

# Choosing the Appropriate Technique\*

\*Or what they forgot to teach you in Statistics class...

# Three Types of Analytical Variables

1. Dependent variable - the variable of interest whose value depends upon the value of another variable.
2. Independent variable - a variable of theoretical importance that is thought to influence the level of a dependent variable.
3. Control variable - a variable that is thought to influence the dependent variable but is not of theoretical importance.

# The Four Levels of Measurement for Variables

1. Nominal – the variable can be placed into categories.
2. Ordinal – the variable has categories that can be ordered from low to high or high to low.
3. Interval – the variable has categories that are equidistant from each other.
4. Ratio – the variable has an absolute zero point.

# Criteria Used to Select Appropriate Statistical Technique

1. Level of measurement of the dependent.
2. Level of measurement of the independent variable.
3. Number of dependent variables.
4. Number of independent variables.
5. Nature of the research question and its associated  $H_{\text{null}}$  and  $H_{\text{alt}}$
6. Assumptions of the test.

# The Good Guys at IDRE...

Choosing the Correct Statistical Test in SAS, Stata and SPSS.pdf - Adobe Acrobat Pro

File Edit View Window Help

Open Create Save Print Mail Comment Tools Fill & Sign Comment

stat > mult\_pkg > whatstat

## What statistical analysis should I use?

The following table shows general guidelines for choosing a statistical analysis. We emphasize that these are general guidelines and should not be construed as hard and fast rules. Usually your data could be analyzed in multiple ways, each of which could yield legitimate answers. The table below covers a number of common analyses and helps you choose among them based on the number of dependent variables (sometimes referred to as outcome variables), the nature of your independent variables (sometimes referred to as predictors). You also want to consider the nature of your dependent variable, namely whether it is an interval variable, ordinal or categorical variable, and whether it is (approximately) normally distributed (see [What is the difference between categorical, ordinal and interval variables?](#) for more information on this).<sup>1</sup> The table then shows one or more statistical tests commonly used given these types of variables (but not necessarily the only type of test that could be used) and links showing how to do such tests using SAS, Stata and SPSS.

Number of Dependent Variables	Nature of Independent Variables	<a href="#">Nature of Dependent Variable(s)</a>	Test(s)	How to SAS	How to Stata	How to SPSS	How to R
0 IVs (1 population)		interval & normal	one-sample t-test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		ordinal or interval	one-sample median	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		categorical (2 categories)	binomial test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		categorical	Chi-square	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>

[http://www.ats.ucla.edu/stat/mult\\_pkg/whatstat/](http://www.ats.ucla.edu/stat/mult_pkg/whatstat/)

# Criteria Used by IDRE in their Chart

1. Level of measurement of the dependent.
2. Level of measurement of the independent variable.
3. Number of dependent variables.
4. Number of independent variables.
5. Assumptions of the test.
6. Independence or dependence of groups or observations.

# One Dependent Variable and No Independent Variables

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	0 IVs (1 population)	interval & normal	onesample ttest
		ordinal or interval	onesample median
		categorical (2 categories)	binomial test
		categorical	Chisquare goodnessoffit

One Dependent Variable and No Independent Variables

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	0 IVs (1 population)	interval & normal	<u>onesample ttest</u>


Business Challenge: Your company makes light bulbs and you advertise that the average lifespan of your “superbulb” is 1500 hours. However, you’ve noticed recently on social media that people are complaining that your bulbs don’t last as long as you claim.

Your Response: You take a random sample of 500 bulbs and measure their average life span to compare it against the 1500 hours claimed.

Your Data Looks like:

Data Case Lightbulb Life in hrs					
1	1200				
2	1650				
3	1525				
4	1475				
5	1380				
.	.				
.	.				
.	.				



One Dependent Variable and No Independent Variables			
			
Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	0 IVs (1 population)		
		ordinal or interval	onesample median

**Business Challenge:** In taste tests, your hamburger chain's "Big Kitty" burger is known as ranking 5<sup>th</sup> in overall flavor across 10 brands. But in recent secret taste tests it seems people are liking your burger better than your top competitors.

**Your Response:** You conduct a taste test with 800 people and have them rank your Kitty Burger and your nine top competitors in terms of taste of burger.

**Your Data Looks like:**

Data Case	Flavor Rank					
1	3					
2	2					
3	5					
4	6					
5	1					
.	.					
.	.					
.	.					

## One Dependent Variable and No Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	0 IVs (1 population)		
		categorical (2 categories)	binomial test

**Business Challenge:** Your company plans to make tennis balls in two colors and your R + D department has predicted that 25% of your customers will pick red balls while 75% will pick green balls.

**Your Response:** You decide to see if R+D is earning its money. You take a random sample of 1,200 people ask them which ball they prefer.

## Your Data Looks like:

Data Case	Ball color					
1	red					
2	red					
3	green					
4	red					
5	green					
.	.					
.	.					
.	.					

One Dependent Variable and No Independent Variables

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	0 IVs (1 population)		
		categorical	Chisquare goodnessoffit

Business Challenge: As manager of the Dead Gulch Casino, you are considering taking wagers on Republican candidates to win the nomination. You think that Trump will get 60% of the vote, Bush 30% and Cruse 10%. But if you're wrong you'll have to eat cat food for the whole year.

Your Response: You randomly survey 1,500 registered voters across the U.S. and ask them which of the three candidates they would vote for.

Your Data Looks like:

Data Case	Candidate					
1	Trump					
2	Bush					
3	Trump					
4	Trump					
5	Cruse					
.	.					
.	.					
.	.					

One Dependent Variable with One Independent Variable  
with two levels  
[independent groups]

**Number of  
Dependent  
Variables**

**Nature of Independent  
Variables**

**Nature of  
Dependent  
Variable(s)**

**Test(s)**

**1**

**1 IV with 2 levels (independent  
groups)**

interval &  
normal

2 independent  
sample ttest

ordinal or  
interval

Wilcoxon  
Mann-  
Whitney test

categorical

Chisquare test

Fisher's exact  
test

One Dependent Variable with One Independent Variable  
with two levels  
[independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 levels (independent groups)	interval & normal	2 independent sample ttest

**Business Challenge:** You are Taylor Swift's research manager. You know that both men and women like her music, but you think there are age differences between the men and women fans and this plays into your new album strategy.

**Your Response:** You randomly survey 1,800 men and women who are Taylor Swift fan club members. You ask each member of the sample their age.

**Your Data Looks like:**

Data Case	Gender	Age				
1	man	26				
2	woman	18				
3	woman	19				
4	woman	16				
5	man	28				
.	.	.				
.	.	.				
.	.	.				

One Dependent Variable with One Independent Variable  
with two levels  
[independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------------	------------------------------------	---------------------------------------	---------

1

1 IV with 2 levels (independent  
groups)

ordinal or  
interval

Wilcoxon  
Mann-  
Whitney test

**Business Challenge:** You are the research director for Huevos Locos – a new Mexican restaurant chain. You are interested how your restaurant is ranked among Hispanics and non-Hispanics.

**Your Response:** You randomly survey 800 Hispanics and non-Hispanics and ask them to rank their top Mexican restaurants including yours.

**Your Data Looks like:**

Data Case	Ethnicity	Huevos Locos Rank				
1	Hispanic	6				
2	Hispanic	4				
3	non-Hispanic	7				
4	non-Hispanic	9				
5	Hispanic	14				
.	.	.				
.	.	.				
.	.	.				

One Dependent Variable with One Independent Variable  
with two levels  
[independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------------	------------------------------------	---------------------------------------	---------

1

1 IV with 2 levels (independent  
groups)

categorical

Chisquare test

Fisher's exact  
test

Business Challenge: You are the research director for National Geographic Wild channel. You want to know if there is a relationship between gender and watching the show “Dr. K Exotic Animal ER” show.

Your Response: You randomly sample 1,400 people from around the United States and ask them if they watch the show.

Fisher's Exact test is useful when < 80% of cells have expected count 5+ or cells with expected count < 1.

Fisher's Exact test has least computational load with 4 cell table but is possible at larger table sizes with more computational load.

## Your Data Looks like:

Data Case	Gender	Watch Dr. K ER?				
1	male	yes				
2	female	yes				
3	female	no				
4	male	no				
5	female	yes				
,	,	,				
,	,	,				
,	,	,				

One Dependent Variable with One Independent Variable  
with two or more levels  
[independent groups]

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 or more levels (independent groups)	interval & normal	oneway ANOVA
		ordinal or interval	Kruskal Wallis
		categorical	Chisquare test



One Dependent Variable with One Independent Variable  
with two or more levels  
[independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 or more levels (independent groups)	interval & normal	oneway ANOVA

**Business Challenge:** As research scientist for HairClub for Dogs you are experimenting with three different drugs to solve the problem of hairless Chihuahuas.

**Your Response:** You apply the drugs to 300 test dogs – 100 dogs for each drug - and count the number of hairs that grow on each dog to see if any of the drugs are effective.

Note that oneway ANOVA will only tell you if one or more of the drugs is effective.

You will need to take advantage of the multiple group comparison techniques available in oneway ANOVA to tell which of the drugs are most effective.

**Your Data Looks like:**

Data Case #	Drug	Number of Hairs				
1	hairitc	1250				
2	furseptic	800				
3	halrothedogprox	460				
4	hairitc	840				
5	furseptic	910				
,	,	,				
,	,	,				
,	,	,				

One Dependent Variable with One Independent Variable  
with two or more levels  
[independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------------	------------------------------------	---------------------------------------	---------

1

1 IV with 2 or more levels  
(independent groups)

ordinal or  
interval

Kruskal Wallis

**Business Challenge:** You are the research scentologist for Aroma360. You've invented a new scent called "Ode de Mouffette" and you want to know if the scent is ranked differently among people from the U.S. West, Midwest and East Coasts among top 10 perfumes.

**Your Response:** You randomly select 1,500 people with 500 from each geography and ask them to rank the top 10 perfumes including Ode de Mouffette.

**Your Data Looks like:**

Data Case #	Region	rank				
1	West	4				
2	Midwest	2				
3	East	1				
4	East	4				
5	Midwest	6				
6	West	3				
7	Midwest	5				
8	East	7				
9	West	8				
10	Midwest	9				

One Dependent Variable with One Independent Variable  
with two or more levels  
[independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------------	------------------------------------	---------------------------------------	---------

1

1 IV with 2 or more levels  
(independent groups)

categorical

Chisquare test

**Business Challenge:** As publisher of Bengal Cat magazine, you are thinking of going digital with your magazine. You suspect that folks from different racial backgrounds may have different preferences for format of the magazine.

**Your Response:** You randomly select 450 people from the U.S. who are Bengal owners and ask them the race they most identify with and their print versus digital reading habits.

**Your Data Looks like:**

Data Case #	Race	Read digital magazine yesterday?				
1	african american	no				
2	white	yes				
3	asian	yes				
4	native american	no				
5	african american	no				
⋮	⋮	⋮				
⋮	⋮	⋮				
⋮	⋮	⋮				

One Dependent Variable with One Independent Variable  
with two levels  
[dependent groups]

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 levels (dependent/matched groups)	interval & normal	paired ttest
		ordinal or interval	Wilcoxon signed ranks test
		categorical	McNemar

One Dependent Variable with One Independent Variable  
with two levels  
[dependent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	<u>Nature of Dependent Variable(s)</u>	Test(s)
1	1 IV with 2 levels (dependent/matched groups)	interval & normal	paired ttest

**Business Challenge:** As an unscrupulous maker of bathroom scales you want to see if your new scale “the fluffmeister” consistently reports under weight.

**Your Response:** You randomly select 1,200 people and weigh them once with your competitor’s scale and then again with the fluffmeister and compare the two weights.

**Your Data Looks like:**

Data Case #	Competitor weight	fluffmeister weight				
1	120	118				
2	155	152				
3	190	187				
4	220	216				
5	119	115				
⋮	⋮	⋮				
⋮	⋮	⋮				
⋮	⋮	⋮				

One Dependent Variable with One Independent Variable  
with two levels  
[dependent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------------	------------------------------------	---------------------------------------	---------

1

1 IV with 2 levels  
(dependent/matched groups)

ordinal or  
interval

Wilcoxon signed  
ranks test

**Business Challenge:** As the research director for Absolutely Fabulous vodka, you want to compare your brand against five other vodka brands in a taste test. You think that the more snozzled they get, the better your brand will taste.

**Your Response:** You randomly select 450 people and offer them five brands to rank on taste. The drinks are offered so that they get Absolutely Fabulous vodka both as a first and last taste choice each time. You then compare the rankings for your brand.

**Your Data Looks like:**

Data Case #	AF vodka first taste rank	AF vodka last taste rank				
1	4	2				
2	3	1				
3	5	1				
4	3	2				
5	4	3				
1	1	1				
1	1	1				
1	1	1				

One Dependent Variable with One Independent Variable  
with two levels  
[dependent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 levels (dependent/matched groups)		
		categorical	McNemar

**Business Challenge:** As head of security for the National Security Agency you rely upon drug tests to make sure your employees are clean. But you are unsure of how good your drug test vendor's product is.

**Your Response:** You randomly select 350 NSA employees and have them submit to a drug test. Five minutes later, you administer the test again and compare the results to see if they are consistent.

**Your Data Looks like:**

Data Case #	Drug test round 1	Drug test round 2				
1	negative	negative				
2	negative	positive				
3	negative	negative				
4	positive	negative				
5	positive	positive				
⋮	⋮	⋮				
⋮	⋮	⋮				
⋮	⋮	⋮				

One Dependent Variable with One Independent Variable  
with two or more levels  
[dependent/matched groups]

**Number of  
Dependent  
Variables**

**Nature of Independent  
Variables**

**Nature of  
Dependent  
Variable(s)**

**Test(s)**

**1**

**1 IV with 2 or more levels  
(dependent/matched groups)**

interval & normal

oneway repeated  
measures  
ANOVA

ordinal or  
interval

Friedman test

categorical

repeated measures  
logistic  
regression



Business Challenge: As head of R+D for Absolutely Fabulous vodka, you're a smart guy. You also make a hangover remedy to help out your "best" customers. You want to know how your hangover cure works against two other cure brands over time.

Your Response: You randomly select 1,450 people over age 21 and have them drink a fifth of AF vodka. At 8am the next morning they perform a skills test. They then take one of three hangover cures (including yours) and ten minutes later they take the skill test again. The skill test scores range between 1 and 100 on the test.

One Dependent Variable with One Independent Variable  
with two or more levels  
[dependent/matched groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 or more levels (dependent/matched groups)	interval & normal	<u>oneway</u> repeated measures ANOVA

Your Data Looks like:

Data Case #	Hangover cure	skill test 1	skill test 2			
1	AF No Blitz	70	75			
2	Headbanger	66	63			
3	NoHork	33	18			
4	AF No Blitz	88	92			
5	Headbanger	53	64			
6	AF No Blitz	70	75			
7	Headbanger	66	63			
8	NoHork	33	18			
9	AF No Blitz	88	92			
10	Headbanger	53	64			

One Dependent Variable with One Independent Variable  
with two or more levels  
[dependent/matched groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------------	------------------------------------	---------------------------------------	---------

1

1 IV with 2 or more levels  
(dependent/matched groups)

ordinal or  
interval

Friedman test

**Business Challenge:** You are the research manager for Suds and Duds laundry detergent. You want to know if your TV commercial is going to have an impact on attitudes toward your brand.

**Your Response:** You randomly select 550 consumers and ask them to rank 8 popular laundry detergents including Suds and Duds on cleaning power. Then you show them the Suds and Duds TV commercial and ask them to rank the detergents again on cleaning power.

**Your Data Looks like:**

Data Case #	Suds and Duds pre-rank	Suds and Duds post-rank				
1	5	4				
2	3	4				
3	7	3				
4	8	2				
5	6	3				
⋮	⋮	⋮				
⋮	⋮	⋮				
⋮	⋮	⋮				

One Dependent Variable with One Independent Variable  
with two or more levels  
[dependent/matched groups]

idre  
INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 IV with 2 or more levels (dependent/matched groups)		
		categorical	repeated measures logistic regression

**Business Challenge:** As the maker of SneezeAway, a popular allergy medicine, you want to know how your medication affects the chances of a patient sneezing ten or more times a day compared to other brands.

**Your Response:** You randomly select 2,200 allergy sufferers and administer one brand to them each day for five days. You ask them to report whether or not they sneezed 10 or more times each day.

**Your Data Looks like:**

Data Case #	Day 1	Day 2	Day 3	Day 4	Day 5	Brand	
1	yes	yes	yes	no	no	SneezeAway	
2	yes	yes	no	no	yes	Snuffles	
3	yes	no	no	no	no	SniffNoMore	
4	yes	yes	yes	yes	no	SneezeAway	
5	yes	yes	yes	yes	no	Snuffles	

# One Dependent Variable with Two or More Independent Variables [independent groups]

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	2 or more IVs (independent groups)	interval & normal	factorial ANOVA
		ordinal or interval	ordered logistic regression
		categorical	factorial logistic regression

## One Dependent Variable with Two or More Independent Variables [independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	2 or more IVs (independent groups)	interval & normal	factorial ANOVA

**Business Challenge:** As the research director of Dominators Pizza, you want to see if there are any effects for using online coupons, gender and ethnicity in the number of pizzas you order.

**Your Response:** You survey 1,450 randomly selected people and ask them how many pizzas they eat in a month, their gender, do they use online coupons and their ethnicity.

## Your Data Looks like:

Data Case #	Number Pizzas	use online coupon?	Sex	Ethnicity
1	5	yes	male	Hispanic
2	1	yes	male	non-Hispanic
3	0	no	female	Hispanic
4	8	yes	female	Hispanic
5	15	no	male	non-Hispanic
6	3	no	male	Hispanic
7	2	yes	female	non-Hispanic
8	4	no	male	Hispanic

## One Dependent Variable with Two or More Independent Variables [independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	2 or more IVs (independent groups)	ordinal or interval	ordered logistic regression

**Business Challenge:** In the intelligence community, daily risks are often categorized in “buckets”, often low, medium and high risks. As an analyst for a risk management company, you are analyzing the risk of a terrorist attack on any given day. You think the risk assessment on any given day may depend upon the direction of the NYSE index and the direction of a price of a barrel of oil.

**Your Response:** You collect the official daily risk assessment each day for two years, along with the daily closing NYSE index and the price of oil that day.

## Your Data Looks like:

Data Case #	Risk Assessment	Closing NYSE index	Oil Price
1	low	up	up
2	low	up	down
3	high	down	down
4	medium	up	down
5	low	down	up

## One Dependent Variable with Two or More Independent Variables [independent groups]

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	<u>Nature of Dependent Variable(s)</u>	Test(s)
1	2 or more IVs (independent groups)		
		categorical	factorial logistic regression

**Business Challenge:** Your product StainAway competes with other products to remove stains from clothing. You want to see how well it works against its competitors under varying conditions including prewash and type of stain.

**Your Response:** You purchase 450 identical white shirts and stain them with three different stains and either prewash or don't prewash them and see if the stain is removed.

### Your Data Looks like:

Data Case #	Stain Removed?	Brand	Stain Type	Prewash?
1	yes	StainAway	Blood	yes
2	no	SpotOff	Grass	no
3	no	CleannBrite	Gravy	no
4	yes	StainAway	Blood	yes
5	yes	SpotOff	Gravy	no

One Dependent Variable with One  
Interval Level Independent Variable

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 interval IV	interval & normal	correlation
		interval & normal	simple linear regression
		ordinal or interval	nonparametric correlation
		categorical	simple logistic regression



One Dependent Variable with One Interval Level Independent Variable

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------	---------------------------------	---------------------------------	---------

1

1 interval IV

interval & normal

correlation

**Business Challenge:** You make Kaboom brand gasoline. It has a secret mileage increasing ingredient – nitroboom – and you want to find some evidence that your secret ingredient increases vehicle mileage and if so, how strong is the relationship.

**Your Response:** You randomly select 450 drivers and each is given a free tank of gas with randomly varying amounts of nitroboom in it. They drive until the tank is empty and report the mileage they got.

**Your Data Looks like:**

Data Case #	Ounces of Nitroboom	Mileage			
1	20	320			
2	45	218			
3	64	265			
4	18	305			
5	44	288			

One Dependent Variable with One  
Interval Level Independent Variable

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 interval IV	interval & normal	simple linear regression

**Business Challenge:** As a dairy cooperative you are trying to establish the strength of the relationship between the amount of the feed additive Milkorama added to your cows feed and the amount of milk each cow produces and predict how much milk would be produced at different standard levels of Milkorama.

**Your Response:** You randomly select 250 cows across all of your farms and assign farmers a random amount of Milkorama to their cow feed. You then measure the milk production of each cow.

**Your Data Looks like:**

Data Case #	Ounces of Milkorama	Gals Milk			
1	10	20			
2	14	22			
3	8	15			
4	17	18			
5	45	32			

One Dependent Variable with One Interval Level Independent Variable

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------	---------------------------------	---------------------------------	---------

1

1 interval IV

ordinal or interval

nonparametric correlation

Business Challenge: As immigration satisfaction officer, you want to see if there is a relationship and if so how strong is that relationship between satisfaction of personal experience at port of entry and the hours of training an immigration officer receives.

Your Response: A random selection of 120 entrants are given the opportunity to hit a smiley face, a neutral face or sad face button after leaving their encounter with the immigration officer at the border.

Your Data Looks like:

Data Case #	Type of Face	Hours training			
1	smiley	25			
2	smiley	22			
3	sad	10			
4	sad	12			
5	neutral	18			

One Dependent Variable with One Interval Level Independent Variable

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------	---------------------------------	---------------------------------	---------

1

1 interval IV

categorical

simple logistic regression

**Business Challenge:** As customer retention director for TransTupolev airlines, you are trying to see if household income is related to the probability of people who fly your airline joining your frequent flyer program.

**Your Response:** You select 550 random passengers and acquire their household income figures, along with whether or not they are members of your frequent flyer program.

**Your Data Looks like:**

Data Case #	Member of Program?	Household Income in thousands			
1	yes	25			
2	yes	32			
3	no	55			
4	yes	65			
5	no	18			

One Dependent Variable with One or More  
Interval and/or Categorical Level Independent Variables

**Number of  
Dependent  
Variables**

**Nature of Independent  
Variables**

**Nature of  
Dependent  
Variable(s)**

**Test(s)**

**1**

**1 or more interval IVs and/or 1  
or more categorical IVs**

interval & normal

multiple  
regression

analysis of  
covariance

categorical

multiple logistic  
regression

discriminant  
analysis

# One Dependent Variable with One or More Interval and/or Categorical Level Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 or more interval IVs and/or 1 or more categorical IVs	interval & normal	multiple regression

**Business Challenge:** As the analyst for the Beaver Creek Power Cooperative, you are trying to better understand factors that drive residential power consumption. You think it has to do with size of house, urban or rural setting, number of occupants and average temperature set in the house. You also want to make a predictive residential model for power consumption.

**Your Response:** You select 1,550 random residences, get their power consumption for each day in the year, and collect data on square footage of the house, get temperature settings from their thermostat each day, note urban or not urban location and number of household members

Your Data Looks like:

Data Case #	Power (kw)	Sq Foot	Temp Set	Urban?	HH members	
1	30	2500	74	yes	1	
2	36	3200	76	yes	2	
3	44	3600	72	no	4	
4	18	1200	70	no	3	
5	64	5500	66	no	6	

One Dependent Variable with One or More  
Interval and/or Categorical Level Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
-------------------------------------	------------------------------------	---------------------------------------	---------

1

1 or more interval IVs and/or 1  
or more categorical IVs

interval & normal

analysis of  
covariance

Business Challenge: As an exercise professional, you are investigating the role of exercise in cholesterol levels among Hispanics and non-Hispanics. You also know that gender and age are factors.

Your Response: You randomly select 1,200 respondents, obtain their cholesterol levels, whether they exercise regularly or not, their gender and age

Your Data Looks like:

Data Case #	Cholesterol	Exercise Regularly	Gender	Age	
1	220	yes	male	28	
2	300	no	male	36	
3	178	yes	female	44	
4	155	yes	female	64	
5	256	no	male	22	

# One Dependent Variable with One or More Interval and/or Categorical Level Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	1 or more interval IVs and/or 1 or more categorical IVs		
		categorical	multiple logistic regression

**Business Challenge:** As the savvy owner of the Semi-Hard Rock Casino, you want to understand how age, marital status and household income affect the probabilities of visiting a casino in the last 6 months. You also want to build a predictive model for direct mail purposes.

**Your Response:** You randomly select 1,250 respondents and ask them if they have visited a casino in the last 6 months. You also record their age, household income and marital status.

Your Data Looks like:

Data Case #	Vist last 6 months?	Married?	Age	Hhinc	
1	yes	yes	54	32	
2	yes	no	42	64	
3	no	no	21	38	
4	yes	yes	26	44	
5	no	no	53	29	



One Dependent Variable with One or More  
Interval and/or Categorical Level Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	<u>Nature of Dependent Variable(s)</u>	Test(s)
1	1 or more interval IVs and/or 1 or more categorical IVs		
		categorical	
			discriminant analysis

**Business Challenge:** As the segmentation specialist for Blue Cow ice cream, you have created a market segmentation on the past 6 months customer data base. Now 6 months later, you have many new customers you want to classify into your segmentation system without changing the system itself.

**Your Response:** You gather data for the same variables that drove the original cluster solution and use discriminant analysis to predict each customer's cluster membership.

Your Data Looks like:

Data Case #	Cluster membership	sex	married?	Favorite ice cream flavor	
1	6	female	yes	vanila	
2	4	male	no	chocolate	
3	2	male	yes	Cherry Garcia	
4	?	female	yes	raspberry	
5	?	male	no	coffee	

## Two or More Dependent Variables with Varying Number and Nature of Independent Variables

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
2+	1 IV with 2 or more levels (independent groups)	interval & normal	oneway MANOVA
	2+	interval & normal	multivariate multiple linear regressions
	0	interval & normal	factor analysis

## Two or More Dependent Variables with Varying Number and Nature of Independent Variables

idre  
UCLA INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
2+	1 IV with 2 or more levels (independent groups)	interval & normal	oneway MANOVA

**Business Challenge:** As research director for the Forrest Gump tutoring centers, your students aren't doing so well and you suspect your clients are from "the bottom of the barrel". You want to compare your program to two other tutoring programs to see if there are any differences in the skill level of entering students in math, science and reading.

**Your Response:** You gather math, science and reading skill scores for entering students across the three tutoring programs.

Your Data Looks like:

Data Case #	Tutoring Program	math score	science score	reading score
1	Forrest Gump	220	315	240
2	ShinyKids	310	410	250
3	Einsteins	180	270	360
4	Forrest Gump	390	230	315
5	ShinyKids	440	390	235

## Two or More Dependent Variables with Varying Number and Nature of Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
2+			
	2+	interval & normal	multivariate multiple linear regressions
	0		

**Business Challenge:** As the research director for Corporal Mills cereal, you are investigating the sales of three different variations of the same cereal. Each variation has different levels of ingredients and processes and you want to assess how each affects characteristics of the cereal. You may also want to compare the cereals across some of these characteristics.

**Your Response:** You collect data from 600 boxes of cereal on the ingredient levels, processes as well as the characteristics of the three cereals – sweetness, crunchiness, amount of “cereal dust” and number of broken pieces.

### Your Data Looks like:

data case #	cereal	sweet	crunch	dust (gr)	broken piece	process	sugar (oz)
1	Scoobies	180	112	355	80	baked	12
2	Nerds	170	88	50	22	fried	20
3	Blinkies	440	235	230	36	boiled	15
4	Scoobies	158	118	310	72	baked	11
5	Blinkies	394	212	244	30	boiled	14

## Two or More Dependent Variables with Varying Number and Nature of Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
2+			
	0	interval & normal	factor analysis

**Business Challenge:** As the campaign manager for Hillary Clinton, you need to define nationalism and patriotism constructs to measure these feelings in the voting population.

**Your Response:** Your search of the literature uncovered factors such as thinking of your country as the best, imagining other countries trying to imitate you, strong feelings of loyalty to country, etc. You run a factor analysis on a set of 7 point scaled questions about these and then use factor analysis to produce constructs of nationalism and patriotism.

Your Data Looks like:

data case #	proud	best	imitate	feelings	loyalty	duty	
1	6	5	4	7	5	7	
2	4	2	1	5	5	6	
3	5	4	3	3	6	7	
4	7	6	3	4	4	4	
5	2	4	2	6	4	5	

# Two Sets of Two or More Dependent Variables with No Independent Variables

**Number of  
Dependent  
Variables**

**Nature of Independent  
Variables**

**Nature of  
Dependent  
Variable(s)**

**Test(s)**

**2 sets of 2+**

**0**

interval &  
normal

canonical  
correlation

## Two Sets of Two or More Dependent Variables with No Independent Variables

idre INSTITUTE FOR DIGITAL RESEARCH AND EDUCATION  
UCLA

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
2 sets of 2+	0	interval & normal	canonical correlation

**Business Challenge:** As the maker of BlueGreenPurple nutritional supplement, you want to measure the strength of the relationship between the perceived product benefits with how people who take your product feel.

**Your Response:** You collect data on 700 randomly selected customers and ask how they think your product should make people feel as well as how they think your product makes them feel. Potential benefits such as energy, stamina and alertness are measured.

### Your Data Looks like:

data case #	prod energy	self energy	prod stamina	self stamina	prod alert	self alert	
1	6	5	4	7	5	7	
2	4	2	1	5	5	6	
3	5	4	3	3	6	7	
4	7	6	3	4	4	4	
5	2	4	2	6	4	5	

# Summary

1. Selecting the appropriate statistical technique takes into account a large number of factors.
2. There are a plethora of statistical techniques available to use
3. There are oftentimes more than one way to analyze the data
4. It's important to get a good look at the data and think carefully about the business decision at hand before selecting a statistical procedure