When to Use a Particular Statistical Test

Central Tendency

Univariate Descriptive

Mode
• the most commonly occurring value
ex: 6 people with ages 21, 22, 21, 23, 19, 21 - mode = 21

Median
• the center value
• the formula is \( \frac{N+1}{2} \)

ex: 6 people with ages 21, 22, 24, 23, 19, 21
line them up in order from lowest to highest
19, 21, 21, 22, 23, 24
and take the center value - mode = 21.5

Mean
• the mathematical average
• the formula is \( \frac{\sum X}{N} \)

ex: mean age = age of person one + age of person two + age of person three, etc./number of people

Variance
• a measure of how spread out a distribution is
• it is computed as the average squared deviation of each number from its mean

Standard Deviation
• how much scores deviate from the mean
• it is the square root of the variance
• it is the most commonly used measure of spread
**Bi- and Multivariate Inferential Statistical Tests**

**Differences of Groups**

**Chi Square**
- compares observed frequencies to expected frequencies

ex: Is the distribution of sex and voting behavior due to chance or is there a difference between the sexes on voting behavior?

**t-Test**
- looks at differences between two groups on some variable of interest
- the IV must have only two groups (male/female, undergrad/grad)

ex: Do males and females differ in the amount of hours they spend shopping in a given month?

**ANOVA**
- tests the significance of group differences between two or more groups
- the IV has two or more categories
- only determines that there is a difference between groups, but doesn’t tell which is different

ex: Do SAT scores differ for low-, middle-, and high-income students?

**ANCOVA**
- same as ANOVA, but adds control of one or more covariates that may influence the DV

ex: Do SAT scores differ for low-, middle-, and high-income students after controlling for single/dual parenting?

**MANOVA**
- same as ANOVA, but you can study two or more related DVs while controlling for the correlation between the DV
- if the DVs are not correlated, then separate ANOVAs are appropriate

ex: Does ethnicity affect reading achievement, math achievement, and overall scholastic achievement among 6th graders?

**MANCOVA**
- same as MANOVA, but adds control of one or more covariates that may influence the DV
Does ethnicity affect reading achievement, math achievement, and overall scholastic achievement among 6th graders after controlling for social class?

**Relationships**

**Correlation**
- used with two variables to determine a relationship/association
- do two variables covary?
- does not distinguish between independent and dependent variables

**Multiple Regression**
- used with several independent variables and one dependent variable
- used for prediction
- it identifies the best set of predictor variables
- you can enter many IVs and it tells you which are best predictors by looking at all of them at the same time
- in sequential regression the computer adds the variables one at a time based on the amount of variance they account for

**Path Analysis**
- looks at direct and indirect effects of predictor variables
- used for relationships/causality

**Group Membership**

**Logistic Regression**
- like multiple regression, but the DV is a dichotomous variable
- logistic regression estimates the odds probability of the DV occurring as the values of the IVs change

What are the odds of a suicide occurring at various levels of alcohol use?
<table>
<thead>
<tr>
<th>Statistical Analyses</th>
<th>Independent Variables</th>
<th>Dependent Variables</th>
<th>Control Variables</th>
<th>Question Answered by the Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># of IVs</td>
<td>Data Type</td>
<td># of DVs</td>
<td>Type of Data</td>
</tr>
<tr>
<td>Chi square</td>
<td>1</td>
<td>categorical</td>
<td>1</td>
<td>categorical</td>
</tr>
<tr>
<td>t-Test</td>
<td>1</td>
<td>dichotomous</td>
<td>1</td>
<td>continuous</td>
</tr>
<tr>
<td>ANOVA</td>
<td>1 +</td>
<td>categorical</td>
<td>1</td>
<td>continuous</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>1 +</td>
<td>categorical</td>
<td>1</td>
<td>continuous</td>
</tr>
<tr>
<td>MANOVA</td>
<td>1 +</td>
<td>categorical</td>
<td>2 +</td>
<td>continuous</td>
</tr>
<tr>
<td>MANCOVA</td>
<td>1 +</td>
<td>categorical</td>
<td>2 +</td>
<td>continuous</td>
</tr>
<tr>
<td>Correlation</td>
<td>1</td>
<td>dichotomous or continuous</td>
<td>1</td>
<td>continuous</td>
</tr>
<tr>
<td>Multiple regression</td>
<td>2 +</td>
<td>dichotomous or continuous</td>
<td>1</td>
<td>continuous</td>
</tr>
<tr>
<td>Path analysis</td>
<td>2 +</td>
<td>continuous</td>
<td>1</td>
<td>continuous</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>1 +</td>
<td>categorical or continuous</td>
<td>1</td>
<td>dichotomous</td>
</tr>
<tr>
<td>Research Question</td>
<td>Number and type of DV</td>
<td>Number and type of IV</td>
<td>Covariates</td>
<td>Test</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>nominal or higher</td>
<td>1 nominal or higher</td>
<td>chi square</td>
<td></td>
</tr>
<tr>
<td></td>
<td>continuous</td>
<td>1 dichotomous</td>
<td>t-test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>continuous</td>
<td>1 categorical</td>
<td>one-way anova</td>
<td></td>
</tr>
<tr>
<td></td>
<td>continuous</td>
<td>2+ categorical</td>
<td>factorial anova</td>
<td></td>
</tr>
<tr>
<td></td>
<td>continuous</td>
<td>1 categorical</td>
<td>one-way ancova</td>
<td></td>
</tr>
<tr>
<td></td>
<td>continuous</td>
<td>2+ categorical</td>
<td>factorial ancova</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2+ continuous</td>
<td>1 categorical</td>
<td>one-way manova</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2+ continuous</td>
<td>2+ categorical</td>
<td>factorial manova</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2+ continuous</td>
<td>1+</td>
<td>factorial manova</td>
<td></td>
</tr>
<tr>
<td>Degree of relationship</td>
<td>continuous</td>
<td>1 continuous</td>
<td>bivariate correlation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2+ continuous</td>
<td>multiple regression</td>
<td>linear combination to predict the DV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+ continuous</td>
<td>2+ continuous</td>
<td>path analysis</td>
<td></td>
</tr>
<tr>
<td>Prediction of group membership</td>
<td>dichotomous</td>
<td>2+ nominal or higher</td>
<td>logistic regression</td>
<td></td>
</tr>
</tbody>
</table>
When trying to decide what test to use, ask yourself the following...

Am I interested in...?
- **description** (association) - correlations, factor analysis, path analysis
- **explanation** (prediction) - regression, logistic regression, discriminant analysis
- **intervention** (group differences) - t-test, anova, manova, chi square

Do I need longitudinal data or is cross-sectional data sufficient for my purpose?  
Do my hypotheses involve the investigation of change, growth, or the timing of an event?  
If longitudinal data is necessary, how many data points are needed?  
(We do not cover these techniques in this class, but your major advisor can direct you to the appropriate procedure.)

Is my dependent variable nominal, ordinal, interval, or ratio?
- **nominal** - chi square, logistic regression
- **dichotomous** - logistic regression
- **ordinal** - chi square
- **interval/ratio** - correlation, multiple regression, path analysis, t-test, anova, manova, discriminant analysis

Do I have moderating or mediating variables?
- A **moderating** relationship can be thought of as an *interaction*. It occurs when the relationship between variables A and B depends on the level of C.  
  A=marital satisfaction  
  B=risk of divorce  
  C=amount of resources

![Diagram showing moderating relationship]

When resources are low, marital satisfaction doesn’t affect divorce, but at high resources, marital satisfaction predicts a greater risk of divorce.

A **mediating** relationship can be thought of as an *intervening* relationship. It is one in which the path relating A to C is mediated by a third variable (B).

We all know that older drivers, up to a point, are safer than younger drivers. But I'm sure that we
don’t think that the aging (some would say deterioration) of the body, or the mere passage of
time, somehow leads to safer driving. What happens, as all right thinking people will agree, is
that age leads to wisdom, and wisdom leads to safer drivers. Hence “wisdom” is the mediating
variable that explains the correlation between age and safe driving. (Forget the part about the
decline in driving related to being way old, that creates a curvilinear relationship and we’re not
going there.)

Leerkes and Crockenberg (1999) were studying the relationship between how a new mother was
raised by her own mother 20+ years before (A=maternal care) and the new mother's level of self-
efficacy as a mother (C=self-efficacy). The idea being that if your mother showed high levels of
maternal care toward you, you would feel more confident of your ability to mother your own
child.

Indeed, the correlation between Maternal Care and Self-Efficacy was .272, which is significant at
$p < .01$. But Leerkes expected that this relationship was mediated by self-esteem, such that if you
had good maternal care, you will have good self-esteem (B=self-esteem), and if you have good
self-esteem, that will, in turn, lead you to have high self-efficacy.

When A (maternal care) and C (self-efficacy) are entered into the regression equation, A predicts
C. When B (self-esteem) is added to the equation, if B predicts C and the A-C relationship
declines in value, that’s a mediating/intervening relationship.