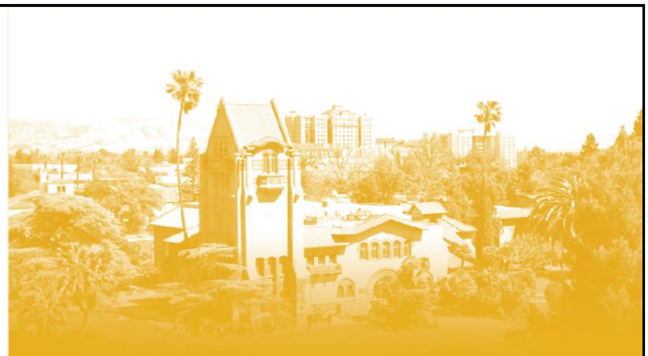
SJSU SAN JOSÉ STATE UNIVERSITY

Hypothesis Testing



1

SJSU SAN JOSÉ STATE UNIVERSITY

Agenda

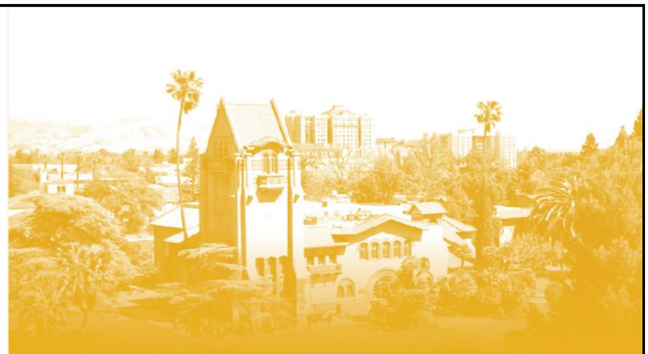
- Introduction to Hypothesis Testing
- Steps of Hypothesis Testing
- Types of Errors, Significance Levels & Power
- Examples, Tests, and Practice Problems

2

2

SJSU SAN JOSÉ STATE UNIVERSITY

Introduction to Hypothesis Testing



3

SJSU SAN JOSÉ STATE UNIVERSITY

What is Hypothesis Testing?

- A statistical method to make or test claims decisions about population parameters.
- Uses sample data to test assumptions.
- Central to scientific reasoning, quality control, and evidence-based decision-making.
- Compares hypotheses using probabilistic reasoning.

- Hypothesis test is also known as test of significance.

4

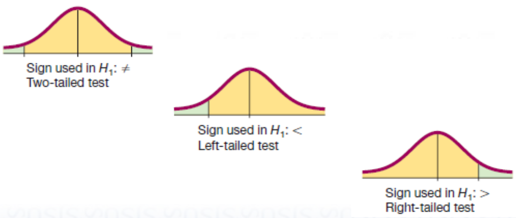
4

SJSU SAN JOSÉ STATE UNIVERSITY

Null and Alternative Hypothesis

The first step is to create the two possibilities in hypothesis testing:

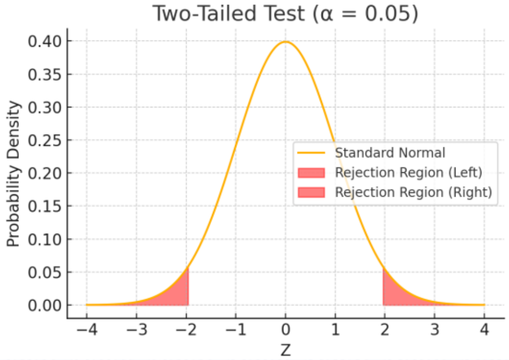
- Null hypothesis (H_0): Default claim or "Status Quo"
  - A statement that the value of a population parameter is equal to some claimed value.
- Alternative hypothesis (H_A or H_1): The claim that we seek evidence
  - A statement that the parameter has a value that somehow differs from the null hypothesis.
- Types:
  - Two-tailed: H_A: $\mu \neq \mu_0$
  - Left-tailed: H_A: $\mu < \mu_0$
  - Right-tailed: H_A: $\mu > \mu_0$



SJSU SAN JOSÉ STATE UNIVERSITY

One-Tailed vs. Two-Tailed Tests

- Two-tailed: Difference in either direction
- One-tailed: Directional difference



SJSU SAN JOSÉ STATE UNIVERSITY

Examples of Hypotheses

- Lifespan of LED bulbs:
  - H_0: $\mu \geq 1000$ hrs
  - H_A: $\mu < 1000$ hrs
- Drug effectiveness:
  - H_0: μ = placebo mean
  - H_A: $\mu \neq$ placebo mean
- Defect rate:
  - H_0: $p \leq 0.05$
  - H_A: $p > 0.05$

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Defining the Null and Alternative Hypothesis

Consider the claim that a medical procedure designed to increase the likelihood of a baby girl is effective, so that the probability of a baby girl is $p > 0.5$.

Define the null and alternative hypotheses.

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Can Dogs Smell Bladder Cancer?

- A study by M. Willis et al. considered whether dogs could be trained to detect if a person has bladder cancer by smelling his/her urine.
- 6 dogs of varying breeds were trained to discriminate between urine from patients with bladder cancer and urine from control patients without it.
- The dogs were taught to indicate which among several specimens was from the bladder cancer patient by lying beside it.
- Once trained, the dogs' ability to distinguish cancer patients from controls was tested using urine samples from subjects not previously encountered by the dogs.

9

9

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Can Dogs Smell Bladder Cancer?

- The researchers blinded both dog handlers and experimental observers to the identity of urine samples.
- Each of the 6 dogs was tested with 9 trials. In each trial, one urine sample from a bladder cancer patient was randomly placed among 6 control urine samples.
- Outcome: In the total of 54 trials with the 6 dogs, the dogs made the correct selection 22 times.
- The dogs were correct for 22/54 ≈ 41% of the time ➔ not fabulous
- If the dogs just guessed at random, they were only expected to be correct for 1/7 ≈ 14% of the time
- Is this difference (41% vs 14%) surprising?

10

10

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Can Dogs Smell Bladder Cancer?

Let p be the probability that a dog makes the correct selection on a given trial.

- Null hypothesis (H_0): $p = 1/7$

  "There is nothing going on." ➔ "null" means "nothing surprising is going on".

  The dogs just guessed at random ➔ lucky to make more correct selections than expected.

- Alternative hypothesis (H_A or H_1): $p > 1/7$

  "There is something going on."

  Dogs can do better than random guessing

SJSU SAN JOSÉ STATE UNIVERSITY

Weighing Evidence Using a Test Statistic

After defining H_0 and H_A, the next step of hypothesis testing is to weigh the evidence ➔

How likely the observed data could have occurred if H_0 was true?

- If the observed result was very unlikely to have occurred under the H_0, then the evidence raises more than a reasonable doubt in our minds about the H_0.

The **test statistic** is a summary of the data that best reflects the evidence for or against the hypotheses.

For the example, the test statistics that is chosen:

X = the total number of correct selections in the 54 trials

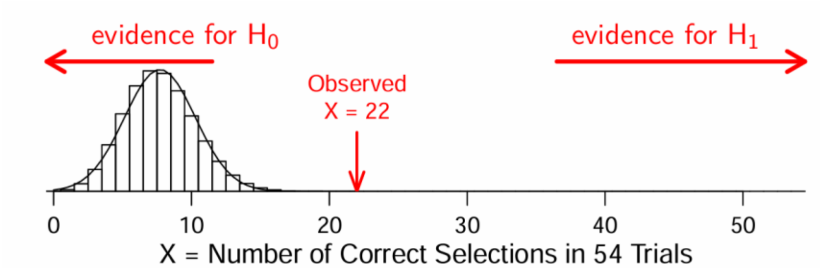
A larger X value is a stronger evidence for H_A and against H_0

Distribution of the Test Statistic Under H_0

For the "Dogs Smell Cancer" example, if H_0 is true, then

$$X \sim \text{Binomial}(n = 54, p = 1/7)$$

which implies



Example: Can Dogs Smell Bladder Cancer?

- What's the conclusion of this example?
- Any evidence that dogs have some ability to smell bladder cancer?
- If so, how practical is it?

SJSU SAN JOSÉ STATE UNIVERSITY

Steps in Hypothesis Testing

- Formulate H_0 and H_A
- Choose significance level α
- Select appropriate test statistic (e.g. z, t, etc)
- Compute the test statistic and p-value
- Decision: **Reject** or **Fail** to reject H_0

Parameter	Sampling Distribution	Requirements	Test Statistic
Proportion p	Normal (z)	$np \geq 5$ and $nq \geq 5$	$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$
Mean μ	t	σ not known and normally distributed population *or* σ not known and $n > 30$	$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$
Mean μ	Normal (z)	σ known and normally distributed population *or* σ known and $n > 30$	$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
St. dev. σ or variance σ^2	χ^2	Strict requirement: normally distributed population	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

15

15

SJSU SAN JOSÉ STATE UNIVERSITY

Test Procedure & Rejection Region

A test procedure is specified by the following:

- a test statistic
- a rejection region

The null hypothesis H_0 will be rejected if & only if the test statistic falls in the rejection region.

A sensible rejection region is of the form

$X \geq k$ for some cutoff k

and the test procedure is reject H_0 if $X \geq k$

16

16

SJSU SAN JOSÉ STATE UNIVERSITY

Type I and Type II Errors

In a hypothesis test, we make a decision about which of H_0 or H_A might be true, but our decision might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type I Error*
	H_1 true	*Type II Error*	✓

- Type I Error (α): rejecting H_0 when it is true.
- Type II Error (β): failing to reject H_0 when it is false.

17

17

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Type 1 and 2 Errors

Consider the claim that a medical procedure designed to increase the likelihood of a baby girl is effective, so that the probability of a baby girl is $p > 0.5$.

What are the Type I and Type II Errors?

18

18

SJSU SAN JOSÉ STATE UNIVERSITY

Interpretations of Type 1 and 2 Errors

Type 1 and Type 2 errors are different sorts of mistakes and have different consequences:

- Usually H_0 is the status quo, thing we generally believe to be true
- If H_0 is not rejected → the status quo is fine. No action needs to be taken
- Rejecting H_0 means something we used to believe is overturned. It might be a scientific breakthrough (e.g. discovery of a new drug).
- A Type 1 error introduces a false conclusion that can lead to a tremendous waste of resources before further research invalidates the original finding.
- A Type 2 error represents a missed opportunity for scientific progress
- Type 2 errors can be costly as well, but generally go unnoticed

19

19

SJSU SAN JOSÉ STATE UNIVERSITY

Significance Level of Hypothesis Testing

- The Significance Level α for a hypothesis test is the probability value used as the cutoff for determining when the sample evidence constitutes significant evidence against the null hypothesis H_0.

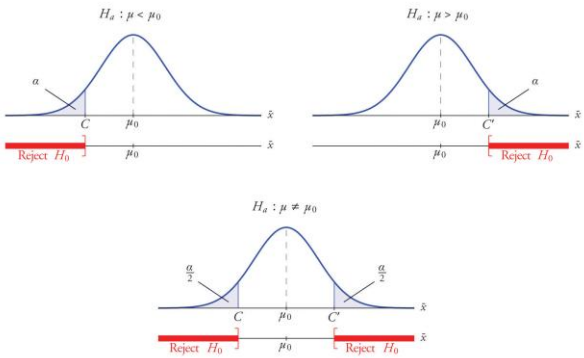
$$\text{Significance Level } \alpha = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) = P(\text{Type 1 error})$$

- The probability of rejecting the null hypothesis H_0 when it is actually true (Type I error).

20

SJSU SAN JOSÉ STATE UNIVERSITY

Significance Level of Hypothesis Testing



- Common choices of α are 0.01, **0.05**, 0.10.
- Controls how "strict" the test is in ruling out random chance (Smaller α means stricter criteria for significance)

21

SJSU SAN JOSÉ STATE UNIVERSITY

Different Significance Levels

- Using a different value for α can change whether we reject H_0 or not.

Example: A cigarette company claims that the average nicotine content is 1.5 mg/cigarette.

$$H_0: \mu = 1.5$$
$$H_A: \mu > 1.5$$

Significance Level α	Rejection Region	Conclusion for z = 2.10	Significance level preferred by
.05	$z_1 > 1.645$	Reject H_0	Anti-tobacco activists
.025	$z_2 > 1.96$	Reject H_0	Anti-tobacco activists
.01	$z_3 > 2.33$	Fail to reject H_0	Tobacco company
.005	$z_4 > 2.58$	Fail to reject H_0	Tobacco company

22

22

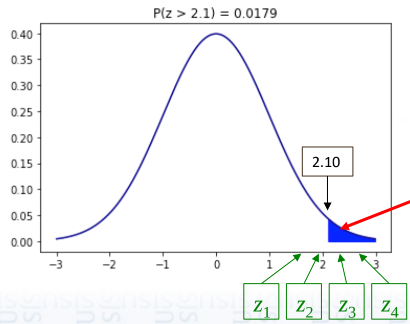
SJSU SAN JOSÉ STATE UNIVERSITY

Different Significance Levels

Significance Level α	Rejection Region	Conclusion for z = 2.10	Significance level preferred by
.05	$z_1 > 1.645$	Reject H_0	Anti-tobacco activists
.025	$z_2 > 1.96$	Reject H_0	Anti-tobacco activists
.01	$z_3 > 2.33$	Fail to reject H_0	Tobacco company
.005	$z_4 > 2.58$	Fail to reject H_0	Tobacco company

H_0: $\mu = 1.5$
H_a: $\mu > 1.5$

The tobacco company can claim that they failed to reject H_0. But the company probably won't tell you what significance level it used for testing.



This area 0.0179 is the smallest α at which we can reject H_0. Any smaller area would not include 2.10

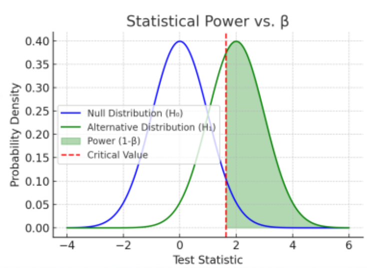
SJSU SAN JOSÉ STATE UNIVERSITY

Power of Hypothesis Testing

- The **Power** for a hypothesis test is the probability correctly rejecting the null hypothesis H_0 when the alternative hypothesis is true H_A.

Power = 1 - P(Type II error) = 1 - β

- The probability of avoiding a Type II error (Correction rejection)
- Higher power means a greater ability to detect true effects.
- Power increases with larger sample sizes and lower variability.
- Higher α ➔ higher power



SJSU SAN JOSÉ STATE UNIVERSITY

Controlling Type 1 Errors

How to Control Type I Error:

- Set a Smaller Significance Level α
  - Common values: 0.05, 0.01, 0.001
  - Example: Use α = 0.01 to reduce false positives in critical medical trials
- Use Two-Tailed Tests When Appropriate
  - Less aggressive than one-tailed, splits α between both tails
- Ensure Assumptions of the Test Are Met
  - Incorrect assumptions (e.g., non-normality) can inflate error rates
- Apply Multiple Comparison Corrections (when testing many hypotheses)

25

25

SJSU SAN JOSÉ STATE UNIVERSITY

Controlling Type 2 Errors

How to Control Type II Error:

- Increase Sample Size n
  - Reduces standard error → increases test sensitivity → lower β.
- Increase Effect Size (if feasible)
  - Sometimes through better instrumentation, experimental design, or stronger treatments.
- Choose a Higher Significance Level α
  - Increases power, but also increases Type I error risk ➔ trade-off must be justified
- Use More Powerful Tests
  - Parametric tests (t-test, z-test) are more powerful than non-parametric ones when assumptions are satisfied

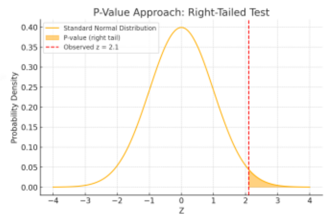
26

26

SJSU SAN JOSÉ STATE UNIVERSITY

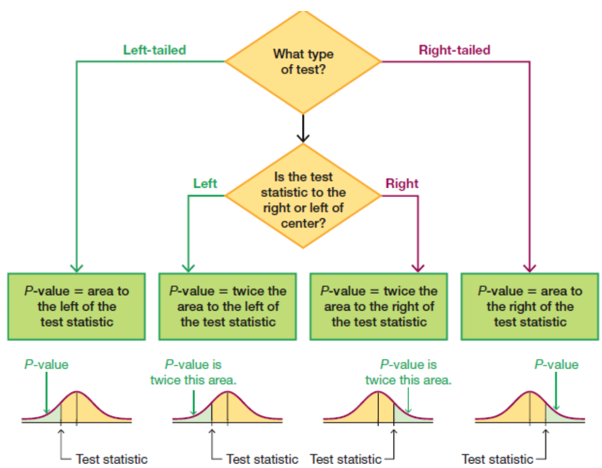
The P-Values and Hypothesis Testing

- The p-value of test is the probability of getting a value of the test statistic that is at least as extreme as the test statistic obtained from the sample data, assuming that the null hypothesis H_0 is true.
- It quantifies the "strength" of the evidence against H_0 (helps us decide whether to reject H_0)
  - A small p-value (usually less than the significance level, such as 0.05) indicates strong evidence against H_0.
  - A large p-value suggests weak evidence against H_0 (fail to reject it).
- Compare the p-value to the significance level α:
  - If p-value $\leq \alpha$: Reject H_0 (evidence supports H_A) at level α.
  - If p-value $> \alpha$: Fail to reject H_0 (evidence is insufficient to support H_A).
- The p-value is the smallest level of significance at which H_0 can be rejected.



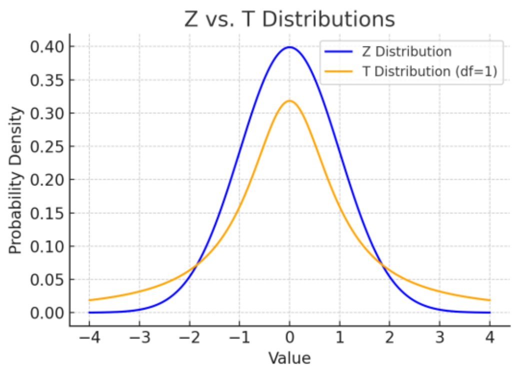
SJSU SAN JOSÉ STATE UNIVERSITY

Finding P-Values



Recall z vs t Tests

- z-test: Known σ or large n
- t-test: Unknown σ, small n
- Use *t-distribution* for small samples



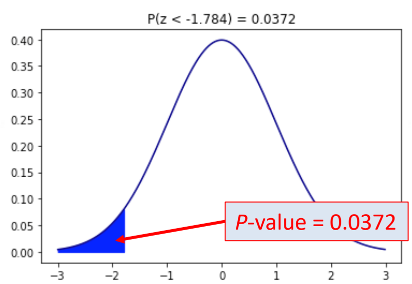
29

Example: P-Value

Given that:

H_0: $\mu = 125$
H_A: $\mu < 125$

It's reported that the p-value = 0.0372. What's the smallest α at which we can reject H_0?



30

30

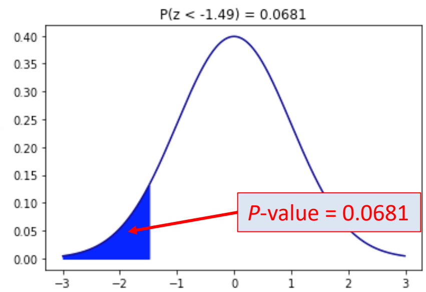
SJSU SAN JOSÉ STATE UNIVERSITY

Example: P-Value

Given that:

H_0: $\mu = 10$
H_A: $\mu < 10$

Suppose that a z test for testing results in the test statistic z = -1.49. What's the smallest α at which we can reject H_0?



SJSU SAN JOSÉ STATE UNIVERSITY

Example

A company claims that their light bulbs last an average of 1,000 hours. You suspect the true mean lifespan is less than 1,000 hours. You take a sample of 40 bulbs and find that the sample mean and standard deviation are 980 hours and 50 hours, respectively. The significance level is 0.05. Prove or disprove your suspicion.

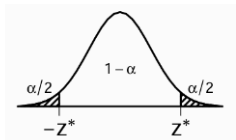
SJSU SAN JOSÉ STATE UNIVERSITY

Relationship Between Confidence Intervals and 2-Sided Hypothesis Testing

For a two-sided test:

H_0: $\mu = \mu_0$
H_A: $\mu \neq \mu_0$

The following are equivalent:



- p-value > α (and hence H_0: $\mu = \mu_0$ is not rejected at level α)
- |z-statistic| = $|(\bar{X} - \mu_0)/SE| < z^*$, where z^* is a value such that
- μ_0 is in the $100(1-\alpha)\%$ confidence interval for μ:

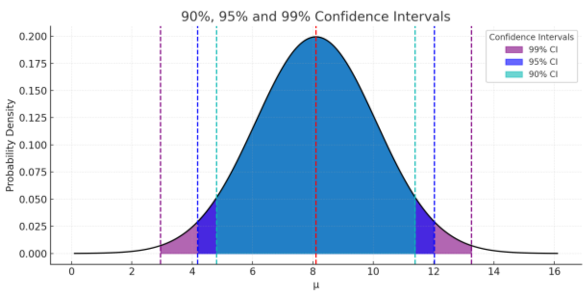
$$\bar{X} - z^* \cdot SE < \mu_0 < \bar{X} + z^* \cdot SE$$

SJSU SAN JOSÉ STATE UNIVERSITY

Relationship Between Confidence Intervals and 2-Sided Hypothesis Testing

Suppose in a study, we have the following

- 90% CI for μ: (4.81, 11.39)
- 95% CI for μ: (4.18, 12.02)
- 99% CI for μ: (2.95, 13.25)



Then

- H_0: $\mu = 4$ is rejected at 5% level but not at 1% level (2-sided p-value is between 1% and 5%)
- H_0: $\mu = 4.5$ is rejected at 10% level but not at 5% level

SJSU SAN JOSÉ STATE UNIVERSITY

Discussions on Hypothesis Testing

- Rejecting H_0 doesn't mean we are 100% sure that H_0 is false. We might make Type 1 errors. Setting a significance level just guarantee we won't make Type 1 error too often.
- P-value is not P(H_0 is true | data) but it is P(data | H_0 is true).
- Another mistake is to conclude from a high p-value that the H_0 is probably true.
- If our p-value is high, can we conclude that H_0 is true?
  - No, we could make a Type 2 error when failing to reject H_0
  - Moreover, unlike Type 1 error rate is controlled at a low level, Type 2 error rate is usually quite high. It is quite often that H_0 is not true but the data fail to reject it.
- When we fail to reject H_0, often it just means the data are not able to distinguish between H_0 and H_A (because the data are to noisy, etc)

35

35

SJSU SAN JOSÉ STATE UNIVERSITY

Summary: Hypothesis Testing Procedure

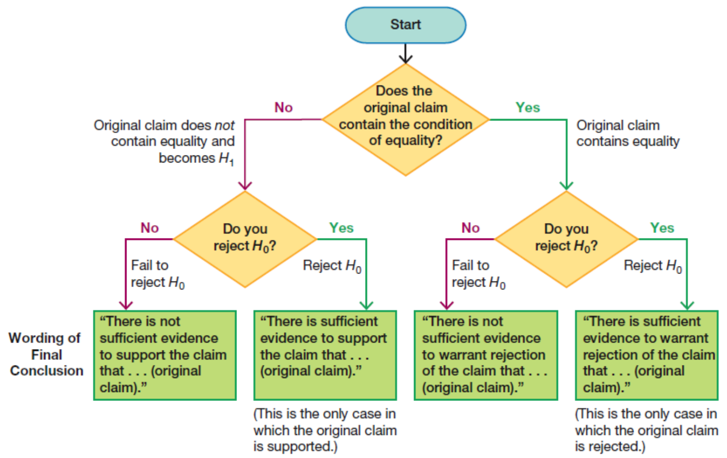
- Start with a null hypothesis (H_0) that represents the status quo or default claim.
- State an alternative hypothesis (H_A) that represents our research question, i.e. what we're testing for.
- Collect data and often summarize the data as a test statistic, which is usually a measure gauging whether H_0 or H_A are more plausible.
- Determine the sampling distribution of the test statistic assuming H_0 is true.
  - If the test statistic is too far away from what the H_0 predicts, then reject the H_0 in favor of the H_A.
- Choose a significance level α = maximal P (Type I error) that we can tolerate.
- Set the rejection region based on the significance level.
- Reject H_0 if the test statistic falls in the rejection region, and do not reject otherwise.

36

36

SJSU SAN JOSÉ STATE UNIVERSITY

Hypothesis Tests: Wording of Final Conclusion



37

37