**World Library and Information Congress:
70th IFLA General Conference and Council**

**22-27 August 2004
Buenos Aires, Argentina**

Programme: http://www.ifla.org/IV/ifla70/prog04.htm

Code Number: **175-E**
Meeting: **139. Library Theory and Research**
Simultaneous Interpretation: **-**

Why Research Methods for Librarians?

Wallace Koehler
Valdosta State University
wkoehler@valdosta.edu

cc[1] Wallace Koehler 2004 Some Rights Reserved

Abstract

This document provides a basic introduction to the why's and on a very limited basis the how's to the performance of research in librarianship. A number of subjects are address including different philosophical approaches, different statistical approaches, and the tools needed to "do" good research.

Introduction

Why Research Methods for Librarians? There are several possible answers to this question. The three most important are:

1. You may be called upon do research or you may decide to pursue a PhD or other research degree. The course provides you with a very rudimentary background to research methods.

[1] Creative Commons Copyright Attribution-NonCommercial-ShareAlike 2.0, for definition see http://creativecommons.org/licenses/by-nc-sa/2.0/

2. You will almost certainly read and use the research findings of others. A good grounding in research methods will help you evaluate the work of others.
3. Many librarians will be called upon to assist patrons in their research. In order to provide that service, you will need an understanding of the what, where, and how of your patrons' pursuits.

There are different approaches to doing research - hence the notion of the Toolkit

Toolkit

The Initial Tool is Judgment.

Research is needed. There are many questions that must be answered: But what kind? What is called for?

What resources do I have to do the research? How much time? How many people? What skills? What equipment, hardware and software?

Who is the research designed for?

Qualitative or Quantitative

Basic Concepts

Elementary Qualitative Concepts

The Case Study

Elementary Quantitative Concepts

Introduction
- Central Tendency
- Value Comparisons
- Testing for Difference
- Testing for Correlation

Quantitative Tools
- Bibliometrics
- Data Mining
- SPSS

Experimentation or Observation

Some research is conducted by setting up replicatable experiments. Others are best done through observation. Much of what the librarian does is "observation," but some research is "experimental".

For example, you might want to know if you can increase library use with open or closed stacks. As difficult as it might be, you might try one system, then the other to determine which increases (decreases) use.

There are numerous different approaches to research, some based on philosophy or theory, others on mundane matters as data type. There are measurement and concept issues.

Library Research Models

How to do research is one set of issues. How libraries are used to do research is another. See Thomas Mann for some valuable insights.

Philosophy of Science and Epistemology

Epistemology could be defined as the philosophy or the study of knowledge. How, it asks, can we distinguish between adequate, valid, or true explanations and inadequate, invalid, or false explanations? This leads necessarily to the question of how can we build better models or experiments or hypotheses to test scientific theory.

Or Epistemology is Scientific Methodology

This leads us to the next set of questions – what is knowledge? Knowledge has been defined in many ways. One definition has it as "integrated information." And integrated information – aka knowledge – is used to make decisions (Hayes 1992).

A Platonic definition of knowledge suggests that it is an imprecise but increasingly more precise reflection of reality. As knowledge increases through observation so do our understandings of the "real world." With such a definition, we appreciate the need for research. But that research must be conducted using appropriate scientific protocol.

The Platonic definition may be too restrictive, for it tends not to take cognitive structures (or what Kant termed categories) into consideration. These categories may include space, time, and causal variables. Like Platonic epistemology, the Kantian retained the idea of an absolute, static knowledge or set of concepts.

Pragmatic epistemology holds that knowledge models reality. All models are simplistic representations of reality. They are not reality itself. A valid, true, or accurate model provides approximations of "real" outcomes.

Or Research is a Path to Knowledge

But we continue to beg the question. How do we test the model or a set of propositions against reality? According to the coherence theory of truth we cannot test propositions against reality; but we test one set of propositions against a more generalized set of propositions. Thus we may ask if a theory or observation is consistent with a more generalized and accepted set of explanations.

The correspondence theory, on the other hand, holds that tests of hypotheses are not made against generalized explanations, but rather against some objective reality.

Elementary Quantitative Concepts for the Librarian

Who ever it was who said: "There are lies, damn lies, and statistics" said it all. Go to the library, find all the works on lying with statistics. Run out of time?

We are not "about" lying with statistics. This purpose of this block of instruction is to inculcate an appropriate discipline and appreciation of data collection and analysis. The lying part should come in the interpretation. So much for tongue in cheek.

Our purpose here is to expose you to the complexities of the process and to teach you an appreciation of "doing it right." I cannot stress strongly enough that the desired outcome is a beginning understanding of the complexity of the process. I am not here to teach statistics, survey methods, etc. This course is designed to explore the research process as it is appropriate for the librarian and the information scientist. We will see things some of you will promptly forget immediately after comps. Others of you may be bitten by the bug, and will seek further enlightenment.

I want you to come away with the ability to interpret and evaluate the work of others and to be able to plan and formulate fairly simple research projects of your own. Professional librarians are no longer "makers and parkers" of books, if ever they were. Librarians are managers, politicians, supervisors, and advocates. Sometimes a little statistics comes in handy when wearing these hats.

Simplified Version

We will delve into a whole new language, a whole new jargon for most of you. This page might be treated as a glossary or dictionary. I will attempt to define a variety of terms in fairly simple and straight forward ways. Remember please, whenever the previous statement is made, the end result is often somewhat unsatisfactory and incomplete.

Data Types

- Statistics are population measurements, numbers that describe an underling data set.
- Statistical data need not be numbers.
- Statistical data come in four degrees of complexity. These are nominal, ordinal, interval, and ratio.

Nominal

Nominal data are information about which one cannot make relational statements - except to not that they are different. Examples are gender, race, and other categorical data. Consider exterior walls of houses. Some are adobe, some brick, some clapboard, and so on. One cannot say a brick house is twice as homey as a wooden one. In effect, there are no objective relationships. There may well be subjective ones -- brick is prettier than ______ take your pick.

Ordinal

Ordinal data are information that can be classified relative to similar ordinal data. Takes grades for example: "A" indicates outstanding performance, "B" good, "C" OK, "D" mediocre, and "F" the kiss of death. There is a relationship expressed among the values, and the directionality of difference is expressed. But it is not possible to measure the degree to which those values differ.

Interval

Interval data are information where both the direction and the amount of difference are expressed, but no other relationships can be. The classic example is temperature. If today it is 10 degrees C outside and tomorrow 20 degrees, is it twice as hot today than yesterday? How about if we express the same temperature in Kelvin? It is 360° K today against 350° K yesterday.

Ratio

Ratio data have the same characteristics as interval, with the added advantage that ratio relationships can and are expressed. If I earn $100 000 per year and you earn $200 000, not only do you earn twice as much as I do (but I should change jobs). We can express magnitude, distance, and ratio.
Understanding and recognizing these differences is critical for the statistician. First, some statistical tests are appropriate for different types of data. As we will see, in bivariate and multivariate situations, there can be a mix of data types. It is necessary or desirable to select the appropriate test given data types.

Sidebar

> Statistics are frequently misapplied. Consider the all important Grade Point Average.
> How is it generated? Students receive final grades in each of their courses. At most universities, these take the form "A" thru "F". To generate the GPA, each of these letters is assigned a value, usually a range from 4 for A to 0 for F.
>
> Consider first the process that results in individual course grades. We all know these are based on ONLY interval level criteria - Right!
>
> We know that A's are better than B's etc. But how much better. By definition, one point better. All A's are equal and so on. What does that point represent?
>
> Records Offices then average (take the mean of) these letters rendered as numbers to generate Grade Point Averages. GPA's then follow you for a long time. But they are based on unsound statistical assumptions. You cannot take the mean of ordinal data.
>
> Some will argue that the GPA is robust. Robustness is a term applied to a statistic that "appears" adequate to describe data "above its station." Some might argue (and do) that the GPA is interval appearing or interval like. If so, a statistic - like the mean - can be employed robustly.

Populations and Samples

We can generally discriminate between two different classes of data: (1) populations or universes and (2) samples. Consider the Year 2000 US Census. About what does it try to collect data? The population of the United States. How many people does it try to count. For the simple version, the answer is everyone inside the borders (OK quibble some). And from whom does the complex version seek responses? One household in six.

STOP and Think. What is the level of analysis for the US Census? The individual or the household? And what is the target population for the general census -- the entire population, which is the universe of possible respondents. What is the target for the "long version"? A sample of one household in six. The sample frame calls for the Census to collect data for both from households.

The Census Bureau would prefer not to have to query every household in the United States. Why do they? Because there is a Constitutional mandate that they do. There are few if any other surveys with a Constitutional or legal mandate to "get them all."

Surveying the Universe

Collecting data, surveying the entire population is easy to conceive, but often times difficult to achieve. Say I wanted to survey IFLA 2004 participants and only IFLA 2004 participants. Is that acceptable? Sure.

Given its size, how hard would it be for me to survey all of you?

Not very difficult. How about everyone in Argentina? A bit more difficult -- and I would have to define "everyone in Argentina."

How about all of Atlanta, Georgia, the United States, all Republicans, Democrats, nudists, dog owners, nudist Democrat dog owners? Suddenly the magnitude of the problem of finding, identifying, surveying all members of the target class becomes monumental.

However, given enough time and resources anything can be done. But time and resources are not infinite. Suppose I am a candidate for public office. Suppose I want to know voter reaction to some campaign promise I have just prevaricated. Suppose I want to know the answer NOW. How much time would my pollsters have to generate that data?

Consider cost. The 1990 Census cost $2.6 billion. The Congress budgeted in 1998 $1.027 billion to plan the 2000 Census. In the words of the Everett Dirkson (Senator from Illinois) "A million here, a million there. Pretty soon it adds up to real money." $1 billion is real money. Most surveys do not have budgets even five orders of magnitude lower than $1 billion.

Cost alone precludes surveying everything and everyone.

Sampling the Universe

What is a sample - statistically speaking? Sampling is that process of selecting some proportion of the universe, based on some set of assumptions, to represent the universe or population. Sampling creates all sorts of methodological headaches. How big a sample, what to sample, when to sample, where to sample. Say I was interested in nudist Democrat dog owners. How much sense would it make if I were to limit my sample to Republican cat owners and haberdashers? How might my sample be distorted if I were to acquire the membership list of the Association of Democrat Nudists or the Association of Nudist Dog Owners (are these nude dogs or nude owners?) and survey their members.

We will talk in this class about sampling bias and statistical bias. Sampling bias incorporates all conditions that inherently skew the sample away from representing the population. As suggested already, sampling bias can be introduced in all sorts of ways.

What, where, and how you sample will depend upon the resources available to you. Remember always the first law of data collection:

Sample as large a group as possible. Collect data at as great a level of complexity as you can. Once the data are collected, you can always aggregate or simplify. You can NEVER disaggregate or make more complex.
. The operant word in the "law" is possible . If you have $1,000,000,000.00 and a staff of thousands much more is possible than if you have what is in your pocket and a staff of one to do the work.

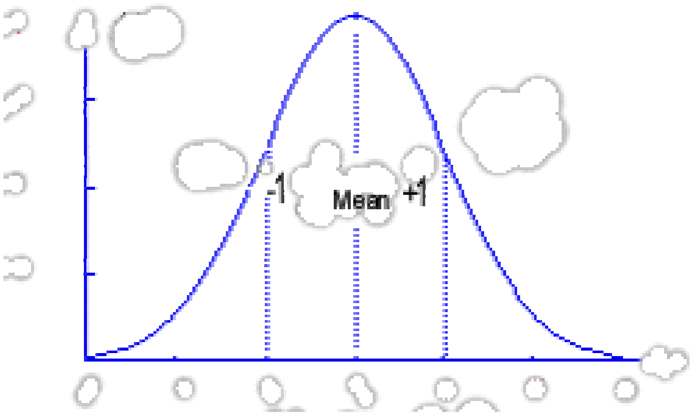
Parametric and Nonparametric statistics

Parametric and Nonparametric statistics are also known as distribution and distribution free tests. Some statistical tests require us to be aware of and sometimes correct for inherent characteristics or parameters. Others are parameter free: Hence parametric and nonparametric tests.

What are examples of parameters - means, standard deviations and the like. These values are often standardized for one of two purposes: (1) to allow comparison between two or more distributions or (2) to normalize the distribution.

We will introduce the idea of normalization here, then do not much with it in this course. Some parametric tests require that the distribution of the sample approach a "normal distribution" or resemble the standard bell shaped curve of statistical fame. Remember algebra 1? Remember x=a. If you do the same thing to both sides of the equation, the equation remains true --
If x=a, then: $x/2=a/2$, $\log_x=\log_a$, $xE^{15}=aE^{15}$, and so on. Same is true in manipulating curves in statistics. The relationship among the numbers remains while the curve is "adjusted." Then the parametric statistics are more precise. Tests of significant difference and tests of covariance based on differences or comparisons of means and similar measures are parametric.
Those that do not rely on such distributions are nonparametric.

The Bell Shape Curve- or Normal or Gaussian Distribution



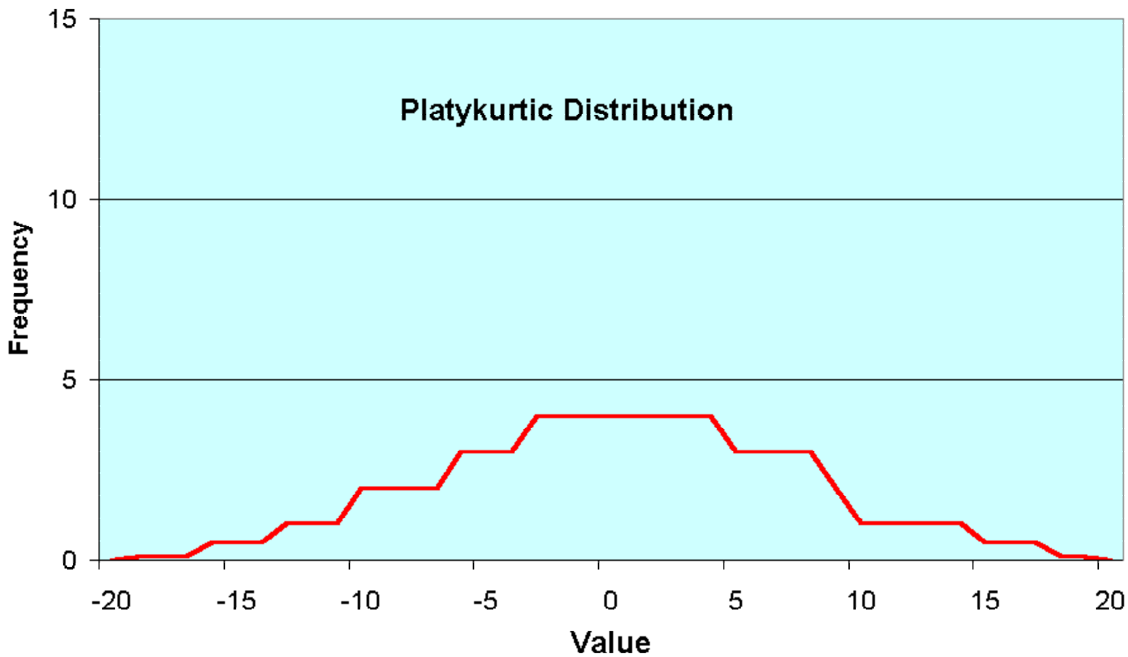
In a normal distribution, about 68% of all responses are found plus/minus one standard deviation from the mean, 96% are found within two standard deviations and 99.7%within three standard deviations. Anything outside plus or minus three standard deviations is rare indeed. But they are not unheard of.

Distribution "Shapedness"

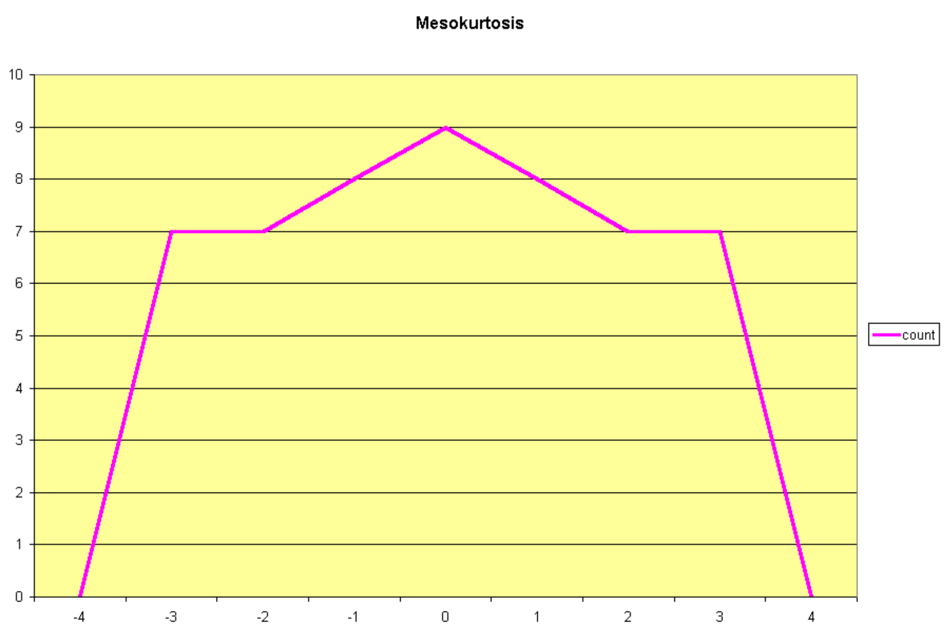
"One-hump" variations

We will also speak of kurtosis and skewedness of distributions. Kurtosis has to do with the sharpness or flatness of the peak - the size of the tails of the distribution. To expand vocabularies, a distribution with "large" tails is called leptokurtic, those with small tails platykurtic, and those that are normally distributed are mesokurtic. Negative values are associated with platykurtic distributions, positive values with leptokurtic.

Platykurtic Distribution

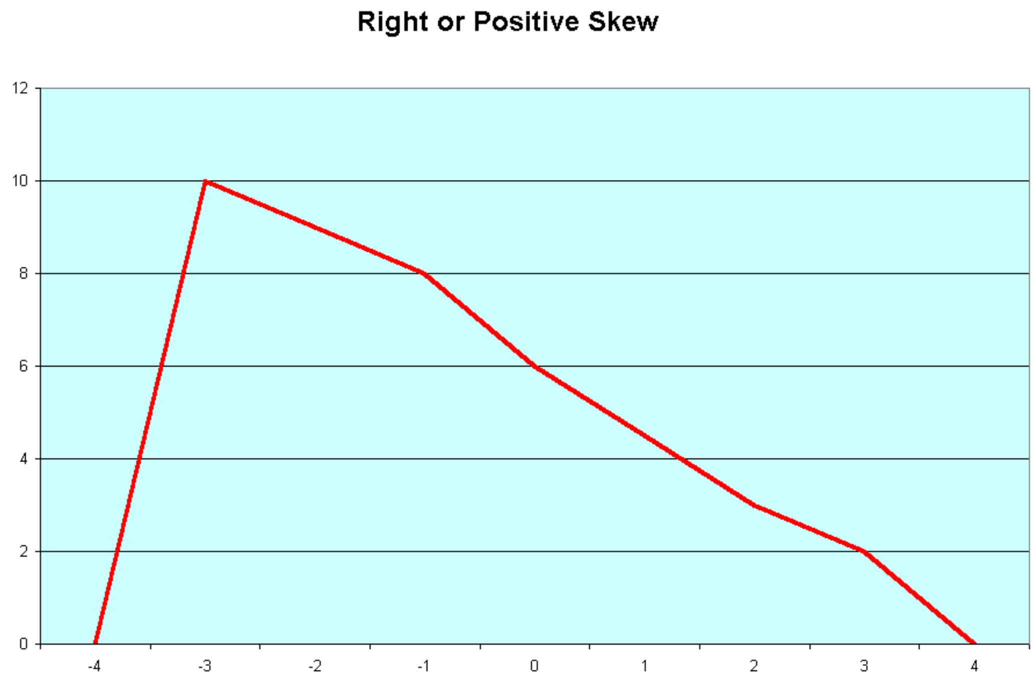


Leptokurtic Distribution

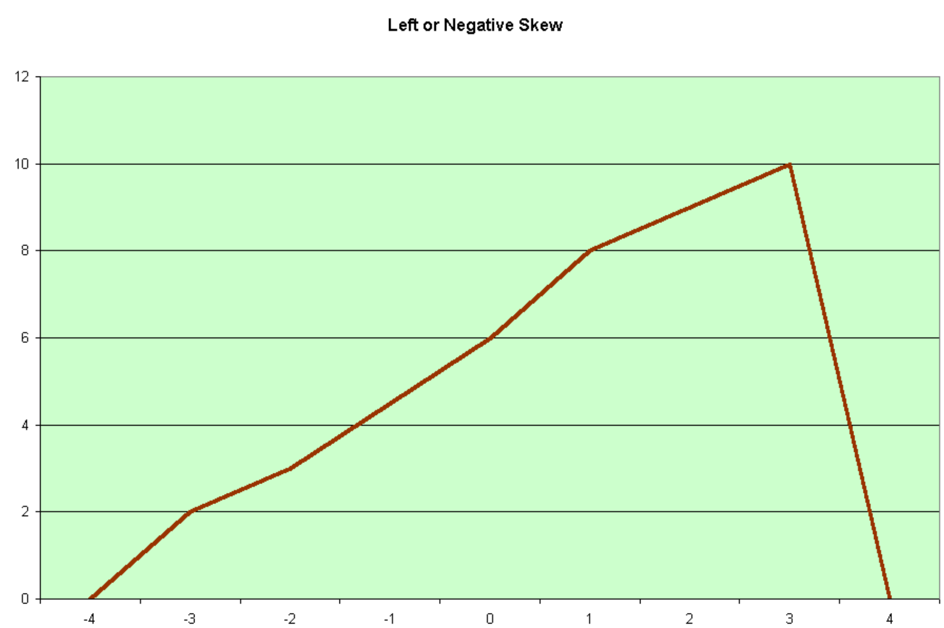


Skewedness has to do with the right/left placement of the curve.

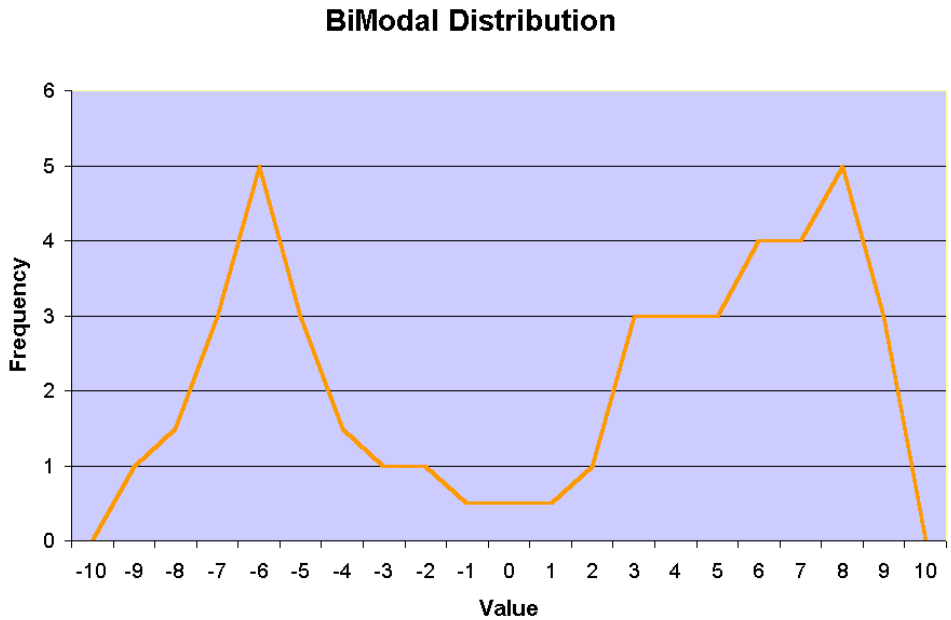
Right or Positive Skew



Left or Negative Skew



"Multi-hump" variations



In "statistics-world" all distributions are normal. In the real world, distributions can take all sorts of forms and shapes. This phenomenon is neither good nor bad, but the good statistician is aware of his/her data distributions and responds accordingly.
The most common variation is probably a multi-modal distribution. A multi-modal distribution is one in which the most frequently occurring value occurs more than once.

Tests of Central Tendency: Averages, Range, and Standard Deviations

Central Tendency describes the distribution of a sample or a population around its center. We use a variety of statistics to accomplish this end. These include averages, range, and standard deviations. There are others, but we will begin with these.

Example of means and standard deviations.

Take a look at samples A-1 and A-2, in green, below. Let's start with the range. RANGE is defined sometimes as the difference between the lowest value and the highest. And sometimes it is defined as the (highest value less the lowest value) minus 1. Take A-1. The highest value is "3" the lowest "1". The range can be calculated as R=3-1=2 OR as R=(3-1)-1=1.

The logic behind the second calculation is that there is but one number separating the highest and lowest values. What problem do we encounter calculating range the second way with A-2?
Range gives us the dispersion of the sample, but tells us little about the distribution of the values within the sample or population.

Note the following two examples A-1 and A-2. What are the modes, medians, and means for both examples? In this case they are "2" Why. The mode is 2 because it is the most frequently occurring value. The median is 2 because beginning from lowest to highest, 2 is the middle value. The mean is 2 because the sum of all values divided by the number of values is also 2. This of course does not usually hold true.

Note however, that even though all "averages" are 2, the standard deviations are different. Why?

	A-1	A-2
	1	2
	1	2
	1	2
	2	2
	2	2
	2	2
	2	2
	3	2
	3	2
	3	2
MEAN	2	2
STD DEV	0.816	0.000

Different Values and their comparison
The table below contains values -- let us say for height. The first column provides those values in inches, the second in centimeters. NOTE that the cm column was calculated by multiplying the first by 2.54.

If I were to ask you to comment on the two sets of values, what conclusion could you bring. If you knew one was inches, the other centimeters? If you did not have the knowledge? If you did not have the knowledge, what could you conclude from a comparison of other sets of means, standard deviations?

	in	cm
1	60	152.40
2	61	154.94
3	62	157.48
4	63	160.02
5	64	162.56
6	65	165.10
7	66	167.64
8	67	170.18
9	68	172.72
10	69	175.26

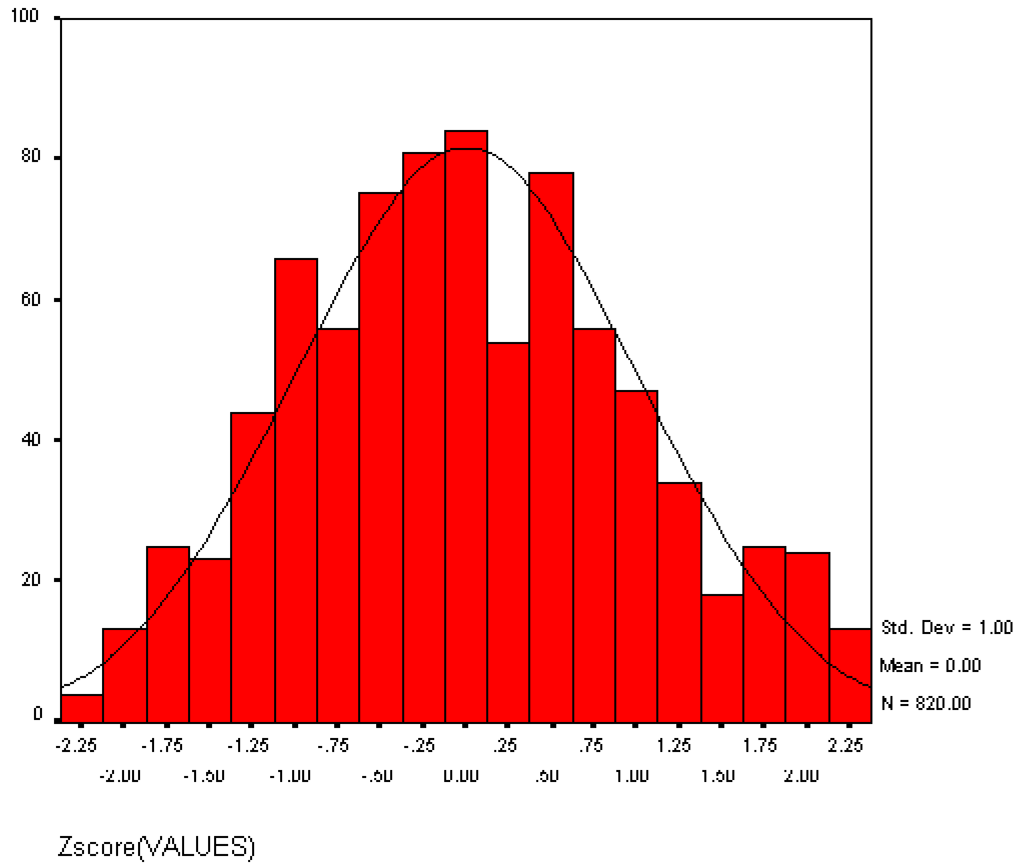
Mean	64.5	163.83
Std Dev	3.028	7.690
Z-scores		
	in	cm
1	-1.486	-1.486
2	-1.156	-1.156
3	-0.826	-0.826
4	-0.495	-0.495
5	-0.165	-0.165
6	0.165	0.165
7	0.495	0.495
8	0.826	0.826
9	1.156	1.156
10	1.486	1.486
Mean	0	-1.5E-15
Std Dev	1.000	1.000

Now consider standardizing the data using Z-scores. First, how are Z-scores calculated?

$Z_i = (a_i - x)/\text{stddev}$

Which means, each individual Z-score is calculated by subtracting the mean (x) from each individual value (ai) and dividing that value by the standard deviation for the sample. Why then, might you ask, are the Z-score mean values for "inches" and "cm" different. Remember, the cm calculations are based on inches. The table was created in Excel, and a rounding error has occurred. Note that the Z inches mean is "0", and both standard deviations are 1.000. The mean Z for cm is -1.5E-15. Or -1.5^{-15} For those of you who are not math inclined, that reads minus 1.5 to the minus fifteenth power -- or mighty close to zero.

The following curve represents a normal distribution. It is one I "made up." The bars represent actual data points and each bar represents all values within .25 standard deviations. The data were purposefully generated to be "unbalanced," that is to say it is not a "perfect" distribution. Few distributions in life ever will be.



Take note of what is going on here. Where are the interval values? What represents ordinal values? Can we conjure up nominal values?

Testing for Difference

We also concern ourselves with two general classes of bivariate tests: tests of differences and tests of "togetherness" (covariance or correlation). This section addresses difference.

Let's pretend that we are beings from the fourth planet from the star Alpha Librarian. Like Captain Kirk, we are on a scientific mission that brings us to the third planet from a sun known locally as Sol. It is alleged that Sol-3 supports intelligent life. It is our assignment to investigate those life forms.

Let us also assume that our resources are limited. We can only beam up two individuals. The first we capture from a clan known as the National Basketball Association, the other from a collective called kindergarten. Now, we have two specimens. One specimen towers over the other. One is 36 inches (1 meter) tall, the other 84 inches (2.75 meters). On initial inspection we find that the smaller one has all the appendages that the taller one has.

They are similar in all but for height. We must generate some initial hypotheses. Remember our sample. Given that sample, just how confident (whoops, I've introduced a statistical term here) are you about your conclusions?

1. Human beings are on average 60 inches (1.5 meters) tall.
2. They come in two types. The alpha type is shorter and always lacks an appendage. The beta type is taller and always has an appendage.
3. Both alpha and beta humans like to play games.
4. What else?

More realistically, remember SATs, GREs? I seem to remember that 400 is "average" for both (if not, play along).

Who takes the SAT and the GRE? SATs are taken by American high school students who intend to attend college. It is not taken by all high school students. It is not taken by all people in the appropriate age cohort. Therefore the SAT is not a measure of all 16 to 18 year olds in the US. It is a measure of those who believe they are college bound.

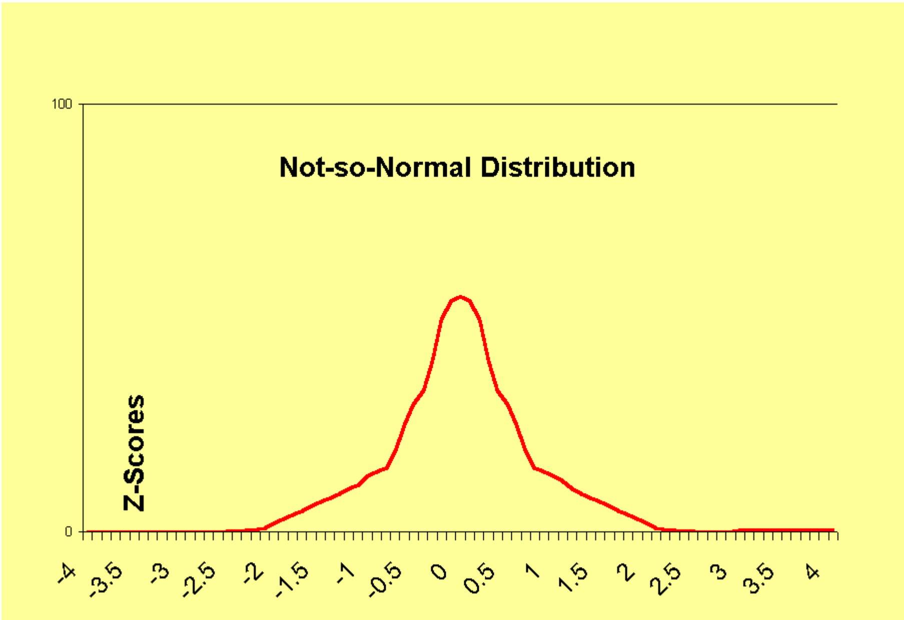
Similarly, the GRE does not measure all college seniors. It is taken by those students who anticipate that they will attend graduate school. Those headed for business, law, or medical may very well not take the GRE, they have tests of their own.

The point of all this is to emphasize that one must always be aware of the population one is drawing from and comparing to. Now back to the original example. How typical of the human race are the two groups from which our alien friends sampled?
What if we knew something more, and realized that kindergarten and the NBA are two different populations of human beings?

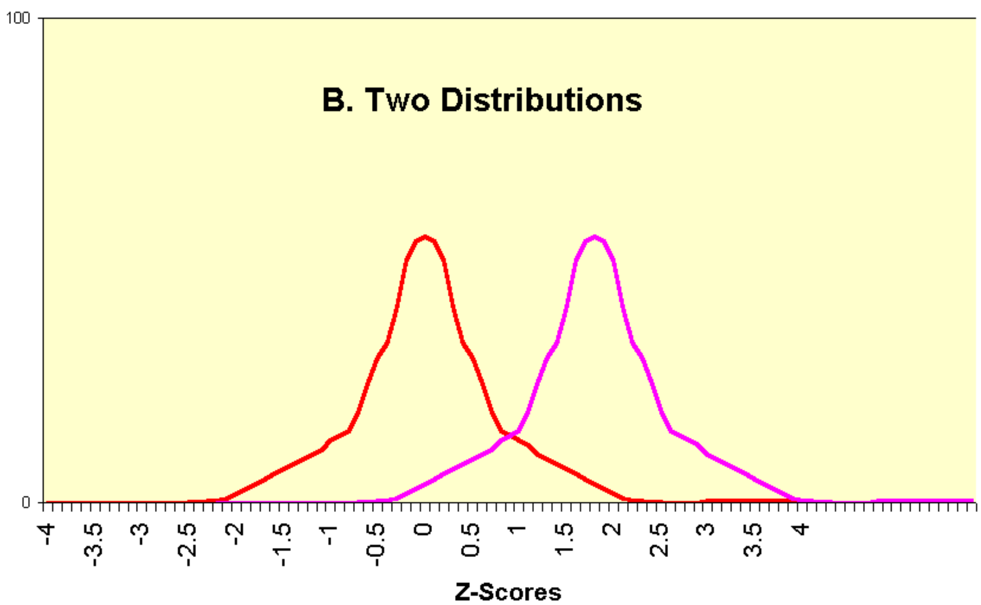
We can perform a number of bivariate and (as it happens) multivariate tests to determine if two samples are alike or different from one another.

Parametric Tests of Significant Difference

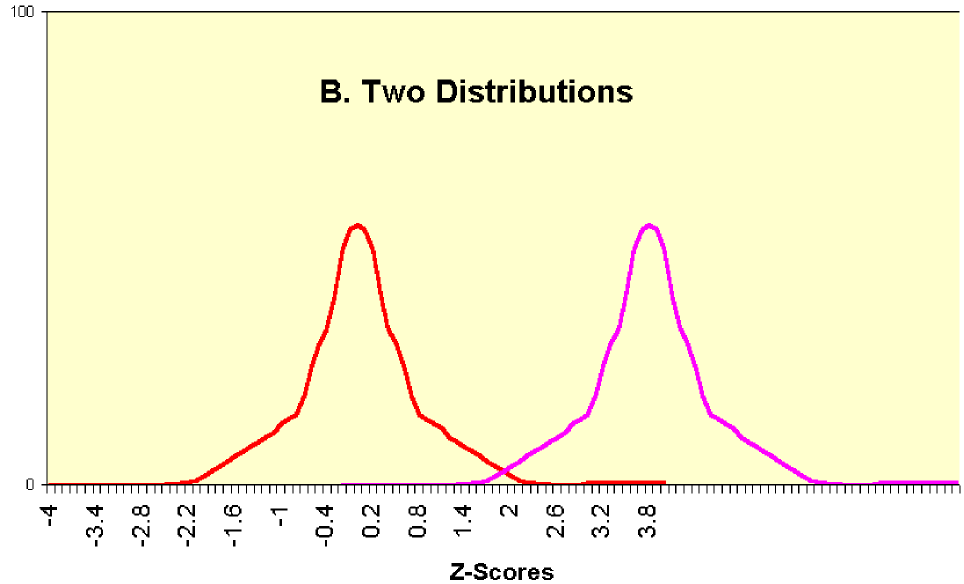
The three figures below show (1) the "normal" distribution for a single sample, (2) two distributions, and (3) two distributions but more distant from one another.



100
Not-so-Normal Distribution
Z-Scores
0
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4



100
B. Two Distributions
0
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
Z-Scores



The distributions we are examining require labels or titles. Consider the distributions in "red." They may represent the universe or population of individuals or they may represent a sample from that population or universe. The lavender lines might represent separate samples. Consider graphs B and C. We may have population data (red line) and we may wish to know whether our sample is representative of the population. Or, we may have two samples, and we wish to know whether our two samples were drawn from the same population or not. And we wish to know whether our conclusions that they are either the same or different have statistical significance.
As we will see, there are all sorts of factors that affect the conclusion that one group is different from another. Remember our aliens. Do you think a sample of one kindergartner and one basketball player can tell us very much?

We refer to sample size as "n". Our aliens have an n of one for each sample. It's not enough. To put it simply, the bigger the "n" the better off we are.

So what do you think about distributions B and C below? Which do you think is more likely to be judged statistically different from the other? Why?

Statistical tests

We concern ourselves with two parametric tests of significant difference: Student's t and ANOVA (analysis of variance). There are many more parametric tests, but we will not concern ourselves with them.

ANOVA is based on the F distribution (observed variation of independent group means)/(expected variation of group means) ANOVA can be used for more than two groups (variables). The important ANOVA is the F-statistic or test of the difference of group means. There are three elements that drive the F-statistic: (1) the size of the difference between group means (see figures B and C above - which is greater and which is less?), (2) sample sizes for the groups, and (3) variances.

The t-test is very similar. It is based on the t-distribution and applies to two groups. It is a variant of ANOVA. Both ANOVA and the t-test generate a statistic (F or t). That statistic alone is not sufficient to determine statistical significance.

The second element is a calculation of degrees of freedom, a number based on sample sizes or number of groups (d.f. = k-1 for between-group tests where k=the number of groups and d.f. = n-k for within groups, where n=sample size.)

The F or t together with the two sets of degrees of freedom calculations gives us a probability level that can reject the null hypothesis. Thus, if the F- or t-statistic is equal to or greater than the threshold value, the null hypothesis can be rejected.
That is usually but not always what we want. The idea of null hypothesis and probability are discussed in the probability section.

I am not going to show examples of the calculation of ANOVA or Student's-t. There are plenty of these in text books, in the SPSS manual, and on the WWW. What I want you to be able to do is to interpret the numbers once a program provides them to you. If it helps you to work though a example, see me, or check out any of the many resources available.

Nonparametric Tests

Nonparametric tests are employed to test for differences between samples or populations for nominal and ordinal level data; where one does not have the luxury of means, standard deviations, and other such stuff. In effect what is done is to compare the expected to what is found. For example, suppose our aliens above know that human populations are more or less 50% female and 50% male. Suppose also that by some quirk of fate they were to beam up all English football hooligans (let's say there are 5243 of them). What do you suppose the division of the sexes is there -- 99% male, 1% female perhaps. Because gender (assessed phenotypically) is a nominal variable, how might we determine whether our hooligans are --by gender distribution -- representative of the human race?

We might do a 2x2 table:

	Female	Male
Hooligans	99 1%	5144 99%
All Humans	5.2 billion 52%	4.8 billion 48%

At a glance, do you think the distribution of hooligans by gender begins to approach the human gender distribution?

How would you go about demonstrating that the two groups are "different" or the "same"? OK, what do we expect? About 50% of each gender. What do we have - 1% female, 99% male.

Another example. Say our aliens want to know which is taller, South America men or women. Say too that adult South American women are on average 5'4" (1.62 m) and men 5'8" (1.72 m). Of course our aliens do not know this. Say they have a stick that is 5'6" (1.67 m) long. They take their stick all around the continent and compare it to the inhabitants they encounter. They score their findings as above (taller) or below (shorter). Ties are scored as ties. Our aliens also base their assessment of gender on inspection. They are not quite sure how to determine adulthood, so they ask. No response or "goo goo" is assumed to mean non-adult.

Once back at the mother ship, they pool their data. To their amazement, they find that half of the sample is taller and half shorter than the stick. But the sample distributes as follows:

Group 1	Above	Below
Female	200 20%	800 80%
Male	800 80%	200 20%

Not to be outdone, a second group of alien scientists also sends out a group with a stick. But this time, the stick is 15' in length. Their findings are just a little different:

Group 2	Above	Below
Female	0 0%	1000 100%
Male	0 0%	1000 100%

Group 1 concludes that American men and women differ by height, group 2 does not reach the same conclusions. Why? Could it somehow be related to the arbitrary choice of the measure? One group defined "tall" as over 1.67m , the other as over 4.57 m. Is there anything wrong with either definition?

Stop for a minute. Go back over the discussion. What other potential sources of error are there is devising these observations?

What was the research question? Something like which is taller, Southh American adult men or women.

(1) How did they determine who is "South American" and who is not. Not stated. Possible error?
(2) How did they determine gender? By observation. How close an inspection did our aliens make and what criteria did they use to discriminate between groups. Not stated. Possible error?
(3) How did they determine that the subjects were adults? They asked. What was the question they asked? In what language? Was it phrased the same each time? What are the criteria for determining adulthood and did all subjects agree with those criteria? Not stated. Possible error?

Add to that:

(4) Are we certain that all sticks used by Group 1 investigators were exactly the same length? By Group 2?
(5) Did any of the investigators "cheat" either by "eye-balling" subjects rather than measuring them or by making up results? It happens.

Lesson Learned

When you write a proposal or write up the research report, you will typically include a section entitled "Methodology" In this section, you will describe how the research was devised and carried and how you addressed possible sources of error. Sometimes it is necessary to accept a certain amount of error and you must acknowledge the potential damage that error might create. You must also acknowledge what, if anything, you have done to correct for it.

Almost all research has "something" wrong with it. The key to being a good researcher is knowing when and what that "something" is and when it is fatal.

Remember also the Cheshire Cat in *Alice in Wonderland*. Any word or concept means what you define it to mean. Define your terms. So long as you use the term consistently thereafter you are "safe."

But also remember as you consume research findings not to confuse your definition with that used in the research.
Statistical tests

There are several nonparametric statistics for determining differences. We will limit ourselves here to chi-square (χ^2).

$$\chi^2 = \text{Sum of } ((O-E)^2 / E)$$
where O = observed,
and E = expected.

Now, let's do one. Here is the Group 1 human height test. We already know the observed values, as outlined in the table. What are the expected values? We need to sum both rows and columns. Female and male row total are 1000, as are above and below column totals (see how convenient these things are when you make them up!?!)

To derive the Female above figure, we multiply the total observed (1000) for the cell times the row total (1000), then divide by the total number of observations giving us (1000*1000)/2000 = 500.As it happens, because we set it up that way, the expected value for each of the four cells is 500 (one of the conveniences of making these things up).

Group 1	Above	Below
Female	200 20%	800 80%
Male	800 80%	200 20%

	Observed	Expected	
Fa	200	500	$(-300)^2/500= 9000/500 = 45$
Fb	800	500	$(-300)^2/500= 9000/500 = 45$
Ma	800	500	$(-300)^2/500= 9000/500 = 45$
Mb	200	500	$(-300)^2/500= 9000/500 = 45$
			180

χ^2 =180. That's nice, so what?

We must next calculate the degrees of freedom of the table. The formula for degrees of freedom is

$$d.f.= (r-1)(c-1),$$
where r=row and c=column.

In our case we have four rows and two columns, or (4-1)(2-1)=3.

Again, so what?

Consult the section on probability. But, to put it in simple terms, one can reject the null hypothesis (that males and females do not differ in terms of height) for a table with three degrees of freedom at the p=0.05 when χ^2 = 7.82, at p=0.01 when χ^2 = 11.35, and at p=0.001 when χ^2 = 16.27. In our case, χ^2 = 180. We easily reject the null hypothesis (that they are the same) which leaves us with the conclusion that male and female height is statistically significantly different.

That's fine. If our question been either that men and women differ in height or that men are taller than women. What if our hypothesis had been "Women are taller than men"?

In the end, the interpretation of statistics depends very much on the question asked, then and only then on how the values generated by any given test relate to the question.
Note that "p" values are published in tables in any statistics textbook. They are also provided as part of printout from most statistical programs, including SPSS.

Tests of Correlation

We will concern ourselves with two general classes of bivariate tests: tests of differences and tests of "togetherness" (covariance or correlation). This page addresses correlation. Our concern as always in this course is with bivariate tests. There are a large number of multivariate tests for correlation. That is for the more advanced and adventurous of you.

The concept of correlation tends to bewilder many students in the beginning. We will try to clarify these issues.

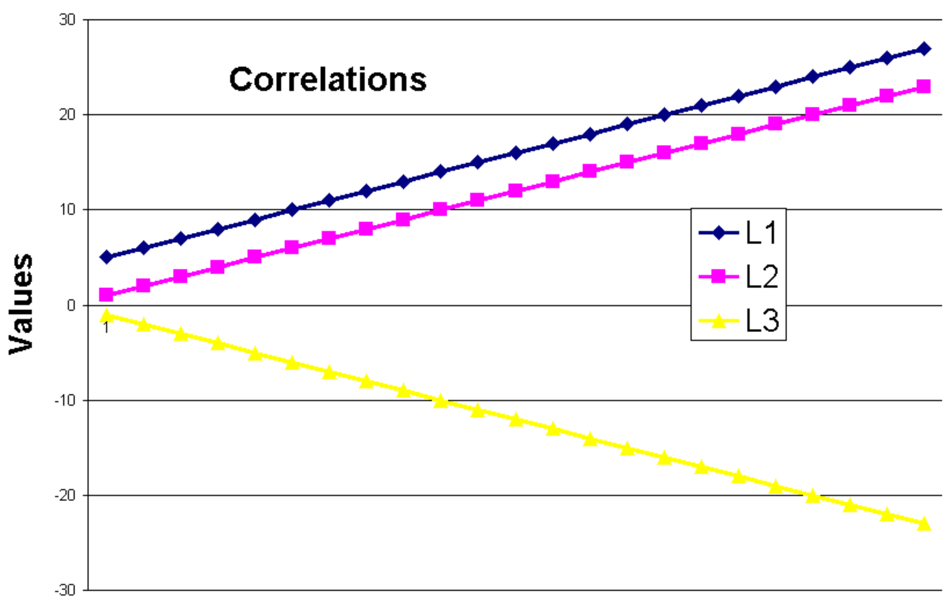
First, what are statistics. Simply put, they are a mathematical expression of the relationship between one or more groups of numbers. Those numbers, presumably, can represent "reality."
One set of relationships is difference. We usually have little difficulty grasping the "difference" concept. Medical doctors as a group earn more per year than librarians do as a group. Men have a different life expectancy than women.

There is a different set of relationships among as well. These are concerned with the "movement" of data. As one set of data changes, so does a second. Consider. If you consume 1500 calories a day, your weight will remain constant. If you consume fewer than 1500 calories, you will loose weight and if you consume more, you will gain weight. Moreover, the greater the number calories you consume, the more weight you gain. The fewer the calories, the more weight you lose. There is a correlation between calorie consumption and weight gain and loss.

This relationship between variables may be either positive or negative - do not impute good or bad here. In a positive relationship as one increases so does the other; and as one decreases so does the other. In a negative relationship, as one increases, the other decreases and vice versa. An example of a negative relation (with a desired outcome) could be as more children are vaccinated (increase) against childhood diseases, the fewer the number of children who die from childhood diseases (decrease).

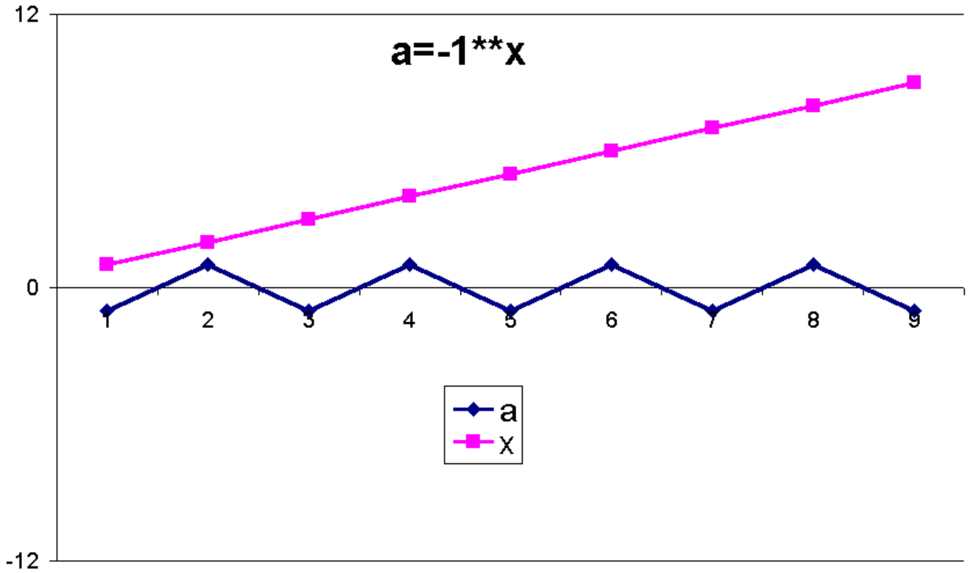
The statistics we use to measure these phenomena are both parametric and nonparametric depending on the underlying quality of the data. As a general rule, the coefficients (values) generated range between +1 and -1. A correlation of +1 indicates that as one value increases, so does the other and ALWAYS by the same amount. Likewise a value of -1 indicates that as one value increases, the other decreases and ALWAYS by the same amount. A value of zero tells us that there is no relationship between the variation of one value and of another.
We can express these simply with two equations: a=b a=b-x, and a=(-b).

These might be graphed as follows:



where L1=a=b, L2=a=b-x, and L3=a=(-b).

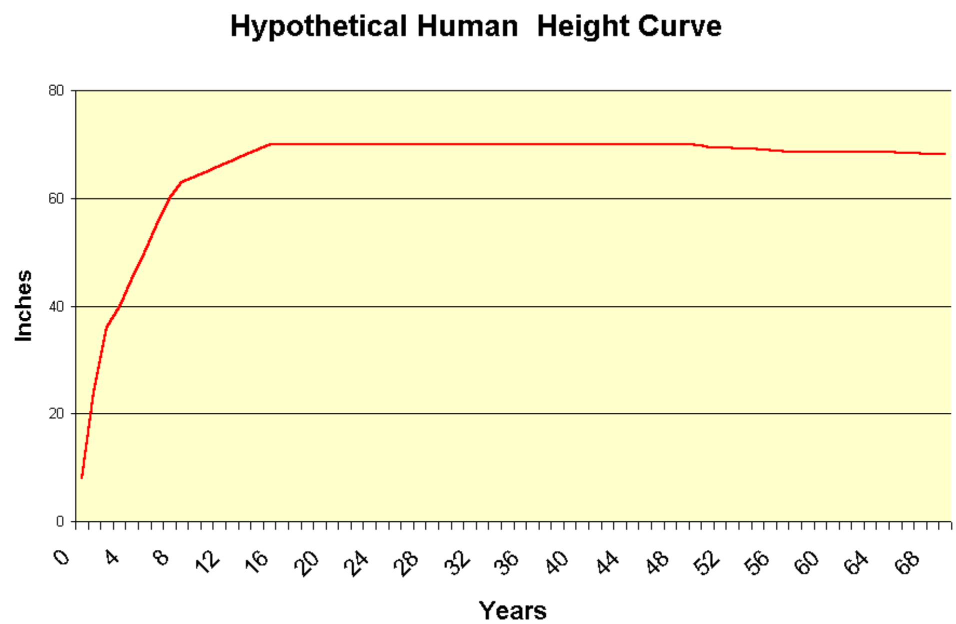
Unfortunately we do not live in a linear world. Consider another simple equation a= -1x



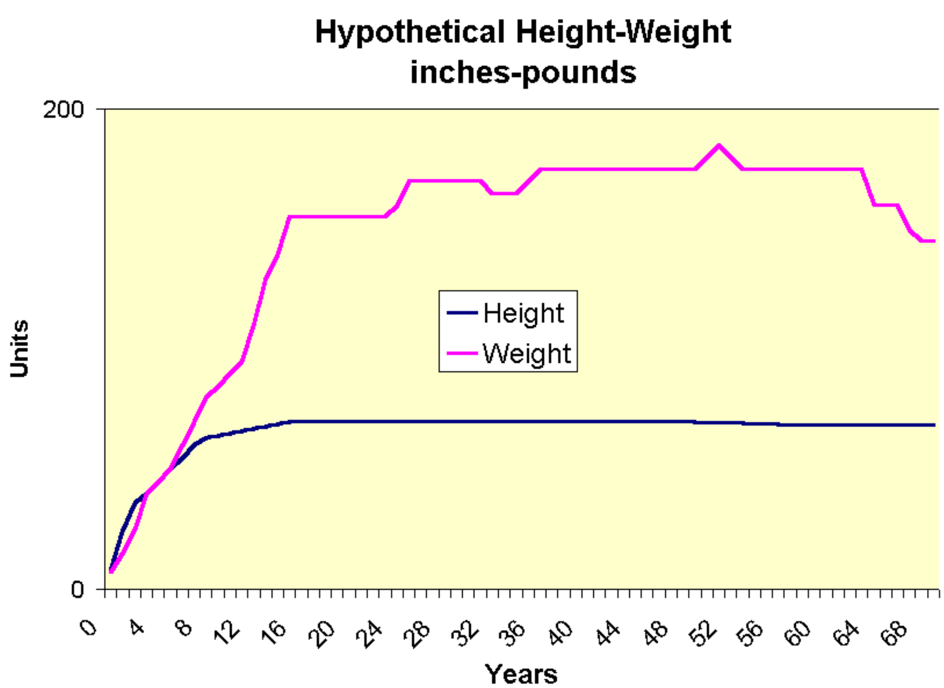
Note the linear increase in the value of x, and the very unlinear behavior in a. The linear correlation between a and x is very small, essentially non existent. However, we know that there is a definite non-linear, mathematical, systematic relationship between the two.

Be warned. Just because, at first blush it seems that there is no relationship between variables, it does not mean that there isn't one. ALWAYS examine your data. Get to know them very well. Graph them. Explore them.

Let us now look at "real life" relationships. Just as mathematically we can have non-linear relationships, so we can in real life. First, let us look at height of individuals over time. Once we are born, we grow rapidly for a while, we stop growing, and if we live long we begin to shrink. This is demonstrated in the following made-up figure:



As we age, our height does not continue to increase. Now what about the dreaded weight as a function of height. The taller we become, we the heavier we become. What happens when we stop growing (upward that is)? Again consider the following possibility:



What can we say about the relationship between height and weight between the ages of 0 and 12? 20 to 50? After 50?

Is it legal to break the data apart and explore fragments? You bet it is, and when you are limited to linear statistics in a non-linear world, you have little choice but to do so.

Statistical tests

We have several tests available to us from SPSS.
For interval data we will us Pearson's r
For ordinal data and interval ordinal data we will use Spearman's rho and Kendall's tau
For nominal data we will use eta and lamda.
There is an excellent table in your text that describes what is appropriate when. Find it. Internalize it.

Quantitative Tools

See a Powerpoint presentation on bibliometrics at
http://books.valdosta.edu/mlis/IFLA04/bibliometricwp_files/frame.htm

See a Powerpoint presentation on Data Mining at
http://books.valdosta.edu/mlis/IFLA04/Data%20Miningwp.htm

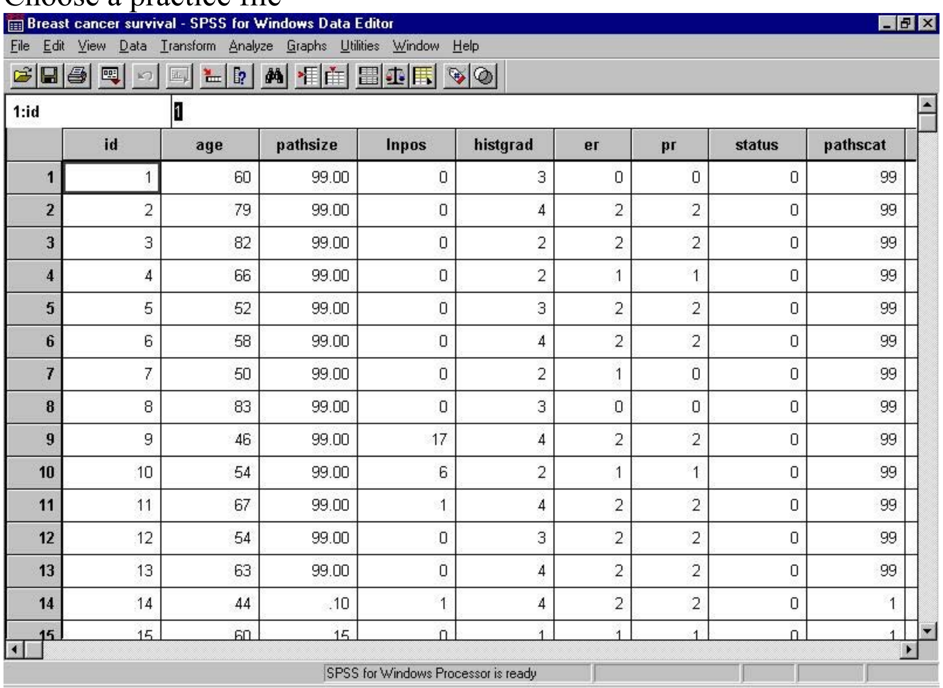
Using Statistics

Once we have gained an understanding of what statistics do and cannot do; and once we have learned how to interpret the numbers we can generate; a next step is to learn to use the many computer packages available to us. There are any number of programs available to the social scientist. The two more comprehensive programs are SAS and SPSS. Just because I was first trained on SPSS I prefer it over SAS. Both are fine programs.

SPSS and the other programs are tools. They can be used correctly and incorrectly to produce meaningful results and junk.

Opening SPSS

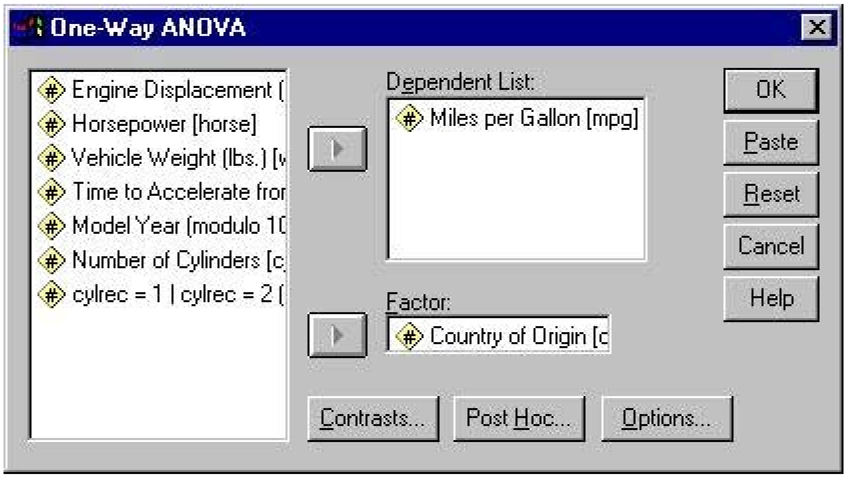
Choose a practice file



	id	age	pathsize	lnpos	histgrad	er	pr	status	pathscat
1	1	60	99.00	0	3	0	0	0	99
2	2	79	99.00	0	4	2	2	0	99
3	3	82	99.00	0	2	2	2	0	99
4	4	66	99.00	0	2	1	1	0	99
5	5	52	99.00	0	3	2	2	0	99
6	6	58	99.00	0	4	2	2	0	99
7	7	50	99.00	0	2	1	0	0	99
8	8	83	99.00	0	3	0	0	0	99
9	9	46	99.00	17	4	2	2	0	99
10	10	54	99.00	6	2	1	1	0	99
11	11	67	99.00	1	4	2	2	0	99
12	12	54	99.00	0	3	2	2	0	99
13	13	63	99.00	0	4	2	2	0	99
14	14	44	.10	1	4	2	2	0	1
15	15	60	.15	0	1	1	1	0	1

SPSS Exercises

Tests for Difference
Test 1
Pull up the "cars" practice data set.



Say we are interested in whether these is a difference among cars in their gas mileage (mpg) and their place of manufacture or nameplate nationality (origin).

First question. Which is the dependent variable, which the independent variable? I would assume that origin would be the independent variable, and that mpg the dependent variable. How might we then pose the question? Cars from different places have different gas mileage (two tail) or cars from place-a have lower mpg than cars from place-b.

How can we test these hypotheses? What is the null hypothesis?

What kind of data do we have? mpg is interval/ratio, ordinal, or nominal?

As the data are the actual miles per gallon data (e.g. 20 miles per gallon), comparisons and ratios can be calculated, we have interval data.

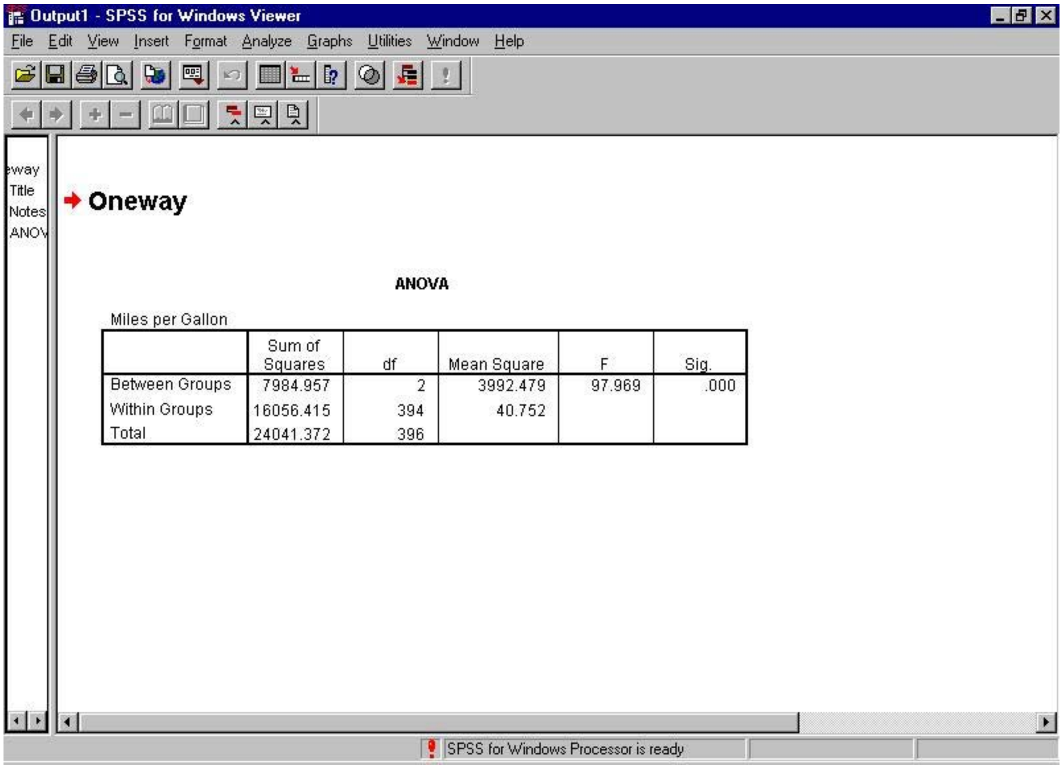
origin is interval/ratio, ordinal, or nominal?

Individual countries and regions (US, Europe, Japan) are listed. These are not continuous nor are they rank ordered. They are nominal.

How many categories do we have? What tests are available to us where the independent variable is nominal and the dependent variable is interval. There are two (t-test, ANOVA). Which allows more than two categories? ANOVA.

We set it up this way: Analyze, Compare Means, One-Way ANOVA

and the results are:



ANOVA

Miles per Gallon

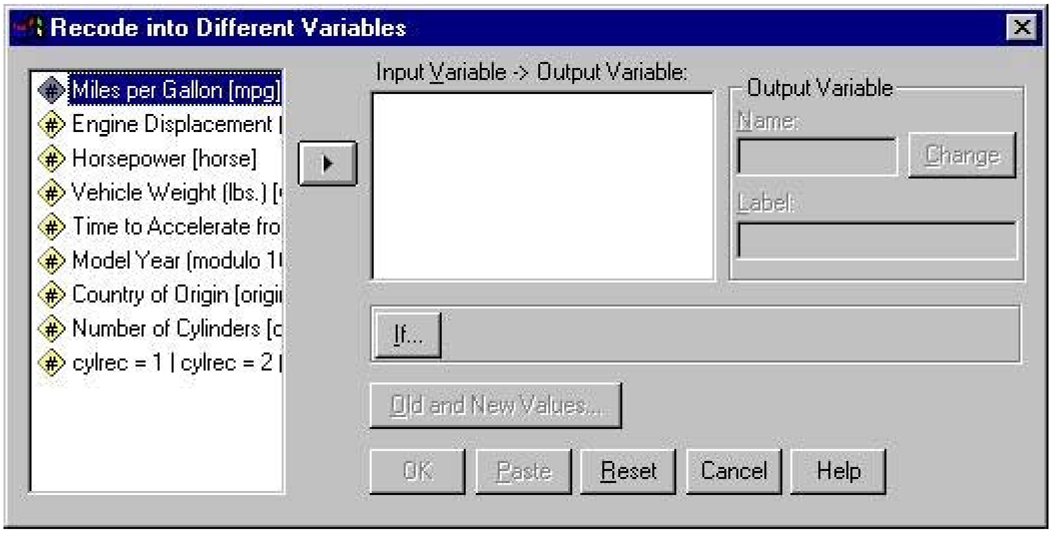
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7984.957	2	3992.479	97.969	.000
Within Groups	16056.415	394	40.752		
Total	24041.372	396			

Test 2

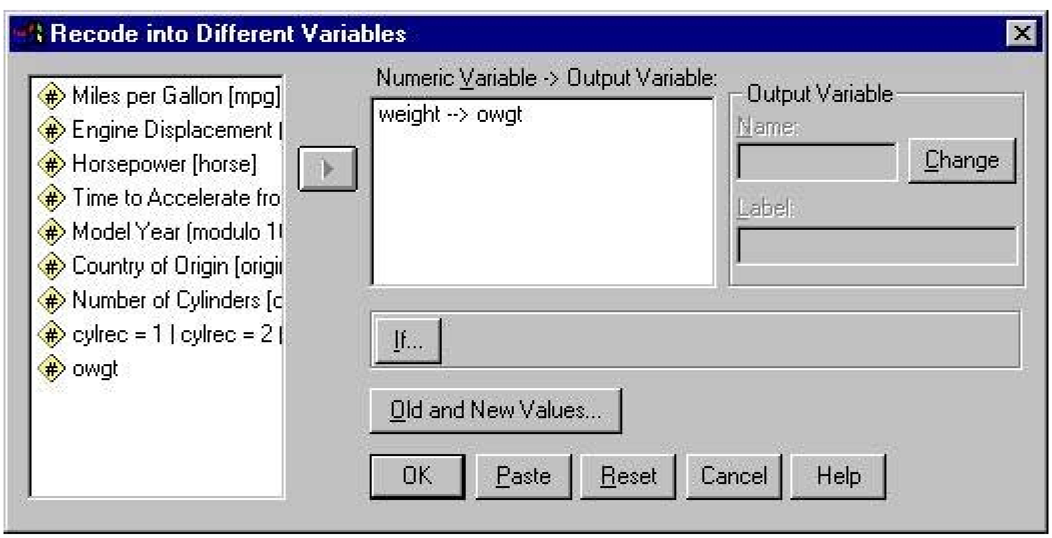
Say, we believe that cars from different places have different weights. Say further, we want to differentiate between heavy, medium, and light weight cars. Who is to say what's heavy, medium, and light weight. Perhaps the industry, perhaps you. For the sake of argument, (and off the top of my head) heavy cars are defined as those weighing 4000 lbs or more, medium between 2001 and 3999 lbs, and light 2000 or fewer lbs.

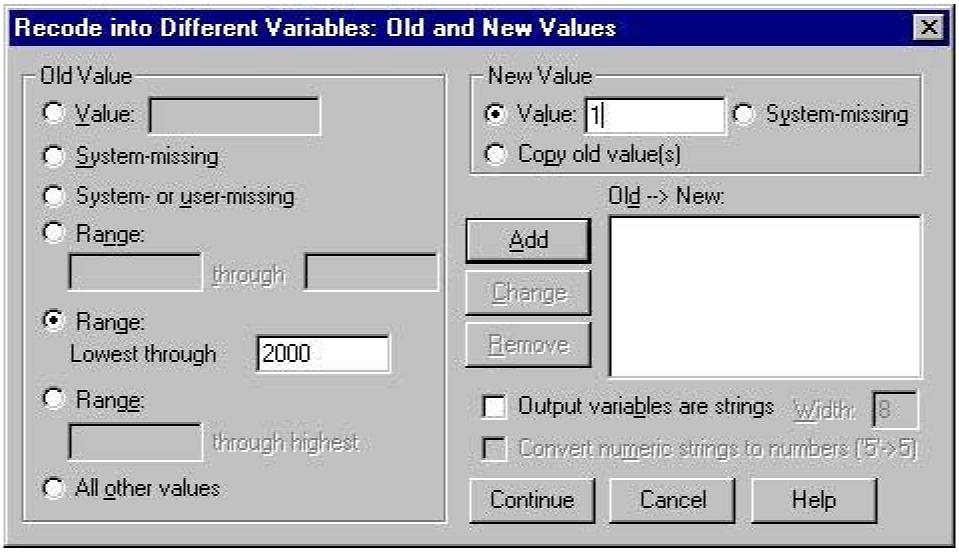
We have no such variable in our data set, but we do have the actual weight of the cars. We can make the new variable using the "Transform, Recode, New Variable" feature. We will create a new variable "Owgt" with three continuous, ordinal, categories.

Start by acquiring the following box (Transform, Recode, New Variable)

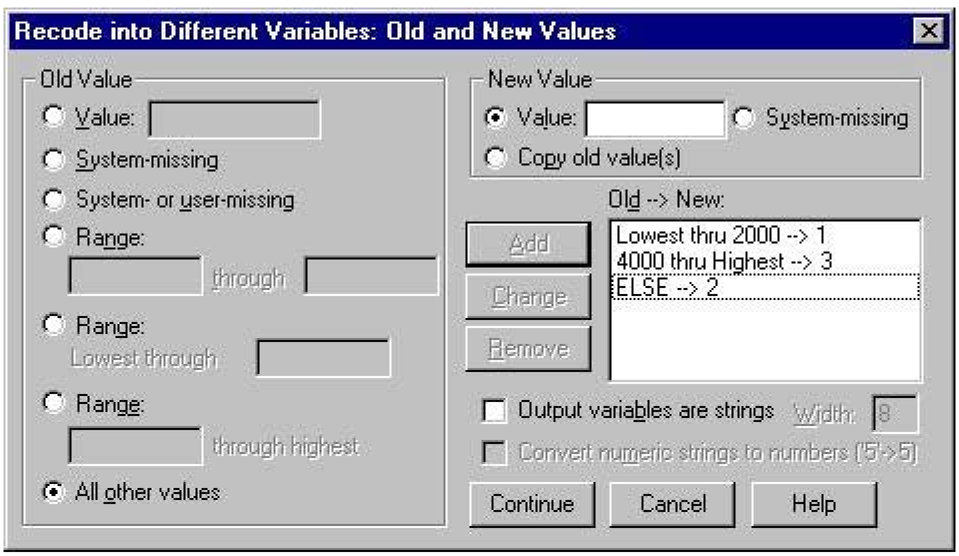


Select Vehicle weight and move it in as the Input Variable
Type the new variable name in the Output Variable Box: then click "Old and New Values..." Since we are dealing with a range of values, we will, step by step, stipulate those values. To achieve the low weight range, we select lowest through 2000, and arbitrarily give it the value 1, the lowest of the values we will assign. Then click the Add button.

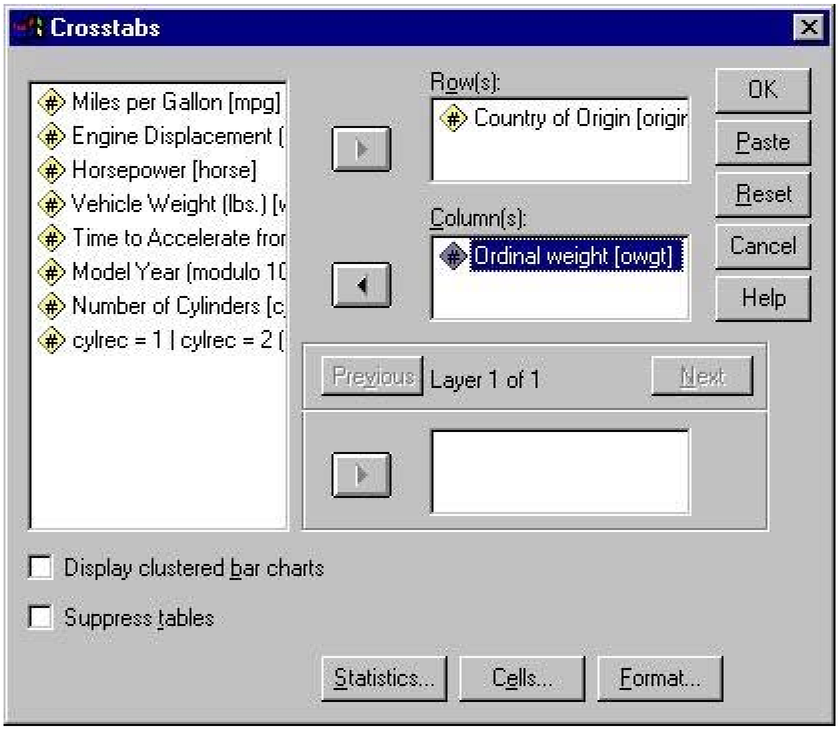




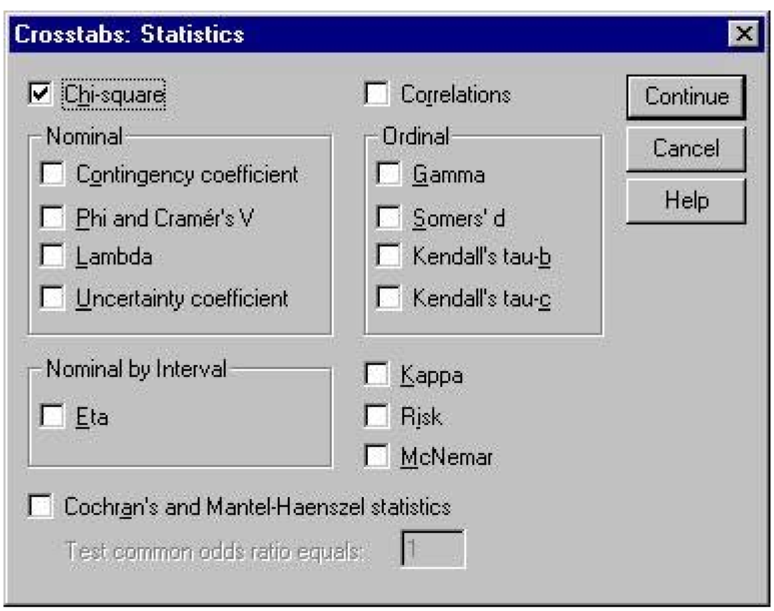
Continue with the other two values:



Click Continue. The screen will change, click Change, OK.
You have just created a new variable according to your established criteria. Now we want to test these two variables for difference -- place of origin by ordinal weight. We have but one choice- chi-square.
To do a chi-square, we select Analyze, Descriptive Statistics, Crosstabs .
We place one variable in rows, the other in columns.

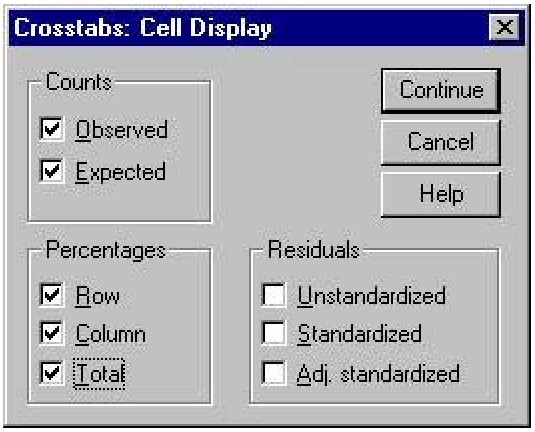


We will select the chi-square statistics after clicking on the Statistics box:

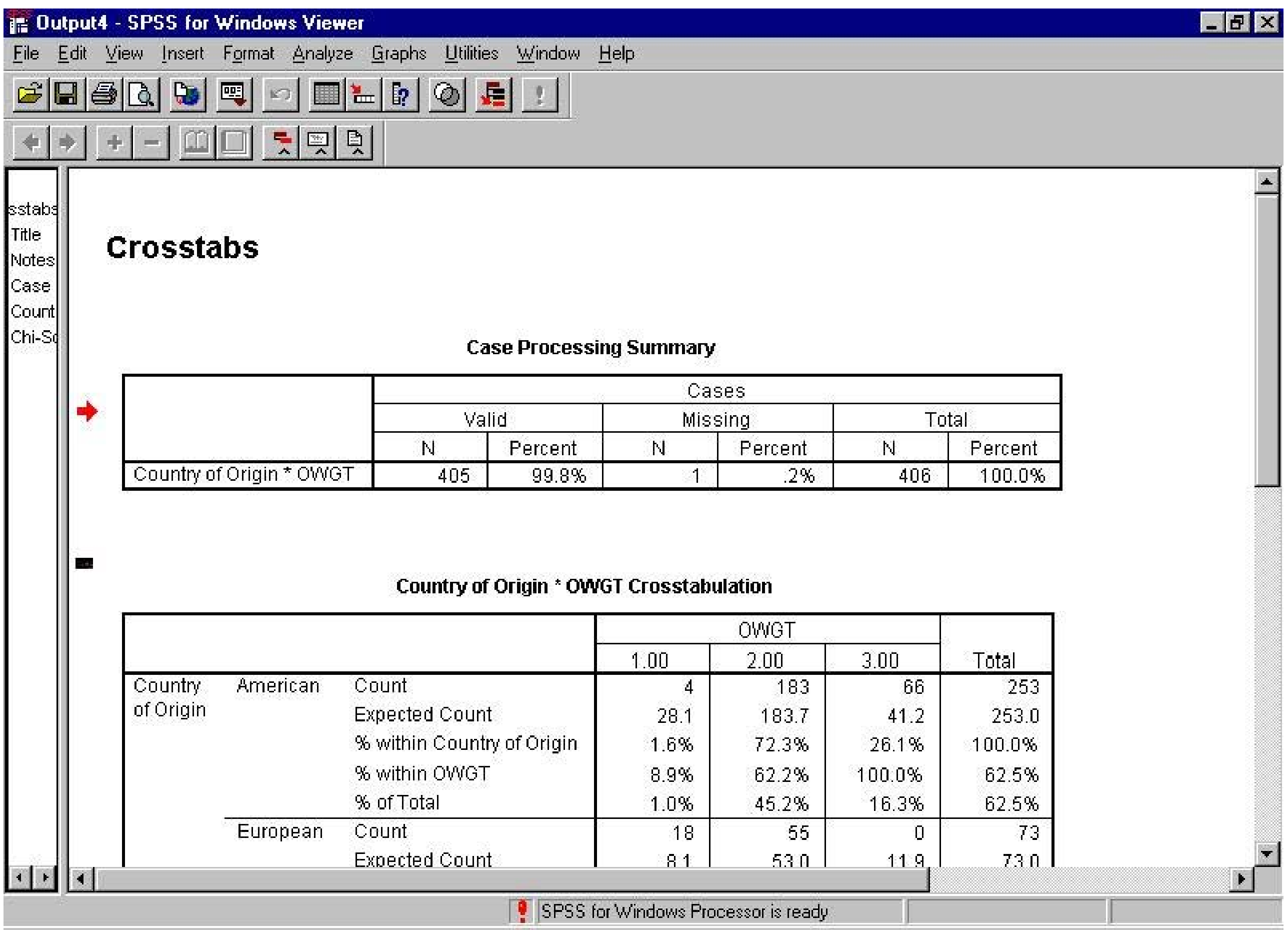


Then press Continue.

Click Cells
You can control the print out by ticking the boxes of your choice. The following will be very busy. Experiment with the output format you find easiest and most meaningful for interpretation.



The above results in three "boxes"

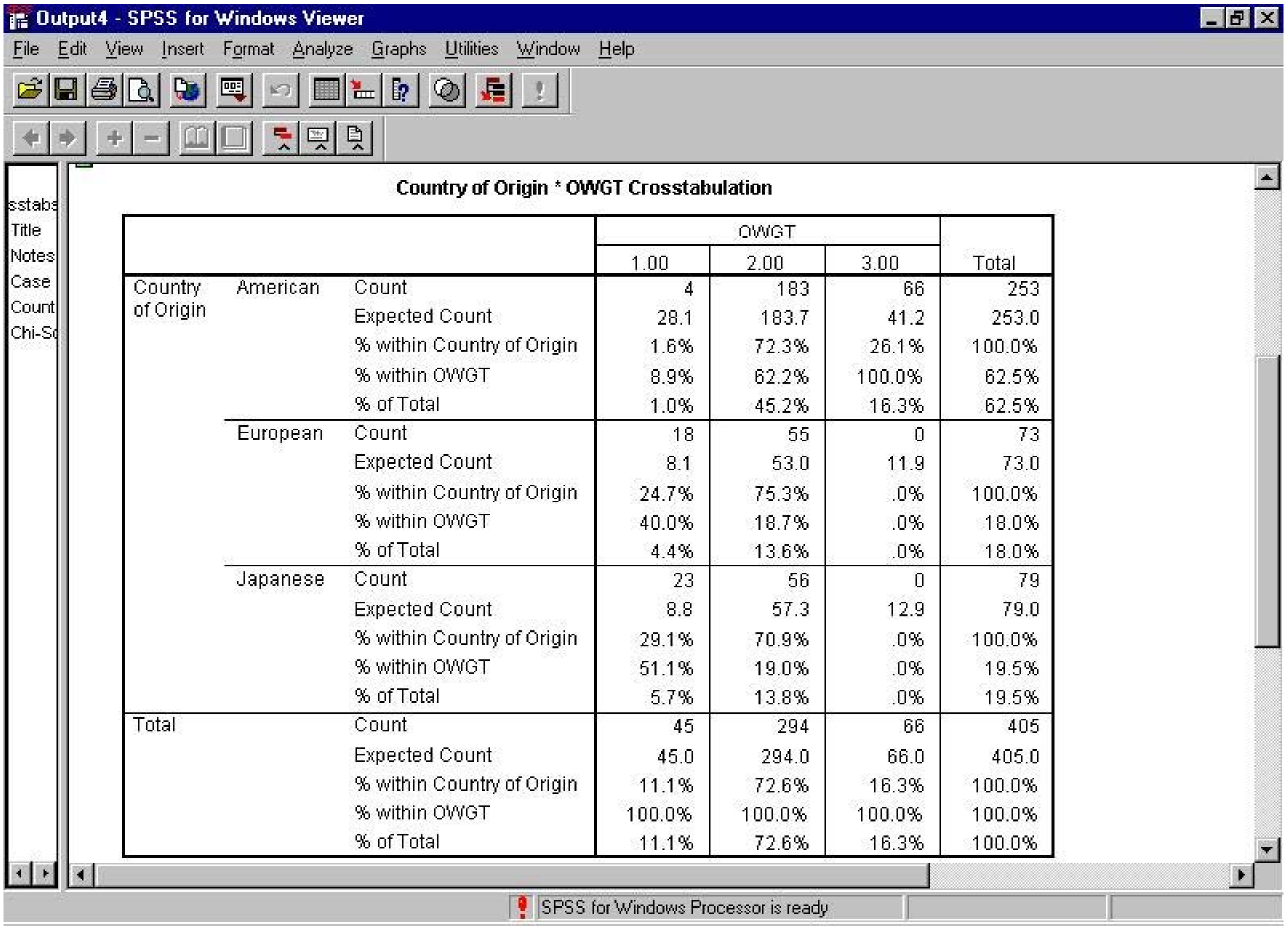


Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Country of Origin * OWGT	405	99.8%	1	.2%	406	100.0%

Country of Origin * OWGT Crosstabulation

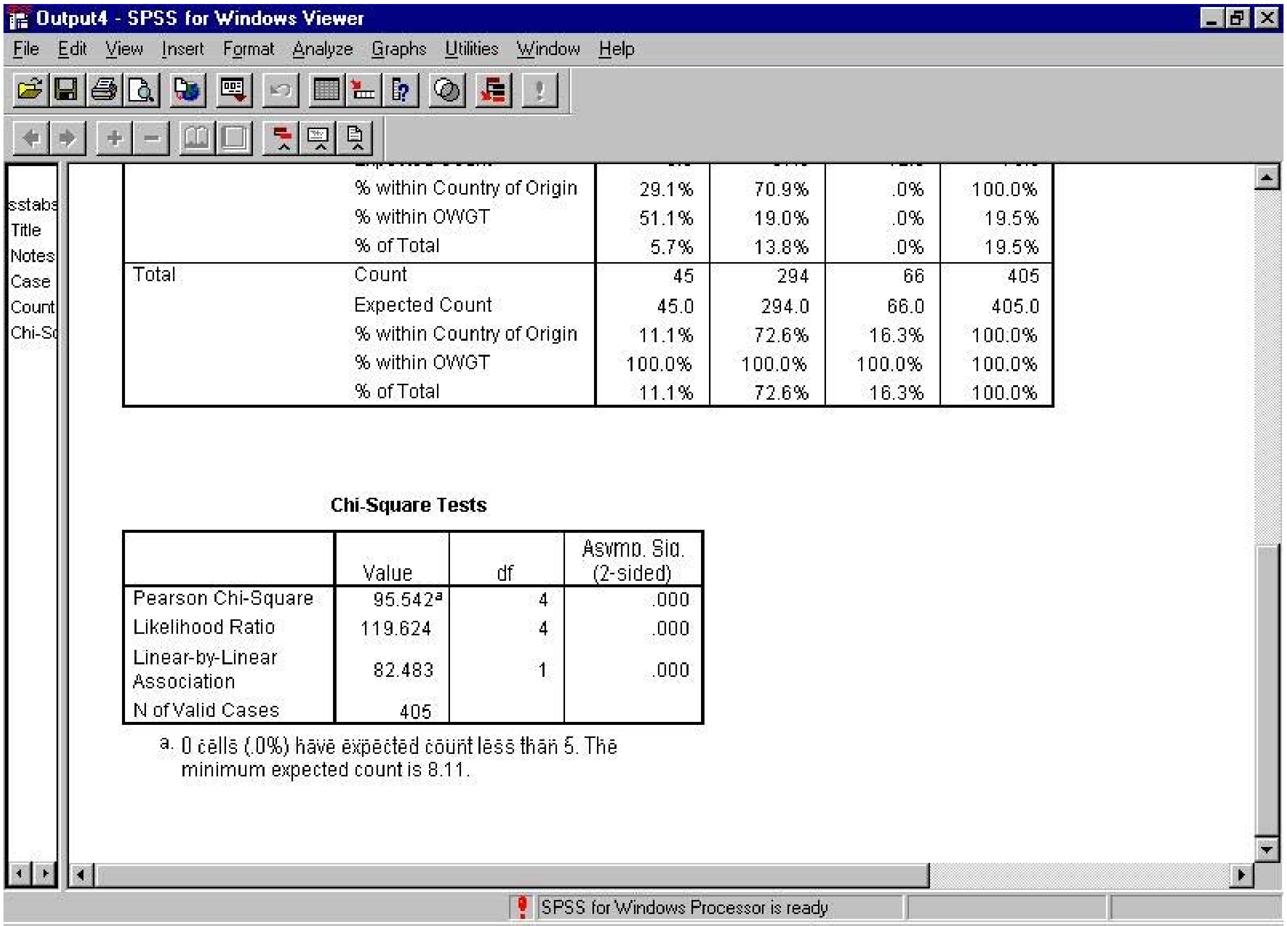
			OWGT			
			1.00	2.00	3.00	Total
Country of Origin	American	Count	4	183	66	253
		Expected Count	28.1	183.7	41.2	253.0
		% within Country of Origin	1.6%	72.3%	26.1%	100.0%
		% within OWGT	8.9%	62.2%	100.0%	62.5%
		% of Total	1.0%	45.2%	16.3%	62.5%
	European	Count	18	55	0	73
		Expected Count	8.1	53.0	11.9	73.0



Country of Origin * OWGT Crosstabulation

			OWGT			
			1.00	2.00	3.00	Total
Country of Origin	American	Count	4	183	66	253
		Expected Count	28.1	183.7	41.2	253.0
		% within Country of Origin	1.6%	72.3%	26.1%	100.0%
		% within OWGT	8.9%	62.2%	100.0%	62.5%
		% of Total	1.0%	45.2%	16.3%	62.5%
	European	Count	18	55	0	73
		Expected Count	8.1	53.0	11.9	73.0
		% within Country of Origin	24.7%	75.3%	.0%	100.0%
		% within OWGT	40.0%	18.7%	.0%	18.0%
		% of Total	4.4%	13.6%	.0%	18.0%
	Japanese	Count	23	56	0	79
		Expected Count	8.8	57.3	12.9	79.0
		% within Country of Origin	29.1%	70.9%	.0%	100.0%
		% within OWGT	51.1%	19.0%	.0%	19.5%
		% of Total	5.7%	13.8%	.0%	19.5%
Total		Count	45	294	66	405
		Expected Count	45.0	294.0	66.0	405.0
		% within Country of Origin	11.1%	72.6%	16.3%	100.0%
		% within OWGT	100.0%	100.0%	100.0%	100.0%
		% of Total	11.1%	72.6%	16.3%	100.0%

and



	% within Country of Origin	29.1%	70.9%	.0%	100.0%
	% within OWGT	51.1%	19.0%	.0%	19.5%
	% of Total	5.7%	13.8%	.0%	19.5%
Total	Count	45	294	66	405
	Expected Count	45.0	294.0	66.0	405.0
	% within Country of Origin	11.1%	72.6%	16.3%	100.0%
	% within OWGT	100.0%	100.0%	100.0%	100.0%
	% of Total	11.1%	72.6%	16.3%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	95.542[a]	4	.000
Likelihood Ratio	119.624	4	.000
Linear-by-Linear Association	82.483	1	.000
N of Valid Cases	405		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.11.

And what might we conclude?
American cars are heavier than European and Japanese cars perhaps?
Do you suppose that this weight differential might account for our findings in the first test - that mpg and country of origin are related?

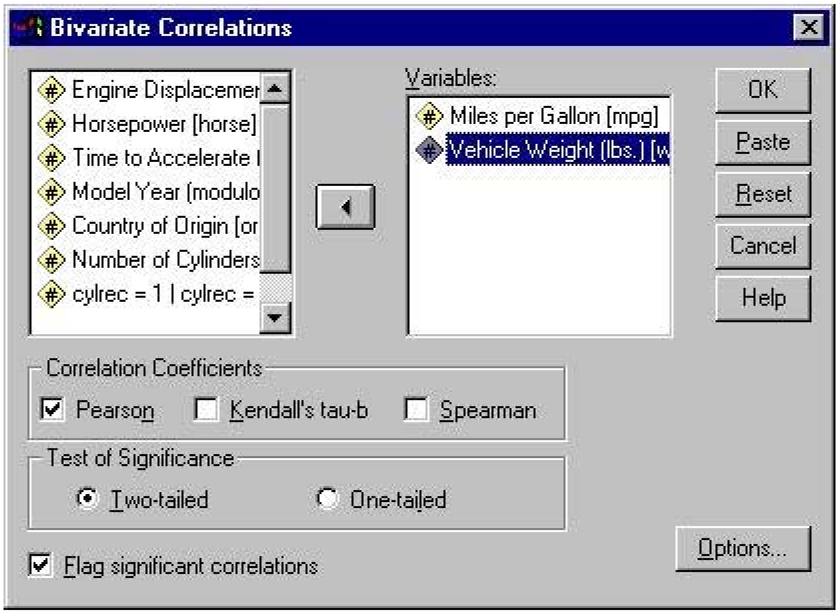
Tests of Correlation or Covariance
Let us assume we want to test the proposition in the Cars database that as weight increases, mpg decreases. Again, what kind of data to we have? Both variables are interval/ratio, so which test of correlation will we select?

We have a number of choices. The one test available to us designed for two interval variables is Pearson's r.

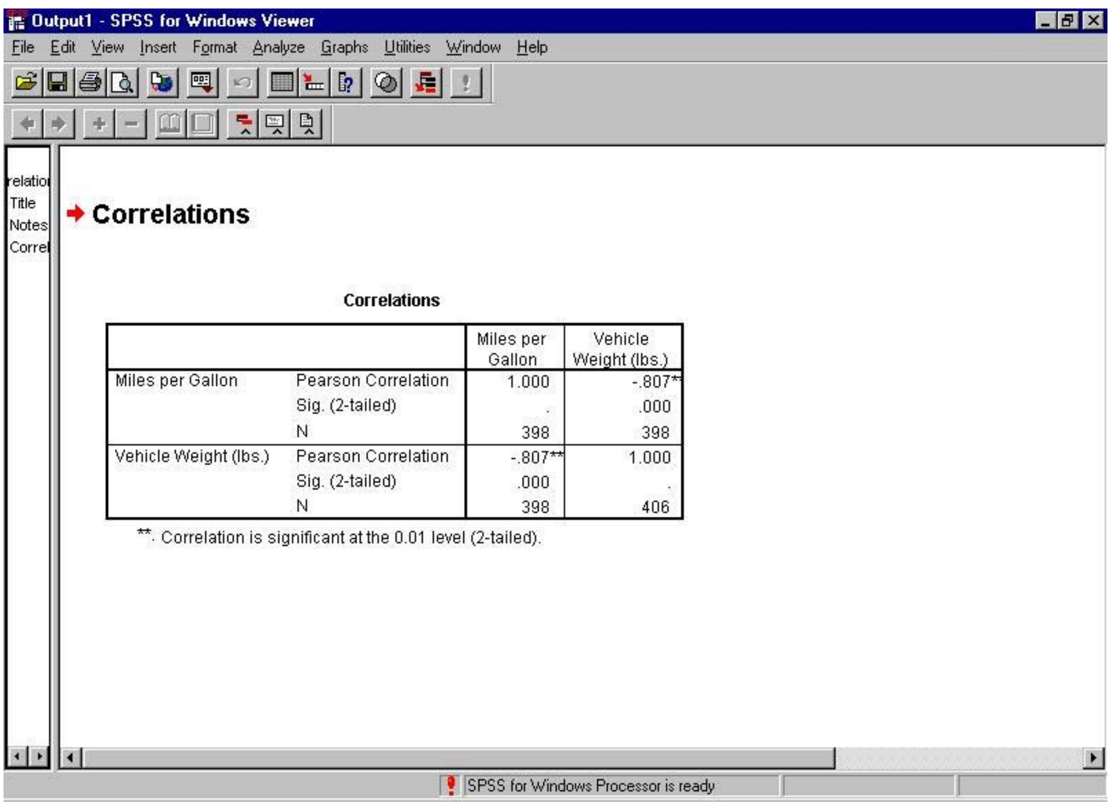
So, open SPSS, open the Cars database.

Select *Analyze, Correlate, Bivariate*

Move the two appropriate variables over and make sure the Pearson box is ticked (the default). You can add additional variables if you wish. The program will generate a matrix of correlations.



and the following results:

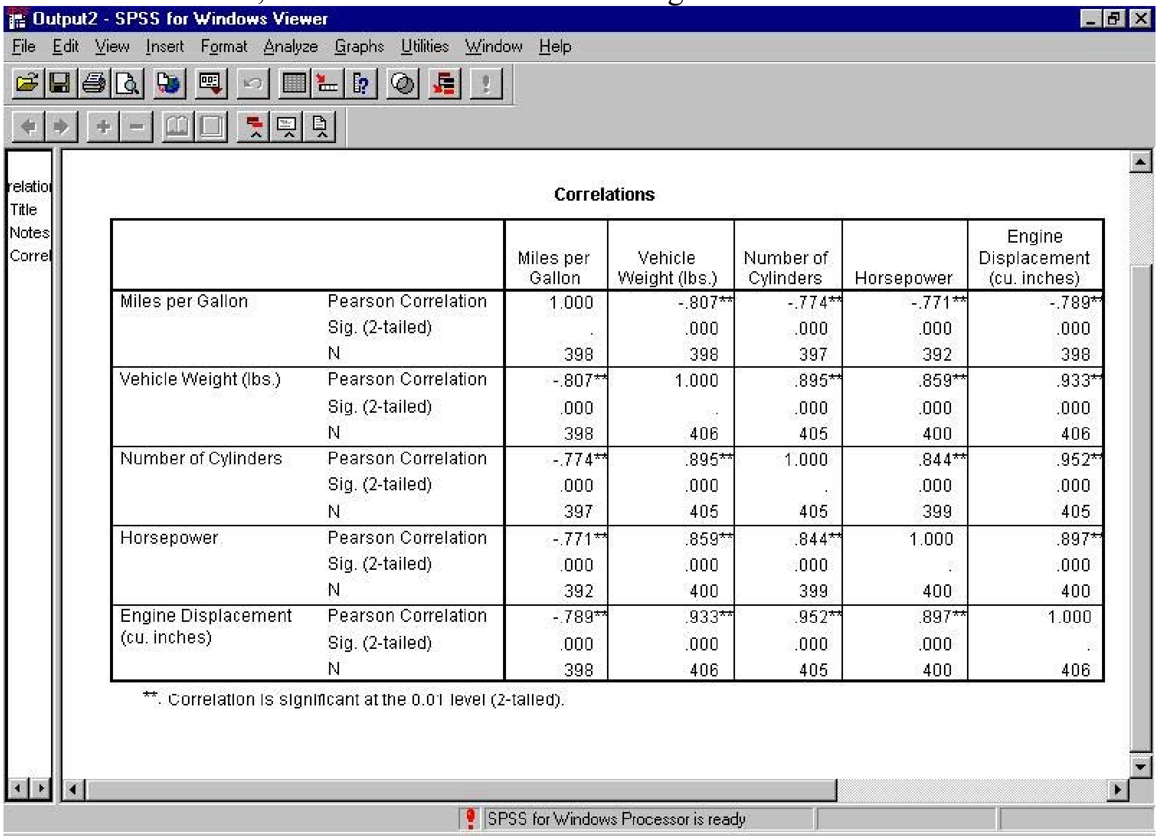


Correlations

		Miles per Gallon	Vehicle Weight (lbs.)
Miles per Gallon	Pearson Correlation	1.000	-.807**
	Sig. (2-tailed)	.	.000
	N	398	398
Vehicle Weight (lbs.)	Pearson Correlation	-.807**	1.000
	Sig. (2-tailed)	.000	.
	N	398	406

**. Correlation is significant at the 0.01 level (2-tailed).

Just for the fun of it, more variables and the resulting matrix:



Correlations

		Miles per Gallon	Vehicle Weight (lbs.)	Number of Cylinders	Horsepower	Engine Displacement (cu. inches)
Miles per Gallon	Pearson Correlation	1.000	-.807**	-.774**	-.771**	-.789**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	N	398	398	397	392	398
Vehicle Weight (lbs.)	Pearson Correlation	-.807**	1.000	.895**	.859**	.933**
	Sig. (2-tailed)	.000	.	.000	.000	.000
	N	398	406	405	400	406
Number of Cylinders	Pearson Correlation	-.774**	.895**	1.000	.844**	.952**
	Sig. (2-tailed)	.000	.000	.	.000	.000
	N	397	405	405	399	405
Horsepower	Pearson Correlation	-.771**	.859**	.844**	1.000	.897**
	Sig. (2-tailed)	.000	.000	.000	.	.000
	N	392	400	399	400	400
Engine Displacement (cu. inches)	Pearson Correlation	-.789**	.933**	.952**	.897**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.
	N	398	406	405	400	406

**. Correlation is significant at the 0.01 level (2-tailed).

As you can see in the matrix, there are a number of positive and negative correlations, all of which are statistically significant.

One way to get a handle on these is to talk through what each means. Take the correlation for miles per gallon and engine displacement for example. As the engine size (displacement) in cubic inhes increases, the gas mileage expected from that engine decreases. Or there is a positive relationship between horsepower and engine displacement -- the bigger the engine, the more powerful it is.

Experimentation or Observation

Some research is conducted by setting up replicatable experiments. Others are best done through observation. Much of what the librarian does is "observation," but some research is "experimental". For example, you might want to know if you can increase library use with open or closed stacks. As difficult as it might be, you might try one system, then the other to determine which increases (decreases) use. There are numerous different approaches to research, some based on philosophy or theory, others on mundane matters as data type.

Different Strokes/Different Research Methods

This section is concerned with two different sets of concepts.

1. First, do different disciplines do research differently. By this, I mean, are there conditions inherent in any given discipline that would predispose it to approach its questions in fundamentally different ways. And, if so, if a research method is legitimate for one discipline is it equally legitimate for another?

2. Different disciplines call upon different resources to support their research. There may be significant overlap. It has been said that in political science the "givens" of economics are variables; but economics it is the other way around. What does that mean to you?

Heuristics

Heuristics might be defined as applying "common knowledge," "rules of thumb," or intuition to the solution of problems. Take a look at http://cogsci.ucsd.edu/~batali/108b/lectures/heuristic.html for some examples of heuristics in action.
For an application of heuristics, see Jakob Nielsen, How to Conduct a Heuristic Evaluation at: http://www.useit.com/papers/heuristic/heuristic_evaluation.html

Experimentation

One methodological approach used in the physical, life, and social sciences is experimentation. Reproducible results are the sine qua non of experimentation. Within acceptable parameters, experiment done by one researcher in one place at one time can be reproduced by the same or another researcher in another place at another time, all other things being equal. Isaac Newton's apple experiment for example works everywhere: Drop an apple and it falls to the ground.
Wait a minute, we can postulate conditions where the apple would in fact not fall to the ground.

What acts on the apple to "cause" it to fall? Gravity. If one were to do the apple experiment in a zero gravity environment, the apple most likely would not fall to the ground. The gravity part is the "all other things being equal" part of the experimental design.

Feminist Research Methods

From the librarian's perspective, do Feminist Research Methods differ from other disciplines, or is it a matter of resources? http://libweb.uoregon.edu/subjguid/women/womenst.html

Historiography

Historiography is the way "historians 'do' history." For an explanation of different approaches to historical explanation, see http://www.pvhs.chico.k12.ca.us/~bsilva/ib/histo.html
The science of historical research is the demonstration that explanations of historical events demonstrate a common thread - and in that sense historical research might be considered a set of case studies from which we induce theory. Historiography is also concerned with the selection of appropriate sources and an understanding of the historical and social context that documents were created in.

It is said that history is written by the victorious. There is an interesting plaque found on a wall in Frederickton, New Brunswick in Canada extolling the virtues of one of the early city fathers, one Benedict Arnold. Arnold is viewed in a very different light with Canadian eyes than with American eyes. Why?

Legal Research Methods

Legal research is an exercise in Hermeneutics. Much of legal research is concerned with document search and interpretation. It is concerned not with the discovery of "scientific fact" but rather with correspondence with human made norms and behaviors. Furthermore, while it is true that "scientific knowledge" can be revised as new understandings are learned, the law can in fact be changed, reversed, overturned by human beings in their courts and their legislatures. Legal research is therefore a practice searching for legal rather than scientific "truths."

In a common law context we must be concerned with case law research as well as statute and administrative law and precedent. We must be concerned with legislative intent.

Second, what is good law in one jurisdiction may not be good law in another. This is particularly true of case law in the American context.

How legal research has changed. See Daryl Teshima, Keeping Current in Electronic Legal Research Methods. http://www.lacba.org/lalawyer/tech/legalresearch.html

Visual Research Methods

Study using images. Marcus Banks argues that the visual arts, like anthropology, uses images as a research resource. (See http://www.soc.surrey.ac.uk/sru/SRU11/SRU11.html):
Thus visual anthropology and visual sociology proceed methodologically by making visual representations (studying society by producing images), by examining pre-existing visual representations (studying images for information about society), and by collaborating with social actors in the production of visual representations.

Banks refers to proxemics ,choreometrics and kinesics, and conversation analysis. What are these methodologies. What implications for librarians here?

On Measurement

There are measurement and concept issues. We are concerned with developing measures for research projects that meet four criteria. These are:

1. Reliability – is it repeatable

2. Validity – does it measure that which we seek to measure
3. Exhaustiveness – captures the concept completely
4. Mutual exclusivity – each observation meets or captures only one criteria/variable in the research. Thus no ambiguity.

On Language and Measurement

Cross-linguistic research can cause all kinds of problems. In fact, cross-cultural research using the same language can be equally challenging.

Consider the following phrases: “I am hungry!” <<j’ai faim!>> “¡tengo hambre!” „Ich habe Hunger!” Aside from the fact that there are punctuation and notation differences, what else is different? In English, “I am hungry” implies a state of being while in the three other languages, it is a condition. The French, German, and Spanish phrases translate as “I have hunger”.

In English we might ask: “Give me a cup of coffee.” In French, it is <<Donnez-moi du café>> or “give me of the coffee.” In French, the idea is that one is drawing from an abstract “pool” of “all coffee” whereas in English we focus on the specific object of our interest – some finite number of litres of a brown hot liquid.

In English, nouns have no gender. In French they may be masculine or feminine. In German and Spanish they may be masculine or feminine but they may also be neuter. In all four languages, there are feminine, masculine, and neuter pronouns (she, he, it/hers, his, its), but in Finnish and Chinese pronouns have no gender.

Citation

Research projects carry citations. Why? There are many reasons. As we saw above, citation analysis is an area of research librarians and information scientists engage in. Librarians may be asked to perform bibliometric analysis to support a faculty member's request for promotion or tenure or a university, company, or country's claim of preeminence. See a somewhat humorous (I hope) paper I published taking a look at one problem with bibliometric analysis (Koehler 2002). What are citations for? There have been a number of studies that point to a variety of reasons writers have for citation. Remember as you evaluate citations to consider the source. If yesterday I cited Karl Marx and V.I. Lenin and today switched those citations to Groucho and John, might you reach somewhat different conclusions on the value of my work? Could it depend in part on the subject of my work? If I were doing a study of popular culture in twentieth century western society, would one set of citations be better than the other; what about nineteenth and twentieth century socialist thought?

We suggest here that there are at least seven citation types:

Affirmation and Explanation. Affirmation is used to provide authority for any given statement of fact or theory. It gives credence to specific issues of fact or theory. It differs in that from situation and validation in that these are concerned with the general placement of the work and not its specifics. "It says so in the encyclopedia, therefore it is so." is a statement of affirmation and of explanation. We often resort to citation to authority, be it an encyclopedia, textbook, holy book, or other authoritative resource to affirm or refute. Consider that until Galileo, astronomers and astrologers held that the earth was the center of the universe. How authoritative are they? The Flat Earth Society argues today that the earth is not a sphere, but a flat plane. How authoritative are they?

Refutation. Refutation is the flip side of affirmation. It is commonly found in legal arguments where it is often necessary to demonstrate a full command of the legal argument. Thus, one may find legal arguments bolstered by citations to affirmation, but also with citations to argument to the opposite.

Demonstration, Review, and Furtherance. Citations may be used to explain the current state of the art, to outline what is understood. They are used to set the stage against which the current work is placed. Thus, these citations may point to the incompleteness or inadequacy of existing theory or our grasp of fact.

Justification. Citations for the purpose of justification give authority to the purpose of the knowledge product. They explain, as does the previous reason, the place the research takes within a given school of thought. Rather than placing the provenance of the work, they argue the incompleteness of what has come before, the contribution of the work presented to our state of understanding. In effect, they document the inadequacy of earlier argument. This argument can take several forms: (1) Changed circumstances; (2) New circumstances; (3) New or reinterpreted evidence or data; (4) New or changed theory; (5) Superior analytical tools; (6) Changed interests or viewpoints. Historians speak of revisionist and traditional history. Traditionalists agree with their antecedents, revisionists find other explanations for events.

Historians also address different issues. It has been said that history is written by the victors. Much of what we consider "history" are accounts of wars and elites. Historians and now Herstorians take different perspectives. These may include among others the impact of technology on society (see The Social History of the Bicycle), the importance of ideas or individuals, or the contribution of non-elites and women on society and local economies (see The Midwife's Tale). Henry Kissinger argued in his PhD dissertation that Clauswitz was a pivotal figure in European history. This single individual remade the face of a continent. The counter argument might be, if not Clauswitz, someone else. Or, it was someone else.

Similarly, Doron Swade (The Cogwheel Brain: Charles Babbage and the Quest to Build the First Computer) argues that Charles Babbage's ideas had little or no impact on the modern development of computers. John von Neumann, he states, found the key concepts described by Babbage *de novo*. Well, maybe. Should we then ignore Babbage and credit von Neumann alone. What of Lady Lovelace?

Data Source. Frequently the data we analyze are taken from some other source. We must acknowledge the data provider and we must also validate the source of that data. Conclusions derived from data taken from "more" credible sources are likely to be accepted than conclusions from less credible sources. This may occur in a least two ways. First, it may reflect on the quality of the resource studied. If, for example, we were to seek to document the evolution of information science through the pages of a single journal, it would be more credible to analyze *JASIST* rather than, say, the American Political Science Review. Or, if I were to try to perform an analysis of international economic trends, it might be more useful to take data from the *UN Statistical Yearbook* than from *Mad Magazine*.

Salutation. Salutation has a recognition function. It serves to recognize major antecedents to the theory or to the discipline in general. For example, Soviet writers were often "required" to recognize the works of Marx and Lenin in all their works. We also tend to recognize major contributors, thus it is not uncommon for information scientists to cite Vannevar Bush at least once in their major works. Salutation is honorific and may not be needed in order to establish the validation or situation of the work.

Situation and Validation. By "situation" I mean citations place a given work or argument within a given theory or belief set. Validation is similar. It provides a given work with its provenance.

In addition, the Web may have opened up a new set of citation types. These are not strictly speaking "new." Before the WWW, they were far too obscure to find and document on a systematic basis. These might include:

Pedagogical citations. We now have much broader access to the syllabi of others. Is the inclusion of the works of others more or less important in syllabi than in research articles. Research articles may cite these works for the above reasons. Syllabi provide them in large part to train future researchers and practitioners.
Resource lists. We have seen a proliferation of on-line "libraries." These are of the "my favorite links" and "important resources" ilk. While bibliographies are prepared in traditional formats, they have become commonplace on the WWW. Is there a qualitative difference between "traditional" citations and these?

Most works contain a mixture of each of the citation types. The proportion of citation types will vary with the work at hand. By and large, most citation analyses do not take the citation type into account. For example, I (and others) was once co-cited with King Solomon. Does that mean that my work rises to Biblical standards? Well, hardly - at least not in her eyes. The author -- Bella Haas Weinberg -- was using Solomon as a framework within which to place subsequent work - mine included. Her point is that there is very little new under the sun, that what some analysts were seeking as new phenomena were not so new after all. She most certainly was not claiming my work as Solomonesque. In either case, however, the co-citation analyst would probably treat both types of citations as the same.

Why don't we use greater sensitivity in these co-citation analyses? First, it can be a very subjective undertaking. Take for example a scholarly article you might chose to critique.Analyze, reflect on the purpose of each of the citations. Is it always clear why the citation was made?

Second, many studies have their basis in the iSi Science, Social Science, and Humanities and Arts Citation Indexes. It is not the purpose of SCI or SSCI to distinguish among different citation types. These indexes capture citation data without qualitative evaluations. There is therefore an inherent limitation build into any citation analysis using iSi data. These data are further limited because until the recent advent of the Web of Science it was only citations to first authors that could easily be explored. Second and subsequent authors could only be studied through much more laborious processes.

As you have already seen, the level of analysis of bibliometric research can be an individual. The individual is often generalized to departments, disciplines, institutions, and countries. That is to say, the same kind of analysis you perform in the bibliometric exercise can be brought to bear on, let us say, VSU, Georgia, universities, the United States. Often, comparative studies are performed. Which researcher, department, university, or country is more productive? These issues as frequently studies as science/technology indicators and may make up an important component in establishing science-technology policy or in the location of scarce economic and research resources.

Sound like economics and politics? It is. Let us also remember that these approaches are neither perfect nor infallible. Take for example, the White and McCain article. While they very carefully circumscribe their sphere of analysis, one of the "faults" of their methodology is that it does not capture the impact, the influence of all of the known information science leaders in the universe. Mortimer Taube and Jesse Sherra are noticeably missing from their list. Why? Because authors writing within the proscribed period of study in the journals they analyze did not cite Taube or Sherra often enough so that they were included in the White and McCain map. Does this mean that Taube, Sherra, or the many other thousands of information scientists who were not cited at a level sufficient to be included in the map are unimportant? Or does it mean something else?
The level of analysis in bibliometric research can also be "knowledge products." What is meant by" knowledge product," a term I believe was first used by Francis Narin? In the broadest sense of the term, it means any intellectual output. That output can be quite ephemeral. Or it might be "somewhat" permanent. Note, I say "somewhat permanent." Let us recognize that nothing that human beings might produce is ever forever (so far as we know). Some things are more permanent than others. Consider art forms. The Mayan and Egyptian pyramids have been with us for centuries. We have but fragments of the literary record of early Chinese or Greek civilizations. We know almost nothing about the Etruscans.

Ephemera need not be limited to the erosion of time. Consider those arts and practices that are inherently ephemeral. The spoken work, Navaho sand painting, London sidewalk chalk artists, students writing messages on the sidewalks, a musical performance.

Typically, we as librarians have been more concerned with objects with a bit more permanence than sidewalk scribblings. And so it is with bibliometricians. Books, journals, articles, newspapers, magazines, can all be the subject of bibliometric analysis. In addition, bibliometricians can also focus on patents, trademark and copyright registrations, legislation, court decisions, and so on.
For those of you considering law librarianship, case and legislative Shepardization is of extraordinary importance. In a similar vein, the medical professions are concerned with "gold standards" or best practices. Shepardization is the determination whether a given court decision or legislation still represents "good law." In a common law system such as our own, court decisions create precedent. Earlier decisions guide later ones. But earlier decisions can subsequently be reversed or revised by higher or later courts or amended by legislation.

Consider the following two US Supreme Court decisons: Marbry v Madison and Penderghast. Which is still good law? When was each decided? What was decided. And why is one still in force while the other is not? It is profoundly important that you as a law librarian get all such questions right.

This same is true in medicine and other professions. As a medical librarian you might not only be asked to research procedures and treatments, but you might also be called upon to research current (and sometimes former) best practices or the state-of-the-art. What is therapeutic phlebotomy? Is it ever indicated today? How has it been used in the past? Is it appropriate in the treatment of osteoarthritis, manic-depressive disorders, myocardial infarctions, polycythemia? Do you think inaccurate findings on your part might compromise patient care or expose your patrons or yourself to legal liability?

As we saw above, citation analysis is an area of research librarians and information scientists engage in. Librarians may be asked to perform bibliometric analysis to support a faculty member's request for promotion or tenure or a university, company, or country's claim of preeminence. See a somewhat humorous (I hope) paper I published taking a look at one problem with bibliometric analysis (see http://books.valdosta.edu/mlis/IFLA04/citenight.pdf).

Library Research Models

How to do research is one set of issues. How libraries are used to do research is another. See Thomas Mann for some valuable insights. Library Research Models

Thomas Mann, in his important and useful book *Library Research Models* provides us with descriptions of seven different models of library research. His book is not on how to research library issues, but rather how to use the library to do research. It is, I believe, a guide for librarians to assist others and to understand others as they - the others - seek to use the library and librarians to meet their research needs. This book is therefore not written specifically for a research methods course as such, but it is very useful in understanding how librarians and libraries contribute to the research process.

One final note. Mann's *Library Research Models* should be required reading for any library student. As such, it should be an integral part of each "information worker's" library. Consider buying it and keeping it.

Mann's seven models are:
1. The Specific Subject Model
2. The Traditional Library Science Model
3. Type-of-Literature Model
4. Actual-Practice Model
5. Principle of Least Effort
6. Computer Workstation Model
7. Method-of-Searching Model

Consider the description he provides. How do your own practices correspond with what Mann describes. How would adoption of one model or another inhibit or enhance the research experience? Those of you who have had library experience will have encountered examples of each. Be prepared to discuss what you have encountered in the field and what your responses where.

What do you gain from the experiences of others? How "scientific" are anecdotal accounts? Would a focus group approach assist in this process?

References

Hayes, Robert M. "Measurement of Information." In Pertti Vakkari and Blaise Cronin, *Conceptions of Library and Information Science*. London: Taylor Graham, 1992

Koehler, W. “Nightmares in Citation Analysis” Sessio Taurino, *Reference and User Services Quarterly 42*, 1 (Fall 2002): 41-2.

Lester, June and Wallace Koehler *Fundamentals of Information Studies* . NY: Neal-Shulman, 2003.

General Resources

Zalta, Edward N., ed. *Stanford Encyclopedia of Philosophy.* Available: http://plato.stanford.edu/

12 August 2004