1996

Ministatistics for Minitab release 9.1: a simple explanation of how to analyze data using the Minitab program on the RIT VAX computer system

Morton Isaacs

Ministatistics for MINITAB RELEASE 9.1

A Simple Explanation of How To Analyze Data Using the MINITAB©
Program On the RIT VAX Computer System

MORTON ISAACS

PROFESSOR OF PSYCHOLOGY

ROCHESTER INSTITUTE OF TECHNOLOGY

Last Revised August 1996

TABLE OF CONTENTS

- - - - -

WHAT'S IT ALL ABOUT?

THE MEANING OF PROBABILITY

Suppose your friend tells you that she can read minds; you're skeptical about that claim so she says she'll prove it to you -- "Think of a color, either blue or red, mark it down on a paper and I'll know what you wrote". You don't believe her, so you hide the paper from her and write "Blue". She concentrates a moment and then says, "It was 'Blue'." Was she just lucky or was something other than chance at work here?

The best answer is that you don't know yet; after all, she only had two choices -- either right or wrong -- so she'd be correct 1-out-of-2 times (50%) by chance alone. What now? You say to her, "I'll bet you can't do it again" and she says, "Oh, yes I can". You mark down Red, she thinks a moment, and again she gets it right. You're a bit impressed, but after all, you say to yourself, by luck alone she'd get it right twice in a row 1-out-of-4 times (25%) anyway.

"Once more", you say. And again she's correct, and again, and… Wait a minute. How often would she have to be right before you'd say, "O.K., what's the trick? I'm pretty sure now that it's not just luck or chance; something else is involved." Suppose she was correct 20 times in a row? Suppose she was correct 19 times but wrong once; would you think it's pure luck? Correct on 18 and missed on 2? Correct 15, missed 5? Correct 10 and missed 10?

There should be a point somewhere 50% success and 100% success (10 out of 20, or 20 out of 20) where you'd feel that the balance had swung from it being “just chance” to it being “something more than chance”. Notice I said "something more than chance"; even if it's not just luck, she hasn't proved telepathy even if she gets 20 out of 20 correct. Maybe she watched your pencil move, or maybe someone is in back of you sent her a signal, or some other trick was involved. Still, the first thing you have to decide is whether the results are likely just to be due to chance alone or not. Only if results are not likely to be due to chance alone should other explanations be offered for any scientific psychological findings!

EVERYTHING IN THIS BOOKLET IS SET UP TO TRY TO HELP YOU DECIDE WHETHER YOUR RESULTS WERE LIKELY TO BE DUE TO CHANCE ALONE OR TO SOMETHING ELSE!

\- - - - -

TWO CRUCIAL CONCEPTS -- SIGNIFICANCE & IMPORTANCE

SIGNIFICANCE: Guessing correctly by luck alone on one trial with two choices occurs 50 times out of 100 (50% or 0.500); on two trials 25 times out of 100, or 0.250. When does "luck alone" change into "perhaps something else"? Is it when the odds are 1 in 10 that it would occur by luck alone, or 1 in 20, or 1 in 100, or 1 in 1000? You could use any of these odds (or any other sets of odds), but in most psychological and medical research…

NOT SIGNIFICANT: If a result could happen by chance alone MORE THAN 5 TIMES OUT OF 100 ($p > 0.0500$) -- random factors cannot be ruled out as the cause of any differences found in the data.

SIGNIFICANT: If a result could happen by chance alone 5 TIMES OUT OF 100 OR LESS (p ¾ 0.0500) -- you can be reasonably sure that differences are due to something other than chance.

VERY SIGNIFICANT: If a result could happen by chance alone LESS THAN 1 TIME OUT OF 100 (p ¾ .0100) -- you can be strongly certain that differences are due to something more than chance.

--> NOTE: MINITAB prints "p = 0.0000" when it should print "p < 0001" (which is very significant.) In your report, it's better to report this as $p < 0.0001$

IMPORTANCE: A very common question is: If a result is "significant" does that mean it is "important", and "not significant" mean "unimportant"? The answer is NO! NO! NO!

Significance is a statistical term derived mathematically from the data.

Importance is a judgment you or others make about the value of the results.

For example, if one brand of aspirin over many trials was shown to get into your blood stream in an average of 22 seconds, and a second brand in an average of 36 seconds, this difference might turn out to be very significant (not very likely to be due to chance alone). However, the difference of 14 seconds might not be important to you, particularly if the second brand costs a lot more money than the first, or has bad side effects!

On the other hand, suppose you do research and find no significant difference between males and females in their attitude toward women rights. The lack of attitude difference might be extremely important to you, particularly if you had previously felt that males would be more opposed to female rights.

- SIGNIFICANCE states how likely it is that a difference would have occurred by chance alone.

• IMPORTANCE is the meaning you attribute to the finding of significance or nonsignificance.

- - - - -

TYPE OF SCALE

• NOMINAL data. Some data groupings are only names; before the data is collected, there is no reason to rank one group higher or lower than any other group. This type of data is called Nominal data. There is no ordering or ranking in this data, so the categories can be listed in any order that you wish. If I wanted to know which car is most preferred, before I collect data there is no mathematical reason to list Ford before or after Isuzu. Car names, cigarettes, gender, race, religion, are all Nominal data.

• ORDINAL data. Some information can be put in an order even before the data is collected; we use comparison words like "more than", or "longer than", or "better than". This type of data is called Ordinal data. If I want to know degree of liking for RIT on a scale from 1 to 5, then I know ahead of time that a person who rates his liking as "3" is showing more liking than if he had rated it "2", and less than if he had rated it "4". Any data collected using instructions like "Rank these pictures in terms of beauty", or "Put a 1 next to the RIT major you most prefer, a 2 next to the next most preferred, and so on", creates this kind of data. Terms like "Put in order", "Rank", or "List first, second, and third..." indicates you're dealing with Ordinal data.

• INTERVAL data. Some scales can not only be ordered but, in addition, the interval or distance between any two adjacent values on the scale are equal. This type of data is called Interval data. If I ask "How much would you pay for this dress?", the amount of money listed is Interval data, since $1 is as far from $2 as $2 is from $3, and so on. Time, length, weight, are all Interval data. Interval data is always ordered, so all tests that can be done on Ordinal Data can also be done on Interval.

PARAMETRIC VS. NONPARAMETRIC

In psychological research, often it is not clear whether data is from a distribution that forms a bell-shaped curve and is Interval (Parametric) or doesn't fit that description (Nonparametric). Psychologists usually treat IQ scores or rating scales as if they were Interval data from bell-shaped distributions even though they're really not, so that more powerful Parametric statistical tests can be used on them. Whether this is right or wrong is a matter for argument among statisticians. Most Parametric tests have similar, slightly weaker Nonparametric tests for use on the same data.

At your level, it probably is better to use a nonparametric test if one is available unless you're pretty sure that the data is Parametric. If in real doubt, ask your teacher which to use.

FOUR EASY STEPS TO SUCCESS

(1) Enter your data into a VAX datafile and name it MINITAB.DATA as described below.

(2) Transfer the data from the file MINITAB.DATA into the MINITAB program as described.

(3) Select and run minitab tests on your data to get results for your report as described.

(4) Print results from the file MINITAB.RESULTS in which they've been saved as described

STEP 1: ENTER YOUR DATA INTO A VAX DATAFILE AND NAME IT MINITAB.DATA

(Things that the computer types will be printed in PLAIN CAPITAL LETTERS.

Things that you type into the computer will be shown in BOLD CAPITAL LETTERS.)

• Getting An Account: To use the RIT VAX system you need a User Name and a Password. As of today, RIT/NTID will give them FREE to any registered student. To get one, go to the Ross building (Building #10, back and to the right of the College of Business), to the main office of Information Systems and Computing (ISC, Rm 10-A365) with your RIT/NTID I.D. and ask for a computer account. You get one generally by the next day.

• Finding A Computer Terminal: They're all over! There are terminals in Bldg 7 (Fine Arts), Bldg 10 (Computer Sciences), Bldg 12 (Business), and in many other places around campus. Ask around; any terminal connected to RIT's computer system is O.K. You can also use a modem by telephone to contact RIT's system; ask at the computer center for instructions on how to do it.

• Common Errors Using A Computer: Many persons make these errors using computers:

The lower-case letter "l" (el) and the number "1" (one) are two different things to a computer. The same holds true for the letter "O" (oh) and the number "0" (zero). If you have trouble with your data, check to make sure that you didn't type a letter instead of a number by error.

You must press the RETURN (or ENTER) key after you have entered each line of your data or the computer will wait forever for you to do so. The RETURN key tells the computer you are ready for it to respond. If nothing happens after you've typed something into the computer, you probably forgot to press the RETURN key. I'll remind you by printing <RET> when you must press it.

• Starting: Press the BREAK key or the RETURN Key on the keyboard to begin.

1. Computer types LOCAL> or some similar word.

2. Type CONNECT RITVAX <RET>.

3. Computer types USERNAME: type the Username you were given <RET>.

4. Computer types PASSWORD: type the Password you were given <RET>.

5. Computer types welcome message of some sort, and then displays a Menu. Hold down the key marked "CONTROL" (sometimes abbreviated CTRL) on your keyboard and while holding the key down, press the "Z" key on your keyboard. This is written "Ctrl-Z"; you'll be using it often; please remember how to do it.

6. The menu disappears and a dollar-sign ($) shows, meaning the computer is ready. You can now begin.

- - - - -

AN EASY EXAMPLE TO FOLLOW...

You've collected information on 4 survey questions (let's call them Ques 1, 2, 3, and 4) and wish to test if the average scores on each of those questions of the male subjects differ significantly from those of the females, you also want to know if the subject's year in college made a difference in the scores, and if hearing vs. hearing/impaired made a difference.

First take a sheet of paper and put the variables you're interested in (Ques 1, 2, 3, 4, Gender, College Year, and Hearing level) across the top of the paper; these VARIABLES are your COLUMNS. Put numbers for each of your subjects down the left hand side of the paper: the SUBJECTS are your ROWS. It's helpful to draw lines down between your columns and across between your rows (or use graph paper).

Next enter all the data from your subjects onto the paper; your sheet should be similar to this:

NOTE: If you don't have info for a subject in a column, put a star (*) on the sheet--DO NOT LEAVE ANY BLANK SPACES ON THE SHEET! In the example, subjects 3, 5, and 6 omitted some information in their responses.

Subj.	Ques 1	Ques 2	Ques 3	Ques 4	M/F	Year	HI/H
1	8	7	8	6	M	3	HI
2	5	3	5	3	F	4	H
3	7	8	8	8	F	*	H
4	1	1	1	2	M	3	HI
5	8	3	4	5	M	1	*
6	7	*	3	3	F	2	H

Now substitute numbers for any letters you've written in the columns; the computer only accepts numbers. In our example, you might code the data as follows:

CODING SHEET

Columns 1-4: Actual scores (no code)

Column 5: Female=1, Male=2

Column 6: Each number is year in college

Column 7: Hearing-Impaired=1, Hearing=2

Your coded sheet should look like this:

8	7	8	6	2	3	1
5	3	5	3	1	4	2
7	8	8	8	1	*	2
1	1	1	2	2	3	1
8	3	4	5	2	1	*
7	*	3	3	1	2	2

At this point, you are ready to enter the data into the VAX as a datafile. Type E MINITAB.DATA <RET>. (Leave a space between the "E" and the "MINITAB.DATA"). The computer responds:

%RITPRE-E-FILNOTFND, could not find file "USER7:[xxxxx]MINITAB.DATA;"-RMS-E-FNF, file not found. Enter <RETURN> to Create, New filename, or <CTRL/Z> to Quit New Filename:

Since you want to create a file called MINITAB.DATA, just press <RET> again. There'll be a pause, then the screen clears and the cursor (bouncing square) will appear in front of the words [END OF

FILE].

Enter Subject #1's data by typing the entire first row of numbers, leaving a space between each of the numbers. Press the <RETURN> key and continue with the second row, and so on for all your subjects. When you're finished, check to see that each of the rows and columns are exactly same length. If not, you've skipped a number or forgotten to put a "star" for missing data, or messed up in some other way!

You can correct any errors you make while entering your data as follows:

To MOVE the cursor -- You can move within what you've typed by using the four arrows on the bottom of the keyboard to move up, down, or sideways within the data. The blinking square (cursor) shows where you're at.

To DELETE numbers -- To erase any error, move the cursor past the unwanted number, then press the X WORD/CHAR Key at the top right of the main keyboard (DELETE or BACKSPACE key on some terminals) to erase one character on each press. Delete and then type in the correct information; the cursor will move over.

To ADD numbers -- Move the cursor where you want to add data and just type it in. The rest of the data will move over to make room for what you're adding.

When you have finished entering your data, you must save it so that you can call it into MINITAB. To do this, use Ctrl-Z (hold down the CONTROL Key and WHILE HOLDING IT, press the letter "Z"). The computer types ... LINES WRITTEN TO FILE... and gives you a $ on the screen. That means you've saved the data.

If you want to change or add to your data file at any time, type E MINITAB.DATA <RET>. The file will be placed on your screen just as it was when you last saved it. Make any changes you wish; when you again press Ctrl-Z , the computer will keep and use the latest version (and one backup version) for you.

STEP 2: TRANSFER THE DATA FROM THE FILE MINITAB.DATA INTO THE MINITAB PROGRAM

Now that you have entered the data into the VAX, you must transfer it into the MINITAB program to run your tests. To do this, first count the number of columns you have in your datafile. There are 7 in the example (Ques1, Ques2, Ques3, Ques4, Sex, Year, and Hearing Level). Record the number of data columns. You'll need this later.

NOTE: Only the words in BOLD CAPITALS need to be typed for MINITAB commands. I've put in words in plain lower case print to make the commands clearer.

Type MINITAB <RET>. The computer will pause, type some stuff, and then print MTB>.

Type READ the data from the VAX file 'MINITAB.DATA' into C1-C___ <RET> as many columns as you need. Don't forget to put a single quote before and after the file name. The program will indicate that your data has been read in. If there is an error, check to see if you left any blanks in your datafile, forgot to put single quotes around the file name, etc.

Type OUTFILE 'MINITAB.RESULTS' <RET> Don't forget to put a single quote before and after the file name. This will automatically save all the information printed on your screen to a VAX datafile called MINITAB.RESULTS. You'll be able to view your test results, put labels on them if necessary, and print them out when you're finished. If you fail to do this, your test results will not be saved for you, and you may have to run all your tests over again.

NAMING COLUMNS (VARIABLES) IN MINITAB

Your data will be easier to understand if you name your columns (variables) rather than just leaving them C1 or C2 as Minitab does. To print names on the columns, type NAME C2 'QUES1' <RET> or whatever you want to call it. Names cannot be more than 8 characters long, no spaces, and you must put a single apostrophe on each side of each name. You can name more than one variable on a single line by leaving a space between each one: e.g., NAME C5 'SEX' C6 'COLL-YR' C7 'HEAR-LVL' <RET>

\- - - - -

FIGURES, GRAPHS, AND TABLES

There are two types of information that you work with in research: Descriptive and Analytic data.

• Descriptive Data: For descriptive data, figures, graphs and tables are excellent. For example, if you want to show how many males and females chose red, blue, black, or yellow when asked to name their favorite color, a chart would be an excellent way to display this information. Decide on what you want to show and print it using a graphing/drawing/painting program on a p.c., do it by hand, or use the graphs and tables in Minitab; ask the lab assistant for help. However you do them, make sure that you label both the horizontal and vertical axes of any figure or graph with a descriptive name, and that you put a title on it that tells the reader what you're describing. Always refer to the figure, graph or table in the RESULTS section of your paper and briefly mention there what it is intended to show.

----->Note: It is NOT useful to make a figure, graph, or table of data using each subject's score; instead, use tables, charts or figures to show summary data of groups of subjects, comparing each groups' responses by displaying them on the same chart or table.

• Analytic Data: For data analysis, you MUST use one of the statistical tests described in this booklet (or any other legitimate test) and list as a probability score the likelihood that any difference or relationship you found was due to chance, . A graph or figure or chart does not do this by itself.

DESCRIPTIVE TABLES OR GRAPHS OR PLOTS ALONE ARE NOT SUFFICIENT FOR A GOOD REPORT. YOU MUST ALSO TEST THE SIGNIFICANCE OF ANY DIFFERENCES OR RELATIONSHIPS YOU HAVE FOUND OR YOU HAVE NOT TRULY COMPLETED YOUR RESEARCH PROJECT!!!

\- - - - -

You are now be ready to run statistical tests on your data

STEP 3: SELECT AND RUN MINITAB TESTS ON YOUR DATA TO GET RESULTS

Most often, you'll need to use more than one statistical test to understand your data. You may use the same test on different parts of your data, or different tests on the same data to get different information. Decide what you want to learn from your data, and mark the questions on paper. Then look over the examples given below for a test description that seems appropriate to answer one of your questions and mark its name down next to the question. When you've identified which tests to use to answer all your questions, run the tests as described. If none of the tests listed seems right for your data, there are other Minitab tests not discussed in this booklet; see your teacher for assistance.

*** IMPORTANT REMINDER: Once you've entered MINITAB, be sure you have typed these commands before running your tests:

MTB > READ 'MINITAB.DATA' C1-C_____ however many columns you have in your data, to read in your data..

MTB > NAME C1 '_________________' C2 '_______________' etc so that all the columns in your data are labeled.

MTB > OUTFILE 'MINITAB.RESULTS' so that the results are saved on the VAX to be printed out later.

MINITAB TESTS DESCRIBED IN THIS BOOKLET

- - - - -

The BINOMIAL TEST

Used with Nominal, Nonparametric Data only

• You used 1 group of subjects -- each subject had only 2_possible choices. You want to see if the subjects' responses differ significantly from chance or from some other expected proportion.

Example 1: You do a survey in which one of your questions is: Do you agree or disagree with this statement: "I like my body image"? You want to see if people answer significantly more one way or the other.

Reasoning: You are considering your subjects as 1 group (you're not separating them into males and females, or old and young, etc.); each subject has only 2 possible choices (agree vs. disagree).If there is no preference for either answer, there is equal likelihood of their choosing either "like" or "dislike". If there is a preference for one choice or the other, you will find that one category or the other has significantly more responses than expected by luck alo ne. Chance would predict a 50/50 split (0.500); you'll be testing if the actual proportion in your data is significantly different from that.

Example 2: You want to see if people have precognition. You take a deck of cards and on each trial, you ask each subject to guess whether the next card will be spades, hearts, diamonds or clubs. You record whether the answer was "right" or "wrong". You are interested in whether people are more accurate than chance in their prediction of the suit of the next card in the deck.

Reasoning: You have 1 group of subjects and each subject has only 2 possible choices (being correct vs. being wrong) . In this case, there is 1 chance for a correct identification out of the four possible suits on each trial by luck alone. The expected outcome by chance in this case is therefore 1 out of 4, or 25% (0.250). You are testing to see if the actual number of correct guesses is significantly different from this.

HOW TO USE THIS TEST

List below where indicated these three numbers for your own data:

K -- the total of "correct" choices OR the total of "wrong" choices, whichever is less. K =______

N -- the total number of trials or the number of subjects in the group, whichever fits N = ______

P -- the probability of the event occurring (if chance, it is 0.50) P = ______

Use these numbers in the below lines

A. Type CDF K=______; <RET> (NOTE: Put a semicolon at the end of the command).

B. Type BINOMIAL N=______ P=______. <RET> (NOTE: Put a period at the end of the command)

A WORKED OUT EXAMPLE

Jan thinks children have significantly more fear of animals than of everything else combined. She asks 20 children, "What is the one thing you fear most?", and categorizes the answers into "Animals" or "Non animals". She finds that 15 list an Animal fear and 5 list a non-animal fear. Is this more than chance variation?

MTB > CDF K=5;

SUBC > BINOMIAL N=20 P=0.50

The computer types on your screen:

K P(X LESS OR = K)

5.00 0.0207

Meaning of the answer: The probability of 15 or more items in one category by chance alone is 0.0207, or only about 2 chances out of 100. Since this is less than 0.05, it is significant; you can be pretty certain that children really do have more animals than non-animal fears, and it's not just random chance that gave you that proportion of answers. Try the same thing with K=4 (16 items out of 20). That shows 0.0059 which would be very significant since 0.0059 is less than 0.01, and you'd be more certain something other than luck was involved.

The CHISQUARE TEST FOR HYPOTHESIS FIT

Used with Nominal, Nonparametric or Ordinal Data

- You used 1 group of subjects and each subject had more than 2 possible choices (5 or less); you want to see if the pattern of the subjects' responses differs significantly from chance or from some other proportion.

Example 1: You want to see if people have color preferences for a new cola you're marketing. You use a food dye and color one batch dark brown, one batch red, and one is left clear. You ask 60 subjects to pick which soda they most prefer. You're interested in whether any color is preferred more or less often than chance.

Reasoning: You have 1 group of subjects (e.g., you're not dividing your subjects into male and female, or young, middle-age, and old for the purposes of this test) and each subject can choose from 3 choices. By chance alone, all should be equally likely to be chosen (33% or .33), so you would expect to find 20 choosing each. If any is chosen significantly more or less than this it would indicate something other than chance at work.

Example 2: As an observer, you record the number of times an elementary school teacher calls on Hispanic males and females compared to non-Hispanic males and females. You want to know if any of the groups are called on more or less often than chance.

Reasoning: The students are members of 1 group; each time she calls on a student, she could call one of the 4 categories. If the class has 5 Hispanic males, 10 Hispanic females, 15 non-Hispanic males, and 20 non-Hispanic females, then 10% of the class is Hispanic male, 20% Hispanic female, 30% non-Hispanic male, and 40% non-Hispanic female. If calling is by chance, we would expect to find that approximate proportion; if the teacher made a total of 200 calls, then by chance alone you'd expect 20 calls of Hispanic males (10% of 200), 40 of Hispanic females (20% of 200), and so on. There's no correct or incorrect call; you're interested in whether the pattern of calling in all four cases is in line with results expected from chance alone.

HOW TO USE THIS TEST

List the number of different categories or choices your subjects can make: Choices _____

Set up this number of Columns to hold the data: if they have 3 choices you need 3 columns, if 4 then 4 columns.

A. Type READ TABLE C20-C____<RET>

B. Type NAME C20 '___________' C21 '____________' <RET> etc. Naming the new columns is a good idea

C. Type ____ ____ ____ ____ etc <RET> Enter your data as the first row

D. Type ____ ____ ____ ____ etc <RET> Enter the expected numbers as the second row

E. Type CHISQUARE C20-C____<RET>

A WORKED OUT EXAMPLE

You feel that the more you conceal your eyes, the more reluctant people will be to talk to you. You ask 80 people at a mall for directions to RIT: 20 while you wear reflective sunglasses, 20 with regular sunglasses, 20 with plain glasses, and 20 with no glasses. You find that, in the above order, the number of people who actually talk to you are 2, 8, 9, and 13. You know that if the type of glasses made no difference, you would expect an equal number of persons to talk to you regardless of what glasses you wore. Since there are 4 conditions and a total of 32 people who talked to you, 8 persons should talk to you in each condition by chance alone. You're testing to see if the actual numbers in each condition are too far from this expected number to be attributable to chance alone. Are the results significant? And if so, what do they mean?

-------> NOTE: Do NOT continue with this test if the average number of persons in any column is less than 5. To find if this is true, divide each column total by the number of rows (10/2, 16/2, 17/2, 21/2). If any result is less than 5, you must combine that column with the one next to it into a single new column (if it makes sense), recalculate the expected # in all the cells and rename the new column before you run the test. The lowest average column (the first) here averages 5, so this example is OK.

MTB > READ TABLE C20-C23

DATA> NAME C20 'REF-SUN' C21 'REG-SUN' C22 'PLAIN' C23 'NO-GLASS'

DATA> 2 8 9 13

DATA> 8 8 8 8

MTB> CHISQUARE C20-C23

2 ROWS READ

Expected counts are printed below observed counts

Ref-Sun Reg-Sun Plain No-Glass Total

1 2 8 9 13 32

5.0 8.0 8.5 10.5

2 8 8 8 8 32

5.0 8.0 8.5 10.5

Total 10 16 17 21 64

ChiSq = 1.800 + 0.000 + 0.029 + 0.595 +

1.800 + 0.000 + 0.029 + 0.595 = 4.849

df = 3

Computer types a table with your data in it. Under that, in our example it prints: Chisq =...4.849 df = 3. Look at the Chi Square table at the back of the booklet using this Chisq number and this df shown on the screen. The table shows that the ChiSq value must BE EQUAL TO OR HIGHER THAN 7.81 to be significant at the 0.05 level for df=3; since this ChiSq is only 4.85, it is not significant. That is the answer.

NOTE: If your answer had been significant, how would you know which condition deviated most from what was expected? The answer is that you can tell which of the boxes contributes most to the significance of the difference by looking at the specific numbers printed after the words "ChiSq =". Each of those numbers tells you how much its individual value differs from chance expectation. Look at the first number (1.80); it refers to the left top box in the table. You had observed 2 persons talking to you in reflective glasses, the computer calculated there should have been 5.0 by chance alone (5.0 is directly under the 2). Similarly, the next number after the Chisq = is 0.00; that refers to the next column, where you entered that 8 talked to you and the computer calculated that 8 should talk to you by chance alone. The largest discrepancy between observed values and expected values therefore, is the number 1.80, in the reflective glasses condition.

If the Total Chisquare had been significant, the largest numbers contribute the most to the difference and that would have been the reflective glasses. However, since your whole Chi Square number is not significant, you can't draw any conclusion from this in the present data set and should report that "No significant difference was found in this data".

Meaning of the Results: Our data haven't shown that wearing glasses makes any difference other than chance in whether people talk to you or not. You therefore should not test any of the conditions further.

The CHISQUARE TEST FOR CORRELATION

Used with Nominal, Nonparametric or Ordinal Data

• You have 2 or more categories of subjects who can make 2 or more possible choices; you want to see if one group's response pattern differs significantly from that of other groups. (This test is sometime used with Ordinal data if the number of responses the subjects can give is limited to a small number -- usually 5 or less).

Example 1: You want to see if there is any relationship between the type of reading that students prefer and their college major. You ask subjects which they like best: Science fiction, Murder mysteries, or Westerns. You also ask them to indicate whether they're majoring in science-related or liberal arts areas.

Reasoning: Subjects are in two nominal variables: Fiction type (Science fiction, Murder Mysteries, or Westerns) and major (Science-Related or Liberal Arts major). You have 3 categories on the first and 2 on the second. You wish to know if knowing the place on one variable helps you predict the other

Example 2: You think family size is related to loneliness at RIT. You ask students for the number of brothers and sisters they have, giving the choices "None", "1 or 2", and "3 or more". You also ask them to indicate whether they are "Very Lonely", "Somewhat Lonely", or "Not Lonely". You want to know if the two are related.

Reasoning: Each subject is placed into the Ordinal variable "Family Size" of 3 categories at the same time as she falls into a place on the other Ordinal variable "Loneliness" of 3 categories. You are interested in seeing if the two are related. Both of these variables are Ordinal since both can be ordered in terms of "more than".

Example 3: You think that men will give more charity than will women. You dress as a street-person and panhandle unselected subjects in downtown Rochester. You record the subject's sex and whether the subject gave "No Money", "Some money but less than $1.00", "Between $1.00 and $5.00", or "Over $5.00".

Reasoning: "Gender" is a Nominal variable with 2 categories. The way the money was recorded is Ordinal with 4 categories. You want to see if gender predicts charity. If we had recorded the exact amount given it would have been Interval data and you should use the GLM test.

HOW TO USE THIS TEST

List the number of different categories or choices your subjects can make: Choices ____

Set up this number of Columns to hold the data: if they have 3 choices you need 3 columns, if 4 then 4 columns.

A. Type READ TABLE C20-C____<RET>

B. Type NAME C20 '___________' C21 '____________' <RET> etc. Naming the new columns is a good idea

C. Type ____ ____ ____ ____ etc <RET> Type in your own first line of data here

D. Type ____ ____ ____ ____ etc <RET> Type in your own second line of data here

(NOTE: If you have a 3rd or 4th line of data, type it in here.)

E. Type CHISQUARE C20-C____<RET>

A WORKED OUT EXAMPLE

Joe wants to see if college students differ from non-college students on the issue of unwanted pregnancies. He writes a short story about a 16-year old unmarried high school girl who is pregnant and undecided about what to do. He asks subjects to give her advice by checking one of these categories: (1) Have an abortion immediately. (2) Have the child but give up for adoption. (3) Have the child but raise by yourself. (4) Have the child and force father to marry you. (5) No Opinion. He gets these answers from 72 college and 105 Non-college students:

Category	Abort	Give Adopt	Raise	Force Mar	No Opinion
Coll	42	10	8	2	10
Non-Coll	30	44	10	6	15

NOTE: You should **NOT** use this test if the average number of persons in any column is less than 5. Therefore, divide each column total by the number of rows before you do anything else: (72/2, 54/2, 18/2, 8/2, 25/2). Since the choice "Force Marriage" averages less than 5 (8/2 = 4), you must combine that column with some other one into a single new column (if it makes sense). In this case, you can combine it with the "No Opinion" column and call the new column "Other".

Category	Abort	Give Adopt	Raise	OTHER
Coll	42	10	8	12
Non-Coll	30	44	10	21

MTB > READ TABLE C20-C23

DATA> 42 10 8 12

DATA> 30 44 10 21

DATA> NAME C20 'ABORT' C21 'ADOPT' C22 'RAISE' C23 'OTHER'

MTB> CHISQUARE C20-C23

2 ROWS READ

Expected counts are printed below observed counts

	ABORT	ADOPT	RAISE	OTHER	
1	42	10	8	12	72
	29.29	21.97	7.32	13.42	
2	30	44	10	21	105
	42.71	32.03	10.68	19.58	

Total	72	54	18	33	177

ChiSq = 5.517 + 6.519 + 0.063 + 0.151 +

3.783 + 4.470 + 0.043 + 0.104 = 20.649

df = 3

Go to the Chi Square Table at the back of the booklet using the Chisq # and df the computer printed (df=3). The chart reads that ChiSq must EQUAL OR BE HIGHER THAN 7.81 to be significant at the 0.050 level, and 11.34 at the 0.010 level. Your number is 20.65, larger than 11.34; it is therefore very significant. That is your answer.

Meaning of the Result: Since your result is very significant, you know that college persons respond in a different pattern from non-college. You can tell which of the responses differed most from chance expectation by looking at the specific numbers after the "ChiSq =". Each of those numbers refers to the cell (box) in the diagram above. The first number refers to the first cell (1, C20), the second number to the second cell (1, C21), and so on. The larger the number, the more that cell value differed from chance expectation. The first number, 5.517, indicates that there is a relatively big difference between the 42 actual College students who suggest "Have Abort" and the 29.29 who would have been expected by chance alone. This difference contributed a lot toward the total Chi Square value of 20.65. The largest difference between observed values and expected values is the next number 6.519; looking up at the table, you can see that 21.97 responses were expected in the column "Give adoption" for college students, but only 10 actually made that choice. The other two alternatives show numbers quite close to those expected by chance alone and are therefore not important. You now know that significantly more college students expressed support for abortion and fewer for adoption than expected by chance alone.

MANN-WHITNEY TWO-SAMPLE RANK TEST

Used with Ordinal Nonparametric or Interval Data

• You use 2 groups of subjects; each subject receives 1 score on a ranking or an ordered scale. You want to see if one group's average ranking or ordering differs significantly from that of the other group.

Example 1: You believe that there may be a difference in the amount of money people tip when a waiter "Hi, my name is _______" compared to when he just says "Hi". You have waiters introduce themselves by name to some customers and not to others, and record the amount of money each customer leaves as a tip. You want to compare whether being introduced to a waiter has a significant effect on the amount of tipping, but you're not sure that the data is parametric since there is one patron who is drunk and leaves a very big tip.

Reasoning: There are 2 groups of subjects ("Introduced to Waiter" and "Not Introduced to Waiter") and each patron is assigned 1 score (the amount of money left). Although amount of money is normally Interval data, very large or very small extremes in the data can give a false result on a Parametric test so you decide to be conservative and treat the data as Nonparametric instead.

Example 2: You want to check how male handsomeness influences sentences for rape. You make up a story of a man picking up a woman in a bar, taking her to his apartment, and eventually raping her. You attach a picture of an ugly man to the story for 20 subjects and a picture of a handsome man to the story for 20 other subjects . You ask each subject to indicate by checking the box what sentence if convicted the rapist should receive to prison: "Less than 1 year", "1-2 years", "2-5 years", "5-10 years", "10-20 years", or "20 years to life".

Reasoning: There are 2 groups of subjects (a group shown the handsome picture; a group shown the ugly picture) and each subject gets only 1 score (the rank of the time in prison assigned the rapist). The scores are Ordinal data since they can be ordered from smaller to larger, but they are not Interval since the distances between the different alternatives are not equal and as well, the last category is indeterminate.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data you want to analyze or UNSTACK: C____.

Which column has the SUBSCRIPTS numbers indicating whether the subject is in Group 1 or 2: C____.

Use those numbers where indicated below:

A. Type UNSTACK C___ C20 C21; <RET> Don't forget to a semi-colon at the end of the command.

B. Type SUBSCRIPTS C____. <RET> Don't forget to put a period at the end of command.

C. Type NAME C20 'LEFT' C21 'RIGHT'<RET> Name the new columns

D. Type PRINT C20 C21 <RET> Check to be sure your data is correct. If not, quit Minitab, and correct it.

E. Type MANN C20 C21 <RET> to get the test results.

A WORKED OUT EXAMPLE

A school psychologist feels that left-handed children are more creative than right-handed children. He makes up a test of creativity, has a group of children take it, and then separates them into those who are left-handed and those who are right-handed. His results are as follows:

(Group 1) Left-handed children: 100 80 76 76 65 60 (N = 6)

(Group 2) Right-handed children: 90 78 72 68 64 24 0 (N = 7)

Since he's not sure the underlying distribution is interval, he is conservative and uses the ordinal Mann-Whitney test. Are the results significant, or likely to be just due to chance?

MTB > UNSTACK C2 C20 C21;

SUBC> SUBSCRIPTS C1.

MTB > NAME C20 'LEFT' C21 'RIGHT'

MTB > PRINT C20 C21

ROW LEFT RIGHT

1 100 90

2 80 78

3 76 72

4 76 68

5 65 64

6 60 24

7 0

MTB > MANN C20 C21

Mann-Whitney Confidence Interval and Test

C20 N = 6 MEDIAN = 76.000

C21 N = 7 MEDIAN = 68.000

POINT ESTIMATE FOR ETA1-ETA2 IS 9.00

96.2 PCT C.I. FOR ETA1-ETA2 IS (-11.99, 59.99)

W = 49.0

TEST OF ETA1 = ETA2 VS. ETA1 N.E. ETA2 IS SIGNIFICANT AT 0.3531

THE TEST IS SIGNIFICANT AT 0.3524 (ADJ. FOR TIES)

CANNOT REJECT ALPHA = 0.05

\

In this example, the computer types some information including IS SIGNIFICANT AT 0.3524 (adjusted for ties).

Meaning of Example: There is a probability of 0.352 that the rankings of the two groups differ only by chance; this is far higher than the .050 level required for significance, so the test result is not significant. This test has not shown that left-handed children are more (or less) creative than right.

The T-TEST WITH INDEPENDENT GROUPS

Used with Interval, Parametric Data

• You use 2 groups of subjects who get 1 score on an Interval scale; you want to see if one group's average score differs significantly from the other group's average score.

Example 1: You want to test if there is a difference in the "helpfulness" of RIT vs. UR students. You ask randomly selected RIT and UR students to help you by writing "a paragraph on any topic you wish, the longer the better". You then measure how many words were written by each of the RIT students and by each of the UR students, to see if students in the two colleges differ in "helpfulness".

Reasoning: There are 2 groups of subjects (RIT students and UR students) and each yields 1 score (number of words written voluntarily). Number of words is Interval data and you think that the distribution is parametric.

Example 2: You wonder if students' gender influences how jealous they are. You give a "Jealousy Test" with scores from 1 to 40 to both males and females and record the results.

Reasoning: There are 2 groups of subjects (males and females) and each subject gets only 1 score (the score received on the Jealousy test). You believe the scores are Interval data and that the distribution falls into a bell-shaped curve.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data you want to analyze or UNSTACK: C____.

Which column has the SUBSCRIPTS numbers indicating whether the subject is in Group 1 or 2: C____.

Use these two numbers where indicated below:

A. Type UNSTACK C___ C20 C21; <RET> Don't forget to put a semi-colon at the end of command

B. Type SUBSCRIPTS C____. <RET> Don't forget to put a period at the end of command

C. Type NAME C20 'NEW' C21 'LECT' <RET> NOTE: Put a single quote before and after each name.

D. Type PRINT C20 C21 <RET> Check to be sure your data is correct. If not, quit Minitab, and correct it.

E. Type TWOS C20 C21

A WORKED OUT EXAMPLE

Marie thinks that a new method of teaching will produce higher grades than straight lecturing. She teaches the same subject matter to 15 students with the new method, and 17 students with the old, and gives identical final exams. The students in each group get the following scores:

(Group 1) New Method: 85, 68, 88, 71, 94, 94, 73, 77, 82, 91, 95, 89, 83, 79, 85. (N=15)

(Group 2) Lecture: 94, 77, 81, 65, 55, 79, 84, 81, 70, 70, 65, 71, 79, 85, 90, 76, 80. (N=17)

Do the two groups differ significantly from each other, and if they do which method appears better?

MTB > UNSTACK C2 C20 C21;

MTB > SUBSCRIPTS C1.

MTB > NAME C20 'NEW' C21 'LECT'

MTB > PRINT C20 C21

ROW NEW LECT

1 85 94

(Note: I've omitted some scores to save space)

17 80

MTB > TWOS C20 C21

TWOSAMPLE T FOR C20 VS C21

N MEAN STDEV SE MEAN

NEW 15 83.60 8.60 2.22

LECT 17 76.59 9.76 2.37

95 PCT C.I. FOR MU C20 - MU C21: (0.3719, 13.65)

TTEST MU C20 = MU C21 (VS NE): T = 2.16 P = 0.039 d.f. = 29

In this example, the computer types some information including P = 0.039.

Meaning of the Results: Since the probability value of 0.039 is less than 0.05 the difference is significant. However, since 0.039 is greater than 0.010, the difference is not very significant. Therefore, there is some chance that random factors alone caused the difference between the groups but it is more likely to be a true difference. The group taught by the new method had a higher MEAN (average) score than the old method group so we can state that the students learned significantly better by the New Method than by Lecture.

WILCOXON MATCHED PAIRS TEST

Used with Ordinal, Nonparametric or Interval Data

• You use 1 group of subjects each of whom takes the SAME TEST 2 times or does the SAME THING two times; you want to see if each subjects' average response has changed significantly between the two events.

Example 1: You believe that appreciation of modern art can be taught. You give a group of college students a group of paintings and ask them to rank them in terms of beauty. Ten of the paintings are modern art and ten are traditional paintings. You then have an art teacher explain what modern artists are attempting to convey with their new techniques. A week later you again ask the students to rank the paintings. You want to see if the rankings of the "modern art" paintings are rated differently the second time compared to the first.

Reasoning: There is 1 group of subjects each of whom ranks the SAME objects 2 times (ranking of the 10 modern paintings the first time and their ranking at the second testing) and you want to see if what you did in between the 2 rankings made a difference in the average value of these scores. Since the data involves ranking, you know that it is Ordinal data.

Example 2: You have 30 persons who have been unemployed for varying lengths of time. You divide them into 15 matched pairs of subjects on the basis of how long they've been unemployed, with each pair having subjects who are as close to each other in attributes as possible. You give one subject in each pair intensive social skill training, and do nothing with the other. You then wait and record the length of time each subject stays unemployed, comparing the persons who received your training to the non-trained ones in each matched pair.

Reasoning: Each matched pair of subjects can be considered as though it were a single subject in a test-retest situation; in other words as though the same subject had 1 test in 2 conditions -- was both exposed to social skill training and not exposed. The two scores therefore are length of time the 2 subjects in each of the 15 pairs remain unemployed; that time should significantly differ if your skill training made a difference. You believe that your scale is Ordinal data, since some people never get employed; there is no number that could be assigned them in an Interval test.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data for the 1st set of scores: C____. Which has the data for the 2nd set of scores: C____.

Use the two numbers were indicated below:

A. Type SUBTRACT C____ from C____ and put the results in C22 <RET> 1st set and 2nd set of scores.

B. Type NAME C____ 'TEST_1' C____ 'TEST_2' C22 'DIFFER' <RET> Use single quotes around each name

C. Type PRINT C____ C____ C22 <RET>

D. Type WTEST C22 <RET>

A WORKED OUT EXAMPLE

Susan has made up a "Fear of Computers" test that she has given to her pre-schoolers, from zero (no fear) to 100 (maximum fear). She then shows them how to play with computer games and lets them play for 1/2 hour. She re-administered the "Fear of Computers" test. Her results are:

SUBJ TEST 1 TEST 2 SUBJ TEST 1 TEST 2

1 60 75 5 73 65

2 85 85 6 20 28

3 40 20 7 35 31

4 27 24 8 65 43

Has she shown that exposure to computer games will lessen the amount of fear on the test?

MTB> SUBTRACT C1 C2 C22

MTB> NAME C1 'TEST_1' C2 'TEST_2' C22 'DIFFER'

MTB > PRINT C1 C2 C22

ROW TEST_1 TEST_2 DIFFER

1 60 75 15

(Note: I've omitted some scores to save space)

8 65 43 -22

MTB > WTEST C22

TEST OF MEDIAN = 0.000000000 VERSUS MEDIAN N.E. 0.000000000

N FOR WILCOXON ESTIMATED

N TEST STATISTIC P-VALUE MEDIAN

C22 8 7 8.5 0.398 -3.750

Meaning of the Results: Since the probability of 0.398 is greater than 0.050, the difference in the test-retest scores is not significant. We therefore can't state that the children showed any less fear after playing with the computer games than they showed before playing with them, since any difference in the scores could have been by chance alone. If the results had been significant, than the values in the "Difference" column would be mostly either all plus or all minus.

The T-TEST WITH CORRELATED GROUPS

Used with Interval, Parametric Data

•You use 1 group of subjects taking the SAME test 2 times; you want to see if the subjects' average responses differs significantly between one test and the other.

Example 1: Before a course in "The Value of Modern Art", you ask students to guess how much money each of 20 paintings sold for. You then teach the course, and ask them again to give monetary guesses as to the sale price of each painting. You want to see if the average evaluation changed due to your course.

Reasoning: There is 1 group of subjects (students in the course) and each subject has 2 scores (money assigned to the painting on the first test and on the second test) . You want to see if what you did in between made a difference in the average value of these scores. You believe that the scores are Interval data and normal.

- - - - -

Example 2: As a school psychologist, you believe that pairing a failing child with a "mentor" will help the child succeed in school. You select 20 children and divide them into two groups: you try as much as possible to match equivalent subjects in each group in terms of their gender, their IQ, and teacher's rating of the child's potential. You then arrange for 10 of these children to meet with mentors once a week and the other 10 receive no mentor. At the end of the school year, you administer an achievement test to all 20 children. You want to see if there is a significant difference in the grades of mentored children vs. non-mentored children.

Reasoning: Each matched pair of subjects can be considered as though it were a single subject in a test-retest situation; in other words as though the "same child" was both given mentoring and not given mentoring. The 2 scores therefore are the grades received by the child in the mentored group vs. the grades received by the matched child in the non-mentored group. You believe that the grades are interval and that the distribution is parametric. In this example, you'd probably be more secure using a Nonparametric test.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data for the 1st set of scores: C____. Which has the data for the 2nd set of scores: C____.

Use the two numbers were indicated below:

A. Type SUBTRACT C____ from C____ and put the results in C22 <RET> 1st set and 2nd set of scores.

B. Type NAME C____ 'TEST_1' C____ 'TEST_2' C22 'DIFFER' <RET> Use single quotes around each name

C. Type PRINT C____ C____ C22 <RET>

D. Type TTEST C22 <RET> to see and save the test results.

A WORKED OUT EXAMPLE

Glenda believes that people's perceptions are affected subtly by hunger. She invites subjects in at lunch-time, gives them a good lunch, and then asks them to guess at objects which are blurred by a frosted glass window. After 8 hours without food, she retests them with the same objects in a different order behind the frosted glass. Here are the results (higher numbers indicate more food related responses):

Subj Fed Hungry Subj Fed Hungry

1 8 12 6 9 9

2 3 4 7 7 14

3 4 9 8 1 12

4 3 11 9 6 14

5 0 6 10 5 11

She wants to test the hypothesis that there will be significantly more food related responses to ambiguous objects when subjects are hungry than when just fed. She believes the data is parametric.

MTB > SUBTRACT C20 C21 C22

MTB > NAME C20 'FED' C21 'HUNGRY' C22 'DIFFER'

MTB > PRINT C20 C21 C22 <RET>

ROW FED HUNGRY DIFFER

1 8 12 4

(Note: I've omitted some scores to save space)

10 5 11 6

MTB > TTEST C22

TEST OF MU = 0.00 VS MU N.E. 0.00

N MEAN STDEV SE MEAN T P VALUE

C22 10 3.60 3.37 1.07 3.37 0.0082

Meaning of the Results: Since the p-value of 0.008 is less than 0.01, the difference in the result is very significant; we can be quite sure that the difference between the before and after eating food responses to the ambiguous image is not due to chance variations alone, but is due to some other factor. Since we've subtracted C20 from C21, there will be a higher Mean score in C22 if the FED condition is higher, and a negative Mean score if the HUNGRY condition is higher. Examine the printout and you'll find the mean is 3.60 and positive; therefore, that would support the hunger variable as the cause of the difference.

SPEARMAN RANK-ORDER CORRELATION

Used with Ordinal Nonparametric or Interval Data

- You use 1 group of subjects taking 2 tests or scales; you want to see whether scores on one are correlated with (useful in predicting) scores on another.

Example 1: You want to see if scores on a "Risk-Seeking" Scale can be used to predict how likely people are to value security in a job over money. You ask subjects to fill out the Risk-Seeking scale and then ask them to check off which of the below different combinations of salary/security they would most prefer for their first position: "Job pays $20,000 yr/no chance of being fired", "Job pays $30,000 yr/ 1 chance in 100 of being fired", "Job pays $40,000 yr/ 1 chance in 50 of being fired", "Job pays $50,000 yr/1 chance in 20 of being fired", "Job pays $60,000 yr/1 chance in 10 of being fired", "Job pays $70,000 yr/1 chance in 2 of being fired".

Reasoning: You have 1 group of subjects with each person having 2 scores on different tests; one on the Risk-Seeking scale and one on the Salary/Security measure. You want to see if there is a correlation being the two scales. You know that the data is Ordinal, since it can be ordered from lowest to highest; it can't be Interval since the distance between the categories is not equal

Example 2: You create 10 stories about people with careers and have experts rank these stories in terms of the amount of success that each one shows. You also get 10 pictures of females in graduation photos and have experts rank them in terms of beauty. You randomize the pictures and stories and ask males to match the stories with the picture they feel fits it best. Your hypothesis is that there will be a negative correlation (high score on amount of success paired with low score on beauty) between the two rankings.

Reasoning: You have 1 group of subjects, each of whom can be assigned 2 scores; the ranking of the success with the ranking of the female's beauty. You want to see if the average rankings of the stories and the pictures correlate. Ranked data is always Ordinal data and must be tested with Nonparametric tests.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data for the 1st set of scores: C____. Which has the data for the 2nd set of scores: C____.

Use the two numbers were indicated below:

A. Type RANK C____ into C20<RET> The 1st set of scores

B. Type RANK C____ into C21 <RET> The 2nd set of scores

C. Type NAME C___ 'ORDER' C____ 'GRADES' C20 'RANK-ORD' C21 'RANK-GRA'<RET>

D. Type PRINT C____ C____ C20 C21 <RET> If your data is not correct, quit Minitab, and correct it.

E. Type CORRELATE C20 C21.

A WORKED OUT EXAMPLE

Example: You believes that the faster a student finishes an exam, the higher will be the grade.

You record the order in which each student hands in a test and the score, and get this:

1st, 65; 2nd, 40; 3rd, 75; 4th, 75; 5th 80; 6th 75; 7th, 80; 8th, 80; 9th, 90; 10th, 85;

11th, 90; 12th, 100; 13th 100; 14th, 90; 15th, 95.

Is the correlation significant? Can you roughly predict a person's score by his order of finishing the exam??

MTB > RANK C1 C20

MTB > RANK C2 C21

MTB > NAME C1 'ORDER' C2 'GRADES' C20 'RANK-ORD' C21 'RANK-GRA'

MTB > PRINT C1 C2 C20 C21

ROW ORD GRA RANK-ORD RANK-GRA

1 1 65 1 2.0

(Note: I've omitted some scores to save space)

15 15 95 15 13.0

MTB > CORRELATE C20 C21

CORR = 0.931

In this example, the computer prints CORR = 0.931. That is the correlation. For correlations, the df is the number of pairs of scores minus 2. We have 15 pairs of scores, so d.f. = 13. You can't use this test with less than 7 pairs of scores, and if you have more than 30 pairs, consult your teacher. Now use the table at the bottom of this page. (The table goes by twos after 10; use the next lower number which is d.f. = 12). The p-value shown under the .05 level is 0.591 and under the .01 level is 0.777; your correlation of 0.931 is greater than 0.777. Therefore, it's p-value is less than .01 and the correlation is very significant.

Meaning of the Result: Since our result is very significant, we know there is a strong relationship between the two sets of data. However, be careful! If you examine your data, you will see that the lowest grades were paired with the earliest finish times, the opposite of the research hypothesis! Correlation only tells you significance; look at the data to see what it means in terms of your hypothesis. The meaning here is that early finishers got lower scores. If you type PLOT C20 C21 <RET> the computer will print a graph of the correlation.

PEARSON PRODUCT-MOMENT CORRELATION

Used with Interval, Parametric Data

• You use 1 group of subjects taking 2 tests or scores; you want to see whether scores on one test are correlated with (used to predict) scores on another test

Example 1: You believe that a man's height will be correlated with the salary he is offered at his first job interview. You collect the height and salary offers of the RIT graduating class and want to see how well they go together.

Reasoning: You have 1 group of subjects, each subject having 2 scores (height and salary offer). The scores of both measures are Interval data (both height and money have equal intervals), and you believe that they will at least roughly fall into a bell-shaped distribution, and you want to see if they correlate with each other.

Example 2: You ask a group of subjects to answer a survey in which one question is, "How much money do you think the average welfare mother with two children receives a month in assistance?" Later in the survey you ask, "How much do you think it takes to support a prisoner in jail for a month?" You want to see if there is a significant correlation between a subject's estimation of welfare expenses and her estimate of jail expenses.

Reasoning: You have 1 group of subjects, each subject making 2 estimates (the amount of money welfare mothers receive and the amount of money jailing costs). You wish to see if the two correlate; if knowing one measure helps predict scores on the other measure. Money is by nature interval data, and you assume that the data will be Parametric (bell-shaped curve).

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data for the 1st set of scores: C____. Which has the data for the 2nd set of scores: C____.

Use the two numbers where indicated below:

A. Type NAME C____ 'HEIGHT' C____ 'SALARY' <RET> Name 1st and 2nd sets of data

B. Type PRINT C____ C____ <RET> If your data is not correct, quit Minitab, and correct it.

C. Type CORRELATE C____ C____

A WORKED OUT EXAMPLE

Example: Martha wonders whether the height of graduating engineer students is related to the salaries that they'll receive in their jobs. She collects this data for 15 students who just graduated and comes up with the following information (each pair of scores represents one student's height and starting job salary): 60", $19000; 66", $15000; 70", $18500; 63", $14800; 66", $16000; 72", $21000; 77", $20000; 69", $17500; 72", $19800; 69", $15900; 72", $20500; 68", $18000; 61", $12800.

Is there a relationship between salary and height -- is low height related to a low starting salary?

MTB > NAME C1 'HEIGHT' C2 'SALARY'

MTB > PRINT C1 C2

ROW HEIGHT SALARY

1 60 19000

(Note: I've omitted some scores to save space)

13 61 12800

MTB > CORRELATE C1 C2

Correlation of C1 and C2 = 0.68

In this example, the computer prints Correlation = 0.68. Look at the table at the end of this booklet for this test. The table asks for the d.f., which for correlations is the number of pairs of scores minus 2. We have 13 pairs of scores , therefore d.f. = 13 - 2, or 11. For the p-value shown under d.f. of 11, the .05 level is 0.553 and under the .01 level is 0.684. Your correlation of 0.68 is more than 0.553 but less than 0.684; it's probability of occurrence by chance falls between 0.05 and 0.01, and the correlation is significant. That is your answer.

Meaning of the Result: According to our data, height is related to starting salary for graduates. Correlation only tells you significance; you must examine the data to see what it means in terms of your hypothesis. You don't yet know if shorter engineers earn more money or if taller ones do. Type PLOT C2 C1 <RET> The computer will then print out a graph which allows you to see what the correlation looks like.

The KRUSKAL-WALLACE TEST

Use with Ordinal Data or Interval Data

- You use more than 2 groups of subjects, each subject giving 1 response to each test or item; you want to see if one group has an average response that differs significantly from that of one or more of the other groups.

Example 1: You get four English teachers to read 20 essays and to rate them from 1 to 10 in terms of their quality. You want to see if one or more of the English teachers differ from the rest in how she rates the papers.

Reasoning: You have more than 2 groups of ratings, each of which has 1 response to each paper. Since the response is in the form of a rating, you don't know that the difference between each number is equivalent or if the data is parametric. Therefore, you decide to be conservative and test it as an ordinal scale.

Example 2: You wish to see if freshmen, sophomores, juniors, or seniors register higher on a test of "Happiness". You give them the test and collect their scores on your scale.

Reasoning: There are more than 2 groups (Freshmen, Sophomores, Juniors, Seniors); each subject gets only 1 score on your happiness scale. You're interested in whether these 4 groups differ among themselves more than expected by chance alone. You've never used the test before so don't know if it's interval; it is ordinal.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data you want to analyze: C____.

Which column has the SUBSCRIPTS numbers indicating which group the subject is in: C____.

Insert the first number after the first C and the second number after the second C in Lines A, B, C below:

A. Type NAME C____ 'EARNINGS' C____ 'FARMS' <RET> The first C is the data, the second the subscripts

B. Type PRINT C____ C____<RET> Check to be sure your data is correct. If not, quit Minitab, and correct

C. Type KRUSKAL for the data in C____ with the subscripts in C____ <RET>

A WORKED OUT EXAMPLE

Example: A social worker feels that migrant workers picking in Farm

#1 are being systematically cheated. She notes the amounts earned by the pickers in that farm and in two other farms over a week's time. She makes no assumptions about the distribution of the population data and therefore uses the nonparametric Kruskal-Wallis test. She enters the data in two columns (I've broken it into 6 columns for convenience in display in this booklet):

Farm # Earnings

1 150

1 148

2 128

3 168

1 185

2 159

1 167

3 134

3 175

2 200

1 133

3 142

2 218

3 139

1 210

3 158

3 145

2 137

1 155

3 128

1 162

2 123

3 145

2 148

MTB> NAME C1 'FARM' C2 'EARNINGS'

MTB > PRINT C1 C2

ROW FARM EARNINGS

1 1 150

(Note: I've omitted some scores to save space)

24 2 148

MTB > KRUSKAL C1 C2

LEVEL NOBS MEDIAN AVE. RANK Z VALUE

1 8 158.5 15.2 1.32

2 7 148.0 11.9 -0.29

3 9 145.0 10.6 -1.01

OVERALL 24 12.5

H = 1.86 d.f. = 2 p = 0.396

H = 1.86 d.f. = 2 p = 0.395 (ADJ. FOR TIES)

In this example, the computer types info, then prints H = 1.86 d.f. = 2 p = 0.396 (p = 0.395 ADJ. FOR TIES)

Meaning of the Results: The probability of the H statistic being 1.86 was found to be 0.396 which is greater than 0.050. The differences between the farms are therefore not significant and you cannot say that one farm has higher or lower average pay than the other farms based on this data, since you can't rule out random factors as accounting for any differences you might find. If H had been high enough, the p-value would have been lower than 0.050 and would have been significant. You should then use the Mann-Whitney test to see which farms had significantly higher or lower pay than the others.

ONEWAY ANALYSIS OF VARIANCE TEST

Used with Interval Data only

- You use more than 2 groups of subjects each giving 1 response on a scale or test to each item; you want to see if one group has an average response that differs significantly from one or more of the other groups.

Example 1: You want to see if there is a difference in the grading practices of three professors all of whom taught Introduction to Psychology in the Spring Quarter. You obtain the number grade they assigned to each of their students. You want to see if the average assigned grade differs more than by chance.

Reasoning: There are more than 2 professors, each assigning 1 score (grade) to a student. You're interested in whether there is any difference in the averages given by the different professors. You believe that the grading is Interval and falls into a bell-shaped distribution.

Example 2: You want to see if different appeals for charity have different results. You go door to door in the dorm asking for money for Cancer research, at one door using an emotional appeal, at another a rational appeal, at another you just ask for money to fight cancer. You wonder which of the techniques is most successful in gathering funds.

Reasoning: You are comparing more than 2 techniques , each of which gets 1 response (money) from each subject. Money lends itself to Interval measurement.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data you want to analyze: C____.

Which column has the SUBSCRIPTS numbers indicating which group the subject is in: C____.

Use these numbers are indicated:

A. Type NAME C____ 'SCORES' C____ 'COLL-YR' <RET> First Column is the data, the second the subscripts

B. Type PRINT C____ C____<RET> Check to be sure your data is correct. If not, quit Minitab, and correct

C. Type ONEWAY the data in C____ with the subscripts in C____ <RET>

A WORKED OUT EXAMPLE

Marilyn has collected these music test scores from students in different years of college. She wants to see which if any of these years differs significantly from the others. Her data is:

Score Year

15 1

24 4

12 2

45 3

18 1

33 2

17 1

34 3

23 4

41 2

42 3

33 1

28 2

39 3

10 1

58 4

17 3

37 2

28 3

55 4

12 1

23 2

23 3

48 4

MTB > NAME C1 'SCORE' C2 'COLL-YR'

MTB > PRINT C1 C2

ROW SCORE COLL-YR

1 15 1

2 24 4

(Note: I've omitted some scores to save space)

21 12 1

Continue? Y

23 23 3

24 48 4

MTB > ONEWAY C1 C2

ANALYSIS OF VARIANCE ON 'SCORE'

SOURCE DF SS MS F p

Note: "Error" is a COLL-YR 3 1662 554 4.15 0.019

statistics term, not ----> ERROR 20 2670 134

something wrong! TOTAL 23 4332

INDIVIDUAL 95 PCT CI'S FOR MEAN BASED ON POOLED STDEV

LEVEL N MEAN STDEV

-----+---------+---------+---------+-

1 6 17.50 8.17 (------*-----)

2 6 29.00 10.49 (-----*------)

3 7 32.57 10.34 (-----*-----)

4 5 41.60 16.92 (------*------)

-----+---------+---------+---------+-

POOLED STDEV = 11.56 15 30 45 60

\- - - - -

Meaning of the Results: The F-Ratio = 4.15, with the probability of p = 0.019. A probability of 0.019 is less than 0.050 but greater than 0.0100 so the differences between the groups are significant [but not very significant]. This means that there is a likelihood that at least one difference among the groups is not due to random factors . You now do a T-Test for Independent Groups between each pair to see which of them have means (averages) significantly differ from each other.

FRIEDMAN TWO-WAY ANOVA BY RANKS

Used with Ordinal, Nonparametric or Interval Data

• You use 1 group of subjects taking the SAME test MORE than 2 times; you want to see if the subjects' average responses differs significantly between any of the tests and the rest.

Example 1: You feel that music influences how well people can memorize. You choose four conditions: hard rock music, jazz music, classical music, and no music, and give the same subject equivalent lists to memorize under each of the conditions. You are interested in whether music makes a difference in the time to memorize lists. Reasoning: There is 1 group of subjects, each subject getting more than 2 scores (one score on each condition, totaling 4) and you want to see if what you did made a difference in the average value of these scores.

Example 2: You time 30 people solving a jigsaw puzzle. You then set up 3 groups, each group having subjects who took roughly equal time to do the puzzle. You give one group a lecture on "using your right mind creatively", the second group some additional puzzles to work on, and the third group watches TV soap operas; afterwards you time each subject on a new puzzle. You're interested in whether what you did in between the testing produced any measurable effects.

Reasoning: Each group of 3 matched subjects can be considered as though they were 1 group of subjects in a Test-Retest method getting more than 2 scores. The length of time the subjects in each of the 3 groups take to do the second puzzle, therefore, should only differ significantly from the matched persons in the other groups only if the in-between conditions had some effect

Enter your data into MINITAB as described on Page 5. Answer these two questions:

Which column has the data you want to analyze: C____.

Which column has the SUBSCRIPTS numbers indicating the first subject grouping: C____.

Which column has the SUBSCRIPTS numbers indicating the second subject grouping: C____.

Use the numbers as indicated below:

A. Type NAME C____ 'ESP_SC' C____ 'GROUP' C____ 'ARTICLE'<RET>

B. Type PRINT C____ C____ C____ <RET> Check to be sure your data is correct. If not, quit Minitab, and correct

C. Type FRIEDMAN for the data in C____ with the first subscripts in C____ and the second in C____<RET>

A WORKED OUT EXAMPLE

You ask people check whether they Disbelieve in ESP strongly, Disbelieve a little, No opinion, Believe a little, or Believe strongly, and place them into 5 groups based on their answers (Belief groups 1, 2, 3, 4, 5). You give random subjects from each group one of four articles: Article 1 attacks the idea of ESP, Article 2 supports it, Article 3 presents pro and con arguments, and Article 4 is not on the topic. You then give an ESP test and are interested in whether there is an interaction between belief level, type of article read, and scores on the test . Since you are not sure ESP scores are normally distributed and you have matched your subjects in the different conditions, you use the Friedman instead of the GLM test.

SCORE GP ART

14 1 1

15 2 1

17 3 1

20 4 1

25 5 1

16 1 2

22 2 2

21 3 2

24 4 2

29 5 2

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

12 1 3

12 2 3

13 3 3

18 4 3

11 5 3

19 1 4

22 2 4

21 3 4

26 4 4

30 5 4

19 1 4

22 2 4

21 3 4

26 4 4

30 5 4

MTB > NAME C1 'ESP_SC' C2 'GROUP' C3 'ARTICLE'

MTB > PRINT C1 C2 C3

ROW ESP_SC GROUP ARTICLE

1 14 1 1

2 15 2 1

3 17 3 1

4 20 4 1

5 25 5 1

6 16 1 2

: (Note: I've omitted some scores to save space)

20 30 5 4

MTB > FRIEDMAN C1 C2 C3

Friedman test of ESP_SC by GROUP blocked by ARTICLE

S = 8.65 d.f. = 4 p = 0.071

S = 8.76 d.f. = 4 p = 0.068 (adjusted for ties)

Est. Sum of

GROUP N Median RANKS

1 4 16.200 5.5

2 4 18.600 10.5

3 4 18.700 11.0

4 4 22.700 17.0

5 4 26.800 16.0

Grand median = 20.600

Meaning of the Results: Since a probability of 0.068 is greater than 0.050, the differences among the groups are not significant (although it's close). As it stands, you can't say that the articles interacted with the subject's beliefs to influence their scores on the test; the variations in the scores could have been due to random factors alone. However, because 0.068 is near to 0.050 you might want to repeat your research with more subjects. If the p-value had been lower then 0.050 and significant, then the medians printed below the S-scores tell you which groups were higher and which lower. You'd do a Mann-Whitney between the groups to find out which differences were significant.

The GENERAL LINEAR MODEL TEST

Used with Interval Data

You use 1 group of subjects and you want to compare the influence of 2 or more independent variables on the subjects' response (1 dependent variable at a time). You believe there may be an interaction between the independent variables that produces significant differences in the subjects' responses

Example 1: You wish to see the influence of prison rehabilitation strategies on parolees' conduct. While they're in prison, you offer some prisoners' work for money and others none. You think the type of prison might influence the outcome, so you run this program in a maximum, a medium, and a minimum security prison. A year after each prisoner is paroled, you contact their employers and collect ratings of Employer Satisfaction toward the parolee on a scale of 1 to 100. You are especially interested in whether there is an interaction effect (e.g., "work for money" might work better at the maximum security prison, while "no work" better at minimum security).

Reasoning: There is 1 group of subjects. For each parolee you have 2 independent variables (Money/No money, Max/Med/Min type prison) and 1 dependent variable (Employer Satisfaction Rating). There may be an important interaction effect of the independent variables.

Example 2: You think that mixing genders on dorm floors will raise Enjoyment of College Life ratings by students. You think the strictness of the Resident Advisor, and the gender of the student who's answering your questionnaire may also influences these ratings and that there could be an interaction effect (e.g., strict advisors on mixed floors might be related to high ratings, while lenient advisors on one-sex floors bring high ratings; females might prefer strict advisors while males like lenient ones). You know the sex of the respondent, have pre-divided the R.A.s into high, medium, and low strictness, and find out which students are on Mixed- and which on Same-sex floors. You obtain ratings from each student on "Enjoyment of College Life".

Reasoning: There is 1 group of subjects, each of which falls into one place on the 3 independent variables (Type of RA, Type of Floor, Gender of Respondent). You want to know if any of these three influence the 1 dependent variable (ratings on the Enjoyment of College Life scale) and whether there might be a significant interaction among the independent variables.

HOW TO USE THIS TEST

Enter your data into MINITAB as described on Page 5. Answer these three questions:

Which column has the data you want to analyze: C____.

Which column has the SUBSCRIPTS numbers indicating the first subject grouping: C____.

Which column has the SUBSCRIPTS numbers indicating the second subject grouping: C____.

A. Type NAME C____ 'PREG' C____ 'SCHOOLS' C____ 'AMTCOUNS'<RET>

B. Type PRINT C____ C____ C____ <RET> Check to be sure your data is correct. If not, quit Minitab, and correct

C. Type GLM for data in C____ = subscripts in C____ and C____ and interaction of C____* C____ <RET>. The actual command typically is GLM C1 = C2 C3 C2*C3, with data in C1 and the subscripts in C2 and C3; the equal sign is required in the command, while the star indicates you want to see if there's a significant interaction.

A WORKED OUT EXAMPLE

You want to see if amount of counseling (once a month, once a week, and daily) have any effect on the number of pregnancies by teenagers in schools. You also think there might be different numbers of pregnancies based upon the economic level of the school, and that there might be an interaction effect if the results of the amount of counseling differ depending on the economic level of the school. Note: If you were only interested in one factor (just the counseling, for example), use the Oneway Analysis of Variance.

You collect the following data from six schools in each sub-category and code them as follows:

ECONOMIC LEVEL OF SCHOOL: Middle-income = 1, Lower-income = 2

AMT OF COUNSELING: Once/month = 1, once/week = 2, daily = 3.

# of Preg Sch Couns

12 1 1

14 1 1

9 1 1

13 1 1

16 1 1

9 1 1

23 2 1

32 2 1

19 2 1

26 2 1

20 2 1

20 2 1

8 1 2

9 1 2

3 1 2

5 1 2

9 1 2

2 1 2

12 2 2

14 2 2

16 2 2

18 2 2

22 2 2

13 2 2

16 1 3

9 1 3

23 1 3

10 1 3

21 1 3

17 1 3

20 2 3

16 2 3

17 2 3

12 2 3

14 2 3

18 2 3

MTB > NAME C1 'PREG' C2 'SCHLS' C3 'AMTCOUN'

MTB > PRINT C1-C3

ROW PREG SCHLS AMTCOUN

1 12 1 1

2 14 1 1

3 9 1 1

4 13 1 1

(Note: I've omitted some scores to save space)

36 18 2 3

MTB > GLM C1 = C2 C3 C2*C3

Factor Levels Values

SCHLS 2 1 2

AMTCOUN 3 1 2

ANALYSIS OF VARIANCE FOR PREG

Source DF SeqSS AdjSS AdjMS F-Ratio P-Value

SCHLS 1 210.25 210.25 210.25 11.00 0.002

AMTCOUN 2 73.56 72.56 36.78 1.92 0.163

SCHLS*AMTC 2 194.00 194.00 97.00 5.08 0.013

ERROR 30 573.17 19.11 19.11

(Note: "Error" here is a stat term, not really an error!)

TOTAL 35 1050.97

Meaning of the Results:

The SCHLS F-Ratio = 11.00 P = 0.002; this probability value is less than 0.010. The difference

in pregnancies due to the types of schools is therefore very significant.

The AMTCOUN F-Ratio = 1.92 P = 0.163; this probability value is greater than 0.050.Any differences

due to amount of counseling received are therefore not significant.

The SCHLS*AMTCOUN (interaction) F-Ratio = 5.08 P = 0.013; this probability value falls

between 0.050 and 0.010. It is significant although not very significant. There is a significant

interaction between the type of school and the amount of counseling on the pregnancy amount.

The School difference was found to be very significant so you can be pretty sure that the economic level of the school makes a difference in the number of pregnancies. You could now do T-Tests for Independent Groups on pairs of schools to find which ones causes the significant difference that was found.

The Amount-of-Counseling differences are not significant; therefore, amount of counseling by itself is not a factor in the number of pregnancies. You should not test further between the means of the different counseling groups, since the variable is not significant and a difference found could be just due to random factors

The Schools by Amount-of-Counseling Interaction is significant. This means that with one amount of counseling teenagers from low-economic schools do better, with a different amount of counseling teenagers from middle-economic schools do better. You have to examine the data to see which is which. In our case, the data appears to indicate that once a month counseling works best for middle class girls and once a week for lower-class girls. Or perhaps try vasectomies on the guys…

==

===

If none of these tests fits what you did, see your professor for help...

STEP 4: PRINTING OUT THE RESULTS FROM THE FILE MINITAB.RESULTS

When you are finished with your MINITAB testing, use CTRL-Z to go back to the VAX. Once you have the $ prompt again, type E MINITAB.RESULTS <RET>. If you followed the instructions and had typed OUTFILE 'MINITAB.RESULTS' at the beginning of the MINITAB session, a copy of everything that appeared on the screen during the session will appear on your VAX screen. Clean up these results by labeling rows and columns in tables or charts; do whatever else will make the meaning of your results clear. Save the changed information by using CTRL-Z.

Computer procedures for printing a VAX-file often change. Ask the lab assistant how to print the VAX-file MINITAB.RESULTS from your account and follow the directions given. After it's printed, examine the print-out; if it is too light to be easily read, ask the assistant to put a new ribbon on the printer and reprint the file. If you're satisfied with the printing, completely finished, and the $ is on the screen, type LOGOUT to close your session.

===
===

SOME USEFUL MINITAB COMMANDS NOT MENTIONED PREVIOUSLY

LET: To correct a single (or a few) errors in data entry while in Minitab, note down the column and the row of the number you want to replace. Suppose you see that you put a 1 instead of a 10 in Column 3, row 4 of your data. Type LET C3(4) = 10 <RET> The parentheses around the row number must be included. After the LET command, Column 3, Row 4 will contain a 10 instead of whatever was there.

CODE: To change data values to other values in one or more columns, decide on what numbers you want changed and what they are to be replaced with. Suppose you've entered the subjects' ages in terms of years in Column 5. You now want to collect all the ages into three groups: 18-25 in one group, 26-50 in another group, and 50-75 into a third group, so you can run a test comparing the effects on test scores of the different age groupings. You want to put this new information into Column 10. Type CODE (18:25) as 1, (26:50) as 2, (51:75) as 3 from column C5 and put the new data into C10<RET>

Actual command --> CODE (18:25) 1 (26:50) 2 (51:75) 3 C5 C10 <RET>

STACK: Just as UNSTACK breaks columns apart, STACK puts columns together. If you want to join Columns C3 and C4 into a new column C10 type STACK C3 onto C4 and put the result into C10<RET>

Actual command --> STACK C3 C4 C10 <RET>

ADD, SUBTRACT, MULTIPLY, DIVIDE: You can do all the arithmetic commands by using the word you want. For example, you could ADD each number in column C3 to its corresponding number in Column C5 and put the results in C20

Actual command --> ADD C3 C5 C20 <RET>

TABLES OF CHI-SQUARE, SPEARMAN, & PEARSON CORRELATION

CHI-SQUARE PROBABILITY TABLE

To be significant, your Chi Square number must be EQUAL TO OR GREATER than the listed number

d.f. 0.05 0.01

1 3.84 6.64

2 5.99 9.21

3 7.81 11.34

4 9.49 13.28

5 11.07 15.09

6 12.59 16.81

7 14.07 18.48

8 15.51 20.09

9 16.92 21.67

10 18.31 23.21

11 19.68 24.73

12 21.03 26.22

13 22.36 27.69

14 23.69 29.14

15 25.00 30.58

16 26.30 32.00

17 27.59 33.41

18 28.87 34.81

19 30.14 36.19

20 31.41 37.57

21 32.67 38.93

22 33.92 40.29

23 35.17 41.64

24 36.42 42.98

25 37.65 44.31

26 38.89 45.64

27 40.11 46.96

28 41.34 48.28

29 42.56 49.59

30 43.77 50.89

If you have more than 30 df -- see your Instructor

SPEARMAN'S RANK ORDER CORRELATION PROBABILITY TABLE

To be significant, your Correlation number must be EQUAL TO OR GREATER than the listed number

d.f. 0.05 0.01

5 1.000 ---

6 0.886 1.000

7 0.786 0.929

8 0.738 0.881

9 0.683 0.833

10 0.648 0.794

12 0.591 0.777

14 0.544 0.715

16 0.506 0.666

18 0.475 0.625

20 0.450 0.591

22 0.428 0.562

24 0.409 0.537

26 0.392 0.515

28 0.377 0.496

30 0.364 0.478

(If your d.f. is more than 30, see your teacher)

PEARSON PRODUCT MOMENT TABLE

To be significant, your Correlation number must be EQUAL TO OR GREATER than the listed number

df	0.05	0.01
1	.997	1.00
2	.950	.990
3	.878	.959
4	.811	.917
5	.754	.874
6	.707	.834
7	.666	.798
8	.632	.765
9	.602	.735
10	.576	.708
11	.553	.684
12	.532	.661
13	.514	.641
14	.497	.623
15	.482	.606
16	.468	.590
17	.456	.575

18	.444	.561
19	.433	.549
20	.423	.537
21	.413	.526
22	.404	.515
23	.396	.505
24	.388	.496
25	.381	.487
26	.374	.478
27	.367	.470
28	.361	.463
29	.355	.456
30	.349	.440