) Obtain 1,000 simple random samples of size n=5 from a normal population with μ=50 and σ=10.

b) Determine the sample mean and sample standard deviation for each of the samples.

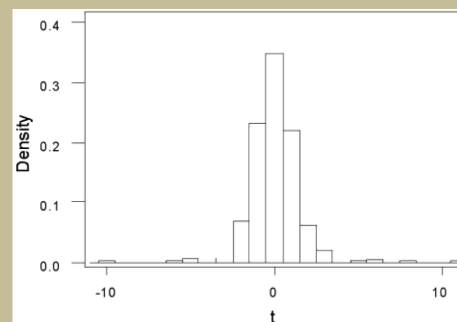
c) Compute $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ and $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ for each sample.

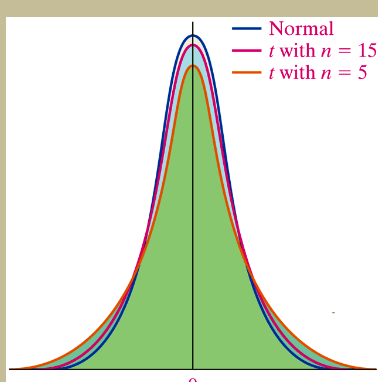
d) Draw a histogram for both z and t.

Histogram for *z*



Histogram for *t*





Constructing a (1-α)·100% Confidence Interval for μ, σ Unknown

Suppose that a simple random sample of size *n* is taken from a population with unknown mean μ and unknown standard deviation σ. A (1-α)·100% confidence interval for μ is given by

Lower bound: $\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

Upper bound: $\bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

Note: The interval is exact when the population is normally distributed. It is approximately correct for nonnormal populations, provided that *n* is large enough.

Constructing a Confidence Interval about a Population Mean

The pasteurization process reduces the amount of bacteria found in dairy products, such as milk. The following data represent the counts of bacteria in pasteurized milk (in CFU/mL) for a random sample of 12 pasteurized glasses of milk. Data courtesy of Dr. Michael Lee, Professor, Joliet Junior College.

Construct a 95% confidence interval for the bacteria count.

Sample	CFU/mL
1	8.96
2	1.95
3	9.22
4	9.41
5	4.14
6	1.79
7	3.63
8	3.49
9	6.95
10	15
11	12
12	0.38

NOTE: Each observation is in tens of thousand. So, 9.06 represents 9.06 x 10^4.

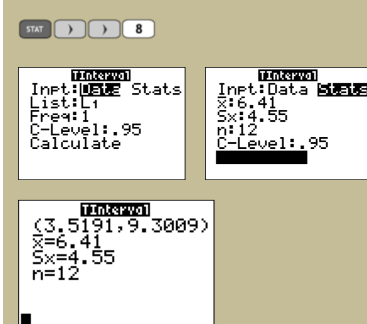
9-22

- $\bar{x} = 6.41$ and $s = 4.55$
- $\alpha = 0.05, \quad n = 12, \quad \text{so } t_{\frac{0.05}{2}} = 2.201$

Lower bound: $6.41 - 2.201 \cdot \dfrac{4.55}{\sqrt{12}} = 3.52$

Upper bound: $6.41 + 2.201 \cdot \dfrac{4.55}{\sqrt{12}} = 9.30$

The 95% confidence interval for the mean bacteria count in pasteurized milk is (3.52, 9.30).



9-24

Confidence Intervals for a Population Proportion

A point estimate is an unbiased estimator of the parameter. The point estimate for the population proportion is $\hat{p} = \frac{x}{n}$ where *x* is the number of individuals in the sample with the specified characteristic and *n* is the sample size.

Sampling Distribution of $\hat{p}$

For a simple random sample of size *n*, the sampling distribution of $\hat{p}$ is approximately normal with mean $\mu_{\hat{p}} = p$ and standard deviation

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ provided } np \text{ and } n(1-p) \geq 15$$

NOTE: We also require that each trial be independent when sampling from finite populations.

Constructing a Confidence Interval for a Population Proportion

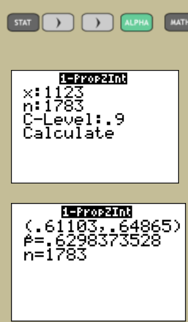
In July of 2008, a Quinnipiac University Poll asked 1783 registered voters nationwide whether they favored or opposed the death penalty for persons convicted of murder. 1123 were in favor.
Obtain a 90% confidence interval for the proportion of registered voters nationwide who are in favor of the death penalty for persons convicted of murder.

Solution

- $\hat{p} = 0.63$
- α=0.10 so $z_{\alpha/2}$=$z_{0.05}$=1.645
- Lower bound:
- Upper bound:

$$0.63 - 1.645\sqrt{\frac{0.63(1-0.63)}{1783}} = 0.61$$

$$0.63 + 1.645\sqrt{\frac{0.63(1-0.63)}{1783}} = 0.65$$



Sample size needed for a specified margin of error, E, and level of confidence (1-α):

$$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$$

Problem: The formula uses $\hat{p}$ which depends on *n,* the quantity we are trying to determine!

A sociologist wanted to determine the percentage of residents of America that only speak English at home. What size sample should be obtained if she wishes her estimate to be within 3 percentage points with 90% confidence assuming she uses the 2000 estimate obtained from the Census 2000 Supplementary Survey of 82.4%?

Solution

- *E*=0.03
- $z_{\alpha/2} = z_{0.05} = 1.645$
- $\hat{p} = 0.824$
- $n = 0.824(1-0.824)\left(\frac{1.645}{0.03}\right)^2 = 436.04$

We round this value up to 437. The sociologist must survey 437 randomly selected American residents.