# The Logic of Statistical Inference--Testing Hypotheses

- Confirming your research hypothesis (relationship between 2 variables) is dependent on ruling out
  - Rival hypotheses
  - Research design problems (e.g. measurement error, non-representative sample), and/or
  - Chance—*sampling error*--the natural tendency of any sample to differ from the population from which it was drawn

# Statistical inference

- The use of *theoretical sampling distributions* to test hypotheses
- *Theoretical sampling* is based on the premise that no relationship exists between the two variables

# Wait--I thought the research hypothesis said there was a difference…!

- **Null hypothesis--**a statement that there is no relationship between two variables of interest. Another way of saying it:
- Any relationship between these variables is only due to *chance*, not a real relationship that exists *in the population* (i.e. sampling error)

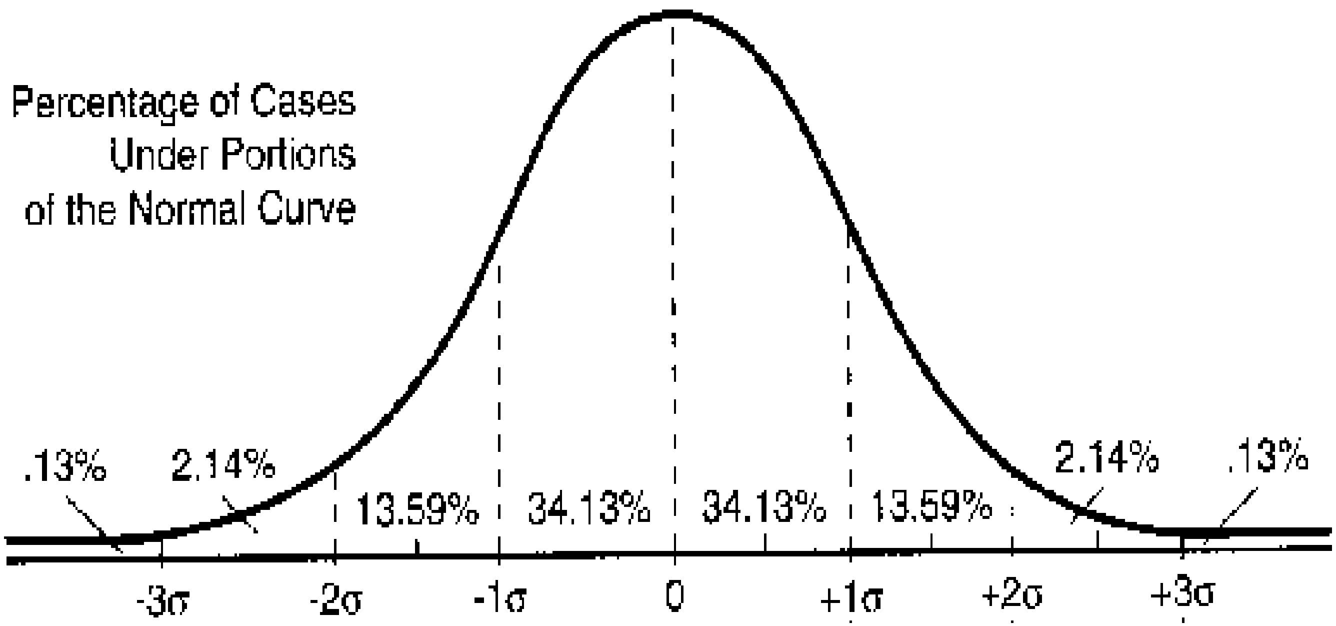
# The opposite of “Null” Is…

- The “research hypothesis”, a.k.a “alternative hypothesis”
- This is the one we worked on last semester, e.g.
  - “There is a difference between these two variables (e.g. “There is a difference in outcomes, comparing the experimental tx and ‘tx as usual’”), OR
  - “The experimental treatment will result in an improved outcome”
- Which one is one-tailed? Two-tailed? (Hint: remember “directional hypothesis”?)

# Where do the “tails” come from?

- From the theoretical sampling distribution
- It’s a frequency polygon that represents the “space” for the entire population of statistics results for a ratio/interval variable
- It shows out of all the “area under the curve,” the tiny probability of rejecting the Null hypothesis in favor of the research hypothesis

# Area under the normal curve



# Statistical inference

- So—inferring whether or not a relationship between variables exists in the population, from your sample, requires disproving or rejecting the Null Hypothesis (“Innocent until proven guilty”)…
  - By calculating (or computing) a test statistic
  - Then locating where the statistic falls in the theoretical sampling distribution, and from that
  - Determining the likelihood (probability) that the statistical result you found is due to chance alone (sampling error)

# What’s a *p* value?

- Probability: the likelihood that an event will occur (# actual events ÷ # possible events)
- How do we use probability in inference testing?
  - To quantify our confidence that our statistical result is not just due to sampling error (chance)
  - To confirm or disconfirm our hypotheses

# The tiny probability…

- If a statistic result falls in the tiny “critical region” then *there is low probability that our results are due to chance alone* meaning that there is a good chance there is *a positive relationship between variables,* and we can reject the Null Hypothesis (Read this aloud a few times)
- What is the cutoff? How tiny is tiny? We set a threshold for the critical region ahead of time, called the “alpha level”
  - $\alpha = .05$ is typical in social sciences research
  - In some cases higher, $\alpha = .10$

# Interpreting the *p* value

- Each statistic result is accompanied by a *p* value
- SPSS gives you the actual *p* value by using the statistic's computation formula and the distribution tables for the statistical test you've chosen
- If your actual *p* value (from SPSS) equals or is smaller than your alpha, then we can say the null hypothesis can be rejected

For example: The experimental group's outcome improved by 10 points, the control group by only 2. Let's say the difference in post-treatment scores has a *p* value of .046. So:

- "The probability is less than 5 in 100 ($p$ = .046) that the difference between the groups is due to chance alone. We can reject the Null (that there is no difference) in favor of the alternative (one-tailed) hypothesis, that treatment outcomes will improve more for the experimental group."

# Summary—the 8 steps to hypothesis testing

1. Identify your independent variable(s)
2. Identify your dependent variable
3. State the Null Hypothesis
4. State the Alternative Hypothesis
5. Identify appropriate statistical test and alpha level
6. Review results (SPSS output)
7. Describe results and decision to reject or not reject Null
8. Discuss results and implications

# Which statistics?

- Using the area under the Normal curve to determine this "critical region" has an important requirement—the data must be "normally distributed" in the population, e.g. when plotted on a frequency polygon the line should follow the normal curve.
- At the very least, the data must be *ratio* or *interval*
- Relevant statistics for these data include t-tests, Anova, and linear regression (all coming up in future exciting classes)

# What about non-ratio/interval data?

- Like *nominal* data, for example
- Nominal level data (gender; satisfied vs. not; ethnicity; receiving services vs. not) have another type of distribution (by definition it is skewed, not symmetric), called the *Binomial Distribution*
- But the basic logic of inference testing is the same
- It requires non-parametric statistical tests (vs. parametric tests for normally distributed data), like The Chi-Square Test of Association

# Chi Square

Also known as:

- Chi Square Test of Association
- Chi Square Test of independence
- $X^2$
- Crosstabulation (Chi Square is one type)

For use with:

- Two or more <u>nominal</u> level variables
- Typically used to describe sample
- Generally does not address causality

# Example of Chi Square Used to Describe Sample (In Red)

| | Condition (Randomly Assigned) | | |
|---|---|---|---|
| | **Experimental** | **Control** | **Difference** |
| Age | **43.5**<br>**SD = 7.8** | **43.5**<br>**SD = 4.7** | **No diff** |
| Gender | | | **No diff** |
| Male | **61%** | **60%** | |
| Race | | | **No diff** |
| White | **37%** | **38%** | |
| Black | **61%** | **62%** | |
| Hispanic | **2%** | | |
| Employed (at least one day per month | **25%** | **19%** | **No diff** |
| Has marketable skill or trade | **66%** | **66%** | **No diff** |
| Current driver's license | **26%** | **11%** | ***P* < .05** |

Excerpted from: Zanis, D. A., Coviello D., Alterman, A. I., & Appling, S. E. (2001). A community-based trial of vocational problem-solving to increase employment among methadone patients. *Journal of Substance Abuse Treatment*, *21*, 19-26.

# Example of Chi Square Used to Show Results

**Table II.** Gender Differences in Social Integration: African American vs. White

| | % women ≥ once per week | | % men ≥ once per week | |
|---|---|---|---|---|
| | African American | White | African American | White |
| Visits by friends | 46.76 | 44.33* | 51.27 | 42.32* |
| Visits to friends | 37.43 | 37.36 | 50.33 | 38.89* |
| Phone close friends or relatives | 82.33 | 89.21* | 68.34 | 75.32* |
| Church, clubs, lodges, other groups | 60.76 | 53.25* | 43.91 | 43.91 |
| Someone to share private feelings, concerns | 82.69 | 88.23* | 81.65 | 84.96* |

*Note:* Differences tested via chi square: (1) African American women vs. White women; (2) African American men vs. White men.
*$p < .05$.

From: Snowden, L. R. (2001). Social embeddedness and psychological well-being among African Americans and Whites. *American Journal of Community Psychology, 29* , 519-536.