Introduction to Hypothesis Testing

"Must you answer every question with a hypothesis?"

Research Questions & Hypotheses

- Research question:
  - A question that the research project sets out to answer
- Hypotheses:
  - Always statements, not questions
  - Statements that we can actually test
  - Usually specify:
    - Population of interest
    - Variable(s)
    - Expected relationship/difference (or lack thereof)

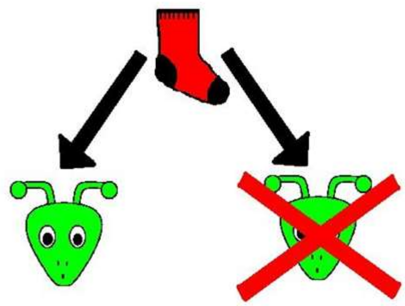
Null Hypothesis

- Null hypothesis = a hypothesis of no relationship or no difference
- We assume that there is no relationship until we have evidence to reject that assumption

Research Hypothesis

- Research hypothesis – the relationship or difference that we think we could find
- Always a statement that there is a difference or a relationship
- This is what we expect without looking at the data (theory, literature, prior research)

Q. Where have all my socks gone?



Research Hypothesis = Extra-terrestrial beings have transported themselves into my house in order to steal my socks.

Null Hypothesis = Aliens are not to blame. There is some other explanation for the disappearing socks.

Two Types of Research Hypotheses

- Non-directional: we expect a difference but have no expectations regarding a specific direction
- Directional: we can specify the direction of the difference

Rules of Writing Hypotheses

- State them in declarative form
- Research hypothesis: a relationship between variables or difference between groups
- Null hypothesis: no relationship or no difference
- Specify your population of interest
- We make hypotheses about populations, not samples → use symbols for population parameters (e.g., μ)
- Be careful with double comparisons:
  - Confusing: White women spend more time on housework than men compared to African Americans
  - Clear: Among Whites, the gender gap in housework is higher than among African Americans

Writing Hypotheses: Example

- Question: How do employed women's incomes differ depending on whether they are mothers in the U.S.?
- H_0: Employed mothers and employed women without children in the U.S. earn the same income.
- H_1 non-directional: Employed mothers and employed women without children in the U.S. differ in their incomes.
- H_1 directional: Employed mothers in the U.S. earn lower incomes than employed women without children.

Hypotheses in Journal Articles

- Null hypotheses are often not explicitly stated
- A study can include multiple research hypotheses, sometimes competing

Notation for Hypothesis Writing

Null hypothesis for comparing two groups:
H0: $\mu_1 = \mu_2$

Alternative/research hypothesis:

H1: $\mu_1 > \mu_2$
or
H1: $\mu_1 < \mu_2$
} directional

or

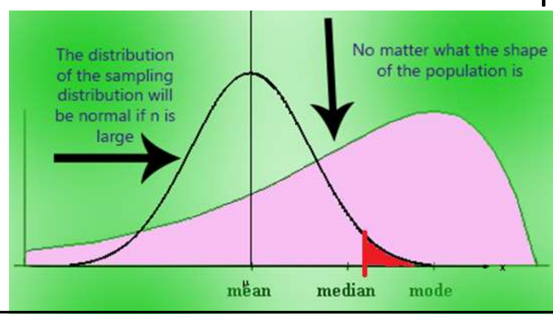
H1: $\mu_1 \neq \mu_2$ } non-directional

Testing Hypotheses: Basic Example

- We know that the average height of women in the U.S. is 64”. In our sample of 100 women enrolled in college, average height is 65.5” and SD=5. Can this be due to the chance or does this mean that women who are in college are taller than American women in general?
- In other words: Does our sample come from a population with a mean of 64” or a different population with another mean?

Combining Central Limit Theorem and Normal Curve Problems

- If we know the population mean, we can find out how likely it is to draw a sample with a specific sample mean
- That would require finding percentage of scores that fall at or above that sample mean



Hypotheses and Chance

- Say, we do not know the "true" population mean, but we want to test whether it's possible that it equals to a certain number → that's our null hypothesis
- From data, we have our sample mean and SD
- We can construct sampling distribution (based on H0 & CLT):
  - mean = hypothesized population mean
  - SEM = sample SD divided by square root of sample size
  - shape = normal
- What are the chances of getting a sample mean like that if the true population mean is what our null hypothesis tells us?

Normal Curve Problem

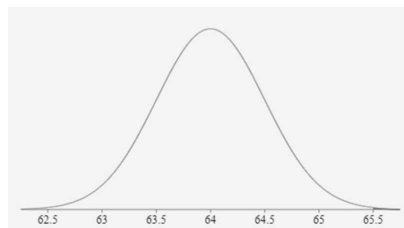
- This is a normal curve problem! We have X (sample mean) and need to calculate the probability of getting a value as large or larger (or as small or smaller)
- If this probability is very low, perhaps the null hypothesis is not a good assumption → reject it

Using Normal Curve for Our Example

- We start by assuming the true population mean of women's height is 64" → that's our null hypothesis, H0: μ=64
- We need to focus on the sampling distribution → it is normal, the mean is as per H0 assumption, and SE=SD/sqrt(n) (and we know SD=5)
- What are the chances of obtaining a sample mean (with n=100) of 65.5 or larger?
- https://goo.gl/forms/ukPXOCBz3EyaNpVm1

Calculations

- We focus on sampling distribution, so its standard deviation is SEM = SD/sqrt(n) = 5/sqrt(100)=0.5
- Z-score for 65.5 = (65.5-64)/0.5 = 3
- We know that a z-score of 3 is fairly rare: in table B1, we find area 49.87; 50-49.87=0.13, so probability of 65.5” or more = .0013



Conclusion

- That probability is called p-value = chances of getting such a sample mean or larger assuming that the population mean is equal to 64”
- The chances are very low→ so we say it’s too unlikely and reject H0
- Women in college appear to be significantly different in terms of their height from the US population of women as a whole

P-value

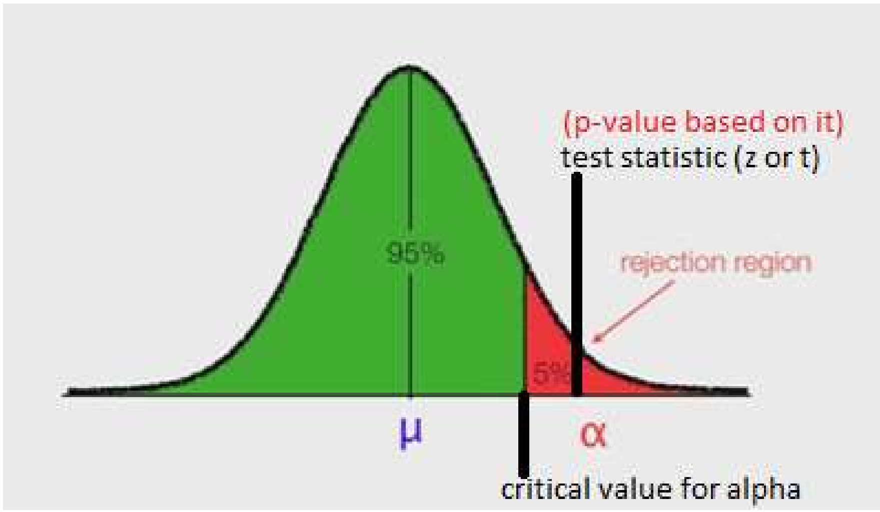
- P-value = probability of obtaining (by chance) an outcome that is as at least as extreme as the one observed (provided our assumptions about population are true)
- The smaller p-value, the more unlikely the outcome

How Unlikely is Too Unlikely?

- Pick a cutoff ahead of time for when to consider something "too unlikely" and reject H0
- We decide how much risk to take when rejecting H0: are we willing to be wrong in 5% of cases? In 1% of cases? In 0.1% of cases?
- Similar to confidence level

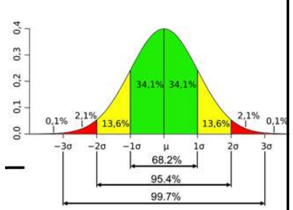
vs probability of error

(alpha)



What is a Test Statistic?

- If H0 is correct, then:
  - Observed value = Expected value + Chance Error
  - Observed = our data, expected = based on H0
  - E.g., 65.5=64+random chance error
- We estimate the size of chance error using standard error as a unit
- Test statistic = how many standard errors away the observed value is from expected



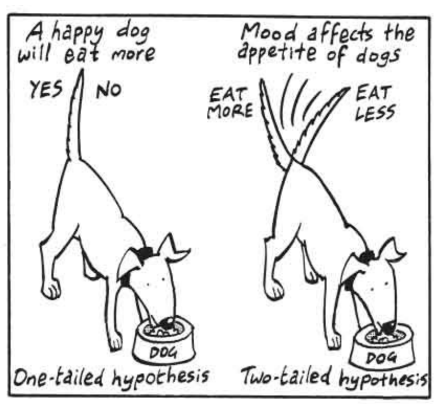
- Therefore, test statistic = (observed – expected)/standard error
- In our example, we calculated z=3 → that was our test statistic

Statistical Significance

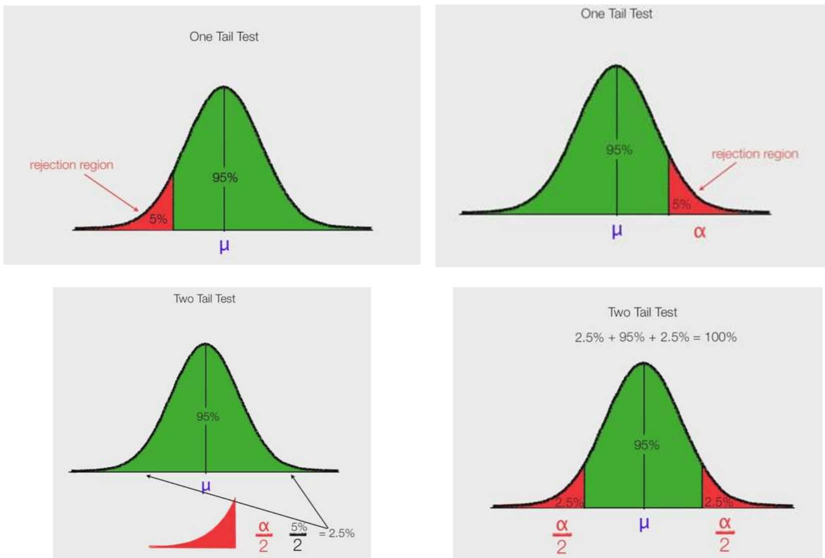
- Statistical significance: if we reject the null hypothesis →the relationship/effect specified in our research hypothesis is statistically significant
- That means we can be reasonably certain it exists in the population
- Level of statistical significance = alpha, probability of Type I error
- Typical levels of significance: 0.05, 0.01, 0.001, less common: 0.10.
- Often indicated as $p < .05$ – that means probability (p-value) of Type I error is less than .05 (so less than 5% chances)

Two-tailed vs One-tailed Test

- If your research hypothesis is non-directional → two-tailed test
- If your research hypothesis is directional → one-tailed test



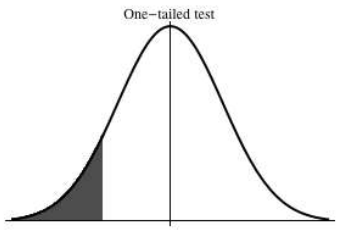
Two-tailed vs One-tailed Test

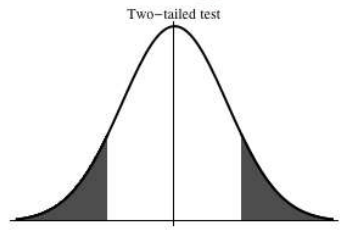


Two-tailed vs One-tailed Test

- Two-tailed test is more conservative → rejection region on each side is smaller → more difficult to reject the null
- We only use one-tailed test if we have <u>prior</u> reasons (theory, literature, prior research) to hypothesize a direction
- But if we do, definitely choose one-tailed test → more powerful

One-tailed vs Two-tailed Test

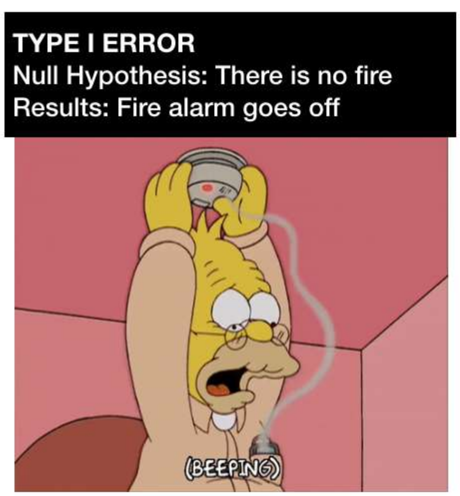


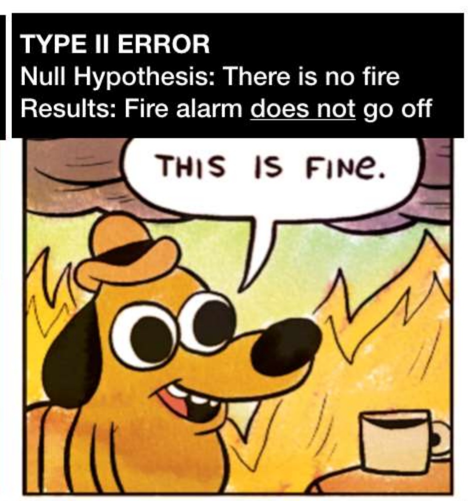


More powerful

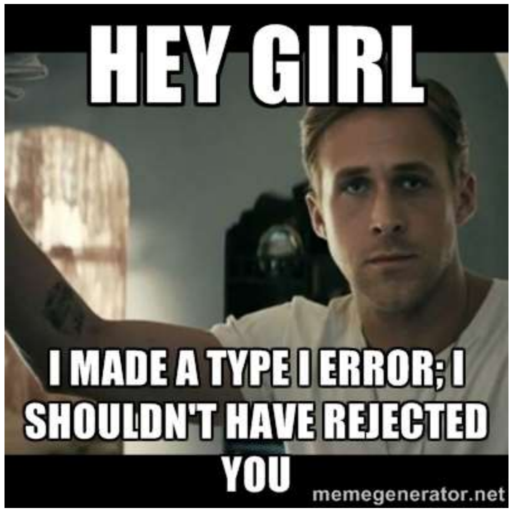
More conservative

Type I and II Errors





Level of Risk When Rejecting H0: "probability of Type I error" = alpha (α)



Type I and Type II Errors

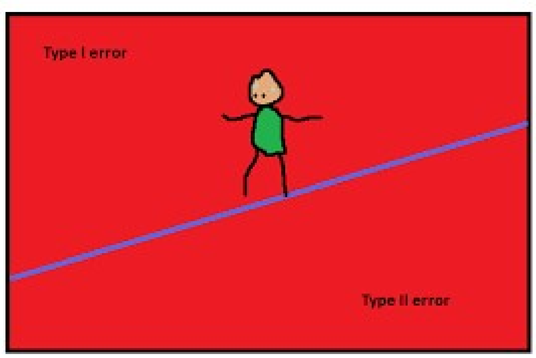
HYPOTHESIS TESTING OUTCOMES	Reality: The Null Hypothesis Is True	Reality: The Alternative Hypothesis is True
Research: The Null Hypothesis Is True	Accurate $1 - \alpha$	Type II Error β
Research: The Alternative Hypothesis is True	Type I Error α	Accurate $1 - \beta$

- α = chance of a false positive
- β = chance of a false negative
- 1 – β = power of a statistical test

Why Can't We Be 100% Right?

- Why can't we have an $\alpha=0$?
- In a given sample, any outcome is possible
- To be 100% certain, we need an infinitely large z score → we can never reject H0
- And if H0 is actually false, and we can't reject it, that is also an error: "Type II error"
- So we settle for having a small chance of Type I error in order to reduce the risk of Type II error

We Need a Balance



So How Do We Make Decisions about Errors?

- We only decide on Type I error (alpha)
- That translates into a specific chance of Type II error, typically larger than the chance of Type I
- That is how we want it – our hypothesis test is more conservative that way
- Typical alpha levels used: 0.05 (most common), 0.01, 0.001, and sometimes 0.10 (especially with smaller samples)

Type II Error: Approximate Size

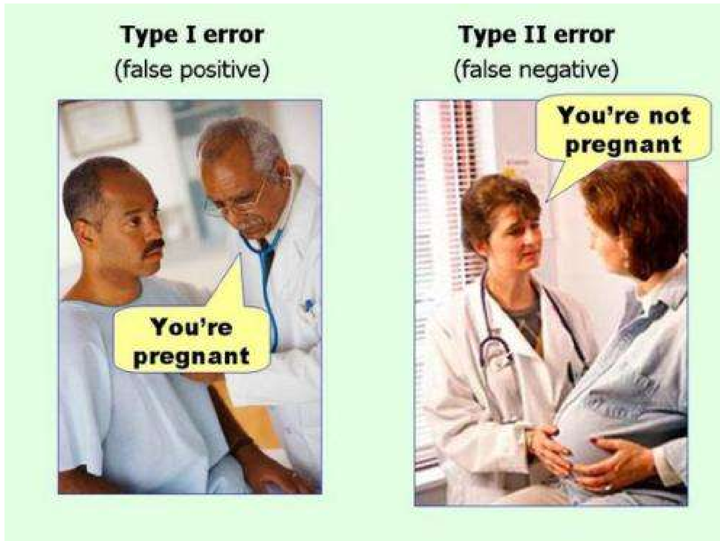
Sample Size	***Type II Error Probability***
<30	very large
30-100	large
101-500	moderate
501-1000	small
1001+	very small

What Affects the Chances of Type I and Type II Error?

- Type II error decreases:
  - if the sample size increases
  - if we allow for a larger chance of Type I error (select larger alpha)
- Type I error:
  - Only depends on the alpha we selected

Once We Are Done with Our Test:

- If we rejected H0, the chance of Type I error = α; the chance of Type II error = 0
- If we failed to reject H0, the chance of Type I error = 0; the chance of Type II error = β
- That is, you can only have a false positive when the decision is positive (YES, reject H0), and you can only have a false negative when the decision is negative (NO, fail to reject H0)


Type I error
(false positive)
You're pregnant
Type II error
(false negative)
You're not pregnant