When to Use a Particular Statistical Test

Univariate Descriptive

Central Tendency

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 form lowest to highest
19, 21, 21, 22, 23, 24
and take the center value - mode =21.5

Mean

- the mathematical average
- the formula is $\sum X/N$

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 6^{th} graders?

MANCOVA

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

ex: 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

ex: Amount of damage to a house on fire and number of firefighters at the fire

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

ex: IVs drug use, alcohol use, child abuse
DV. suicidal tendencies

Path Analysis

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

ex: Child abuse causes drug use which leads to suicidal tendencies.

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

ex: What are the odds of a suicide occurring at various levels of alcohol use?

Statistical Analyses	Independent Variables		Dependent Variables		Control Variables	Question Answered by the Statistic
	# of IVs	Data Type	# of DVs	Type of Data		
Chi square	1	categorical	1	categorical	0	Do differences exist between groups?
***t*-Test**	1	dichotomous	1	continuous	0	Do differences exist between 2 groups on one DV?
ANOVA	1 +	categorical	1	continuous	0	Do differences exist between 2 or more groups on one DV?
ANCOVA	1 +	categorical	1	continuous	1 +	Do differences exist between 2 or more groups after controlling for CVs on one DV?
MANOVA	1 +	categorical	2 +	continuous	0	Do differences exist between 2 or more groups on multiple DVs?
MANCOVA	1 +	categorical	2 +	continuous	1 +	Do differences exist between 2 or more groups after controlling for CVs on multiple Dvs?
Correlation	1	dichotomous or continuous	1	continuous	0	How strongly and in what direction (i.e., +, -) are the IV and DV related?
Multiple regression	2 +	dichotomous or continuous	1	continuous	0	How much variance in the DV is accounted for by linear combination of the IVs? Also, how strongly related to the DV is the beta coefficient for each IV?
Path analysis	2 +	continuous	1 +	continuous	0	What are the direct and indirect effects of predictor variables on the DV?
Logistic Regression	1 +	categorical or continuous	1	dichotomous	0	What is the odds probability of the DV occurring as the values of the IVs change?

Statistics Decision Tree

Research Question	Number and type of DV	Number and type of IV	Covariates	Test	Goal of Analysis
Group differences	nominal or higher	1 nominal or higher		chi square	determine if difference between croups
	continuous	1 dichotomous		*t*-test	determine significance of mean group differences
		1 categorical		one-way anova	
			1+	one-way ancova	
		2+ categorical		factorial anova	
			1+	factorial ancova	
	2+ continuous	1 categorical		one-way manova	create linear combo of Dvs to maximize mean group differences
			1+	one-way mancova	
		2+ categorical		factorial manova	
			1+	factorial mancova	
Degree of relationship	continuous	1 continuous		bivariate correlation	determine relationship/ prediction
		2+ continuous		multiple regression	linear combination to predict the DV
	1+ continuous	2+ continuous		path analysis	estimate causal relations among variables
Prediction of group membership	dichotomous	2+ nominal or higher		logistic regression	create linear combo of IVs of the log odds of being in one group

When trying to decide what test to use, ask youself 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?

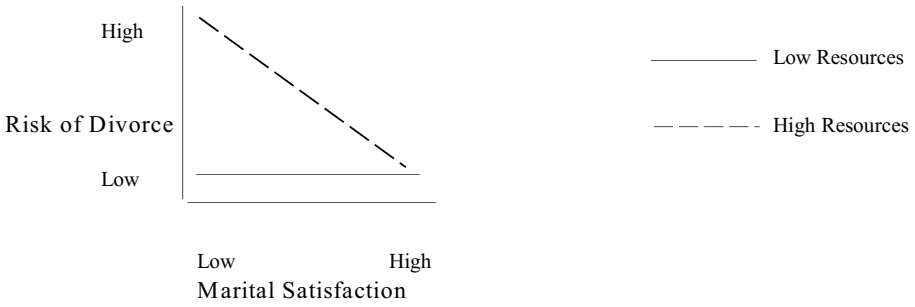
Do my hypotheses involve the investigation of change, growth, or the timing of an event?
If longitudinal data is necessary, how many data points are needed?
(We do not cover these techniques in this class, but your major advisor can direct you to the appropriate procedure.)

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



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

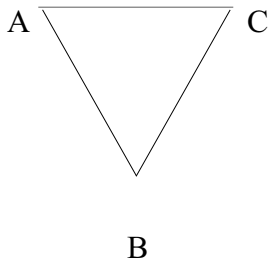
A *mediating* relationship can be thought of as an *intervening* relationship. It is one in which the path relating A to C is mediated by a third variable (B).

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.