Chapter 11, Secs. 1-6

Inferential Statistics with Theoretical Distributions

Sampling Dist'ns of Statistics

- We used statistical data to form confidence intervals using simulation and find p-values for hypothesis tests.
- We "looked" at the data to summarize it.
- That is, we are looking at an appropriate sampling dist'n of the sample statistic to find

  Middle 90%

  or

  " the probability that we get data as extreme or more extreme than our data if Ho is true."

Transition to Theoretical Dist'ns

- Software such as StatKey allows us to look at those approximations of the sampling dist'n **from the data**, **without making assumptions about the pop'n dist'n.**
- The Central Limit Theorem encourages us to look at the sampling dist'n of the sample mean from **the theory, because asymptotically, it relies very little on assumptions about the pop'n dist'n.**

Other Theoretical Dist'ns

- We can, of course, find sampling dist'ns of statistics, using theoretical results from probability, in other ways besides the Central Limit Theorem.
- The derivations of those are either
  - beyond the scope of this course (so we take them as given, such as chi-squared and F dist'ns)
  - or not typically used in applied statistics and so not in the required topics of this course.
  - (Additional information is on the course page).

End of section

Theory: Central Limit Theorem

- What do you remember that it says?
- Do the following pictures help you remember?

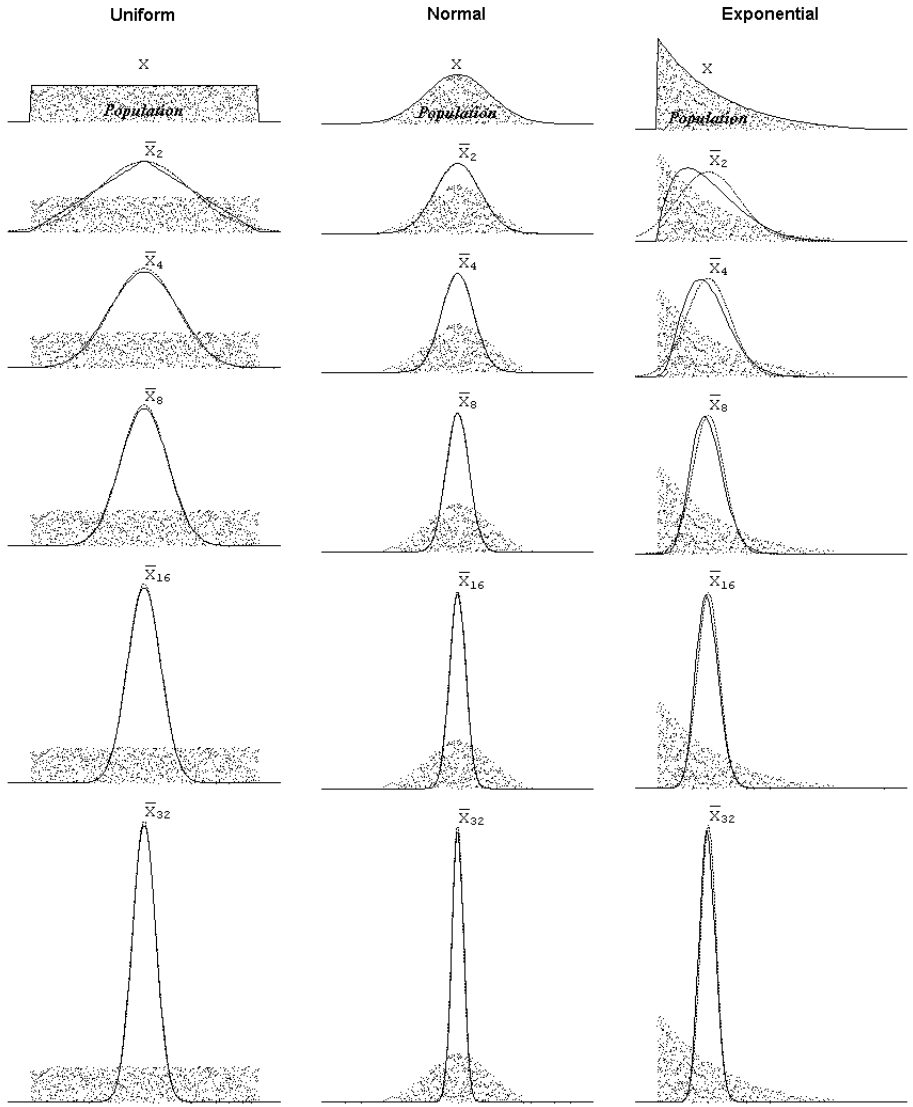
(Links to these are on the course page.)

Illustrations of the Central Limit Theorem

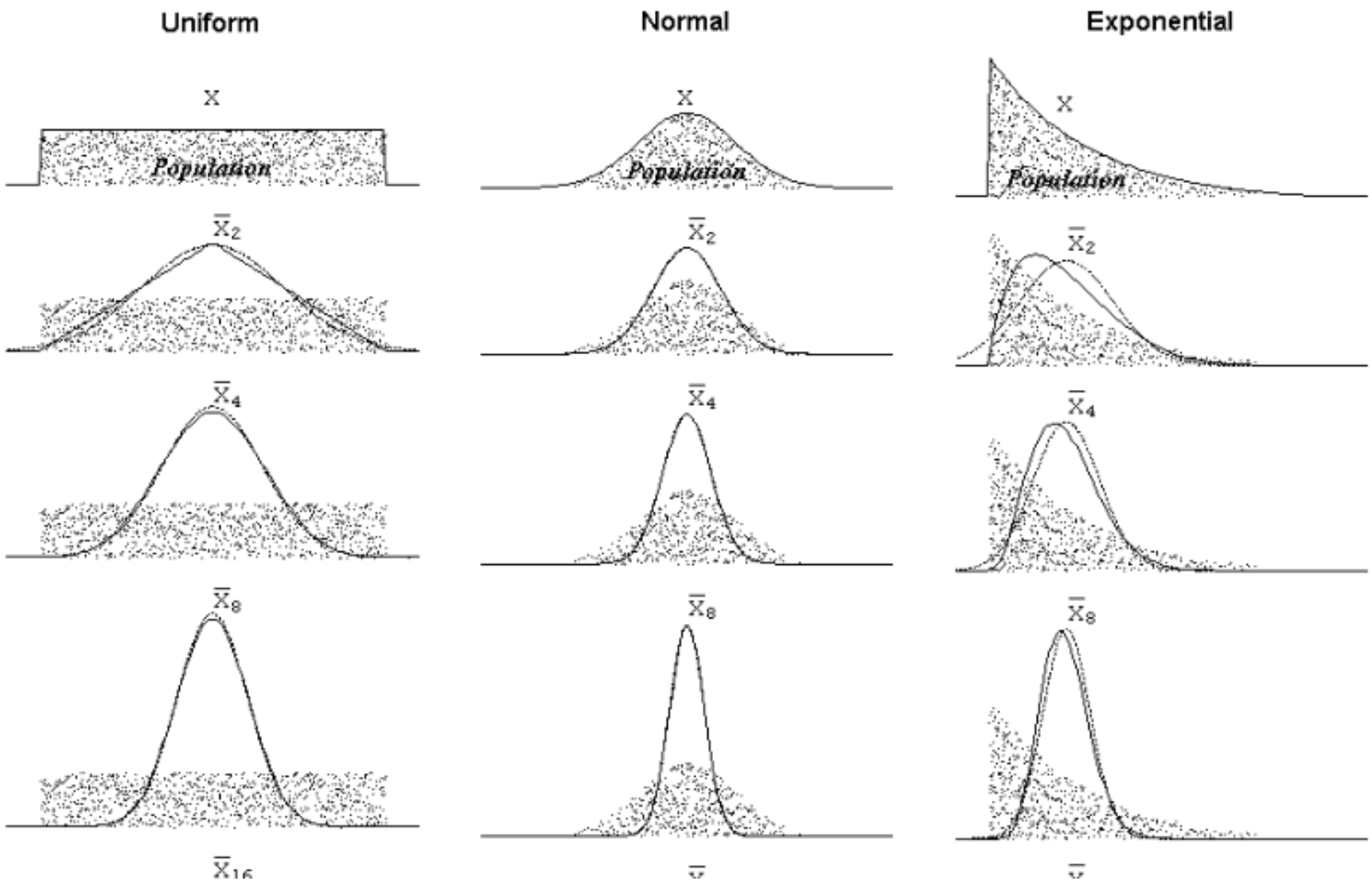
The following three slides illustrate,
for three different populations,
the sampling dist'n of the sample mean
for six different sample sizes: 1, 2, 4, 8, 16, 32.

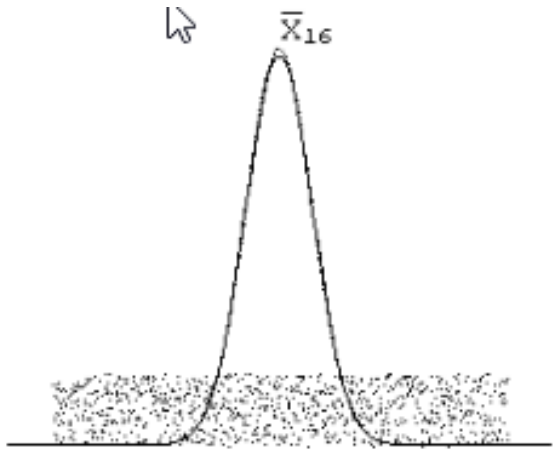
The results for the six sample sizes are graphed on the same scale, illustrating the same center for each and the smaller standard deviation for each.

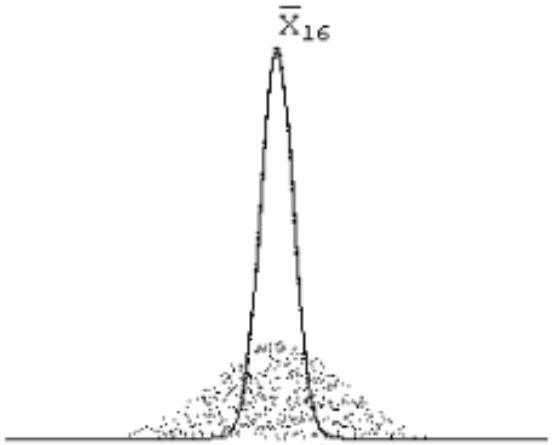
Distributions of means for samples of increasing size from various distributions

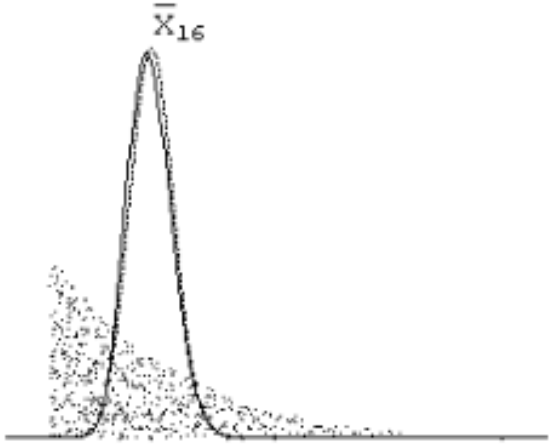


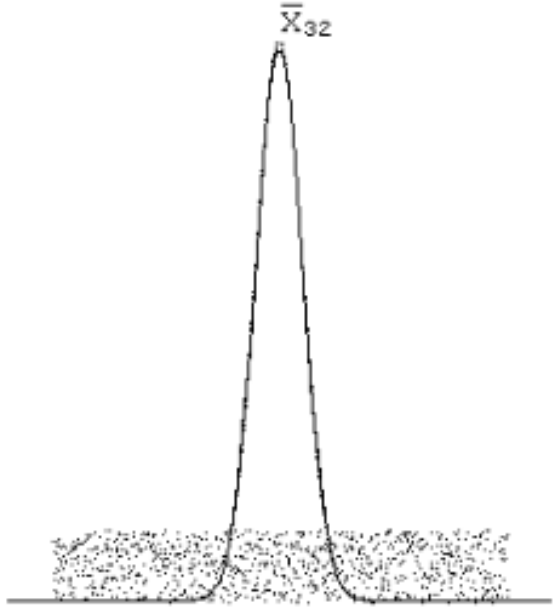
Distributions of means for samples of increasing size from various distributions

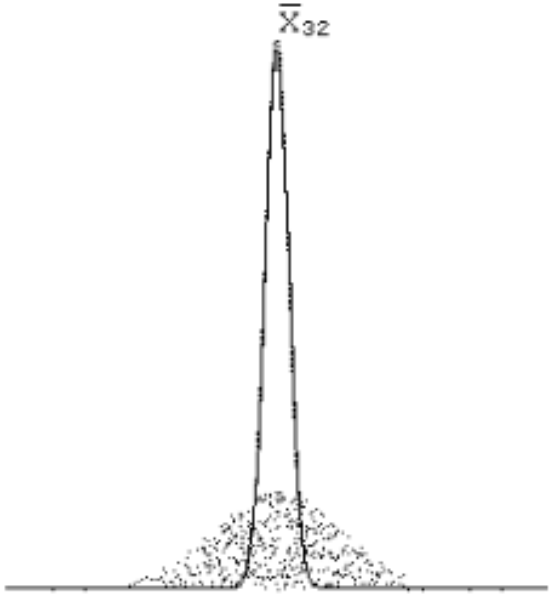


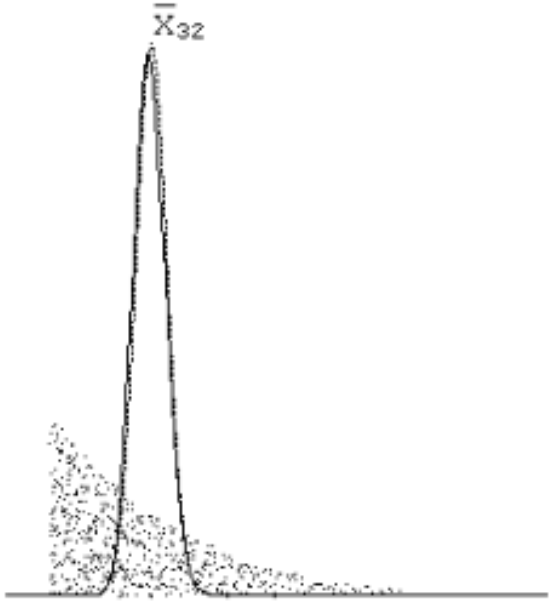

$\bar{X}_{16}$


$\bar{X}_{16}$


$\bar{X}_{16}$


$\bar{X}_{32}$


$\bar{X}_{32}$

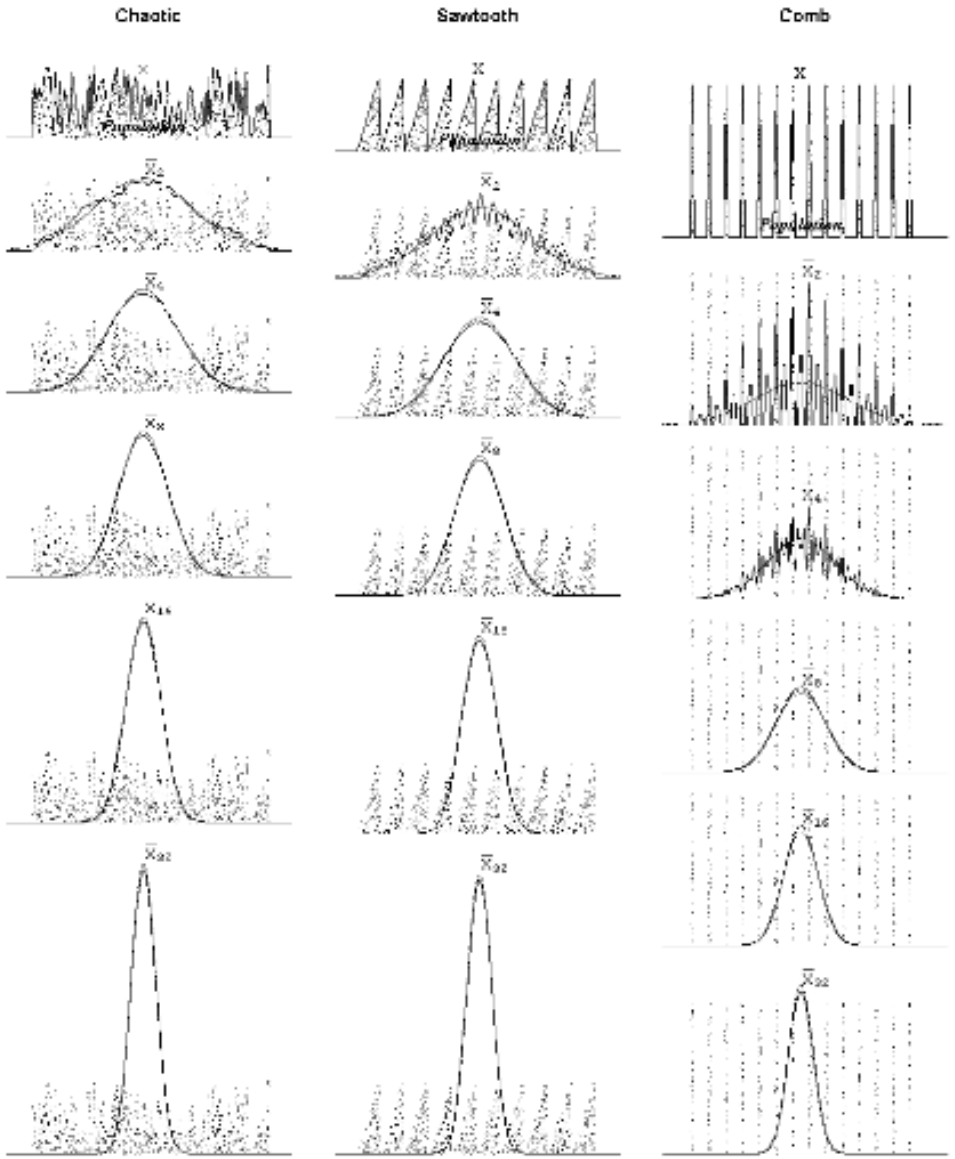

$\bar{X}_{32}$

Other illustrations of the Central Limit Theorem

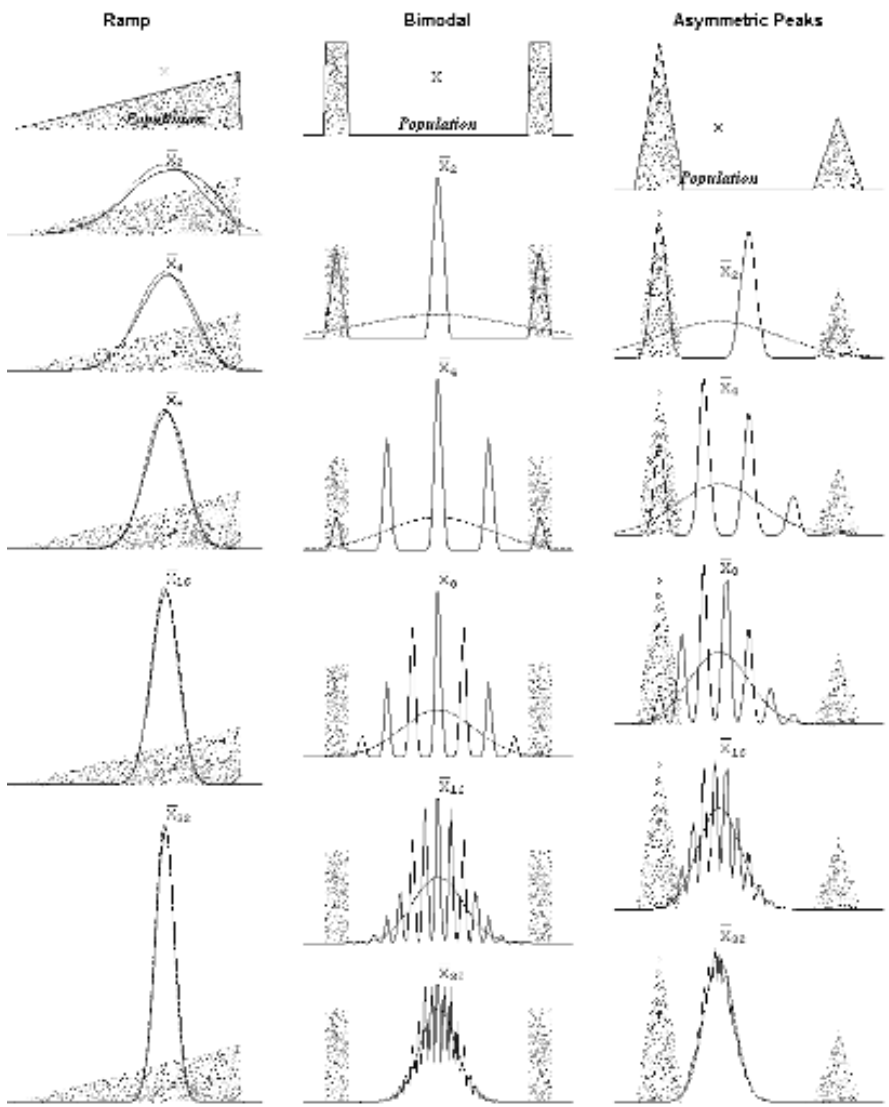
The next two slides illustrate the the sampling dist'n of the sample mean, for samples of sizes 1, 2, 4, 8, 16, 32

from more (unusual) different shapes for the population dist'n.

Distributions of means for samples of increasing size from various distributions



Distributions of means for samples of increasing size from various distributions



Several views of the same illustration of CLT

- The next slide is from a different applet, showing the dist'n of the sample mean for a sample of size 30, from a pop'n with an exponential dist'n.
- Two different "scalings" of the graph are shown:
  - On the left, to illustrate the decrease in spread from original pop'n.
  - On the right, to illustrate the difference in shape from the original pop'n dist'n.)

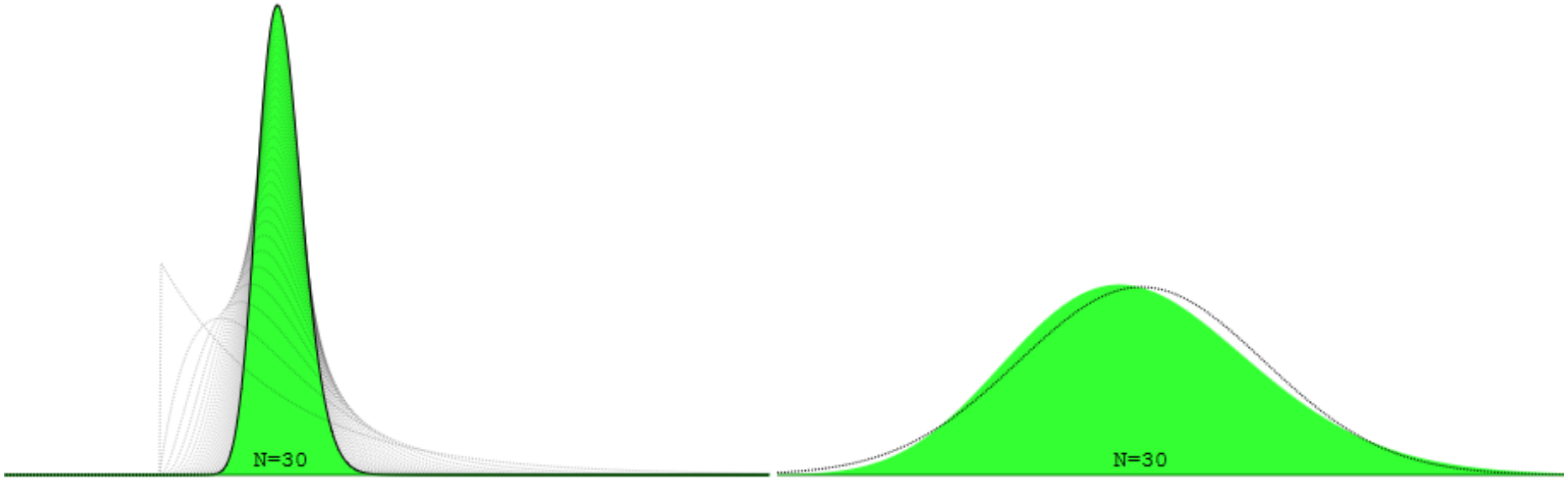
○ Uniform ○ Normal ◉ Exponential ○ Binomial ○ Notch ○ Overlap ○ Triangle ○ Triangles ○ Comb

Sample sizes: ○ 5 ○ 10 ○ 15 ○ 20 ○ 25 ◉ 30 ○ 35 ○ 40 ○ 50 ○ 75 ○ 100 or list: [ ] (e.g., 2,5,10 or 1:50)

Display Height: 300 Width: 500 ☑ Compare shapes ☐ Static image

Exponential distribution: probability distributions of the sample means for increasing sample sizes

Distribution shapes, rescaled to correct for sample size (the dotted line is the limiting normal distribution)



An open-source site -- Problems/suggestions? Notify hunter@ellinger.org

Home

Sampling Dist'ns related to means

- There are several sampling dist'ns associated with estimating the pop'n mean.
- The mathematical derivations start by assuming the pop'n data is normally distributed, but then various mathematical statistics results, mostly focused on the Central Limit Theorem, allow us to extend our work to data from pop'ns that fit the requirements for the CLT (finite mean and variance.)
- We can, and do, **look at the theoretical sampling dist'ns of statistics that we learn about**.
  But, because our usual methods of finding confidence intervals and p-values from them is to use formulas, we often just think in terms of the formulas rather than the overall picture of the sampling dist'n.

Visualizing Sampling Dist'ns of Statistics

- It is important to "think with" pictures of our sampling dist'ns even after we have the theoretical methods available to describe the important aspects of them with various formulas.
- It is also important to "think with" the patterns that the formulas give us as well as with the pictures.

Formulas based on the Central Limit Theorem

- The document of formulas provided on the course web page is one attempt at an "organization" of these ideas so that you

  don't try to think of them all as numerous different "facts" to learn/memorize.

- The document is too long to conveniently show here. It's important for you to print or save it and use it when you need to use the usual applied statistics techniques.
- The document may change somewhat before these lectures are redone. Be assured that the current version is the best we have at the time you're looking at it.

Ilustrations of the Formula Sheets

- The next three slides illustrate the information given about each method.
- The fourth slide has a list of the different methods addressed.

Type of inference	Parameter (or question for HT)	Sample statistic and theoretical sampling dist'n of test statistic	Conditions needed to use theoretical sampling dist'n of test statistic
Inference on one mean	μ	$\bar{X}$ t dist'n	Dist'n normal or CLT applies, meaning $n \geq 30$ approximately
Inference on one proportion	p	$\hat{p}$ normal dist'n	$np \geq 10$ AND $n(1-p) \geq 10$
Inference on two means	$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$ t dist'n	In EACH group, Dist'n normal or CLT applies, meaning $n \geq 30$ approximately.
Inference on two proportions	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$ normal dist'n	In EACH group: $np \geq 10$ AND $n(1-p) \geq 10$

Type of inference	**Sample statistic**	**Test statistic** SE is "standard error. Generally, use software to obtain this. See next page for formulas.	**Theoretical dist'n of Test statistic**	**Degrees of freedom**
Inference on one mean	$\bar{X}$	$t = \frac{\bar{X} - \mu_0}{SE}$	*t*-dist'n	$df = n - 1$
Inference on one proportion	$\hat{p}$	$z = \frac{\hat{p} - p_0}{SE}$	Normal dist'n	Not relevant
Inference on two means	$\bar{X}_1 - \bar{X}_2$	$(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)$	*t* -dist'n	*df* = Smaller of $n_1 - 1$ and $n_2 - 1$

Type of inference	**Sample statistic**	**Confidence Interval formula** SE is "standard error. Generally, use software to obtain this. See last page for formulas.	**Theoretical dist'n of Test statistic**	**Degrees of freedom**
Inference on one mean	$\bar{X}$	$\bar{X} - t^* \cdot SE \le \mu \le \bar{X} + t^* \cdot SE$	*t*-dist'n	$df = n - 1$
Inference on one proportion	$\hat{p}$	$\hat{p} - z^* \cdot SE \le p \le \hat{p} + z^* \cdot SE$	Normal dist'n	Not relevant
Inference on two means	$\bar{X}_1 - \bar{X}_2$	$(\bar{X}_1 - \bar{X}_2) - t^* \cdot SE \le \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t^* \cdot SE$	*t*-dist'n	df = Smaller of $n_1 - 1$ and $n_2 - 1$

Sample size for estimating one proportion	Sample size for estimating one mean
$n = \left(\frac{z^*}{ME}\right)^2 \cdot \tilde{p}(1-\tilde{p})$, where ME is the chosen margin of error and we use $\tilde{p} = 0.5$ or some other value of $\tilde{p}$ if available.	$n = \left(\frac{z^* \cdot \tilde{\sigma}}{ME}\right)^2$, Where ME is the chosen margin of error and $\tilde{\sigma}$ is an estimate of the population standard deviation.

Type of question and type of inference	Standard Error formula for CI	Standard Error formula for HT
One mean	$SE = \frac{s}{\sqrt{n}}$	$SE = \frac{s}{\sqrt{n}}$
One proportion	$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$ Where p_0 is the value in the null hypothesis
Two means	$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Types of Inference currently on Formula Sheets

Inference on one mean
Inference on one proportion
Inference on two means
Inference on two proportions
Inference on the mean of differences from matched pairs study
Inference on the correlation coefficient
Inference on the Slope Coefficient in a regression model
Test of goodness of fit
Test of association of two categorical variables
Analysis of Variance (ANOVA) for difference of means
Analysis of Variance for regression

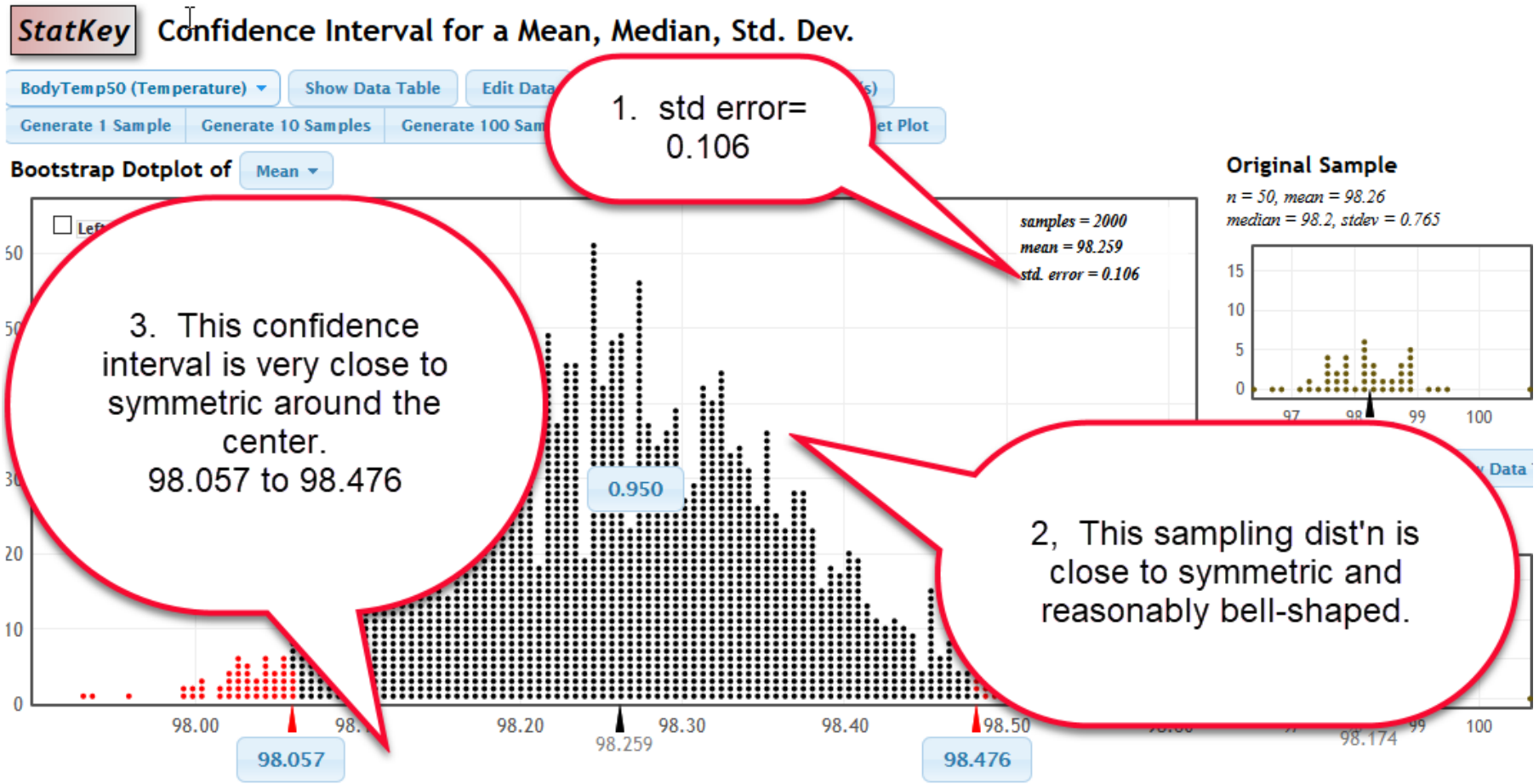
End of section and start of next

Symmetric Confidence Intervals

- We introduced confidence intervals in our simulation-based inference as the "middle" 90% (or whatever percent) of the approximate sampling dist'n. And that's fine.
- In many cases, the dist'n is quite symmetric, which leads us to the idea Confidence Interval:
  - Estimate – Margin of Error (ME)

    to
  - Estimate + Margin of Error (ME)

98.259 minus/plus 2*0.106 = 98.047 to 98.471



Why did I choose 2 to multiply by there?

- You know that the Central Limit Theorem gives us a reason to approximate the sampling dist'n of the mean by a normal dist'n in many cases.
- For a normal dist'n, the middle 95% is bounded by
  center – 1.96 * standard deviation
  and
  center + 1.96 * standard deviation

Why did I choose 2 to multiply by there?

- It's easier to remember the number 2 than 1.96, so a typical "rule of thumb" is "In a normal dist'n, about 95% of the values lie within 2 standard deviations of the center (mean)."
- For this bootstrap dist'n, that was

  98.259 – 2*0.106 and 98.259 + 2*0.106

  98.047 and 98.471
- For this particular bootstrap dist'n generated, the CI was 98.057 to 98.476 (pretty close!)
- I did the bootstrap process two more times and found these intervals:
  98.055 to 98.467, 98.048 to 98.472

Conclusion to this example

- This bootstrap dist'n looks close to normally distributed.
- The 95% confidence interval given by the formula from a normal dist'n (with 1.96 rounded to 2) gives a confidence interval that is very similar to the corresponding bootstrap confidence intervals.

What next?

- Let's investigate a theoretical normal dist'n to see what values (similar to the 1.96, which is approximately 2) give different confidence levels.

- Call those values z* and then our confidence interval is

  Center minus/plus (z* times SE)

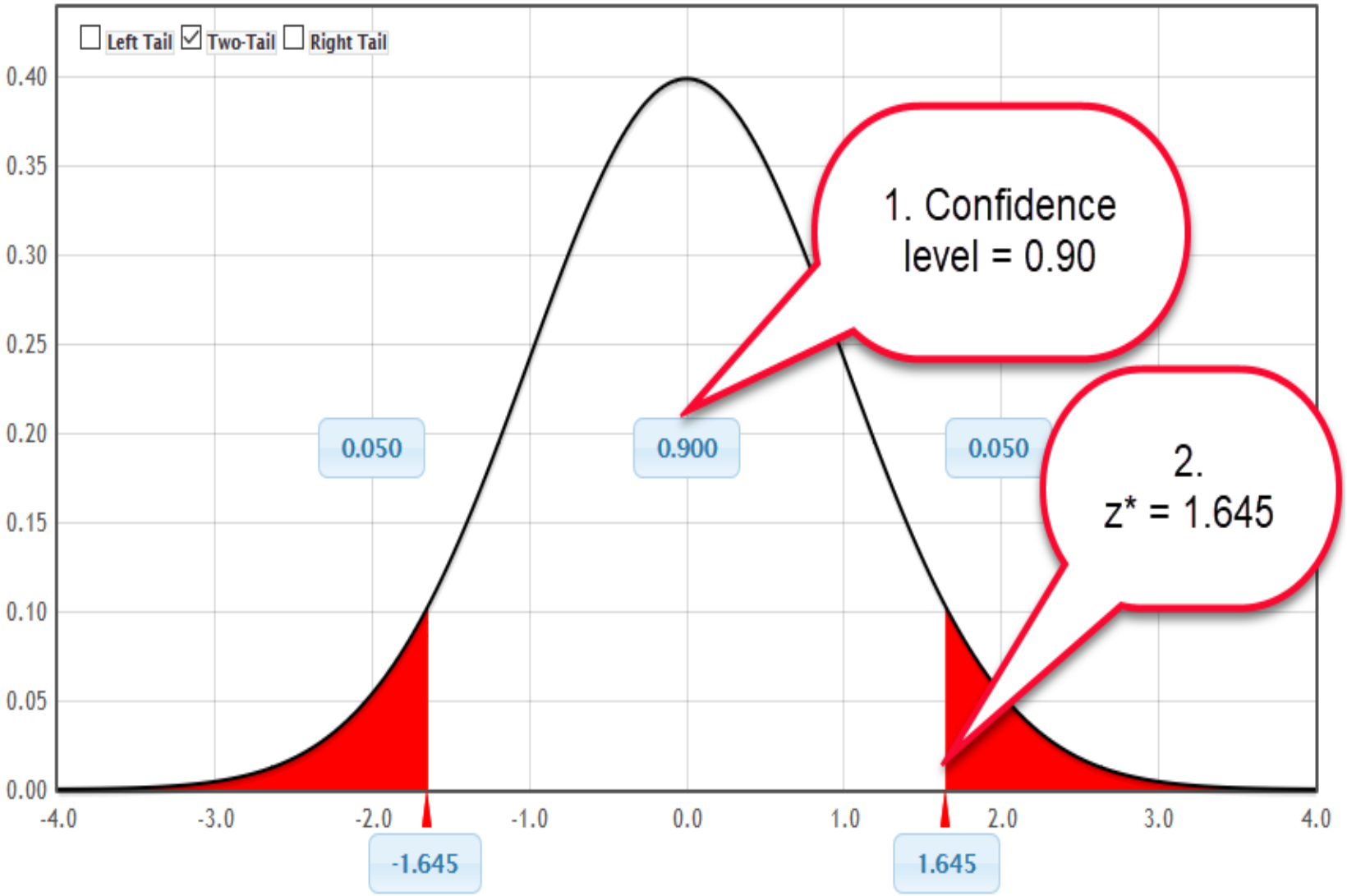
  Center minus/plus Margin of Error

Different confidence levels

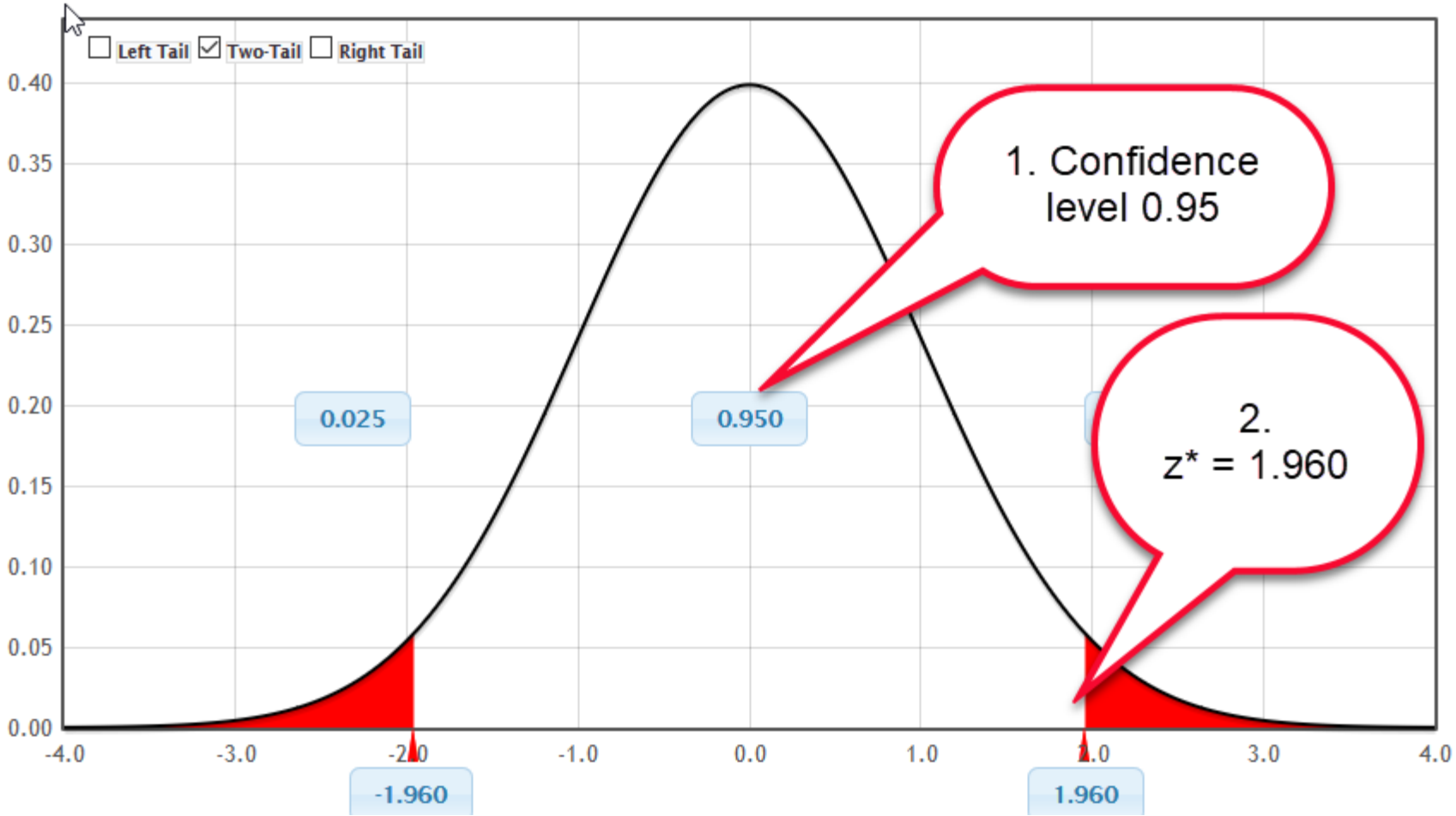
In the next few pictures, we'll see the values (called z*) in a Normal(0,1) distribution to obtain confidence intervals for the population mean for various levels normal dist'n.

(And how you could use the same idea to find this for any confidence level.)

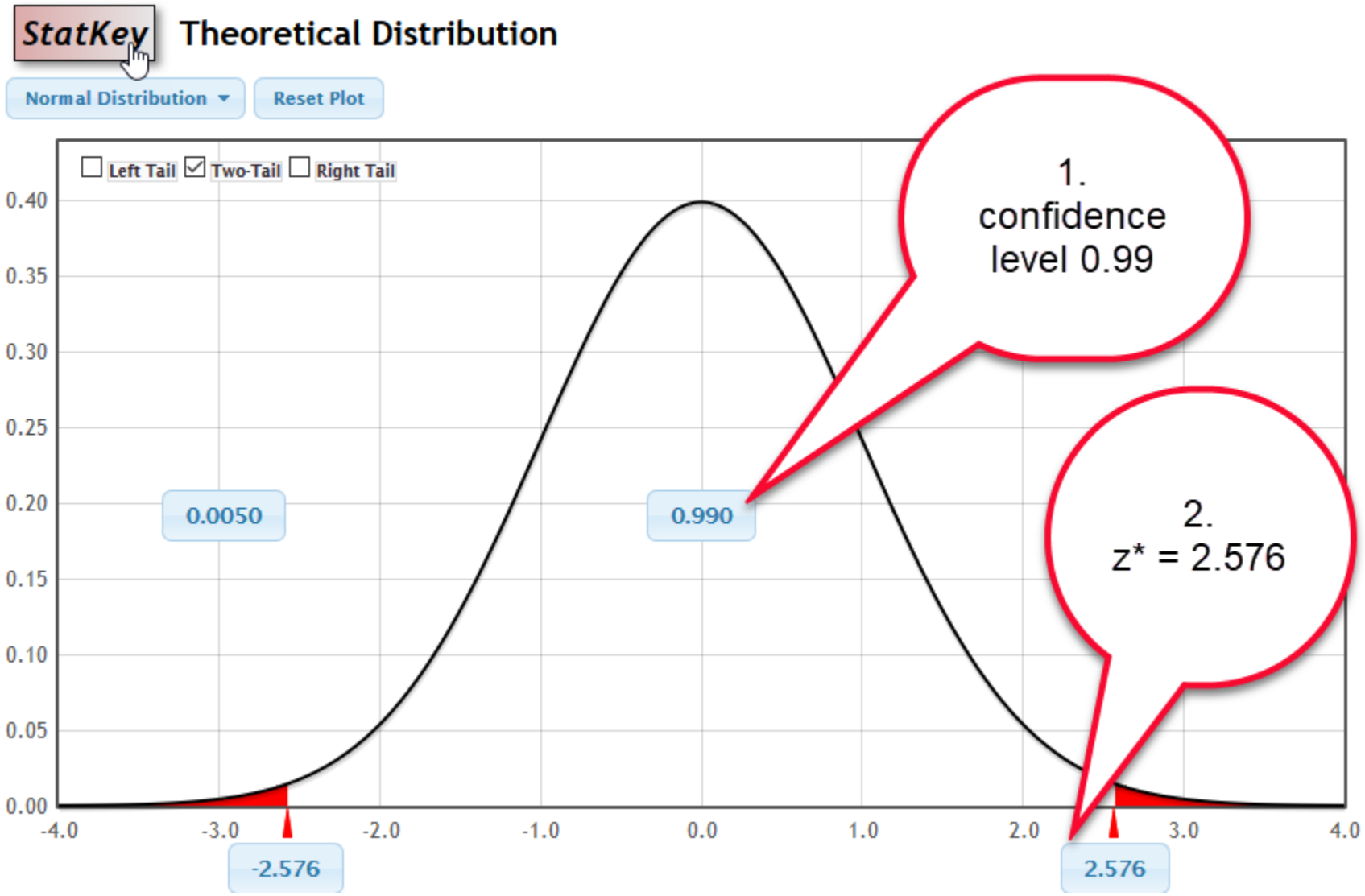
90%


Left Tail
Two-Tail
Right Tail
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
0.050
0.900
0.050
-1.645
1.645
1. Confidence level = 0.90
2. z* = 1.645

95%


Left Tail
Two-Tail
Right Tail
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
0.025
0.950
-1.960
1.960
1. Confidence level 0.95
2. z* = 1.960

99%


StatKey
Theoretical Distribution
Normal Distribution
Reset Plot
Left Tail
Two-Tail
Right Tail
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-4.0
-3.0
-2.0
-1.0
0.0
1.0
2.0
3.0
4.0
0.0050
0.990
-2.576
2.576
1.
confidence
level 0.99
2.
z* = 2.576

Summary

- For forming confidence intervals, we shifted from
  "counting dots" to find the middle __ % of the sampling dist'n of the sample mean

  to
- Using a formula based on the theoretical normal dist'n to find the middle __ % of the sampling dist'n of the sample mean.
- But this part did not address how to find the value called SE in the bootstrap dist'n.

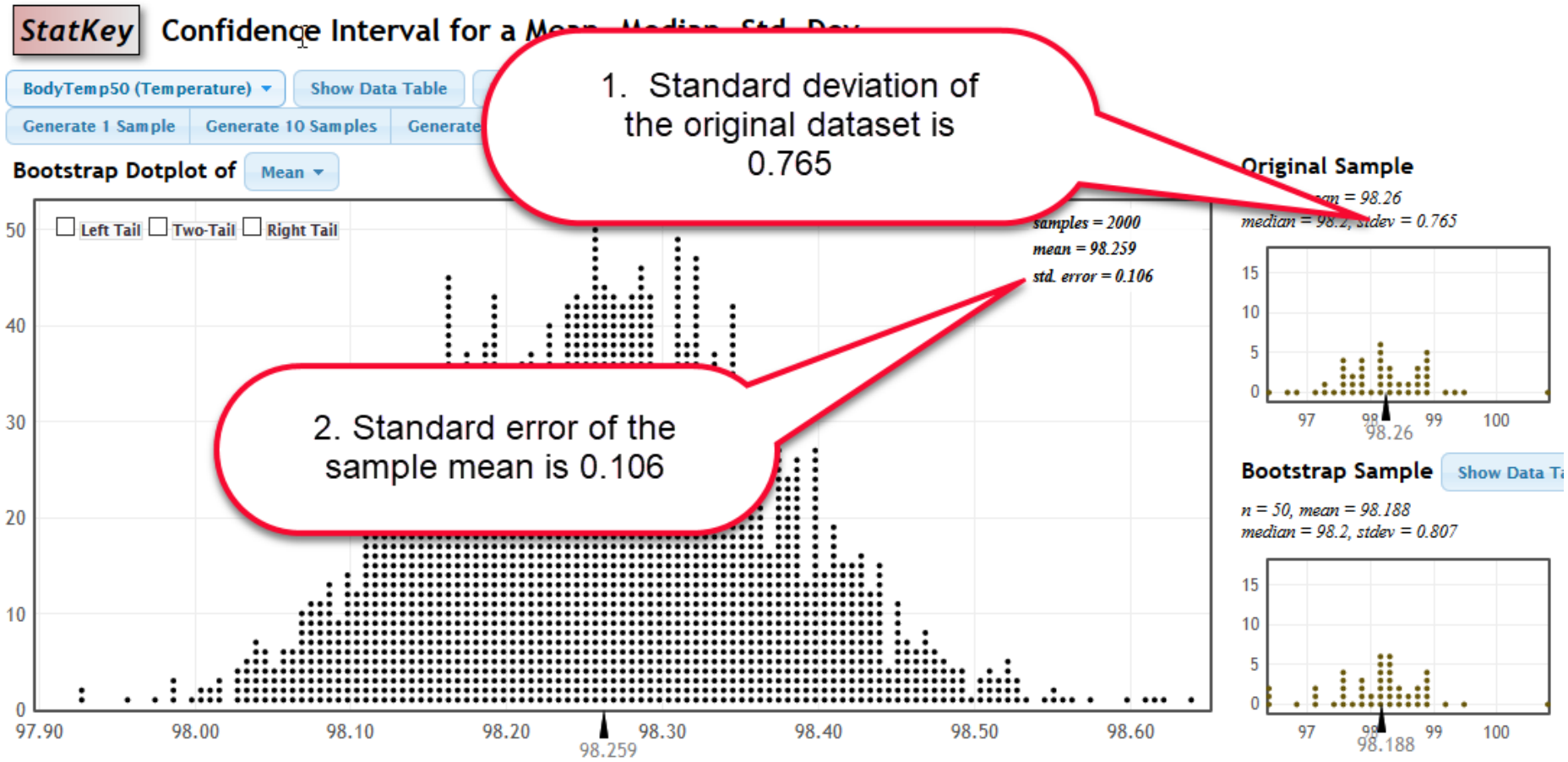
- End of section

Variability in the **Sampling Dist'n**

- Remember that all the distributions we are using for CI's and HT's are sampling dist'ns of sample statistics.
- And the variability in the dist'ns is often discussed.
- We distinguish these by giving different names
  - **standard devation** of the dist'n of the **data**

    and the
  - **standard error** of the sampling dist'n of the **sample statistic**.

    (They are both computed as the standard deviation of a distribution.)


StatKey
Confidence Interval for a Mean, Median, Std. Dev
BodyTemp50 (Temperature)
Show Data Table
Generate 1 Sample
Generate 10 Samples
Generate
Bootstrap Dotplot of
Mean
1. Standard deviation of the original dataset is 0.765
2. Standard error of the sample mean is 0.106
Left Tail
Two-Tail
Right Tail
samples = 2000
mean = 98.259
std. error = 0.106
50
40
30
20
10
0
97.90
98.00
98.10
98.20
98.259
98.30
98.40
98.50
98.60
Original Sample
median = 98.2, stdev = 0.765
15
10
5
0
97
98
99
100
98.26
Bootstrap Sample
Show Data Ta
n = 50, mean = 98.188
median = 98.2, stdev = 0.807
98.188

Formula-based statistical Inference

- One of the major advantages of formula-based (theoretical dist'ns) statistical inference is that we derive formulas for the standard error of the statistics from the standard deviation of the data.
- We can use these formulas to address some questions more efficiently than we can with simulation-based methods.

Find the needed sample size

In the next few slides, see how we use these formulas to find the sample size needed to achieve a confidence interval with a given confidence level and given “margin of error.”

Standard deviation of the dist'n of mean

Assume $X_1, X_2, \ldots, X_n$ are a random sample from a $N(\mu, \sigma^2)$ distribution

var(x̄) = variance of the sample mean

$$\text{var}\left(\bar{X}\right) = \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) == \left(\frac{1}{n}\right)^2 n\sigma^2 = \frac{\sigma^2}{n}$$

Normal: Var(X) = Var(μ + σZ) = σ^2Var(Z) = σ^2

$$\text{std deviation}\left(\bar{X}\right) = \sqrt{\text{var}\left(\bar{X}\right)} = \frac{\sigma}{\sqrt{n}}$$

SE of the dist'n of the mean

Our estimator of this quantity is known as the **Standard Error of the mean.**

$$\text{standard deviation}\left(\bar{X}\right) = \sqrt{\text{var}\left(\bar{X}\right)} = \frac{\sigma}{\sqrt{n}}$$

In our different types of problems, we have different sample statistics. Each of those has a "Standard Error" of the statistic.

In the StatKey bootstrap dist'ns it is called std. error

On our formula sheet it is called SE.

Confidence interval for the mean

The formula we derived from the theoretical normal dist'n of the sample mean is

$$\bar{X} - z^*\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{X} + z^*\left(\frac{\sigma}{\sqrt{n}}\right)$$

Which, on our formula sheet, would be summarized as

$$\bar{X} - z^*(SE) \le \mu \le \bar{X} + z^*(SE)$$

But, in fact, on our formula sheet, it takes into account the approximations and this is done with a t-dist'n

$$\bar{X} - t^*(SE) \le \mu \le \bar{X} + t^*(SE)$$

CI based on the mean

CI for mean based on known σ^2 and sample size n

$$\bar{X} - z^* \left(\frac{\sigma}{\sqrt{n}} \right) \leq \mu \leq \bar{X} + z^* \left(\frac{\sigma}{\sqrt{n}} \right)$$

Margin of error of CI for mean, based on known σ^2 and sample size n

$$ME = z^* \left(\frac{\sigma}{\sqrt{n}} \right)$$

Formula for the sample size

Solve that formula for sample size.

$$n = \left(\frac{z^* \cdot \sigma}{ME} \right)^2,$$

And, when we don't know the pop'n variance σ,
we estimate it with some value that we call $\tilde{\sigma}$,
resulting in the formula from our formula sheet:

$$n = \left(\frac{z^* \cdot \tilde{\sigma}}{ME} \right)^2,$$

Where ME is the chosen margin of error and
$\tilde{\sigma}$ is an estimate of the population standard deviation.

A further step

And, when we don't know the pop'n variance σ,
we estimate it with some value that we call $\tilde{\sigma}$,
resulting in the formula from our formula sheet:

$$n = \left(\frac{z^* \cdot \tilde{\sigma}}{ME} \right)^2,$$

Where ME is the chosen margin of error and
$\tilde{\sigma}$ is an estimate of the population standard deviation.

A further consideration

When we don't know the pop'n standard deviation, but are estimating it (which is, by far, the usual situation) we plug in the sample variance S^2 for the population variance.
We introduce more variability and how much more depends on the sample size.

That leads to the t-dist'n. Important facts:

- There are an infinite number of t-dist'ns, indexed by (integer) degrees of freedom.
- The larger the degrees of freedom, the closer the t-dist'n is to a Normal(0,1) dist'n.
- For n > 30 or so, we might simply use a normal dist'n instead of using a t-dist'n for calculations.

Finding the necessary sample size

- What does this estimation and our moving to the t-dist'n mean for using the formula?
- **Answer:**
  Think of the formula as giving you a "reasonably close" value for the needed sample size.

Conclusion about sample size calculation

- The sort of simulation we did with bootstrap dist'ns on our sample values would be tedious to use
- (compared to this formula-based method)
- to find an equivalent result about the necessary sample size to achieve our goals.

End of section

CLT: Many SE formulas

Type of question and type of inference	Standard Error formula for CI	Standard Error formula for HT
One mean	$SE = \frac{s}{\sqrt{n}}$	$SE = \frac{s}{\sqrt{n}}$
One proportion	$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$ Where p_0 is the value in the null hypothesis
Two means	$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
Two proportions	$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions from the two separate samples	$SE = \sqrt{\frac{\overline{p}(1-\overline{p})}{n_1} + \frac{\overline{p}(1-\overline{p})}{n_2}}$ For testing whether the pop'n proportions are equal. Here $\overline{p}$ is the "pooled" proportion. $\overline{p} = \frac{\text{sum of counts from both samples}}{\text{sum of trials from both samples}}$
Mean of differences from matched pairs data	$SE = \frac{s_d}{\sqrt{n_d}}$ where the subscripts refer to using the differences	$SE = \frac{s_d}{\sqrt{n_d}}$ where the subscripts refer to using the differences
Correlation coefficient	$SE = \sqrt{\frac{1-r^2}{n-2}}$	$SE = \sqrt{\frac{1-r^2}{n-2}}$

CLT: Many SE formulas

- Don’t think too much about the details of those formulas.
- Generally speaking, you should use software to compute these, rather than using the formulas.
- If you’re not using a standard statistical package to do the computations, on the course website you can find some applets or a spreadsheet set up to do these calculations for you.
- This section discusses, generally, how the formulas are derived.

SE for sample mean

Back to the section where we needed

$$\bar{X} - z^* \left(\frac{\sigma}{\sqrt{n}} \right) \leq \mu \leq \bar{X} + z^* \left(\frac{\sigma}{\sqrt{n}} \right)$$

What if we don't know σ^2?

We estimate it by $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

because it can be shown that S^2
is an unbiased estimator of σ^2.

We decide that a reasonable choice is:

$$\text{estimated std deviation}(\bar{X}) = \frac{S}{\sqrt{n}}$$

SE terminology

- Usually (and in this course) the term SE (standard error) is reserved for the **estimated** standard deviation of the sampling dist'n of the statistic.
- It is NOT for the theoretical standard deviation of the sampling dist'n of the statistic.
- That is because we typically want to COMPUTE with the SE value, and, for that, we need it **to not be a function of the parameter(s)** but **to be a function of only of the statistics**.

SE for dist'n of sample mean

So, to match up with the information we saw in the bootstrap dist'n to find a confidence interval for the mean, we use $\text{std error} = \text{SE} = \frac{S}{\sqrt{n}}$

SE Finding the formula

- The formulas for the SE's for other formulas on our formula sheet are derived in similar ways from the variances of the sampling dist'ns of the statistics.
- And, in a similar way to what we did here, the computed variance/standard deviation of the statistic, is a function of the parameter(s).
- To **use** those results, **we must plug in an estimator of each parameter**.

SE What do we plug in?

- This leads to an interesting situation with the proportion problems.
- The appropriate estimator for the parameter p to plug into the SE formula is DIFFERENT depending on whether you are doing a confidence interval or a hypothesis test.
- The assumption in the HT problem is that Ho is true, so we use the value from Ho.
- But we don't make such an assumption in a CI, so, for a CI, we simply plug in the obvious estimator of p.

SE formulas for proportion problems

Type of question and type of inference	Standard Error formula for CI	Standard Error formula for HT
One proportion	$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$ Where p_0 is the value in the null hypothesis
Two proportions	$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions from the two separate samples	$SE = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}$ For testing whether the pop'n proportions are equal. Here $\bar{p}$ is the "pooled" proportion. $\bar{p} = \frac{\text{sum of counts from both samples}}{\text{sum of trials from both samples}}$

HT test statistics for proportion problems

Type of inference	Sample statistic	Test statistic SE is "standard error. Generally, use software to obtain this. See next page for formulas.	Theoretical dist'n of Test statistic	Degrees of freedom
Inference on one proportion	$\hat{p}$	$z = \frac{\hat{p} - p_0}{SE}$	Normal dist'n	Not relevant
Inference on two proportions	$\hat{p}_1 - \hat{p}_2$	$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE}$	Normal dist'n	Not relevant

End of section

Brief discussion of deriving other results

To find the distributions of S^2 and $t = \dfrac{\bar{X} - \mu}{\sqrt{S^2/n}}$

requires mathematical statistics work using transformations of pdfs including some transformations of multivariate pdfs.

Also, to find the distribution of the ratio of the variances (of independent random samples from normal dist'ns)

$$F = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} \sim F_{p,q}$$

requires mathematical statistics work using transformations of multivariate pdfs.

Other considerations

- What if the data don’t meet the conditions for the procedure?
- What if the data are on the “edge” of meeting the conditions for the procedure?

- Should you trust the results?

Other considerations: Partial answer

- The crucial question is not so much whether the data meet the conditions.
- It is whether the sampling distribution of the statistic you will use to make your inference has the distribution that the theory is using.
- Simulation using your sample data available to you.
- Thus you have the capability to explore an approximation to that sampling dist'n.
- Use that to see (in a holistic way) how it differs (or not) from the theoretical dist'n you are expecting.

How should you THINK about these?

- All inference questions for which the sampling dist'n is either normal or a t-dist'n are worked in very similar ways.
- For these, think of the overview of how to do them, and then, as needed, pay attention to anything that is a bit different about the particular type you are working on.
- The two types of chi-squared tests and the ANOVA for means test require different insights. It is important to see examples in addition to those provided here.

What are you expected to know how to DO?

You are expected to do (and interpret the results of) all the inference procedures on our formula sheets

EXCEPT Analysis of Variance in Regression (which is covered in a different course in our program.)

How should you learn to DO these?

- You should be working from an applied statistics book or an elementary statistics book from among those listed on our course web page.
- That book should have examples and exercises of each of these types. (Or if the book you chose to use does not have enough for you, ask for guidance about that.)
- Additional resources are suggested on the course page.

End of section