Unit 3: Foundations for inference
3. Hypothesis tests

Sta 101 – Spring 2019

Duke University, Department of Statistical Science

Dr. Ellison

Slides posted at
https://www2.stat.duke.edu/courses/Spring19/sta101.001/

Outline

1. Housekeeping

2. Main ideas
   1. Another Application of the CLT: 🔍 ⚙ ☺ Use hypothesis tests to make decisions about population parameters
   2. Relationship Between Two CLT Applications: 🔍 👫 ⚙ Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree
   3. Interpretation of Hypothesis Testing:
      1. 🔍 👫 ⚙ Results that are statistically significant are not necessarily practically significant
      2. 🔍 Hypothesis tests are prone to decision errors

Announcements

Coming up…

- Lab Assignment 5 is due **Thursday just before your lab section time.**

- Problem Set 3 is due **Friday 2/22 11:55pm**
- Performance Assessment 3 is due **Sunday 2/24 11:55pm** *(opens today 1-2 hours after class)*

- Readiness Assessment 4 is due **Monday 2/25**
- Peer Evaluations is due **Thursday 2/28 11:55pm** *(part of your participation grade)*

1

Outline

1. Housekeeping

2. Main ideas
   1. Another Application of the CLT: 🔍 ⚙ ☺ Use hypothesis tests to make decisions about population parameters
   2. Relationship Between Two CLT Applications: 🔍 👫 ⚙ Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree
   3. Interpretation of Hypothesis Testing:
      1. 🔍 👫 ⚙ Results that are statistically significant are not necessarily practically significant
      2. 🔍 Hypothesis tests are prone to decision errors

Outline

Have we done frequentist hypothesis testing yet?

Outline

Have we done frequentist hypothesis testing yet?

Yes! Randomization testing is a <u>type</u> of frequentist hypothesis testing!

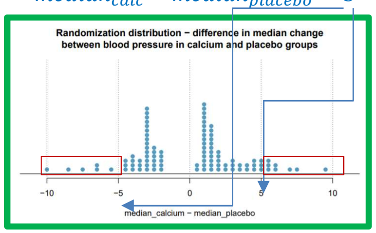
Step 1: Hypotheses

$Ho: Median_{calc} = Median_{placebo}$

$Ha: Median_{calc} \neq Median_{placebo}$

Step 2: Calculate Point Estimate/Sample Statistic

$\widehat{median}_{calc} - \widehat{median}_{placebo} = 5$



Step 3: Calculate p-value =

$$P\left(\begin{matrix}\textit{sample stat that is as}\\ \textit{equal to or "more extreme"}\\ \textit{than the one observed}\end{matrix} \,\middle|\, Ho\ is\ true\right)$$

=0.17

Step 4: Make conclusion

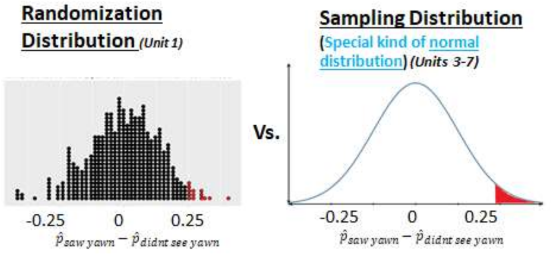
- p-value ≥ 0.05 → Fail to reject null hypothesis. There is not sufficient evidence to suggest the alternative hypothesis.

Outline

What is a big difference between Randomization Hypothesis Testing and CLT-based Hypothesis Testing?

Outline

What is a big difference between Randomization Hypothesis Testing and CLT-based Hypothesis Testing?



Randomization Distribution *(Unit 1)*

Vs.

Sampling Distribution (Special kind of normal distribution) *(Units 3-7)*

To calculate the p-value:

- Randomization Hypothesis Testing uses a randomization distribution
- CLT-based hypothesis testing uses the sampling distribution

---

Outline

What is the general framework for frequentist hypothesis testing (using CLT-based methods)?

---

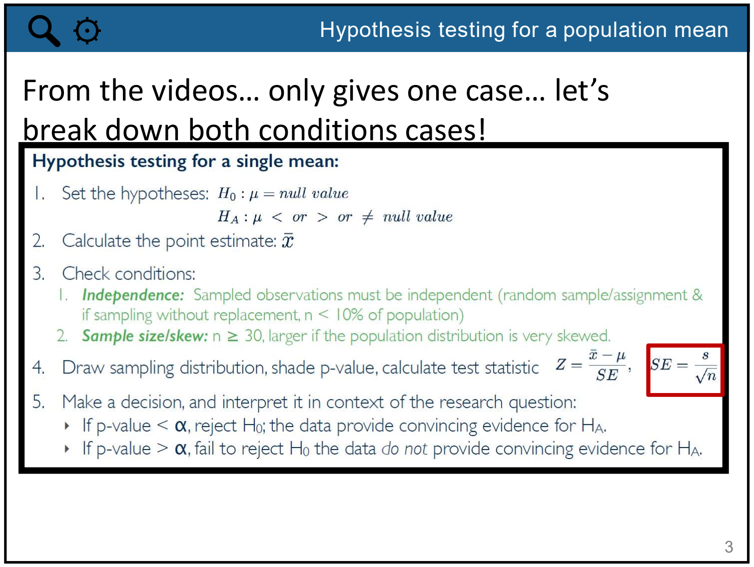
1. Use hypothesis tests to make decisions about population parameters

Hypothesis testing framework:

1. Set the hypotheses.
2. Check assumptions and conditions.
3. Calculate a *test statistic* and a p-value.
4. Make a decision, and interpret it in context of the research question.

Outline

How do we conduct frequentist hypothesis testing (using CLT-based methods) for a population mean?

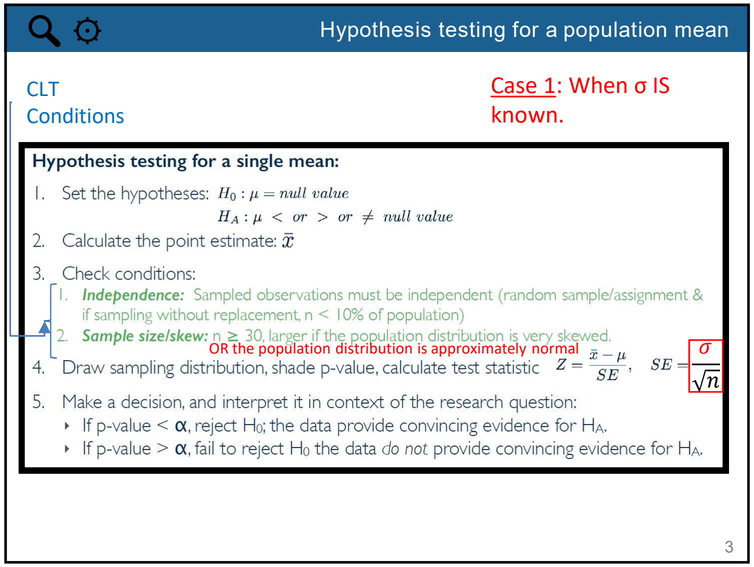


Hypothesis testing for a population mean

From the videos… only gives one case… let's break down both conditions cases!

Hypothesis testing for a single mean:

1. Set the hypotheses: $H_0 : \mu = null\ value$
   $H_A : \mu\ <\ or\ >\ or\ \neq\ null\ value$
2. Calculate the point estimate: $\bar{x}$
3. Check conditions:
   1. ***Independence:*** Sampled observations must be independent (random sample/assignment & if sampling without replacement, n < 10% of population)
   2. ***Sample size/skew:*** n ≥ 30, larger if the population distribution is very skewed.
4. Draw sampling distribution, shade p-value, calculate test statistic $Z = \frac{\bar{x} - \mu}{SE}$, $\boxed{SE = \frac{s}{\sqrt{n}}}$
5. Make a decision, and interpret it in context of the research question:
   - If p-value < α, reject H_0; the data provide convincing evidence for H_A.
   - If p-value > α, fail to reject H_0 the data *do not* provide convincing evidence for H_A.



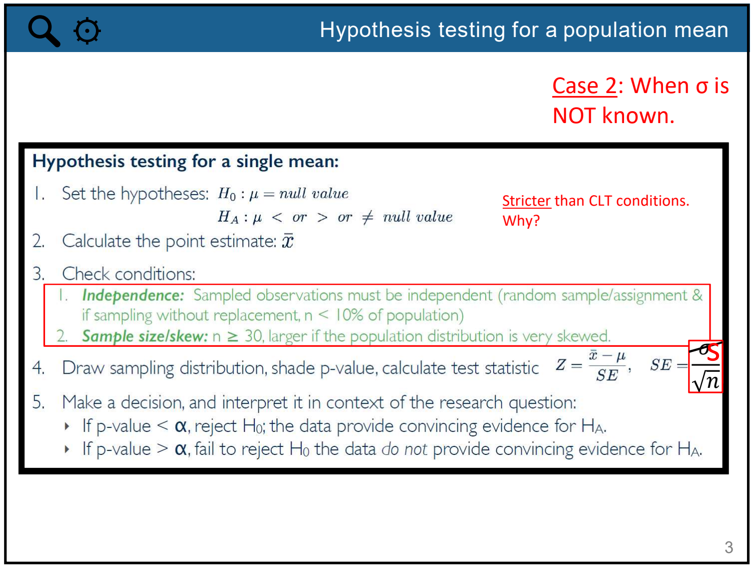
Hypothesis testing for a population mean

CLT
Conditions

Case 1: When σ IS known.

Hypothesis testing for a single mean:

1. Set the hypotheses: $H_0 : \mu = null\ value$
   $H_A : \mu\ <\ or\ >\ or\ \neq\ null\ value$
2. Calculate the point estimate: $\bar{x}$
3. Check conditions:
   1. ***Independence:*** Sampled observations must be independent (random sample/assignment & if sampling without replacement, n < 10% of population)
   2. ***Sample size/skew:*** n ≥ 30, larger if the population distribution is very skewed.
   OR the population distribution is approximately normal
4. Draw sampling distribution, shade p-value, calculate test statistic $Z = \frac{\bar{x} - \mu}{SE}$, $SE = \boxed{\frac{\sigma}{\sqrt{n}}}$
5. Make a decision, and interpret it in context of the research question:
   - If p-value < α, reject H_0; the data provide convincing evidence for H_A.
   - If p-value > α, fail to reject H_0 the data *do not* provide convincing evidence for H_A.



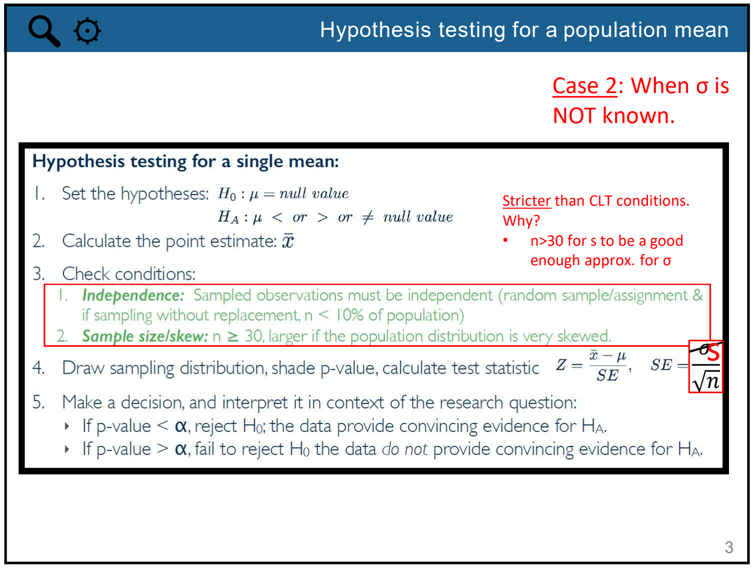
Hypothesis testing for a population mean

Case 2: When σ is NOT known.

Hypothesis testing for a single mean:

Stricter than CLT conditions.
Why?

1. Set the hypotheses: $H_0 : \mu = null\ value$
   $H_A : \mu\ <\ or\ >\ or\ \neq\ null\ value$
2. Calculate the point estimate: $\bar{x}$
3. Check conditions:
   1. ***Independence:*** Sampled observations must be independent (random sample/assignment & if sampling without replacement, n < 10% of population)
   2. ***Sample size/skew:*** n ≥ 30, larger if the population distribution is very skewed.
4. Draw sampling distribution, shade p-value, calculate test statistic $Z = \frac{\bar{x} - \mu}{SE}$, $SE = \boxed{\frac{\cancel{\sigma}\, s}{\sqrt{n}}}$
5. Make a decision, and interpret it in context of the research question:
   - If p-value < α, reject H_0; the data provide convincing evidence for H_A.
   - If p-value > α, fail to reject H_0 the data *do not* provide convincing evidence for H_A.



Hypothesis testing for a population mean

Case 2: When σ is NOT known.

Hypothesis testing for a single mean:

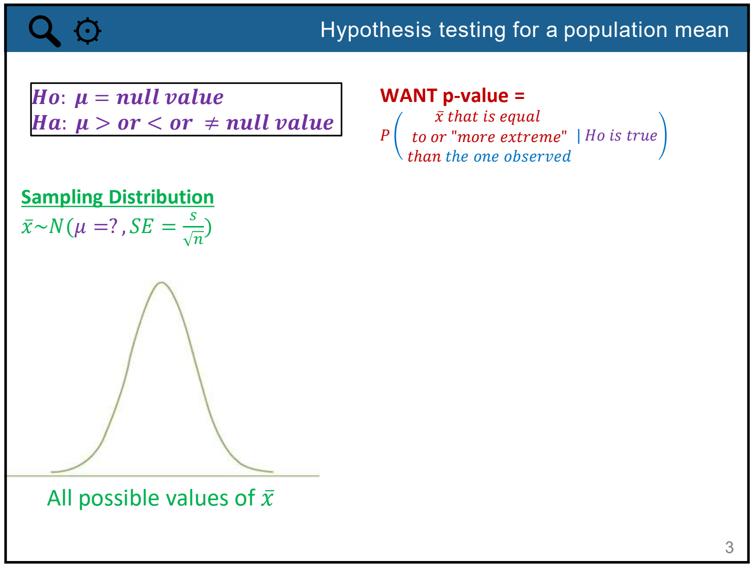
Stricter than CLT conditions.
Why?

- n>30 for s to be a good enough approx. for σ

1. Set the hypotheses: $H_0 : \mu = null\ value$
   $H_A : \mu\ <\ or\ >\ or\ \neq\ null\ value$
2. Calculate the point estimate: $\bar{x}$
3. Check conditions:
   1. ***Independence:*** Sampled observations must be independent (random sample/assignment & if sampling without replacement, n < 10% of population)
   2. ***Sample size/skew:*** n ≥ 30, larger if the population distribution is very skewed.
4. Draw sampling distribution, shade p-value, calculate test statistic $Z = \frac{\bar{x} - \mu}{SE}$, $SE = \boxed{\frac{\cancel{\sigma}\, s}{\sqrt{n}}}$
5. Make a decision, and interpret it in context of the research question:
   - If p-value < α, reject H_0; the data provide convincing evidence for H_A.
   - If p-value > α, fail to reject H_0 the data *do not* provide convincing evidence for H_A.

Outline

☼ 🚻 How do we use a sampling distribution to calculate the p-value for a population mean?



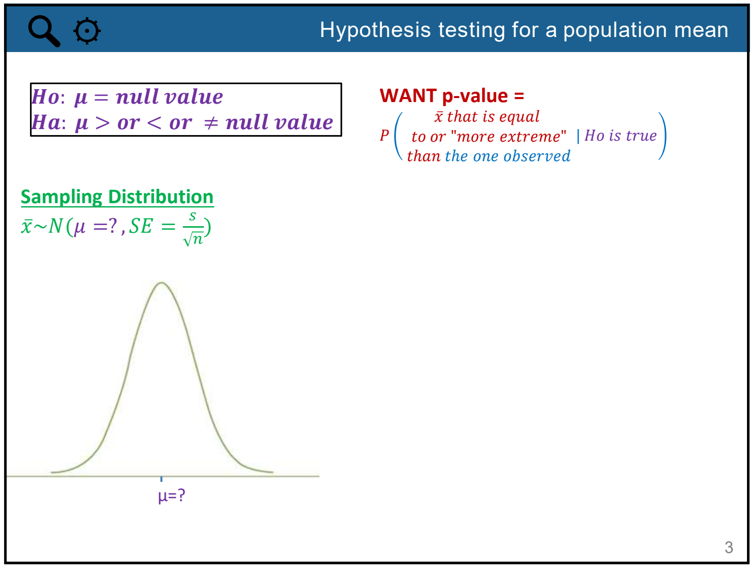
🔍 ☼ Hypothesis testing for a population mean

$Ho: \mu = null\ value$
$Ha: \mu > or < or \neq null\ value$

WANT p-value =
$P\left(\begin{matrix}\bar{x}\ that\ is\ equal\\ to\ or\ "more\ extreme"\\ than\ the\ one\ observed\end{matrix} \middle| Ho\ is\ true\right)$

Sampling Distribution
$\bar{x} \sim N(\mu = ?, SE = \frac{s}{\sqrt{n}})$

All possible values of $\bar{x}$

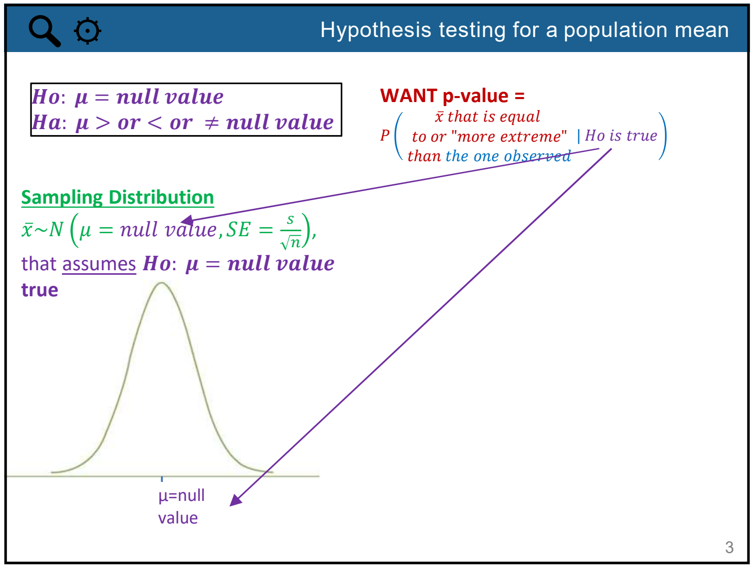


🔍 ☼ Hypothesis testing for a population mean

$Ho: \mu = null\ value$
$Ha: \mu > or < or \neq null\ value$

WANT p-value =
$P\left(\begin{matrix}\bar{x}\ that\ is\ equal\\ to\ or\ "more\ extreme"\\ than\ the\ one\ observed\end{matrix} \middle| Ho\ is\ true\right)$

Sampling Distribution
$\bar{x} \sim N(\mu = ?, SE = \frac{s}{\sqrt{n}})$



🔍 ☼ Hypothesis testing for a population mean

$Ho: \mu = null\ value$
$Ha: \mu > or < or \neq null\ value$

WANT p-value =
$P\left(\begin{matrix}\bar{x}\ that\ is\ equal\\ to\ or\ "more\ extreme"\\ than\ the\ one\ observed\end{matrix} \middle| Ho\ is\ true\right)$

Sampling Distribution
$\bar{x} \sim N\left(\mu = null\ value, SE = \frac{s}{\sqrt{n}}\right)$, that assumes $Ho: \mu = null\ value$ **true**

Outline

☼ What does “more extreme” mean in the p-value definition?

Outline

☼ What does “more extreme” mean in the p-value definition?

Depends on the alternative hypothesis. What types of sample statistics will make us suspect the alternative?

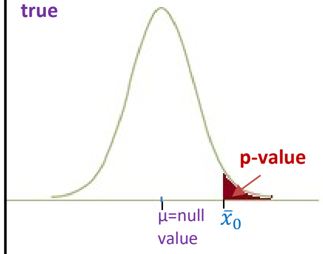
Hypothesis testing for a population mean

Hypotheses
$Ho\colon \mu = null\ value$
$Ha\colon \mu > null\ value$

Sampling Distribution

$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,

that assumes $\boldsymbol{Ho\colon \mu = null\ value}$ **true**



WANT p-value

$$=P\left(\begin{matrix}\bar{x}\ that\ is\ equal\\ to\ or\ "more\ extreme"\\ than\ the\ one\ observed\end{matrix} \mid Ho\ is\ true\right)$$

$\bar{x} \geq \bar{x}_0$ make us suspect the alternative hypothesis.

$$= P(\bar{x} \geq \bar{x}_0 \mid Ho\ is\ true)$$

$$= P\left(Z > \frac{\bar{x}_0 - \mu}{\frac{s}{\sqrt{n}}} \mid \mu = null\ value\right)$$

$$= P\left(Z > \frac{\bar{x}_0 - null\ value}{\frac{s}{\sqrt{n}}}\right)$$

Use z-tables

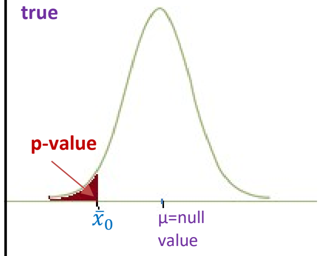
Hypothesis testing for a population mean

Hypotheses
$Ho\colon \mu = null\ value$
$Ha\colon \mu < null\ value$

Sampling Distribution

$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,

that assumes $\boldsymbol{Ho\colon \mu = null\ value}$ **true**



WANT p-value

$$=P\left(\begin{matrix}\bar{x}\ that\ is\ equal\\ to\ or\ "more\ extreme"\\ than\ the\ one\ observed\end{matrix} \mid Ho\ is\ true\right)$$

$\bar{x} \leq \bar{x}_0$ make us suspect the alternative hypothesis.

$$= P(\bar{x} \leq \bar{x}_0 \mid Ho\ is\ true)$$

$$= P\left(Z < \frac{\bar{x}_0 - \mu}{\frac{s}{\sqrt{n}}} \mid \mu = null\ value\right)$$

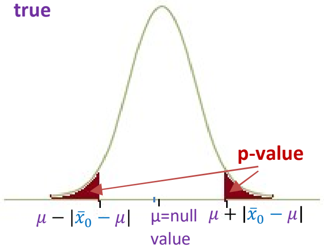
$$= P\left(Z < \frac{\bar{x}_0 - null\ value}{\frac{s}{\sqrt{n}}}\right)$$

Use z-tables

Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu \neq null\ value$

Sampling Distribution
$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $\boldsymbol{Ho: \mu = null\ value}$ **true**



WANT p-value

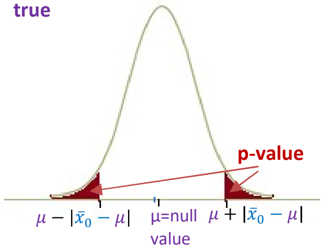
$$=P\left(\begin{matrix}\bar{x}\ that\ is\ equal \\ to\ or\ "more\ extreme" \\ than\ the\ one\ observed\end{matrix} \middle| Ho\ is\ true\right)$$

$\bar{x} \geq \mu + |\bar{x}_0 - \mu|$ or
$\bar{x} \geq \mu - |\bar{x}_0 - \mu|$ make us suspect the alternative hypothesis.

Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu \neq null\ value$

Sampling Distribution
$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $\boldsymbol{Ho: \mu = null\ value}$ **true**



WANT p-value

$$=P\left(\begin{matrix}\bar{x}\ that\ is\ equal \\ to\ or\ "more\ extreme" \\ than\ the\ one\ observed\end{matrix} \middle| Ho\ is\ true\right)$$

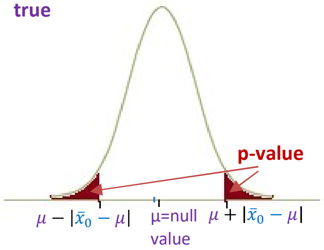
$\bar{x} \geq \mu + |\bar{x}_0 - \mu|$ or
$\bar{x} \geq \mu - |\bar{x}_0 - \mu|$ make us suspect the alternative hypothesis.

$$= P\left(\begin{matrix}\bar{x} \geq \mu + |\bar{x}_0 - \mu| \\ or\ \bar{x} \leq \mu - |\bar{x}_0 - \mu|\end{matrix} \middle| \mu = null\ value\right)$$

Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu \neq null\ value$

Sampling Distribution
$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $\boldsymbol{Ho: \mu = null\ value}$ **true**



WANT p-value

$$=P\left(\begin{matrix}\bar{x}\ that\ is\ equal \\ to\ or\ "more\ extreme" \\ than\ the\ one\ observed\end{matrix} \middle| Ho\ is\ true\right)$$

$\bar{x} \geq \mu + |\bar{x}_0 - \mu|$ or
$\bar{x} \geq \mu - |\bar{x}_0 - \mu|$ make us suspect the alternative hypothesis.

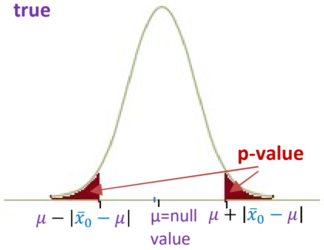
$$= P\left(\begin{matrix}\bar{x} \geq \mu + |\bar{x}_0 - \mu| \\ or\ \bar{x} \leq \mu - |\bar{x}_0 - \mu|\end{matrix} \middle| \mu = null\ value\right)$$

$$= P(\bar{x} \geq \mu + |\bar{x}_0 - \mu| | \mu = null\ value)$$
$$+ P(\bar{x} \leq \mu - |\bar{x}_0 - \mu| | \mu = null\ value)$$

Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu \neq null\ value$

Sampling Distribution
$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $\boldsymbol{Ho: \mu = null\ value}$ **true**



WANT p-value

$$=P\left(\begin{matrix}\bar{x}\ that\ is\ equal \\ to\ or\ "more\ extreme" \\ than\ the\ one\ observed\end{matrix} \middle| Ho\ is\ true\right)$$

$\bar{x} \geq \mu + |\bar{x}_0 - \mu|$ or
$\bar{x} \geq \mu - |\bar{x}_0 - \mu|$ make us suspect the alternative hypothesis.

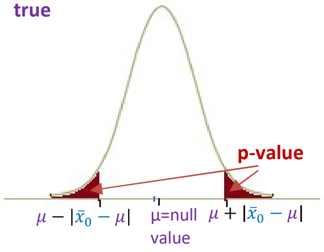
$$= P\left(\begin{matrix}\bar{x} \geq \mu + |\bar{x}_0 - \mu| \\ or\ \bar{x} \leq \mu - |\bar{x}_0 - \mu|\end{matrix} \middle| \mu = null\ value\right)$$

$$= P(\bar{x} \geq \mu + |\bar{x}_0 - \mu| | \mu = null\ value)$$
$$+P(\bar{x} \leq \mu - |\bar{x}_0 - \mu| | \mu = null\ value)$$
$$= 2P(\bar{x} \geq \mu + |\bar{x}_0 - \mu| | \mu = null\ value)$$

Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu \neq null\ value$

Sampling Distribution
$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $\boldsymbol{Ho: \mu = null\ value}$ **true**



WANT p-value

$$=P\left(\begin{matrix}\bar{x} \text{ that is equal} \\ \text{to or "more extreme"} \\ \text{than the one observed}\end{matrix} \;\middle|\; Ho \text{ is true}\right)$$

$\bar{x} \geq \mu + |\bar{x}_0 - \mu|$ or
$\bar{x} \geq \mu - |\bar{x}_0 - \mu|$ make us suspect the alternative hypothesis.

$$= P\left(\begin{matrix}\bar{x} \geq \mu + |\bar{x}_0 - \mu| \\ or\ \bar{x} \leq \mu - |\bar{x}_0 - \mu|\end{matrix} \middle| \mu = null\ value\right)$$
$$= P(\bar{x} \geq \mu + |\bar{x}_0 - \mu| | \mu = null\ value)$$
$$+P(\bar{x} \leq \mu - |\bar{x}_0 - \mu| | \mu = null\ value)$$
$$= 2P(\bar{x} \geq \mu + |\bar{x}_0 - \mu| | \mu = null\ value)$$
$$= 2P\left(Z > \frac{\mu + |\bar{x}_0 - \mu| - \mu}{\frac{s}{\sqrt{n}}} | \mu = null\ value\right)$$
$$= 2P\left(Z > \frac{|\bar{x}_0 - null\ value|}{\frac{s}{\sqrt{n}}}\right)$$

Outline

☼ 👫 How does the test-statistic help us find the p-value?

Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu > null\ value$

WANT p-value

$$=P\left(\begin{matrix}\bar{x} \text{ that is equal} \\ \text{to or "more extreme"} \\ \text{than the one observed}\end{matrix} \;\middle|\; Ho \text{ is true}\right)$$

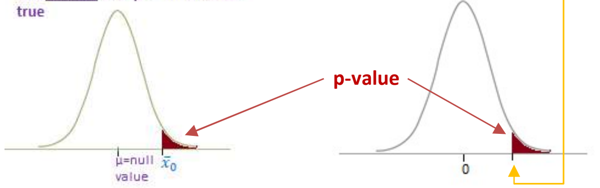
$$= P\left(Z > \boxed{\frac{\bar{x}_0 - null\ value}{\frac{s}{\sqrt{n}}}}\right)$$

Use z-tables

Test Statistic

Sampling Distribution
$\bar{x} \sim N\left(\mu = null\ value, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $\boldsymbol{Ho: \mu = null\ value}$ **true**

Standard Normal Distribution
$Z \sim N(0,1)$



Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu < null\ value$

WANT p-value

$$=P\left(\begin{matrix}\bar{x} \text{ that is equal} \\ \text{to or "more extreme"} \\ \text{than the one observed}\end{matrix} \;\middle|\; Ho \text{ is true}\right)$$

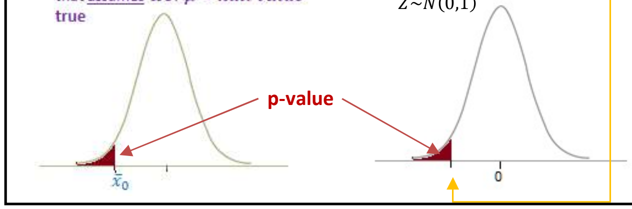
$$= P\left(Z < \boxed{\frac{\bar{x}_0 - null\ value}{\frac{s}{\sqrt{n}}}}\right)$$

Use z-tables

Test Statistic

Sampling Distribution
$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $\boldsymbol{Ho: \mu = null\ value}$ **true**

Standard Normal Distribution
$Z \sim N(0,1)$



Hypothesis testing for a population mean

Hypotheses
$Ho: \mu = null\ value$
$Ha: \mu \neq null\ value$

WANT p-value

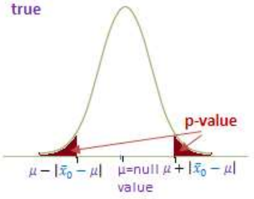
$$=P\left(\begin{matrix}\bar{x}\ that\ is\ equal\\ to\ or\ "more\ extreme"\\ than\ the\ one\ observed\end{matrix}\ \middle|\ Ho\ is\ true\right)$$

Use z-tables

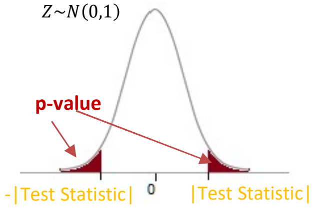
$$= 2P\left(Z > \boxed{\frac{\bar{x}_0 - null\ value}{\frac{s}{\sqrt{n}}}}\right) \qquad = 2P\left(Z < \boxed{\frac{\bar{x}_0 - null\ value}{\frac{s}{\sqrt{n}}}}\right)$$

If $\bar{x}_0 > null\ value$ If $\bar{x}_0 < null\ value$

Sampling Distribution
$\bar{x} \sim N\left(\mu = ?, SE = \frac{s}{\sqrt{n}}\right)$,
that assumes $Ho: \mu = null\ value$ true



Standard Normal Distribution
$Z \sim N(0,1)$

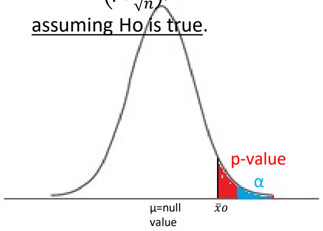


Hypothesis testing for a population mean

From the videos…

5. Make a decision, and interpret it in context of the research question:
   - If p-value < α, reject H_0; the data provide convincing evidence for H_A.
   - If p-value > α, fail to reject H_0 the data *do not* provide convincing evidence for H_A.

Sampling Distribution
of $\bar{x} \sim N\left(\mu, \frac{s}{\sqrt{n}}\right)$,
assuming Ho is true.



Application exercise: 3.2 Hypothesis testing for a single mean

See course website for details.

Outline

How do we interpret the p-value?

Clicker question

Which of the following is the correct interpretation of the p-value from App Ex 3.2?

(a) The probability that average GPA of Duke students has changed since 2001.
(b) The probability that average GPA of Duke students has not changed since 2001.
(c) The probability that average GPA of Duke students has not changed since 2001, if in fact a random sample of 63 Duke students this year have an average GPA of 3.58 or higher.
(d) The probability that a random sample of 63 Duke students have an average GPA of 3.58 or higher, if in fact the average GPA has not changed since 2001.
(e) The probability that a random sample of 63 Duke students have an average GPA of 3.58 or higher or 3.16 or lower, if in fact the average GPA has not changed since 2001.

Clicker question

Which of the following is the correct interpretation of the p-value from App Ex 3.2?

(a) The probability that average GPA of Duke students has changed since 2001. ***-P(alternative hyp)***
(b) The probability that average GPA of Duke students has not changed since 2001. ***-P(null hyp)***
(c) The probability that average GPA of Duke students has not changed since 2001, if in fact a random sample of 63 Duke students this year have an average GPA of 3.58 or higher. ***-P(null hyp|***$\bar{x} \geq 3.58$)
d) The probability that a random sample of 63 Duke students have an average GPA of 3.58 or higher, if in fact the average GPA has not changed since 2001. ***-P(***$\bar{x} \geq 3.58|\boldsymbol{null\ hyp}$)
e) The probability that a random sample of 63 Duke students have an average GPA of 3.58 or higher or 3.16 or lower, if in fact the average GPA has not changed since 2001.
-P($\bar{x} \geq 3.58\ or\ \bar{x} \leq 3.16|\boldsymbol{null\ hyp}$)

Common misconceptions about hypothesis testing

1. P-value is the probability that the null hypothesis is true
   A p-value is the probability of getting a sample that results in a ***test statistic/sample statistic*** *as or more extreme than what you actually observed (and in favor of the null hypothesis) if in fact the null hypothesis is correct. It is a conditional probability, conditioned on the null hypothesis being correct.*

Common misconceptions about hypothesis testing

1. P-value is the probability that the null hypothesis is true
   A p-value is the probability of getting a sample that results in a test statistic/sample statistic as or more extreme than what you actually observed (and in favor of the null hypothesis) if in fact the null hypothesis is correct. It is a conditional probability, conditioned on the null hypothesis being correct.
2. A high p-value confirms the null hypothesis.
   A high p-value means the data do not provide convincing evidence for the alternative hypothesis and hence that the null hypothesis can't be rejected.

☹ Common misconceptions about hypothesis testing

1. P-value is the probability that the null hypothesis is true
   A p-value is the probability of getting a sample that results in a test statistic/sample statistic as or more extreme than what you actually observed (and in favor of the null hypothesis) if in fact the null hypothesis is correct. It is a conditional probability, conditioned on the null hypothesis being correct.
2. A high p-value confirms the null hypothesis.
   A high p-value means the data do not provide convincing evidence for the alternative hypothesis and hence that the null hypothesis can't be rejected.

3. A low p-value confirms the alternative hypothesis.
   A low p-value means the data provide convincing evidence for the alternative hypothesis, but not necessarily that it is confirmed.

Outline

1. Housekeeping

2. Main ideas
   1. Another Application of the CLT: 🔍 ☼ ☹ Use hypothesis tests to make decisions about population parameters
   2. Relationship Between Two CLT Applications: 🔍 👫 ☼ Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree
   3. Interpretation of Hypothesis Testing:
      1. 🔍 👫 ☼ Results that are statistically significant are not necessarily practically significant
      2. 🔍 Hypothesis tests are prone to decision errors

Outline

☼ How do we conduct hypothesis testing with a confidence interval (instead of a p-value)?

🔍 👫 2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

Hypotheses

$Ho: \mu =$ null value

$Ha: \mu \neq$ null value

Fail to Reject Ho when:

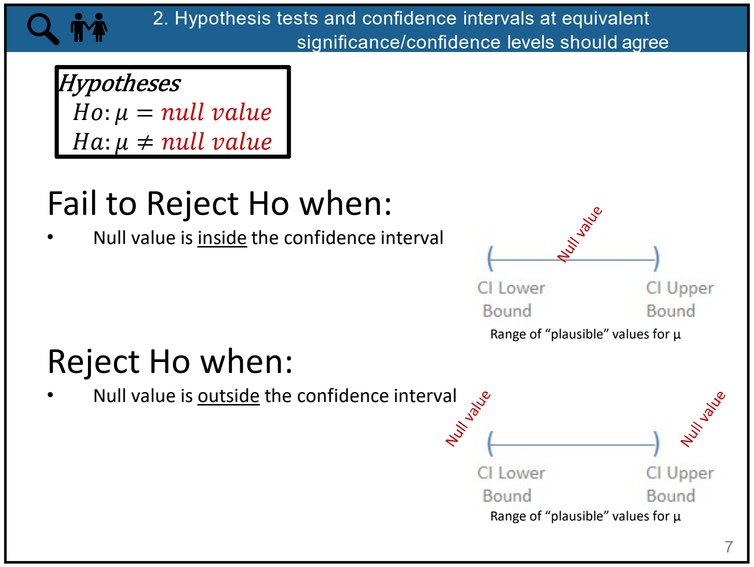
- Null value is ______ the confidence interval

Range of "plausible" values for μ

Reject Ho when:

- Null value is ______ the confidence interval

Range of "plausible" values for μ



2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

Hypotheses
$Ho: \mu = \text{null value}$
$Ha: \mu \neq \text{null value}$

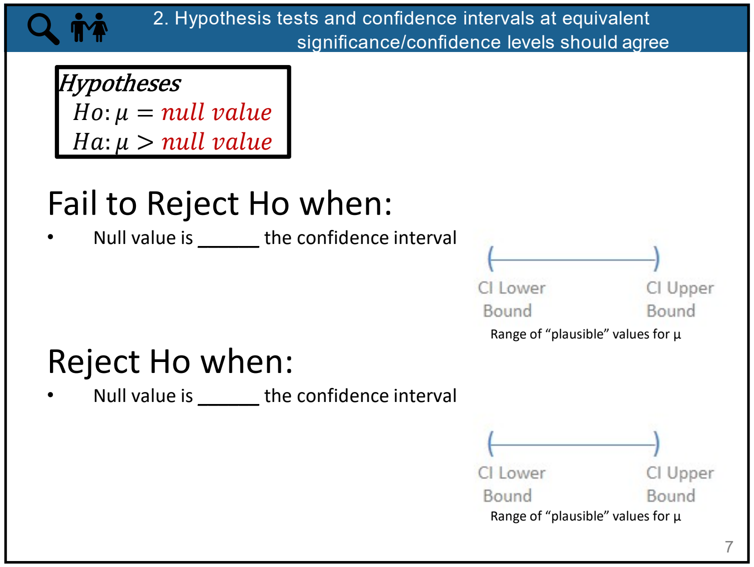
Fail to Reject Ho when:

- Null value is <u>inside</u> the confidence interval

Reject Ho when:

- Null value is <u>outside</u> the confidence interval

7



2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

Hypotheses
$Ho: \mu = \text{null value}$
$Ha: \mu > \text{null value}$

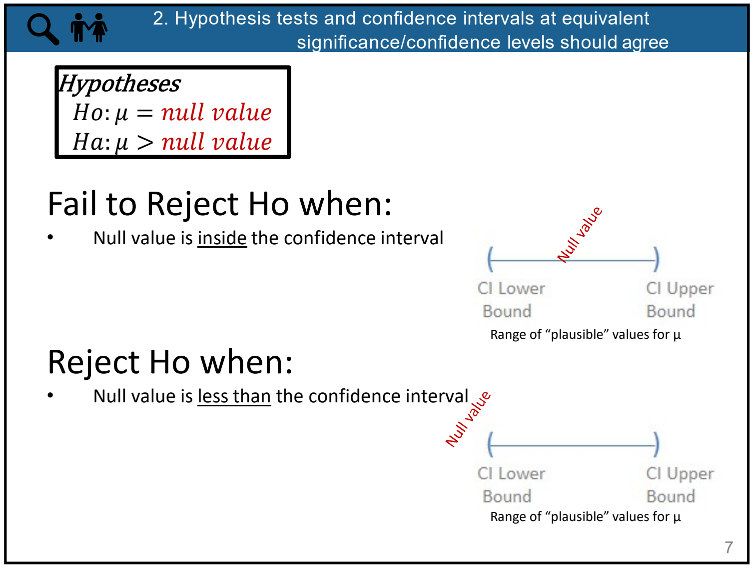
Fail to Reject Ho when:

- Null value is ______ the confidence interval

Reject Ho when:

- Null value is ______ the confidence interval

7



2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

Hypotheses
$Ho: \mu = \text{null value}$
$Ha: \mu > \text{null value}$

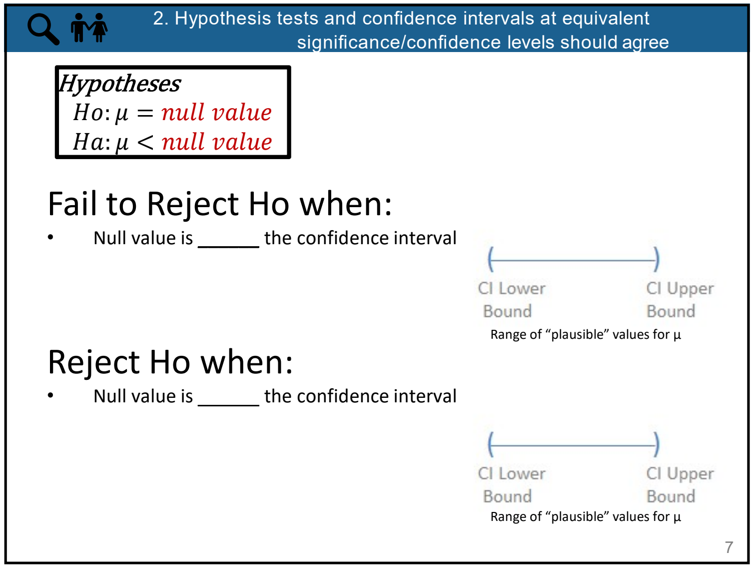
Fail to Reject Ho when:

- Null value is <u>inside</u> the confidence interval

Reject Ho when:

- Null value is <u>less than</u> the confidence interval

7



2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

Hypotheses
$Ho: \mu = \text{null value}$
$Ha: \mu < \text{null value}$

Fail to Reject Ho when:

- Null value is ______ the confidence interval

Reject Ho when:

- Null value is ______ the confidence interval

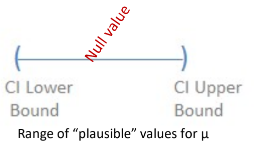
7

2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

Hypotheses
$Ho: \mu =$ *null value*
$Ha: \mu <$ *null value*

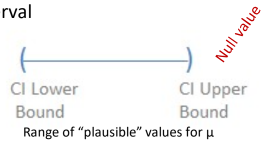
Fail to Reject Ho when:

- Null value is <u>inside</u> the confidence interval



Reject Ho when:

- Null value is <u>greater than</u> the confidence interval



Outline

☼ <u>When</u> do the conclusions of the following agree?

- **Hypothesis testing with a confidence interval**
- **Hypothesis testing with a p-value**

☼ 2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

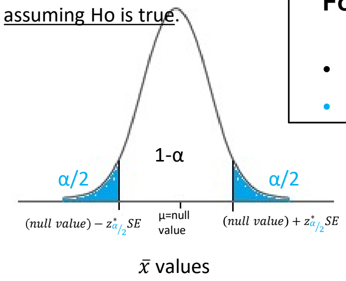
For <u>Two-Sided Test</u>,
Conclusions Agree When:

- Conf. level = (1- α)
- α = (1 - Conf. Level)

☼ 2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

<u>TWO</u> sided Hypothesis Test
with <u>**α** significance level</u>

Sampling Distribution of $\bar{x} \sim N(\mu, SE)$, <u>assuming Ho is true</u>.



Fail to Reject Ho when:

- $(null\ value) - z^*_{\alpha/2}SE < \bar{x} < (null\ value) + z^*_{\alpha/2}SE$
- $\bar{x} - z^*_{\alpha/2}SE < null\ value < \bar{x} + z^*_{\alpha/2}SE$

For <u>Two-Sided Test</u>,
Conclusions Agree When:

- Conf. level = (1- α)
- α = (1 - Conf. Level)

2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

For One-Sided Test, Conclusions Agree When:

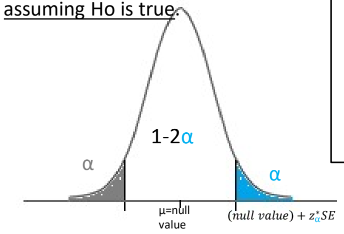
- Conf. Level = (1- 2α)
- α = $\frac{(1-Con\ \ .Level)}{2}$

---

2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree

ONE sided Hypothesis Test with **α** significance level

Sampling Distribution of $\bar{x} \sim N(\mu, SE)$, assuming Ho is true.



$\bar{x}$ values

Fail to Reject Ho when:

- $\bar{x} < (null\ value) + z_{\alpha}^{*}SE$
- $null\ value > \bar{x} - z_{\alpha}^{*}SE$

For One-Sided Test, Conclusions Agree When:

- Conf. Level = (1- 2α)
- α = $\frac{(1-Conf.Level)}{2}$

Clicker question

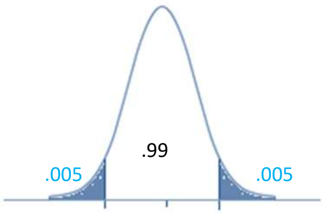
What is the confidence level for a confidence interval that is equivalent to a two-sided hypothesis test at the 1% significance level? *Hint: Draw a picture and mark the confidence level in the center.*

(a) 0.80
(b) 0.90
(c) 0.95
(d) 0.98
(e) 0.99

Clicker question

What is the confidence level for a confidence interval that is equivalent to a two-sided hypothesis test at the 1% significance level? *Hint: Draw a picture and mark the confidence level in the center.*

(a) 0.80
(b) 0.90
(c) 0.95
(d) 0.98
(e) *0.99*



Clicker question

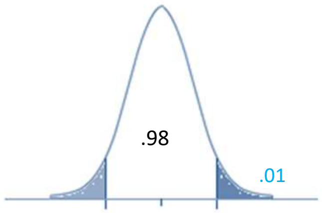
What is the confidence level for a confidence interval that is equivalent to a one-sided hypothesis test at the 1% significance level? *Hint: Draw a picture and mark the confidence level in the center.*

(a) 0.80
(b) 0.90
(c) 0.95
(d) 0.98
(e) 0.99

Clicker question

What is the confidence level for a confidence interval that is equivalent to a one-sided hypothesis test at the 1% significance level? *Hint: Draw a picture and mark the confidence level in the center.*

(a) 0.80
(b) 0.90
(c) 0.95
(d) *0.98*
(e) 0.99



Clicker question

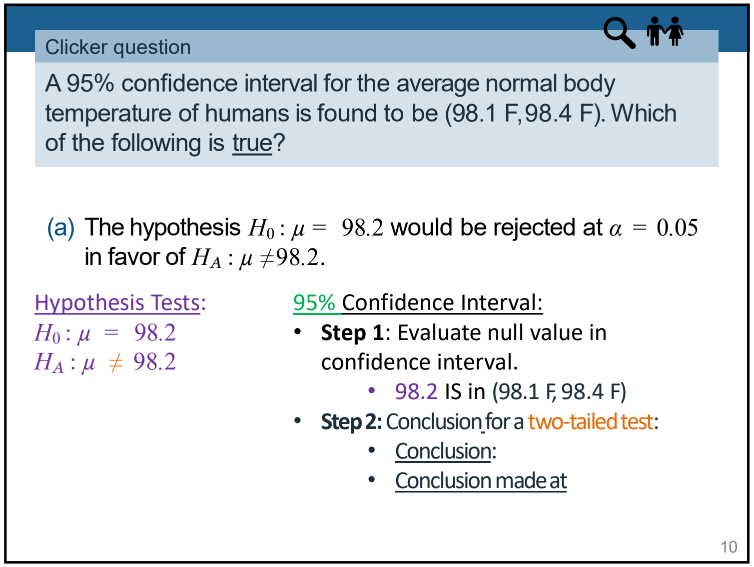
A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F, 98.4 F). Which of the following is true?

(a) The hypothesis $H_0 : \mu = 98.2$ would be rejected at $\alpha = 0.05$ in favor of $H_A : \mu \neq 98.2$.
(b) The hypothesis $H_0 : \mu = 98.2$ would be rejected at $\alpha = 0.025$ in favor of $H_A : \mu > 98.2$.
(c) The hypothesis $H_0 : \mu = 98$ would be rejected in favor of $H_A : \mu \neq 98$ using a 90% confidence interval.
(d) The hypothesis $H_0 : \mu = 98.2$ would be rejected would be rejected in favor of $H_A : \mu \neq 98.2$ using a 99% confidence interval.

Clicker question

A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F, 98.4 F). Which of the following is true?

(a) The hypothesis $H_0 : \mu = 98.2$ would be rejected at $\alpha = 0.05$ in favor of $H_A : \mu \neq 98.2$.
(b) The hypothesis $H_0 : \mu = 98.2$ would be rejected at $\alpha = 0.025$ in favor of $H_A : \mu > 98.2$.
(c) The hypothesis $H_0 : \mu = 98$ would be rejected in favor of $H_A : \mu \neq 98$ using a 90% confidence interval.
(d) The hypothesis $H_0 : \mu = 98.2$ would be rejected would be rejected in favor of $H_A : \mu \neq 98.2$ using a 99% confidence interval.



Clicker question

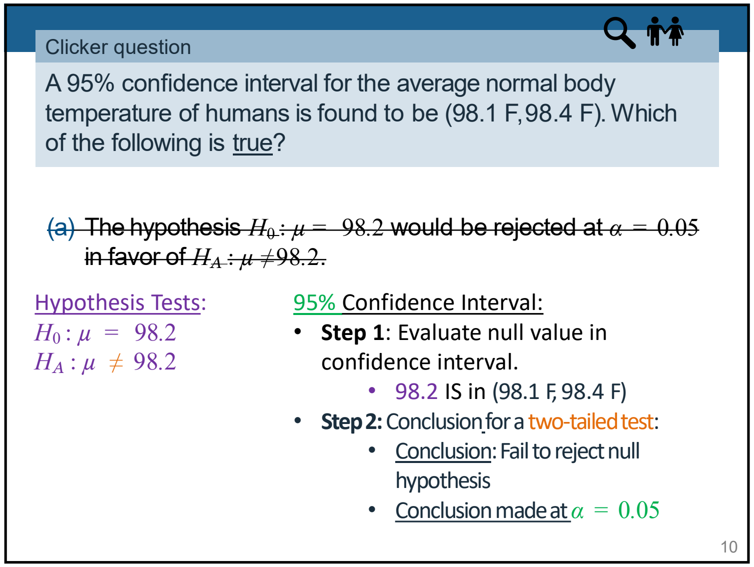
A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F,98.4 F). Which of the following is true?

(a) The hypothesis $H_0: \mu = 98.2$ would be rejected at $\alpha = 0.05$ in favor of $H_A: \mu \neq 98.2$.

Hypothesis Tests:
$H_0: \mu = 98.2$
$H_A: \mu \neq 98.2$

95% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98.2 IS in (98.1 F, 98.4 F)
- **Step 2:** Conclusion for a two-tailed test:
  - Conclusion:
  - Conclusion made at



Clicker question

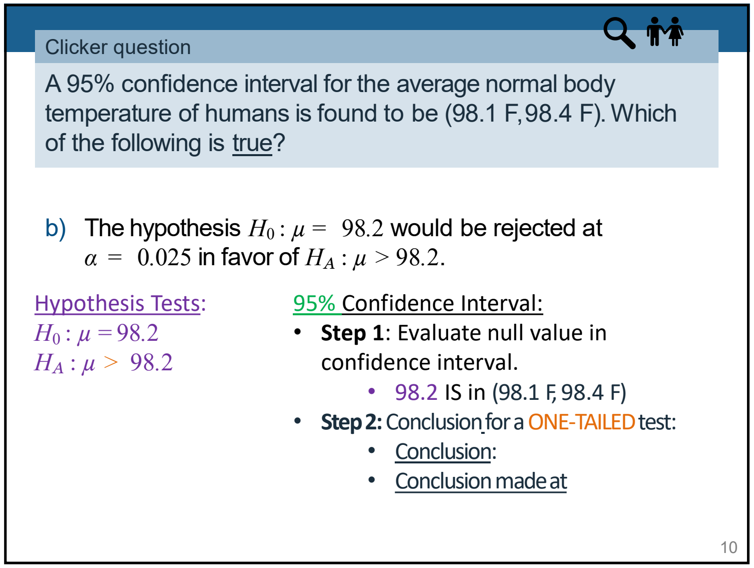
A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F,98.4 F). Which of the following is true?

~~(a) The hypothesis $H_0: \mu = 98.2$ would be rejected at $\alpha = 0.05$ in favor of $H_A: \mu \neq 98.2$.~~

Hypothesis Tests:
$H_0: \mu = 98.2$
$H_A: \mu \neq 98.2$

95% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98.2 IS in (98.1 F, 98.4 F)
- **Step 2:** Conclusion for a two-tailed test:
  - Conclusion: Fail to reject null hypothesis
  - Conclusion made at $\alpha = 0.05$



Clicker question

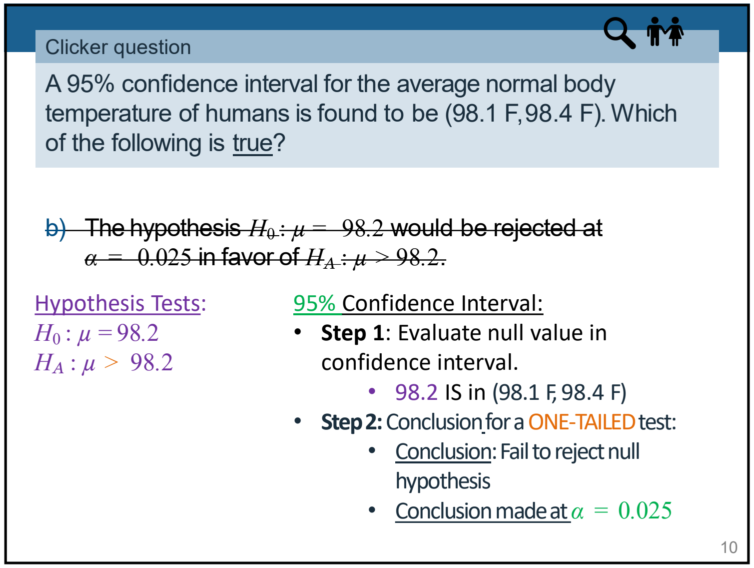
A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F,98.4 F). Which of the following is true?

b) The hypothesis $H_0: \mu = 98.2$ would be rejected at $\alpha = 0.025$ in favor of $H_A: \mu > 98.2$.

Hypothesis Tests:
$H_0: \mu = 98.2$
$H_A: \mu > 98.2$

95% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98.2 IS in (98.1 F, 98.4 F)
- **Step 2:** Conclusion for a ONE-TAILED test:
  - Conclusion:
  - Conclusion made at



Clicker question

A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F,98.4 F). Which of the following is true?

~~b) The hypothesis $H_0: \mu = 98.2$ would be rejected at $\alpha = 0.025$ in favor of $H_A: \mu > 98.2$.~~

Hypothesis Tests:
$H_0: \mu = 98.2$
$H_A: \mu > 98.2$

95% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98.2 IS in (98.1 F, 98.4 F)
- **Step 2:** Conclusion for a ONE-TAILED test:
  - Conclusion: Fail to reject null hypothesis
  - Conclusion made at $\alpha = 0.025$

Clicker question

A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F, 98.4 F). Which of the following is true?

(c) The hypothesis $H_0 : \mu = 98$ would be rejected in favor of $H_A : \mu \neq 98$ using a 90% confidence interval.

Hypothesis Tests:
$H_0 : \mu = 98$
$H_A : \mu \neq 98$

90% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98 IS NOT in (98.1 F, 98.4 F) – 95% CI
  - 98 _______ 90% CI
- **Step 2:** Conclusion for a TWO-TAILED test:
  - Conclusion: ________ null hypothesis with 90% CI

Clicker question

A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F, 98.4 F). Which of the following is true?

(c) The hypothesis $H_0 : \mu = 98$ would be rejected in favor of $H_A : \mu \neq 98$ using a 90% confidence interval.

Hypothesis Tests:
$H_0 : \mu = 98$
$H_A : \mu \neq 98$

90% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98 IS NOT in (98.1 F, 98.4 F) – 95% CI
  - 98 is not in a 90% CI either (90% is more narrow than 95% CI)
- **Step 2:** Conclusion for a TWO-TAILED test:
  - Conclusion: Reject null hypothesis with 90% CI

Clicker question

A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F, 98.4 F). Which of the following is true?

d) The hypothesis $H_0 : \mu = 98.2$ would be rejected would be rejected in favor of $H_A : \mu \neq 98.2$ using a 99% confidence interval.

Hypothesis Tests:
$H_0 : \mu = 98.2$
$H_A : \mu \neq 98.2$

99% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98.2 IS in (98.1 F, 98.4 F)
  - 98 _________99% CI
- **Step 2:** Conclusion for a TWO-TAILED test:
  - Conclusion: ________null hypothesis

Clicker question

A 95% confidence interval for the average normal body temperature of humans is found to be (98.1 F, 98.4 F). Which of the following is true?

~~d) The hypothesis $H_0 : \mu = 98.2$ would be rejected would be rejected in favor of $H_A : \mu \neq 98.2$ using a 99% confidence interval.~~

Hypothesis Tests:
$H_0 : \mu = 98.2$
$H_A : \mu \neq 98.2$

99% Confidence Interval:

- **Step 1**: Evaluate null value in confidence interval.
  - 98.2 IS in (98.1 F, 98.4 F)
  - 98 is also in 99% CI (99% is more wide than 95% CI)
- **Step 2:** Conclusion for a TWO-TAILED test:
  - Conclusion: Fail to reject null hypothesis with a 99% CI

Outline

1. Housekeeping

2. Main ideas
   1. Another Application of the CLT: Use hypothesis tests to make decisions about population parameters
   2. Relationship Between Two CLT Applications: Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree
   3. Interpretation of Hypothesis Testing:
      1. Results that are statistically significant are not necessarily practically significant
      2. Hypothesis tests are prone to decision errors

Outline

What are some issues if our sample size is too small?

What are some issues if our sample size is too large?

---

3. Results that are statistically significant are not necessarily practically significant

Clicker question

All else held equal, will p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$
(b) $n = 10,000$

---

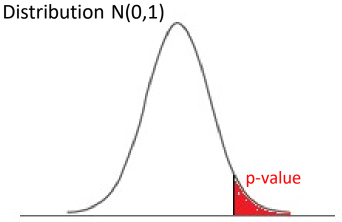
3. Results that are statistically significant are not necessarily practically significant

Clicker question

All else held equal, will p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$
(b) $n = 10,000$

Standard Normal Distribution N(0,1)



Test Statistic

$z = \frac{\bar{x} - null\ value}{\sigma/\sqrt{n}}$

3. Results that are statistically significant are not necessarily practically significant

🔍 ⚙ 👫

Clicker question

All else held equal, will p-value be lower if $n = 100$ or $n = 10,000$?

(a) $n = 100$
(b) $n = 10,000$

Suppose $\bar{x} = 5,\ s = 2,\ H_0 : \mu = 4.5$, *and* $H_A : \mu > 4.5$.

Test Statistic

$$Z_{n=100} = \frac{5-4.5}{\frac{2}{\sqrt{100}}} = \frac{5-4.5}{\frac{2}{10}} = \frac{0.5}{0.2} = 2.5, \quad p\text{-value} = 0.0062$$

$$Z_{n=10000} = \frac{5-4.5}{\frac{2}{\sqrt{10000}}} = \frac{5-4.5}{\frac{2}{100}} = \frac{0.5}{0.02} = 25, \quad p\text{-value} \approx 0$$

As n increases - $SE\downarrow$, $Z\uparrow$, p-value $\downarrow$

Outline

1. Housekeeping

2. Main ideas
    1. Another Application of the CLT: 🔍 ⚙ ☹ Use hypothesis tests to make decisions about population parameters
    2. Relationship Between Two CLT Applications: 🔍 👫 ⚙ Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree
    3. Interpretation of Hypothesis Testing:
        1. 🔍 👫 ⚙ Results that are statistically significant are not necessarily practically significant
        2. 🔍 Hypothesis tests are prone to decision errors

Outline

🔍 What are two ways we can make an error in hypothesis testing?

---

4. Hypothesis tests are prone to decision errors

🔍

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

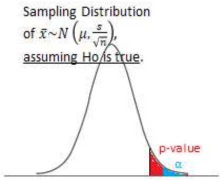
4. Hypothesis tests are prone to decision errors

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		

4. Hypothesis tests are prone to decision errors

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*, α
	H_A true		

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true: α
  - P(Type 1 Error) = α



4. Hypothesis tests are prone to decision errors

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*, α
	H_A true	*Type 2 Error*, β	

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
  - P(Type 1 Error) = α
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.
  - P(Type 2 Error) = β

4. Hypothesis tests are prone to decision errors

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*, α
	H_A true	*Type 2 Error*, β	*Power*, $1 - \beta$

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
  - P(Type 1 Error) = α
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.
  - P(Type 2 Error) = β
- *Power* is the probability of *correctly* rejecting H_0, and hence the complement of the probability of a Type 2 Error
  - Power = 1 - β

4. Hypothesis tests are prone to decision errors

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error,* α
	H_A true	*Type 2 Error,* β	*Power,* $1 - \beta$

Ex: What does a Type 1 and Type 2 error mean given the hypotheses below. Why type of error is worse?

Ho: Defendant is innocent.
Ha: Defendant is guilty.

Summary of main ideas

1. Use hypothesis tests to make decisions about population parameters
2. Hypothesis tests and confidence intervals at equivalent significance/confidence levels should agree
3. Results that are statistically significant are not necessarily practically significant
4. Hypothesis tests are prone to decision errors