



# Inferential Statistics

Katie Rommel-Esham

Education 604



# Probability

- Probability is the scientific way of stating the degree of confidence we have in predicting something
- Tossing coins and rolling dice are examples of probability experiments
- The concepts and procedures of *inferential statistics* provide us with the language we need to address the probabilistic nature of the research we conduct in the field of education

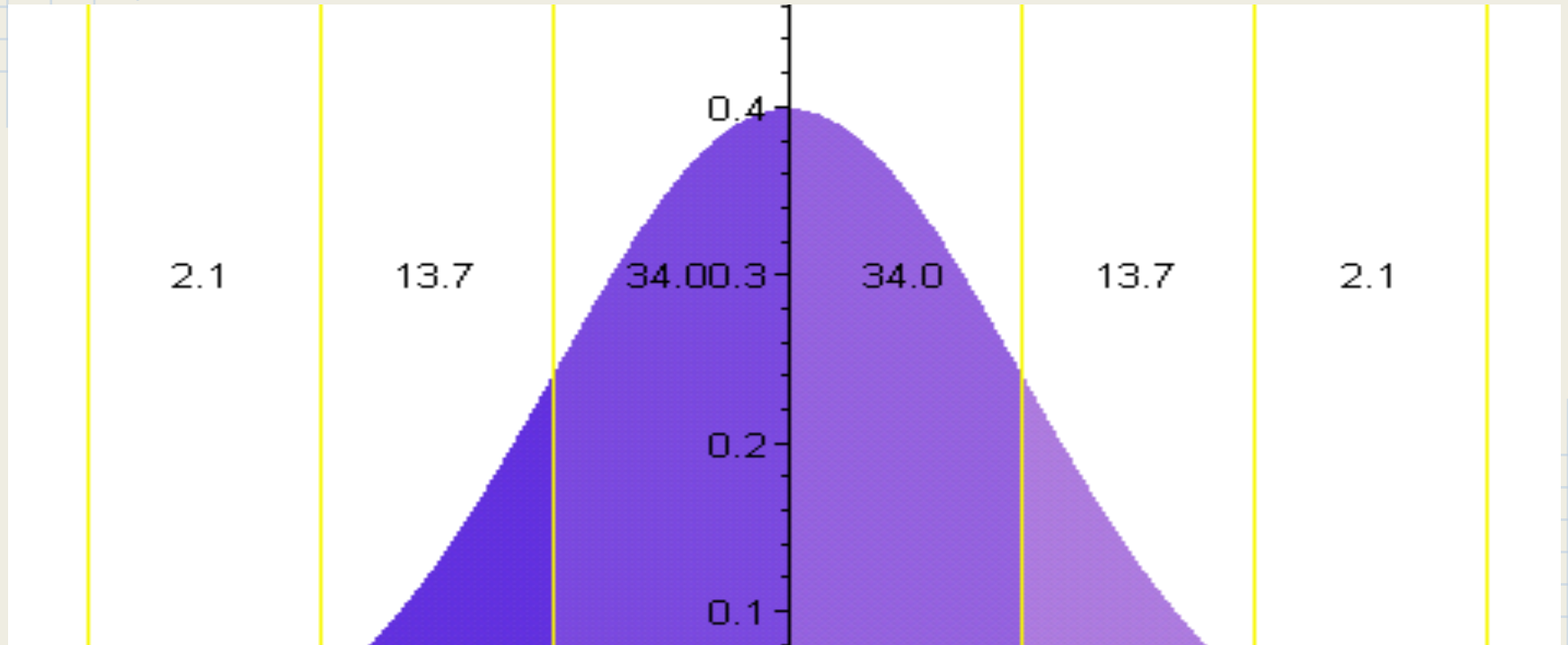
# From Samples to Populations

- Probability comes into play in educational research when we try to estimate a population mean from a sample mean
- Samples are used to generate the data, and inferential statistics are used to generalize that information to the population, a process in which error is inherent
- Different samples are likely to generate different means. How do we determine which is “correct?”

# The Role of the Normal Distribution

- If you were to take samples repeatedly from the same population, it is likely that, when all the means are put together, their distribution will resemble the normal curve.
- The resulting normal distribution will have its own mean and standard deviation.
- This distribution is called the *sampling distribution* and the corresponding standard deviation is known as the *standard error*.

# Remember me?



# Sampling Distributions

- As before, with the sampling distribution, approximately 68% of the means lie within 1 standard deviation of the distribution mean and 96% would lie within 2 standard deviations
- We now know the *probable* range of means, although individual means might vary somewhat

# The Probability-Inferential Statistics Connection

- Armed with this information, a researcher can be fairly certain that, 68% of the time, the population mean that is generated from any given sample will be within 1 standard deviation of the mean of the sampling distribution.

# Hypotheses Revisited

- Research hypothesis: the research prediction that is tested (e.g. *students in situation A will perform better than students in situation B*)
- Null hypothesis: a statement of “no difference” between the means of two populations (*there will be no difference in the performance of students in situations A and B*)



# Why do we need a Null Hypothesis?

- The null hypothesis is a technical necessity of inferential statistics
- The *research hypothesis* is more important than the null hypothesis when conceiving and designing research

# How does all this fit together?

- Researchers use inferential statistics to determine the probability that the *null hypothesis is untrue*
- Recall that if the null hypothesis is untrue, that is if it is *not true that there is no difference*, the most plausible conclusion is that there is indeed a difference
- We never *prove that anything is true*, we only *fail to disprove*

# Levels of Significance

- Used to indicate the chance that we are wrong in rejecting the null hypothesis
- Also called the level of probability or  $p$  level
- $p=.01$ , for example, means that the probability of finding the stated difference *as a result of chance* is only 1 in 100

# Errors in Hypothesis Testing

- A *type I error* is made when a researcher rejects the null hypothesis when it is true
- The probability of making this type of error is equal to the level of significance
- A *type II error* is made when a researcher accepts the null hypothesis when it is false
- As the level of significance increases, the likelihood of making a Type II error decreases

# In summary...

	<i>Accept Null</i>	<i>Reject Null</i>
<i>Null is true</i>	Correct decision: there is no difference	<b>Type I error</b>
<i>Null is false</i>	<b>Type II error</b>	Correct decision: there are differences

# Interpreting Level of Significance

- Researchers generally look for levels of significance equal to or less than .05
- If the desired level of significance is achieved, the null hypothesis is rejected and we say that there is a *statistically significant* difference in the means

# Some notes on $p$ -values

- Acceptable levels of significance are situation specific
- $p = .05$  is fine for most educational research
- $p = .05$  is not an acceptable level if we are considering the error in a test concerning usage of a drug that might cause death

# Parametric Statistics

- Used when the researcher can assume that the *population values are normally distributed, variances are equal, and data are interval or ratio in scale*



# $t$ -tests

- The most common statistical procedure for determining the level of significance when **two** means are compared
- Generates a number that is used to determine the  $p$ -level of rejecting the null hypothesis
- Assumes equal variability in both data sets

# Calculating the $t$ statistic

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1 - \overline{X}_2}}$$

- As the difference between the means increases, and the error decreases, the  $t$ -statistic gets larger
- The denominator represents the standard error of measurement between the means (the amount of error inherent in estimating population means from sample means)

# Using the $t$ Statistic

- A “one-tailed”  $t$ -test is used for directional hypotheses, while a “two-tailed”  $t$ -test is used for nondirectional hypotheses
- A one-tailed  $t$ -test is used when the researcher is fairly confident that a result opposite of the research hypothesis will *not* be obtained
- Once calculated, the  $t$ -statistic is compared to a table of values in order to determine the level of significance ( $p$ -value)

# Using the $t$ Statistic

- A  $t$ -test may also be used when a researcher wants to show that a *correlation coefficient is significantly different from 0* (which would indicate no correlation)
- It may also be used if a researcher is interested in knowing whether the group mean is significantly different from 85, for example, where 85 is used as a *reference value* and not as the mean of another group

# Variations on the $t$

- An *independent samples t-test* is used when the groups have no relationship to one another, as would an experimental group and control group
- You may also encounter literature that references a *dependent sample, paired, correlated, or matched t-test*
- These are used *if the subjects in the two groups are matched in some way*, as they would be matched with themselves in a pretest-posttest situation

# Analysis of Variance (ANOVA)

- Similar to a  $t$ -test, but used when there are more than two groups being compared
- ANOVA is an extension of the  $t$ -test
- Addresses the question "Is there a significant difference between **any** two population means?"

# How ANOVA Works

- Analysis of variance allows a researcher to examine differences in **all** population means simultaneously rather than conducting a series of *t*-tests
- It uses *variances* (rather than means) of groups to calculate a value that reflects the degree of differences in the means

# Interpreting ANOVA

- Produces an *F statistic* (or *F ratio*) which is analogous to the *t*-statistic
- A "1x4 ANOVA" is a one-way (i.e. one independent variable) ANOVA that is comparing four group means



# 1x4 ANOVA

Classroom #1	Math achievement test
Classroom #2	Math achievement test
Classroom #3	Math achievement test
Classroom #4	Math achievement test

# Factorial Analysis of Variance

- A *factorial analysis of variance* is used when there are two or more independent variables being analyzed simultaneously
- A "2x3 ANOVA" indicates that there are *3 groups being compared on 2 variables*

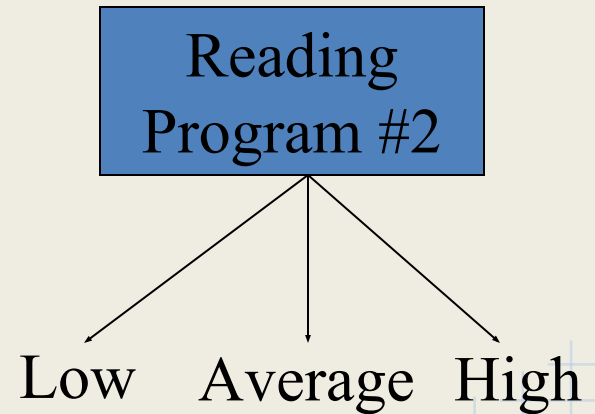
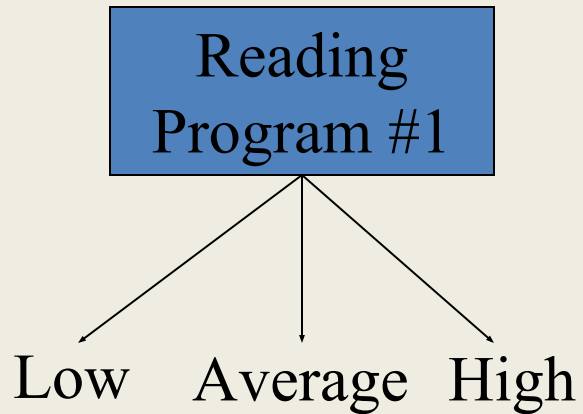
# 2x3 ANOVA

	Reading Program #1	Reading Program #2
Low ability readers		
Average ability readers		
High ability readers		

# Interpretation

- In this example, three groups are being compared on two variables
- A related research question might be “*How do the two reading programs affect the reading achievement of low, average, and high ability readers?*”

# Another way to look at the same thing...



# Then what?

- Once the analysis of variance is complete, we still need to know *where the difference lies*, as it only tells us there *is* a difference in two or more of the means

# *Post hoc* tests

- Statistical tests that tell the researcher which means are different
- Common *post hoc* comparisons include *Fisher's LSD* (least-significant difference), *Duncan's new multiple range test*, the *Newman-Keuls*, *Tukey's test*, and the *Scheffe's test*
- The choice of *post hoc* test depends on characteristics of the data sets

# ANCOVA

- *Analysis of covariance (ANCOVA)* is used when the researcher needs to adjust initial group differences statistically on one or more variables that are related to the independent variable but uncontrolled, and to increase the likelihood of finding a significant difference between two group means



# For example...

- Two groups are pretested, group A's mean is higher. The same two groups are posttested, group A's mean is still higher.
- Is the higher posttest mean due to the fact that group A's pretest mean was higher (i.e. are they “smarter?”)
- ANCOVA *adjusts for these initial pretest score differences*

# Multivariate Analyses

- Used to investigate problems in which the researcher is interested in studying more than one *dependent* variable

# An Example

- “*Attitudes towards science*” is a complex construct that might involve things like enjoying science, valuing science, attitudes towards different branches of science (Earth Science, Biology, Chemistry), lab work, science field trips, etc.
- *Multivariate methods* allow researchers to look at each of these components separately

# Multivariate Tests

Univariate Test	Multivariate Test
<i>t</i> -test	Hotelling's $T^2$
ANOVA	MANOVA
ANCOVA	MANCOVA

# Nonparametric Tests

# Nonparametric Procedures

- If the assumptions associated with parametric procedures are not met, then nonparametric procedures are used
- Most parametric procedures have analogous nonparametric procedures

# Parametric and Nonparametric Analogs

Parametric	Nonparametric
Independent samples t-test	Mann-Whitney U test
Dependent samples t-test	Wilcoxon matched-pairs signed ranks test
One-way ANOVA	Kruskal-Wallis one-way ANOVA of ranks

# Chi-Square

- Nonparametric procedure used *when data are in nominal form*
- It is a way of answering questions about relationship *based on frequencies of observations in categories*



# An Example

- What is the relationship between year in college (freshman, sophomore, junior, senior) and use of campus counseling services?
- Responses to this question will involve a count of how many in each group use the counseling service
- The independent variable is *year in college* which has *four categories*

# Other Procedures You May Encounter

- *Repeated Measures Analysis of Variance* (measure repeated across subjects)
- *Factor Analysis* (how well items are related to one another and form clusters or factors)
- *Path Analysis* (examines correlations with a *hint* of directionality)
- *Multiple Regression* (using multiple factors to increase predictive power)
- *Rasch Analysis* (allows for direct comparison between the difficulty of an item and the probability of a student at any ability level getting it correct)