Fundamentals of Hypothesis Testing

- Upper One-Sided Test
- Lower One-Sided Test
- Two-Sided Test
- Relationship to Confidence Intervals
- Factors Leading to Smaller P-Values

University of Pittsburgh

Motivation: Hypothesis Testing

- **Scenario:** SAT is designed to have normally distributed scores with population mean $\mu = 1000$ and standard deviation $\sigma = 217$.
- **Question:** How can we determine if 1000 is a plausible value for the average SAT score of Pitt students?
- **Solution #1:** Find an ______ of plausible values for what the mean SAT score of Pitt students could be
  - Calculate a ___________________
- **Solution #2:** Determine how ________ the sample mean would be under the __________ that the true population mean SAT score of Pitt students is _____
  - Perform a _____________ – an inferential procedure that tests whether a _________________ for a parameter is plausible

Null and Alternative Hypotheses

- **Null hypothesis:** initial guess made about a parameter

$$H_0 : \mu = \mu_0$$

Stated "H-naught"

Parameter

Hypothesized value of unknown parameter

- **Alternative hypothesis:** the conclusion we come to if the collected evidence indicated the null hypothesis may be false

$$H_A : \mu > \mu_0 \qquad H_A : \mu < \mu_0 \qquad H_A : \mu \neq \mu_0$$

Upper one-sided alternative

Lower one-sided alternative

Two-sided alternative

Note: *μ_0 will be replaced with the hypothesized value in both the null and alternative hypotheses.*

EXAMPLE: NULL AND ALTERNATIVE HYPOTHESES

- **Scenario:** SAT is designed to have normally distributed scores with population mean $\mu = 1000$ and standard deviation $\sigma = 217$. We want to determine if the average SAT score of Pitt students is higher than the national average.
- **Question:** What are the null and alternative hypotheses?
- **Answer:**
  - **Null Hypothesis:** ______________
    - **Interpretation:** "The ____________________________________ the national average of ______."
  - **Alternative Hypothesis:** ______________
    - **Interpretation:** "The __________________________________ than the national average of ______."

EXAMPLE: NEXT STEPS

- **Scenario:** SAT is designed to have normally distributed scores with population mean $\mu = 1000$ and standard deviation $\sigma = 217$. A random sample of 25 Pitt students finds a sample mean SAT score of 1250.
- **Question:** What can we do with this data?
- **Answer:** Three things…
  - Check if the shape of the sampling distribution is _______
    - Necessary to continue with the hypothesis test
  - Calculate the _______________
    - "How many ______________ is 1250 from the hypothesized mean of 1000?"
  - Calculate the _________
    - "How _______ is this sample mean of 1250 if the true population mean SAT score of Pitt students is actually 1000?"

EXAMPLE: SAMPLING DISTRIBUTION

- **Scenario:** SAT is designed to have normally distributed scores with population mean $\mu = 1000$ and standard deviation $\sigma = 217$. A random sample of 25 Pitt students finds a sample mean SAT score of 1250.
- **Question:** What is the sampling distribution of the sample mean?
- **Answer:**
  - **Mean:** $\mu_{\bar{X}} =$ ______
    - In inference, use the _______________ – need to know how unusual the sample mean is relative to what we believe the mean could be
  - **Standard Error:** $\sigma_{\bar{X}} =$ _____ $=$ _____
  - **Shape:** _________
    - Rule of Thumb #__ applies: Original population ___________________

TEST STATISTIC

- **Test statistic:** one criterion used to decide if a hypothesized value is a plausible value for the parameter
  - Measure of how different the statistic is from the hypothesized value of the parameter
  - Calculation changes depending on the type of test

- For testing a population mean, the test statistic is:

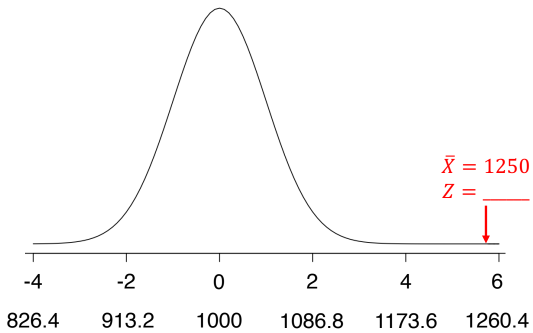
$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

  - Measures the number of standard errors that separate the sample mean and hypothesized mean

EXAMPLE: TEST STATISTIC

- **Summary:** Testing $H_0: \mu = 1000$ vs. $H_A: \mu > 1000$. Sample mean of 25 SAT scores is $\bar{x} = 1250$ and we know $\sigma = 217$.
- **Question:** What is the test statistic?
- **Answer:**

$Z =$ ________________ $=$ _______



  - **Interpretation:** If the ____________________ SAT score of Pitt students ____________, then finding a sample mean of ____ from a sample of ___ students is ________________________ the mean.

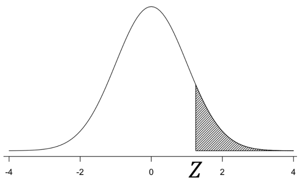
P-VALUE

- **P-value:** the probability of obtaining data as extreme or more extreme than what was originally observed **assuming the null hypothesis (H_0) is true**
  - Calculated in different ways depending on the alternative hypothesis
  - Small p-values are evidence against H_0
    - Indicates the sample mean was unlikely if the hypothesized value is correct
    - Suggests the hypothesized value may not be correct
  - Large p-values are evidence supporting H_0
    - Indicates the sample mean is not terribly unusual if the hypothesized value is correct
    - Suggests that the hypothesized value could be plausible as the true parameter

Calculating P-Values

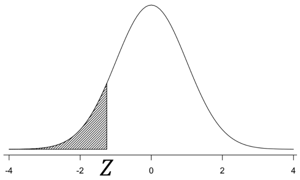
Alternative Hypothesis | **Method of Calculation**

Upper one-sided
$H_A: \mu > \mu_0$



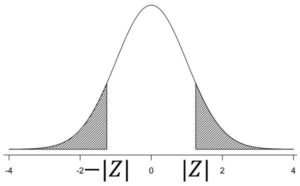
Area in ______ tail above test statistic

Lower one-sided
$H_A: \mu < \mu_0$



Area in ______ tail below test statistic

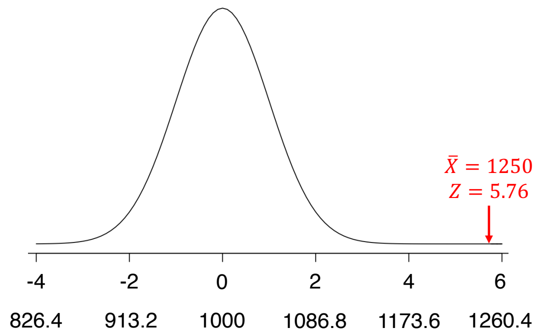
Two-sided
$H_A: \mu \neq \mu_0$



Total area in ______ tails beyond positive and negative values of test statistic

Example: P-Value

- **Summary:** Testing $H_0: \mu = 1000$ vs. $H_A: \mu > 1000$. Sample mean of 25 SAT scores is $\bar{x} = 1250$, yielding a test statistic of $Z = 5.76$.
- **Question:** What is the p-value?
- **Answer:** $p =$ ____________
  $=$ ____________



  - **Interpretation:** If the true population mean SAT score of Pitt students actually is 1000, then the ______________________________________ ____________ from a sample of 25 students is ___________
    - Either this was a _____________ event, or the true population mean is _________

Making a Decision

- If the p-value is **small**, then we **reject the null hypothesis** and side with the alternative hypothesis
  - i.e. The hypothesized value is not plausible as the true parameter
- If the p-value is **large**, then we **fail to reject the null hypothesis** and conclude that the hypothesized value is plausible as the true parameter

- **Question:** How do we make this decision?
- **Answer:** Specify a ______________________ (α), usually 1%, 5%, or 10%
  - If p-value $< \alpha$, then ________ $\boldsymbol{H_0}$ – result is _________________________
  - If p-value $> \alpha$, then ______________ $\boldsymbol{H_0}$.

Example: Interpreting the Results

- **Hypotheses:** $H_0: \mu = 1000$ vs. $H_A: \mu > 1000$.
- **Test Results:** $\bar{x} = 1250$, $Z = 5.76$, and $p = 4.20\times10^{-9}$.

- **Question:** What decision can we make regarding the average SAT score of Pitt students using a 5% level of significance?
- **Answer:** __________ because $p =$ __________________
  - Result is ________________

- **Question:** What is our final conclusion about the mean SAT score?
- **Answer:** The true mean SAT score of Pitt students is ____________ _________________.
  - While we don't know the __________, we are confident it is some value from the ____________________.

One-Sample Z-Test for a Population Mean

Step	Description
Used for	Performing inference on a single unknown population mean when the population standard deviation (σ) is known
Conditions	Shape of sampling distribution of sample mean must be normal *Use a histogram, the sample size, and the Central Limit Theorem to verify*
Test Statistic	$Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ which follows a standard normal distribution
Confidence Interval	$\bar{X} \pm Z \dfrac{\sigma}{\sqrt{n}}$

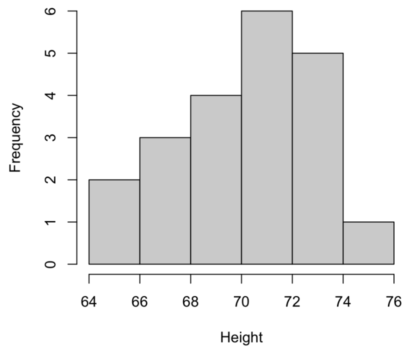
Hypothesis Testing Procedure

- We will cover 8 hypothesis tests this semester. Each is customized to the variable situation, but all follow the same procedure:

Step	Description
	Determine the appropriate test to use
1	Determine the null and alternative hypotheses
	Collect the data
2	Ensure the conditions for performing the test hold
3	Calculate the test statistic
4	Calculate the p-value
5	Calculate a confidence interval that matches the level of significance
6	Write a conclusion using the above results

EXAMPLE: LOWER ONE-SIDED TEST

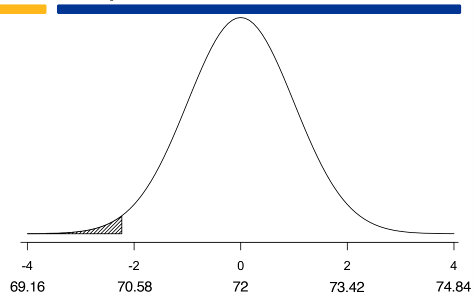
- **Scenario:** Random sample 21 American males found a sample mean height of 70.46 inches. Assume the population standard deviation is 3.25 inches.
- **Question:** Can we conclude that the average height of American males is significantly less than 72 inches using a 10% level of significance (i.e. $\alpha = 0.10$)?



- **Hypotheses:** H_0: ________ vs. H_A: ________
- **Normality Condition:** ___________
  - Rule of Thumb #__: Distribution is ________ skewed, but the sample size is at least ___

EXAMPLE: LOWER ONE-SIDED TEST (CONT.)

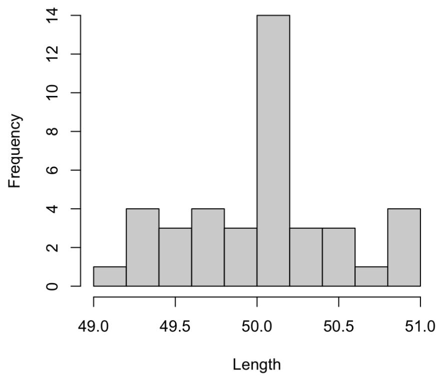
- **Values:** $\bar{x} = 70.46$, $\sigma = 3.25$, $n = 21$



- **Test Statistic:** $Z =$ __________ $=$ ______
- **P-Value:** $p =$ _____________ $=$ ________
- **90% Confidence Interval:** ________________________ $=$ _______________
- **Conclusions:**
  - ___________ because the _________ (_______) is ______ than the level of significance ($\alpha = 0.10$)
  - The _____________ of all American males is ____________________.
  - We are _____________ that the true mean height of all American males is ________________ and _________ inches.

EXAMPLE: TWO-SIDED TEST

- **Scenario:** A particular type of screw must have a length of 50 mm. We want to ensure the machine making them is under control. Assume the population SD for the length of all screws is 0.55 mm.
- **Question:** Can we conclude that the average length of the screws differ from 50 mm using a 1% level of significance?

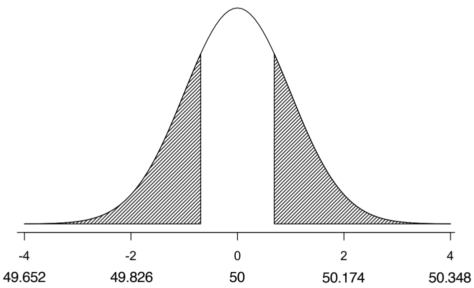


- **Hypotheses:** H_0: ________ vs. H_A: ________
- **Normality Condition:** ___________
  - Shape is mostly _________ with one __________
  - Rule of Thumb #__: Shape of distribution does __________ because sample size is at least ___

EXAMPLE: TWO-SIDED TEST (CONT.)

- **Values:**

```
> screws = read.table("screws.csv", header = TRUE, sep = ",")
> mean(screws$length)
[1] 50.06
> nrow(screws)
[1] 40
```



- **Test Statistic:** $Z =$ __________ $=$ ______
- **P-Value:** $p =$ ______________

  $=$ _________________

  $=$ ______________

  $=$ ________

- **99% Confidence Interval:** _______________________ $=$ _______________

EXAMPLE: TWO-SIDED TEST (CONT.)

- **Conclusions:**
  - ______________ because the p-value (______) is ____________ the level of significance ($\alpha = 0.01$)
  - 50 is a ______________ for the mean screw length of __________________ _______________.
  - We are _____________ that the true mean screw length is ______________ and ______ mm.

- **Important Notes:**
  - When failing to reject H_0, we **never** ________ the null hypothesis.
  - We don't know what the true mean ___________ – only that the hypothesized value _______ be the true value.

RELATIONSHIP BETWEEN HYPOTHESIS TESTS AND C.I.'S

- **Confidence interval:** provides a range of plausible values for an unknown parameter
- **Hypothesis test:** determines if a hypothesized value is plausible

Confidence Interval Result	Interpretation of Hypothesized Value
C.I. contains hypothesized value	Hypothesized value is a ________ estimate of the parameter
C.I. lies entirely above hypothesized value	True parameter is significantly _______ than the hypothesized value
C.I. lies entirely below hypothesized value	True parameter is significantly ____ than the hypothesized value

<u>Note</u>: Minor contradictions can occur with one-sided tests because confidence intervals are two-sided. This is infrequent enough that we won't worry about it.

Example: Relationship Between Hypothesis Tests and C.I.'s

- **Question:** Are the hypothesis tests and confidence intervals from the three examples consistent?
- **Answer:**

	SAT Scores	Male Heights	Screw Length
Hypotheses	$H_0: \mu = 1000$ $H_A: \mu > 1000$	$H_0: \mu = 72$ $H_A: \mu < 72$	$H_0: \mu = 50$ $H_A: \mu \neq 50$
Lev. of Sig.	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$
P-Value	$p = 4.20\times10^{-9}$	$p = 0.0150$	$p = 0.4902$
Conf. Int.	(1164, 1335)	(69.29, 71.63)	(49.84, 50.28)
Consistent?	____: Test rejected H_0 and C.I. is entirely ______________	____: Test rejected H_0 and C.I. is entirely ______________	____: Test failed to reject H_0 and C.I. __________

Benefits and Drawbacks of Smaller P-Values

- **Recall:** We normally like to have narrower confidence intervals as they provide a more precise estimate of the parameter.
- **Question:** Do we always want to attain a smaller p-value?
- **Answer:** ______________
  - **Pitt SAT Scores:** Wanted a _____ p-value to show that Pitt students scored ______ than the national average
  - **Male Height:** P-value really ___________ – nothing _________ was coming from the result as we were just seeing if the average height was less than 72 inches
  - **Screw Length:** Wanted a ______ p-value to show that the machine was ______________ and making screws the ________ length

Factors Leading to Smaller P-values

- **Question:** How can we get a smaller p-value in a hypothesis test?

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

- **Answer:** Need the test statistic to be ____________________
  - _________ difference between sample mean and hypothesized mean – more easily attained by choosing a different __________________
  - _________ standard deviation – Not always _________
  - _________ sample size – Usually the best option, but must be done _________