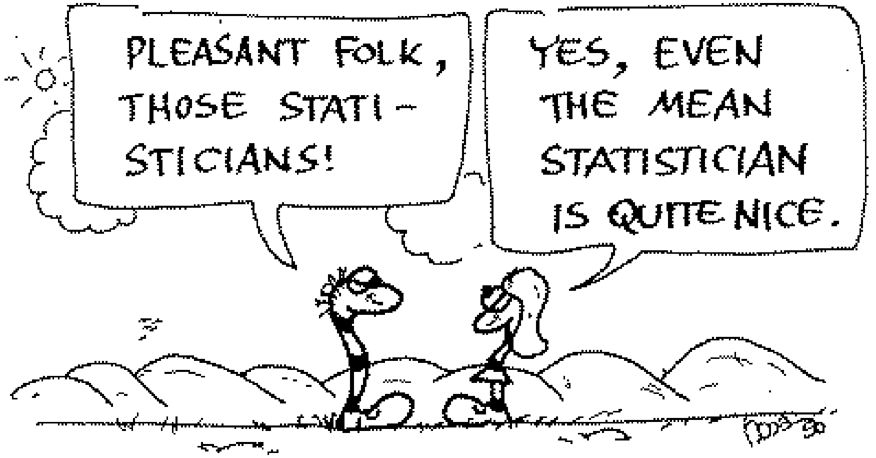
CHAPTER 13

UNDERSTANDING RESEARCH RESULTS: STATISTICAL INFERENCE



LEARNING OBJECTIVES

- ✓ Explain how researchers use inferential statistics to evaluate sample data
- ✓ Distinguish between the null hypothesis and the research hypothesis
- ✓ Discuss probability in statistical inference, including the meaning of statistical significance
- ✓ Describe the *t* test and explain the difference between one-tailed and two-tailed tests
- ✓ Describe the *F* test, including systematic variance and error variance

LEARNING OBJECTIVES

- ✓Describe what a confidence interval tells you about your data
- ✓Distinguish between Type I and Type II errors
- ✓Discuss the factors that influence the probability of a Type II error
- ✓Discuss the reasons a researcher may obtain nonsignificant results
- ✓Define power of a statistical test
- ✓Describe the criteria for selecting an appropriate statistical test

SAMPLES AND POPULATIONS

- **Inferential statistics** are used to determine whether the results match what would happen if we were to conduct the experiment again and again with multiple samples.
  - In essence, we are asking whether we can infer that the difference in the *sample means* reflects a true difference in the *population means*.
  - **Inferential Statistics**
    - Make conclusions on the basis of sample data
  - They give the probability that the difference between means reflects random error rather than a real difference

NULL AND RESEARCH HYPOTHESES

- ✓**Null hypothesis** is simply that the population means are equal—the observed difference is due to random error
  - ✓The null hypothesis states that the independent variable had no effect
  - ✓H_0 - Population means are equal ($H_0 : \mu = \mu$)
- ✓**Research hypothesis** is that the population means are, in fact, not equal
  - ✓the research hypothesis states that the independent variable did have an effect
  - ✓H_1 - Population means are not equal ($H_1 : \mu \neq \mu$)

NULL AND RESEARCH HYPOTHESES

- **Statistical significance**
  - The null hypothesis is rejected when there is a very low probability that the obtained results could be due to random error.
    - $P \leq .05$
  - This is what is meant by **statistical significance**—A significant result is one that has a very low probability of occurring if the population means are equal.
  - More simply, significance indicates that there is a low probability that the difference between the obtained sample means was due to random error.
  - Significance, then, is a matter of probability.

PROBABILITY

- ✓**Probability**: is the likelihood of the occurrence of some event or outcome
  - ✓A key question then becomes:
    - ✓How unlikely does a result have to be before we decide it is significant?
  - ✓A decision rule is determined prior to collecting the data.
  - ✓The **alpha level is the** probability required for significance.
  - ✓The most common alpha level probability used is .05.
    - ✓The outcome of the study is considered significant when there is a .05 or less probability of obtaining the results;
    - ✓that is, there are only 5 chances out of 100 that the results were due to random error in one sample from the population.

SAMPLING DISTRIBUTIONS

- ✓**Sampling distributions**: are based on the assumption that the null hypothesis is true
- ✓**Sample size** – is the total number of observations
  - ✓As the size of the sample increases, one is more confident that the outcome is actually different from the null hypothesis expectation.

THE *t* AND *F* TESTS

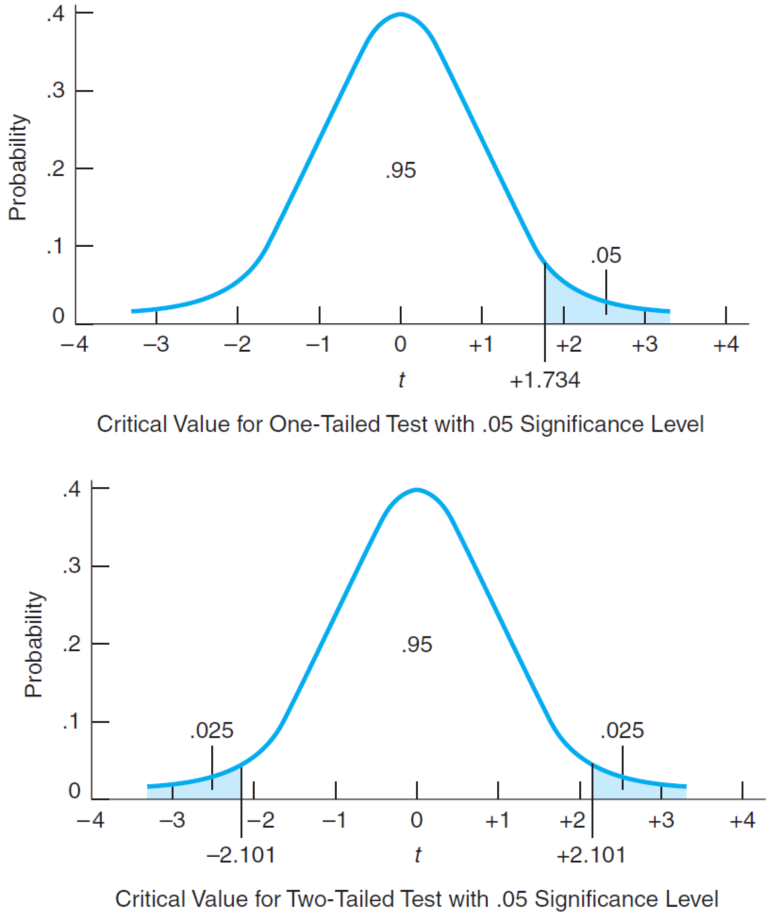
- ✓**t test**: Examines whether two groups are significantly different from each other
  - ✓*t* value - Ratio of two aspects of data
    - ✓Difference between the group means
    - ✓Variability within groups

$$t = \frac{\text{Group difference}}{\text{Within-group variability}}$$

- ✓The ***F* test** is a more general statistical test that can be used to ask whether there is a difference among three or more groups or to evaluate the results of factorial designs

One-tailed versus two-tailed tests

- ✓ In essence, one-tailed tests allow for the possibility of an effect in just one direction where with two-tailed tests, you are testing for the possibility of an effect in two directions – both positive and negative.
- ✓ The two-tailed test can show evidence that the control and experimental groups are *different*, but the one-tailed test is used to show evidence if the experimental group is *better than* the control group.



One-tailed versus two-tailed tests

- One-tailed versus two-tailed tests
  - One-tailed - Critical *t* chosen when research hypothesis specifies a direction of difference between the groups
  - Two-tailed tests - Critical *t* chosen when research hypothesis does not specify a predicted direction of difference
- The benefit to using a one-tailed test is that it requires fewer subjects to reach significance.
- A two-tailed test splits your significance level and applies it in both directions, thus each direction is only half as strong as a one-tailed test (which puts all the significance in one direction) and thus requires more subjects to reach significance

Degrees of Freedom (*df*)

- ✓First, forget about statistics. Imagine you’re a fun-loving person who loves to wear hats. You couldn't care less what a degree of freedom is. You believe that variety is the spice of life.
- ✓Unfortunately, you have constraints. You have only 7 hats. Yet you want to wear a different hat every day of the week.

- ✓On the first day, you can wear any of the 7 hats. On the second day, you can choose from the 6 remaining hats, on day 3 you can choose from 5 hats, and so on.
- ✓When day 6 rolls around, you still have a choice between 2 hats that you haven’t worn yet that week.

Degrees of Freedom (*df*)

- ✓But after you choose your hat for day 6, you have no choice for the hat that you wear on Day 7. You must wear the one remaining hat.

- ✓You had 7-1 = 6 days of "hat" freedom—in which the hat you wore could vary!
- ✓That's kind of the idea behind degrees of freedom in statistics.
- ✓Degrees of freedom are often broadly defined as the number of "observations" (pieces of information) in the data that are free to vary when estimating statistical parameters.

Degrees of Freedom (*df*)

- ✓**Degrees of freedom** (***df***): Number of scores free to vary once the means are known.
- ✓The concept of degrees of freedom is central to the principle of estimating statistics of populations from samples of them.
- ✓When comparing means from two groups, one assumes that the degrees of freedom are equal to $n_1 + n_2 - 2$, or the total number of participants in the groups minus the number of groups.

THE *t* AND *F* TESTS

- ***F* test** or **analysis of variance**:
  - Is an extension of the *t* test.
  - The analysis of variance is a more general statistical procedure than the *t* test.
  - When a study has only one independent variable with *two* groups, *F* and *t* are virtually identical—the value of *F* equals t^2 in this situation.
  - However, analysis of variance is also used when there are more than two levels of an independent variable and when a factorial design with two or more independent variables has been used.

THE *t* AND *F* TESTS

- ✓***F* test** or **analysis of variance**:
- ✓Used when:
  - ✓There are more than two levels of an independent variable (One-Way Analysis of Variance (Between Subjects ANOVA or Within Subjects ANOVA))
  - ✓Factorial design with two or more independent variables has been used (Factorial Design 2 X 2 ANOVA; One-Between-One-Within ANOVA)
  - ✓The *F* statistic is a ratio of two types of variance—systematic variance and error variance (hence the term *analysis of variance*).
    - ✓**Systematic variance**: Deviation of the group means from the grand mean
    - ✓**Error variance**: Deviation of the individual scores in each group from their respective group means
  - ✓Larger *F* ratio can lead to significant results

Effect Size

- ✓Calculating effect size
  - ✓After determining that there was a statistically significant effect of the independent variable, researchers will want to know the magnitude of the effect.
  - ✓Cohen's *d* - Effect size estimate used when comparing two means
  - ✓Effect size *r* – effect size estimate used when computing correlations

t tests

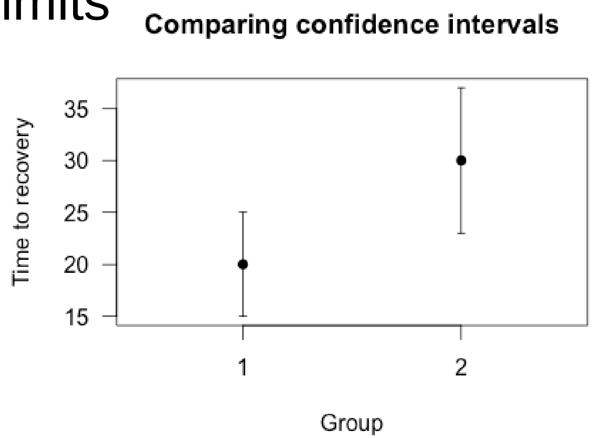
$$d = \frac{M_1 - M_2}{\sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}}$$

correlations

$$\text{effect size } r = \sqrt{\frac{t^2}{t^2 + df}}$$

Confidence Intervals

- **Confidence intervals**
  - An interval of values that defines the most likely range of actual population values.
  - The interval has an associated confidence interval—a 95% confidence interval indicates that one is 95% sure that the population value lies within the range
    - a 99% interval would provide greater certainty but the range of values would be larger.
  - Represented in bar graphs as Vertical I-shaped line bounded by upper and lower limits



Statistical Significance

- ✓Statistical significance
  - ✓People want to be confident that they would obtain similar results if they conducted the study over and over again.
    - ✓**Goal** of the test is to help decide if the obtained results are reliable
      - ✓Significance level (alpha level) people choose indicates how confident they wish to be when making the decision.
      - ✓A .05 significance level says that they are 95% sure of the reliability of their findings; however, there is a 5% chance that they could be wrong

Sample Size & Effect Size

- **Sample Size**
  - Researchers are most likely to obtain significant results when they have a large sample size
  - Larger sample sizes provide better estimates of true population values.
- **Effect Size**
  - Significant results are most likely when the effect size is large
    - Which means that differences between groups are large and variability of scores within groups is small.
      - Example:
        - Group A scores all fall around the group's mean of 5.2
        - Group B scores all fall around the group's mean of 1.7
        - The difference in means between Group A and Group B is large, but the difference in the means within each group are minimal (This indicates small sampling error and high external validity).

DECISION MATRIX FOR TYPE I AND TYPE II ERRORS

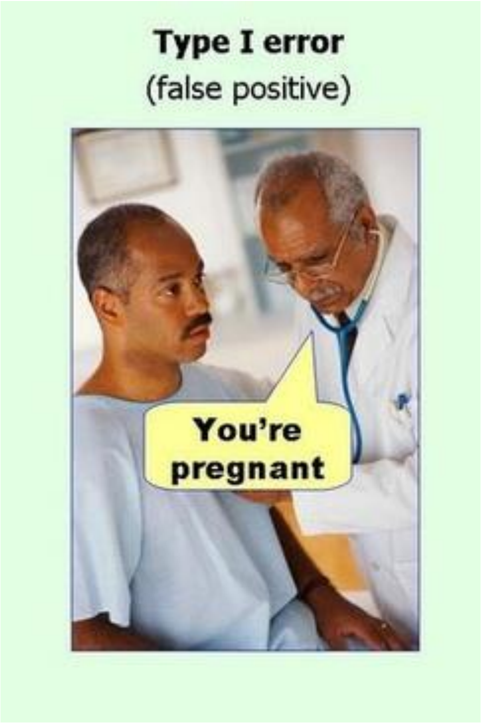
- ✓ The decision to reject the null hypothesis is based on probabilities rather than on certainties.
  - ✓ That is, the decision is made without direct knowledge of the true state of affairs in the population.
- ✓ **Correct Decisions**

1. One correct decision occurs when one rejects the null hypothesis and the research hypothesis is true in the population.
2. The other correct decision is to accept the null hypothesis, and the null hypothesis is true in the population—the population means are in fact equal.

Decision	True State in Population: Null Hypothesis Is True	True State in Population: Null Hypothesis Is False
Reject the Null Hypothesis	Type I Error (α)	Correct Decision ($1 - \beta$)
Accept the Null Hypothesis	Correct Decision ($1 - \alpha$)	Type II Error (β)

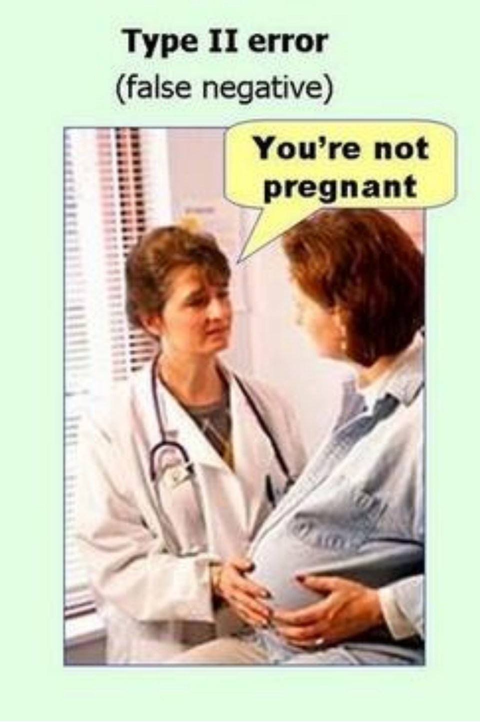
TYPE I ERRORS

- ✓**Type I error** is made when one rejects the null hypothesis but the null hypothesis is actually true.
  - ✓One's decision is that the population means are not equal when they actually are equal.
- ✓Occurs when a large value of *t* or *F* is obtained



TYPE II ERRORS

- ✓A Type II error occurs when the null hypothesis is accepted although in the population the research hypothesis is true.
  - ✓The population means are not equal, but the results of the experiment do not lead to a decision to reject the null hypothesis.
- ✓Related factors
  - ✓Significance (alpha) level
  - ✓Sample size
  - ✓Effect size



DECISION MATRIX FOR A JUROR

- ✓ For example, the use of a decision matrix involves the important decision to convict someone of a crime.
- ✓ If the null hypothesis is that the person is “Innocent” for one, and the true state is that the person is either “guilty” or “innocent,” one must decide whether to go ahead and find the person guilty.

		True State	
		Null Is True (Innocent)	Null Is False (Guilty)
Decision	Reject Null (Find Guilty)	Type I Error	Correct Decision
	Accept Null (Find Innocent)	Correct Decision	Type II Error

SIGNIFICANCE LEVEL

- Researchers traditionally used a .05 or a .01 significance level in the decision to reject the null hypothesis
  - It specifies the probability of a Type I error if the null hypothesis is rejected.
    - If there is less than a .05 or a .01 probability that the results occurred because of random error, the results are said to be significant.
    - In the case of a .05 significance level, one takes a 5% risk that he or she has committee a Type I error (rejecting the null hypothesis when, in fact, it was true).
  - However, there is nothing magical about a .05 or a .01 significance level.
- Significance level chosen and the consequences of a Type I or a Type II error are determined by the use of the results (How close do your results need to be?)

INTERPRETING NONSIGNIFICANT RESULTS

- Although “accepting the null hypothesis” is convenient terminology, it is important to recognize that researchers are not generally interested in accepting the null hypothesis.
  - Research is designed to show that a relationship between variables does exist, not to demonstrate that variables are unrelated.
- Results of a single study can be nonsignificant even when a relationship between variables in the population exist
- A meaningful result can be overlooked when the significance level is very low
- Sample size should be large enough to find a real effect
- Evidence of non related variables should come from multiple studies

KEEP CALM
and
DON’T COUNT
ME OUT JUST
YET!

CHOOSING A SAMPLE SIZE: POWER ANALYSIS

- ✓ An alternative approach is to select a sample size on the basis of a desired probability of correctly rejecting the null hypothesis.
  - ✓ This probability is called the **power** of the statistical test. It is obviously related to the probability of a Type II error.
- ✓ **Power** of statistical test: Determines optimal sample size based on probability of correctly rejecting the null hypothesis

$$\text{Power} = 1 - p\ (\text{Type II error})$$

- ✓ Effect sizes range and desired power
  - ✓ Smaller effect sizes require larger samples to be significant at the .05 level
  - ✓ Higher desired power demands a greater sample size
  - ✓ Researchers usually strive power between .70 and .90 to determine sample size

IMPORTANCE OF REPLICATIONS

- ✓If the results of the means and standard deviations are statistically significant, one concludes that they would likely be obtained over and over again if the study were repeated.
  - ✓This speaks to **Reliability**
- ✓Scientists attach little importance to the results of a single study
- ✓Detailed understanding requires numerous studies examining same variables
- ✓Researchers look at the results of studies that replicate previous investigations

SIGNIFICANE OF PEARSON *r* CORRELATION COEFFICIENT

- ✓Used to describe the strength of the relationship between two variables
  - ✓Both variables have interval or ratio scale properties
- ✓A statistical significance test helps to:
  - ✓Decide the rejection of a null hypothesis
- ✓The null hypothesis in this case is that the true population correlation is 0.00—the two variables are not related.
  - ✓What if one obtains a correlation of .27 (plus or minus)?
  - ✓A statistical significance test will allow one to decide whether to reject the null hypothesis and conclude that the true population correlation is, in fact, greater than 0.00.

COMPUTER ANALYSIS OF DATA

- Statistical analysis software packages make it easy to calculate statistics for any data set
  - SPSS
    - More often used by academics and some businesses
  - SAS
    - More often used by businesses and some academics.
  - SYSTAT
    - Often used by the science community.
    - It was developed by a psychology professor and was sold to SPSS, which later sold it to a company in India, with a headquarters in Chicago.
  - R and MYSTAT
    - R is a free, open-source statistical program used by various audiences
    - MyStat is the free student version of Systat
  - Microsoft Excel
    - Comes with Microsoft Office
    - Has statistical capabilities but they are not user friendly. It is more often used for basic percentage and frequency counts.
    - Statisticians prefer other statistical programs for more complicated analysis

COMPUTER ANALYSIS OF DATA

- ✓ Steps in analysis
  - ✓ Input data into rows and columns
    - ✓ Rows represent cases or each participant's data
    - ✓ Columns contain a participant's score for a specific variable
  - ✓ Properly code data after it is entered
    - ✓ Code categorical variables with numbers
      - ✓ Ex: Male = 1, Female = 2
    - ✓ Calculate variable constructs
      - ✓ Ex: Add up all answers pertaining to narcissism to come up with an overall score for narcissism.
    - ✓ Give Label Names to variables and identify Variable Types (i.e., Nominal, Ordinal, Scale)
  - ✓ Run descriptive analysis and charts to look for data inconsistencies and correct them
  - ✓ Run statistical tests
  - ✓ Interpret output

SELECTING THE APPROPRIATE SIGNIFICANCE TEST

- ✓Variables with:
  - ✓**Ordinal** & **Nominal** scale properties have two or more discrete values.
    - ✓Ordinal & nominal data also known as **Categorical** data and are sometimes called qualitative, discrete, or dichotomous variables.
    - ✓Called Ordinal and Nominal data in SPSS.
  - ✓**Interval** or **Ratio** scale properties have many values
    - ✓Interval & ratio data also known as **Continuous** data and are sometimes called quantitative variables.
    - ✓It is called Scale data in SPSS

RESEARCH STUDYING TWO VARIABLES (Bivariate Analysis)

- ✓ In bivariate analysis, the researcher is studying whether two variables are related.
- ✓ In general, people would refer to the first variable as the independent variable (IV) and the second variable as the dependent variable (DV).

IV	DV	Statistical test
Categorical *Male-female*	Categorical *Vegetarian—yes/no*	Chi-square
Categorical (2 groups) *Male-female*	Continuous *Grade point average*	t test
Categorical (≥3 groups) *Study time (low, medium, high)*	Interval/ratio *Test score*	One-way analysis of variance (Between Subjects ANOVA; Within Subjects ANOVA)
Continuous *Optimism score*	Continuous *Sick days last year*	Pearson correlation

RESEARCH STUDYING MULTIPLE VARIABLES (Multivariate Analysis)

- ✓ In multivariate analysis, the researcher is studying whether three or more variables are related.
- ✓ These research design situations have been described in previous chapters.
- ✓ There are, of course, many other types of designs.

IV	DV	Statistical test
Categorical (2 or more variables)	Continuous	Analysis of variance (Factorial Design 2 X 2 ANOVA; One-Between-One-Within ANOVA)
Continuous (2 or more variables)	Continuous	Multiple regression

LAB

- Labs:
  - Statistical Inference Activity
  - Statistical Decisions Activity
  - Using the Correct Statistical Test Activity
  - (Due before class next week)

"I call it, 'Research Paper Lite.' It contains a third fewer facts, but you'd never know it."