

# One-way ANOVA



Copyright 1998, 2000 Tom Malloy

**This is the text of the in-class lecture which accompanied the Authorware visual graphics on this topic. You may print this text out and use it as a textbook. Or you may read it online. In either case it is coordinated with the online Authorware graphics.**

**PRINT: You may print this web page on your local printer if you wish. Then you can read the hard copy of the lecture text as you look at the Authorware graphics.**

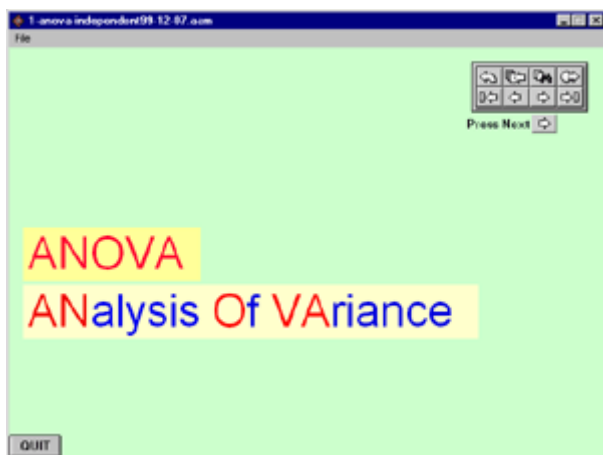
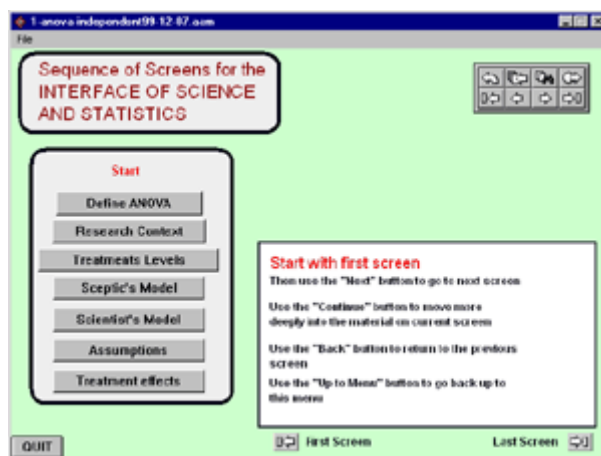
**ANOVA.** Analysis of variance, which is usually shortened to ANOVA, is the most commonly used statistical method for testing hypotheses about 3 or more means. The ANOVA statistic is called the *F-test*, after its developer, Fisher. We use ANOVA when we want to test the null hypothesis ( $H_0$ ) that 3 or more means are drawn from the same population. If we have 2 means, we use the *t-test* which turns out to be just a special case of ANOVA.

Like the *t*, *F* depends on degrees of freedom to determine probabilities and critical values. But there is a difference between *t* and *F* in terms of the degrees of freedom

concept.  $F$  has two different degrees of freedom to calculate. In contrast,  $t$  has only one formula for calculating degrees of freedom.

We will start by describing some kinds of scientific research contexts where you would use ANOVA.

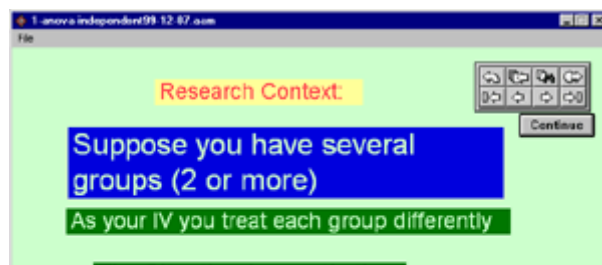
## *The Interface Between Science and Statistics*

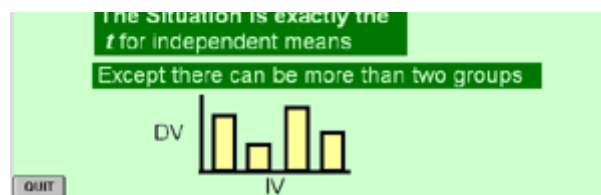


## **ANALYSIS OF VARIANCE**

**ANOVA** stands for analysis of variance. There are many forms of ANOVA, just as there are many forms of t-tests. We are beginning with the simplest form of ANOVA, the "one-way analysis of variance for independent groups."

The one-way analysis of variance for independent groups applies to an experimental situation where there might be more than just two groups. The t-test was limited to two groups, but the Analysis of Variance can analyze as many groups as you want.





**Research Context:**

Suppose you have several groups (2 or more)

As your IV you treat each group differently

**EXAMPLES**

- Each group has a different diet
- Each group is given different memory instructions
- Each group sees a different commercial
- Each group is somehow treated differently

QUIT

**INDEPENDENT GROUPS.** Imagine you would like to do an experiment where there are several independent groups, two or more. The research participants are randomly assigned to the various groups so that we can assume that the groups are independent of each other.

**TREAT EACH GROUP DIFFERENTLY.** Once you have the groups, you treat each group differently. The ways in which the groups are treated differently depends on what your independent

variable is.

**EXAMPLES OF IV'S.** For example, if your IV is Diet, there may be several groups of participants and each group eats a different diet. Or, if your IV is Memory Strategy, the several groups of research participants might receive different memory instructions. Or, if your IV is Type of Advertisement, the several groups might see different political campaign commercials. In general, each of several groups is treated differently in some manner that corresponds to the IV.

This is a generalization of the t-test for independent means in which there are only two groups. In ANOVA there can be three or four groups, or as many groups as you want. So the experimental context for ANOVA is just an extension of one for which the t-independent is appropriate. Keep in mind that, like the t for independent means, we are assuming independent groups. This is an important distinction. In a later lecture we will describe an ANOVA for correlated groups.

**Example**

The DV is weight

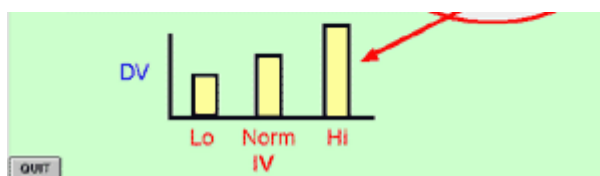
The IV (Treatment) is Type of Diet

- Low Fat
- Normal
- High Fat

These are called Treatment levels or Levels of IV

## Treatment Levels

**DIET EXAMPLE.** Suppose that you want to study the effect of dietary fat on weight when calories are held constant. Suppose your research participants are Congolese males in their late twenty's.



You take your pool of participants and randomly divide them into three groups. All groups will eat 2000 calories per day. The Low Fat group will get only 10% of its 2000 calories from fat. The Normal

group will get 30% of its 2000 calories from fat. The High Fat group will get 50% of its 2000 calories eating from fat. After four months, you measure the weight of all participants.

**INDEPENDENT VARIABLE.** The IV is type of diet. More precisely, the IV is the percent of a person's calories that come from fat.

**DEPENDENT VARIABLE.** The DV is weight in pounds.

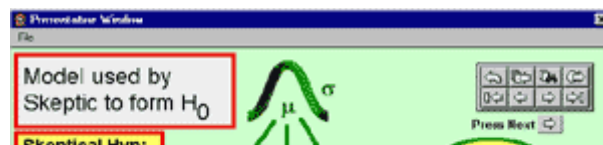
New Jargon: Levels of the IV
<b>LEVELS OF THE IV.</b> In this study the IV has three levels of fat content--low (10%), norm (30%), and high (50%). A different study could have four levels, say 10%, 30% 50% and 60%.
<b>TREATMENT LEVELS.</b> An alternative form of jargon uses the term "treatment levels" instead of "levels of the IV."
In most discussions of ANOVA the term "Treatment" is used interchangeably with "Independent Variable." I will follow the tradition in ANOVA and use the term treatment when referring to the IV. Thus "treatment effects" are to be understood as meaning the "effects of the IV."

## Three Populations?

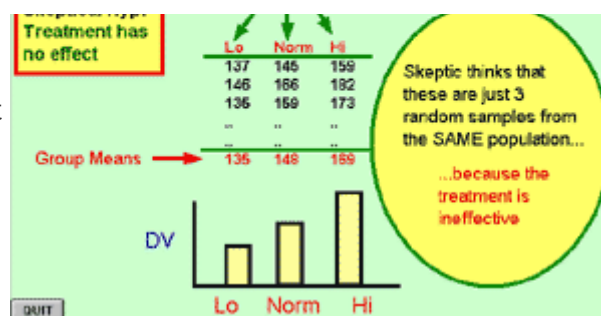
We are once again going to call on your experience with the Double Sample Tool and the Detect Difference Game to understand statistical theory. This theory is just an extension of the logic we laid out in the Independent-t lecture.

In the Detect Difference lecture (based upon the Double Sample Tool and Detect Difference Game) we learned that two groups of data might each be separately drawn from two different populations or they might both be drawn from a single population. We will now extend that logic to three groups of data which might be sampled from three different populations or which might just be three samples from one population.

*Model used by the  
Skeptic to Form  $H_0$*



The skeptic does not believe in the IV; the skeptic does not believe that percent of calories from fat affect body weight. The skeptic believes all that matters is the total number of calories eaten not where the calories come from.

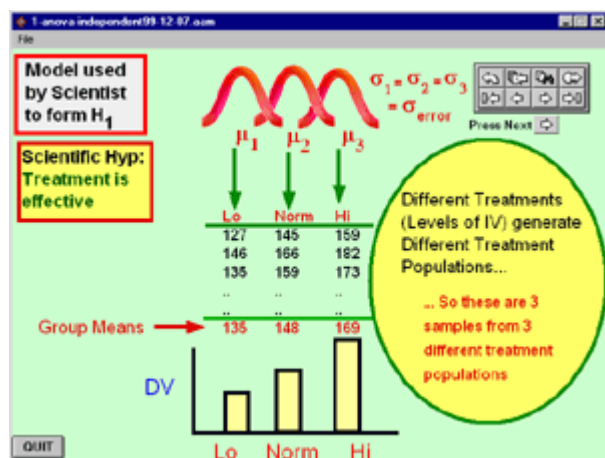


ONE POPULATION. Consequently, since all three groups in your study eat the same number of calories, the skeptic thinks that the weights of participants in all three groups are just three samples from the same population. Any differences between the weights in the groups are due to chance alone.

Based on this model (that we have three samples from one population) the skeptic will form the null hypothesis. We will explicitly state  $H_0$  later in the lecture after we've developed a few more foundation ideas which are essential for understanding  $H_0$ . For now, we just want to point out that the skeptic thinks our IV is ineffective and that our groups are all samples from the same population.

UNTREATED, BASELINE POPULATION. We will refer to the single population in the skeptic's model as the untreated or baseline population.

In summary, the skeptic thinks that the IV does not work and there is only one untreated population from which all your groups are sampled. This single normal population, and most especially its  $\mu$ , will be important in equations for defining treatment effects and treatment effect size later in this lecture.



## Model used by the Scientist to Form $H_1$

As the scientist, you are hypothesizing that even with calories held constant, fat content of diet affects weight.

THREE POPULATIONS. Consequently, you assume that each group is sampled from a different population. That is, you assume that people who eat a low fat

diet weigh less as a population than those who eat a normal diet. In turn, people who eat a normal diet weigh less as a population than those who eat a high fat diet.

**TREATMENT POPULATIONS.** As a scientist you are modeling your research with probability theory. Your IV has three treatment levels. You are assuming that each treatment level in your study corresponds to a population. We refer to these populations as "Treatment Populations." This is nothing complicated. In creating a statistical model of your research, you assume there is a Treatment Population to go with each of your groups.

**THE ALTERNATIVE HYPOTHESIS.** Based on this model (that there are three treatment populations) you will later state H1. But we will get to an explicit statement of H1 further along in the lecture.

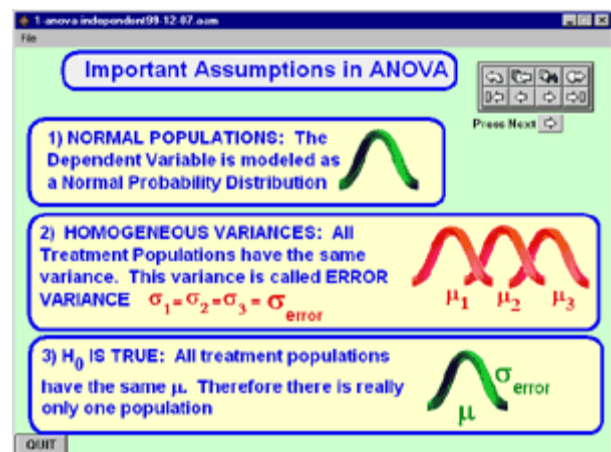
Switching Models
Keep both of these probability models clearly in mind. Visualizing each model in a simple mental picture may help keep them in mind. We will switch back and forth between models as we talk about various topics in this lecture. Having the two models mind and knowing when you are using one rather than the other can make understanding the ideas in this lecture much easier.
Most Notable Switches in Models
Assume Skeptic's model. When we eventually talk about setting up the logic of the test of $H_0$ and when we talk about statistical conclusion validity and significance we will be assuming that the skeptic's model is true.
Assume Scientist's model. When we talk about treatment effects and treatment effect sizes we will be assuming that the scientist's model will be true.

## Assumptions for ANOVA

**NORMALITY ASSUMPTION.** As with most common inferential statistics we assume that the DV can be modeled as a Normal Population.

**HOMOGENEITY OF VARIANCE.** As we've mentioned, we assume that the sigma's of any populations in our model will be equal.

**NULL HYPOTHESIS.** A very general way to test a model is to assume it is true and then examine its predictions. If its predictions are absurd in the light of data, we tend to reject the model. If its predictions are sensible given the known data, we tend to keep the model.



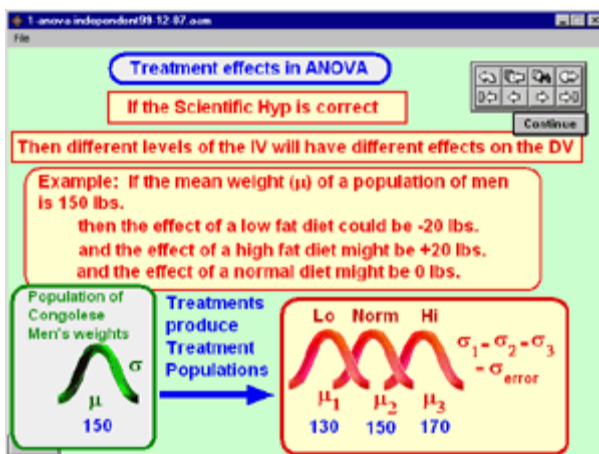


We are testing  $H_0$ . The logic of  $H_0$  testing is to build our statistical model assuming  $H_0$  is true and then examine the consequences of that assumption in light of the sample data. If the sample data are extremely unlikely ( $p < .05$ ) then we reject  $H_0$ . If the sample data are reasonably likely given  $H_0$  ( $p > .05$ ), then we don't reject  $H_0$ .

### BASELINE POPULATION

In order to define treatment effects and effect size we will use the single normal population proposed in the skeptic's model. This is the green population in the various graphics. The mean ( $\mu$ ) of the untreated (or baseline) population will be important in equations defining treatment effect size.

The variance (sigma squared) of the baseline population is typically called error variance. We will discuss why later.



### Treatment Effects

The topic of "treatment effects" should be a review and an extension of what we covered in the Detect Difference and Independent-t lectures.

**ASSUME SCIENTIFIC HYPOTHESIS IS RIGHT.** To think about treatment effects we must switch from the skeptic's model to the scientist's model by assuming that percent of dietary fat does affect

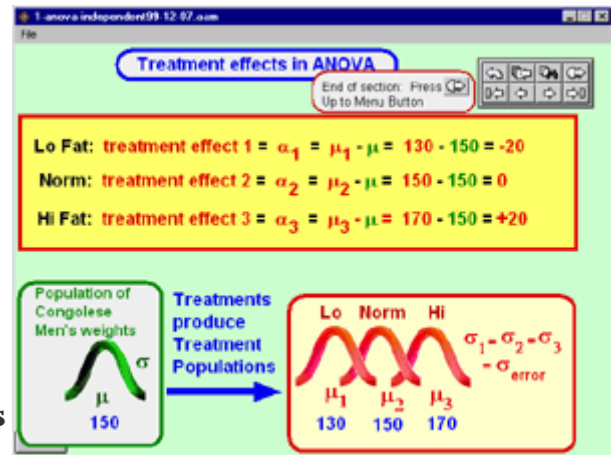
body weight, even with calories held constant. In this case, then a good probability model of the experiment is that the data in the three groups is sampled from three treatment populations (scientist's model).

**TREATMENT EFFECTS.** Assume that the normal weight of Congolese males eating a normal diet is 150 pounds (Green Distribution or Baseline Population). Then (assuming the scientific hypothesis is true) a low (10%) diet should affect weight by lowering it from that baseline. And a high (50%) fat diet should affect weight by raising it from that baseline. And a normal (30%) diet should have no effect on weight because it is the same as the baseline assumption.

## Treatment Effect Sizes

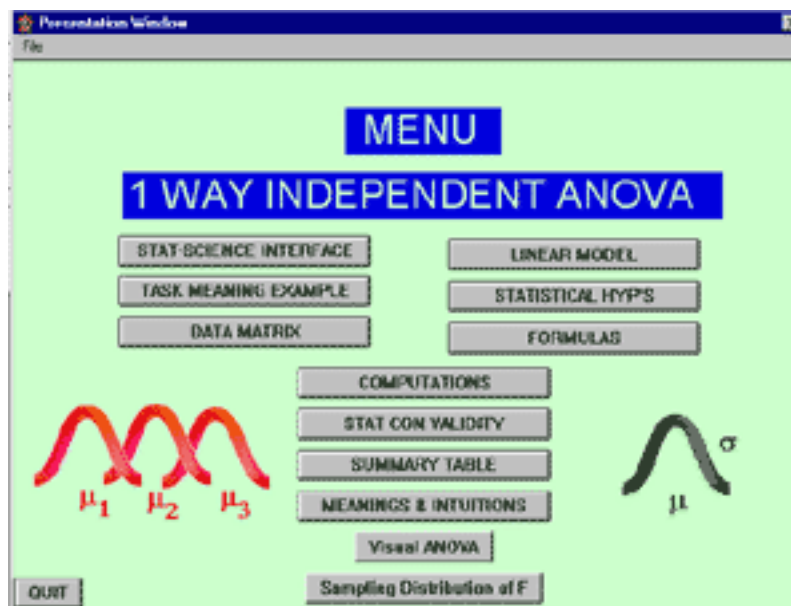
The size of the treatment effects for each treatment population will be the difference between the  $\mu$  of the treatment population and the  $\mu$  of the untreated baseline population shown in green.

We have assumed in our model that the  $\mu$  of the untreated (green) population is 150 pounds. We get the treatment effect size for each treatment population by subtracting the  $\mu$  of the green population from the  $\mu$  of each of the treatment populations.



The current graphic shows explicit examples of calculating treatment effect size. Make these calculations for yourself and write them down in your notes.

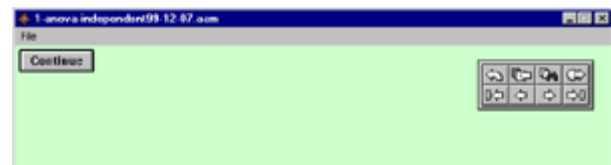
Now let's let all this theory settle for a bit and go on to a scientific example which we will later use to learn how to calculate the F statistic.



## Task Meaning Example

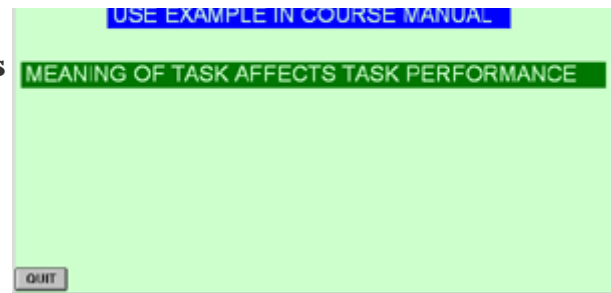
### Details of a Study

**EXAMPLE:** (This paragraph of text is printed in your class notes outline, so



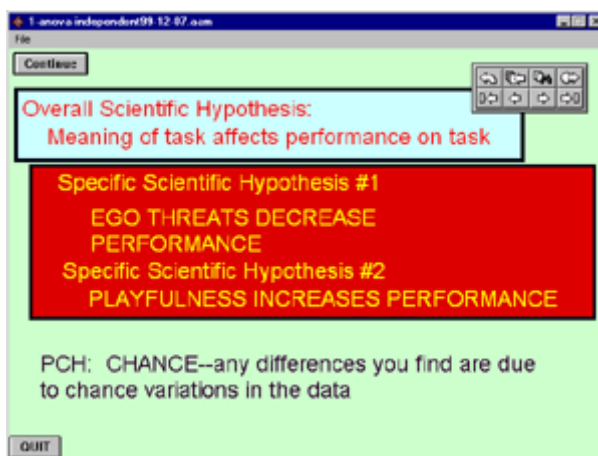


you don't have to write out the details, just understand them.) A scientist thinks that the meaning a person gives to a task will effect performance on that task. More specifically, she believes that ego-involving, stressful, and negative attitudes toward the task will produce poor performance, while fun-oriented, relaxed, and positive attitudes will produce good performance. She takes 30 college students and randomly breaks them into three groups. Two students drop out of one group and one drops out of another group, so she ends up with 27 participants. She gives them geometric puzzles to solve. The Neutral group is given matter-of-fact instructions to solve the puzzles. The Positive group is given instructions which interpret the task as a marketing survey to discover the puzzles which are the most fun to solve. The participants are told to solve the puzzles as best as they are able, to have fun, and finally, to rate how much fun each puzzle is. The Negative group is told that the puzzles indicate basic intelligence in high-school students and that as college students they should be able to solve them all easily. High performance is expected, and any failure to solve the puzzles indicates some problem. Her performance measure is number of errors made in solving the puzzles.



**OVERVIEW.** A researcher is interested in evaluating the general hypothesis that *task meaning affects performance*. More specifically, she wants to evaluate two forms of this general hypothesis. The first hypothesis is that a playful (positive) task meaning will improve performance. The second is that an ego-threatening (negative) meaning will impair performance.

## Experimental Design



As we said, this example deals with how the meaning of a task (IV) affects task performance (DV). This is a very general and an extremely abstract statement of the variables. Consequently, we need some knowledge of how we are going to operationalize the IV and DV so that they have specific, concrete meanings for our study.

**REVIEW OF RESEARCH STUDY.** All participants will solve a series of interesting but very difficult puzzles.

The number of errors they make in solving these puzzles will be the DV. The IV will be defined by the instructions given to three different groups of

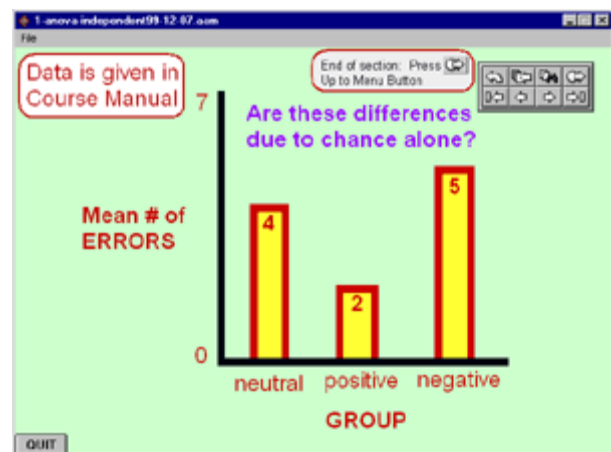
randomly assigned participants. One group is given neutral instructions, one is given positive instructions, and one given negative instructions.

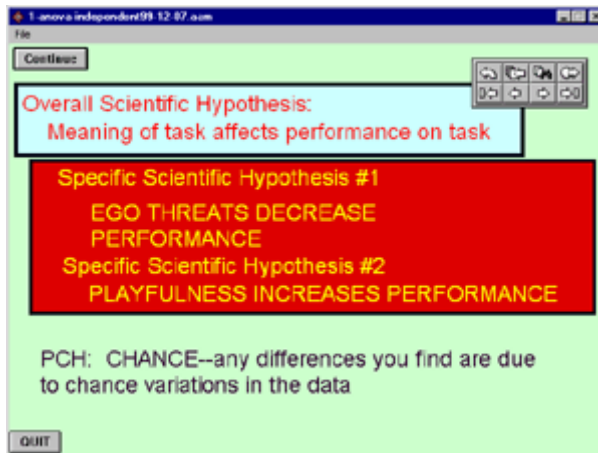
The negative instructions tell the participants (college students) that they are going to solve *"a series of puzzles which have been found to be easy for high school students. This set of puzzles is being developed to predict who should or should not be admitted to college"*. Remember, the puzzles are actually very difficult. Presumably the difficulty of the puzzles will threaten the self-evaluation of participants in that group. As people begin to fail to solve these difficult puzzles, they might well think *"Wow, high school students solved these easily."* Thus, these instructions are likely to make solving the puzzles a negative experience.

The positive instructions given to the participants are: *"We're a marketing company and we want to market some puzzles to people who like difficult puzzles. We want to find the puzzles that are the most fun to do, so please work on the following puzzles and solve each puzzle (if you can). When you're done with the puzzles we want you to rate how fun each puzzle is."* The neutral group instructions are, *"We're studying a series of puzzles, please solve these."* These neutral instructions don't give any particular meaning. That's as much detail as we will go into here on the operational definitions of the IV and DV.

**NOTE:** Remember, it is up to you to critique the operational definitions using logic and common sense. These operations may or may not make any sense. Statistical procedures just take the numbers that come out of the DV operations and analyze them to evaluate the PCH of chance. Whether the study has any merit to it or not is beyond statistics, but it is not beyond the critical thinking of either researchers or consumers of research.

**DATA PATTERN.** The next screen shows the results found by the researcher. Assume that there were seven puzzles and by errors we mean the number of puzzles that are not solved. The researcher's general hypothesis was that the meaning of the task would affect performance. The data pattern fits this general hypothesis since different groups that were given different task instructions seem to perform differently.



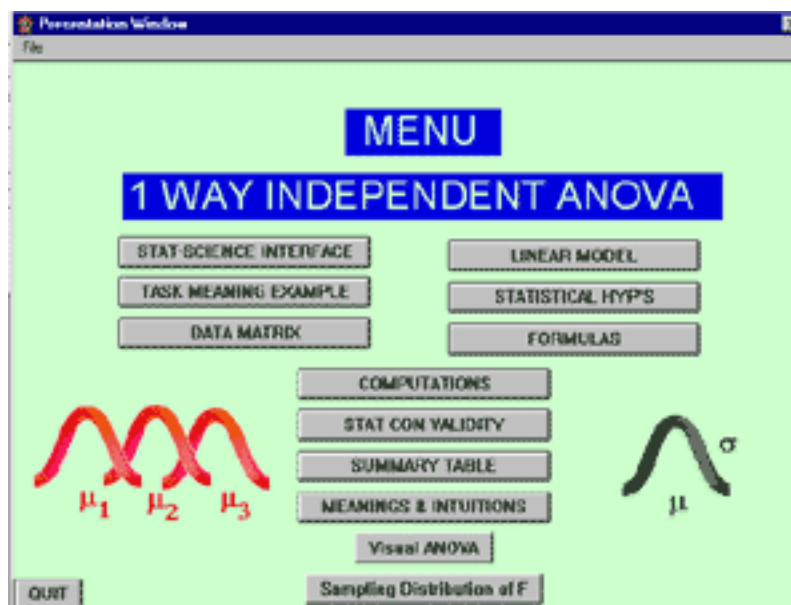


**SCIENTIFIC HYPOTHESES.** The researcher also had some sub-hypotheses. One was that positive instructions, such as playfulness, will improve performance. The data seems consistent with that since that group averaged only 2 errors as compared to the neutral control group at 4 errors. The other sub-hypothesis was that negative (or ego-threatening) task meaning would impair performance. The negative group averaged 5 errors compared to the neutral group's 4, so

the data pattern is consistent with that scientific hypothesis too.

**PCH OF CHANCE.** The skeptic would say, "*Well, the data pattern does fit your hypotheses but maybe the data pattern is entirely due to chance.*" So the first thing that we have to do before we even get into an intelligent conversation about what this all means is to do some statistics to evaluate the PCH of chance. So as usual, statistics don't do anything very smart, they just address a relatively trivial but persistent and very general objection to the results (that they might be due to chance alone).

Now let's let the scientific concepts settle for a bit and go on to the details of formulas and number crunching.



## *The Data Matrix*

The formulas for the Analysis of Variance are long and complicated. To make them manageable, we will break them into small pieces, and then, when we've calculated the pieces, we will put them together into a whole called the ***F ratio***.

Normally, we would use the computer to calculate these long statistical analyses. However, for the homework I will ask you to do one problem requiring a calculation of a one-way ANOVA by hand. If you skip ahead to the "Some Sums" screens, you will be given (relatively) simple, functional instructions for doing the arithmetic involved in ANOVA.

The following set of screens is a quick and incomplete review of the symbols and procedures used in double summation. You may or may not need this review. The way that the data matrix looks should look familiar if you worked on the optional double summation class materials earlier.

## OPTIONAL REVIEW OF DATA MATRIX SYMBOLS

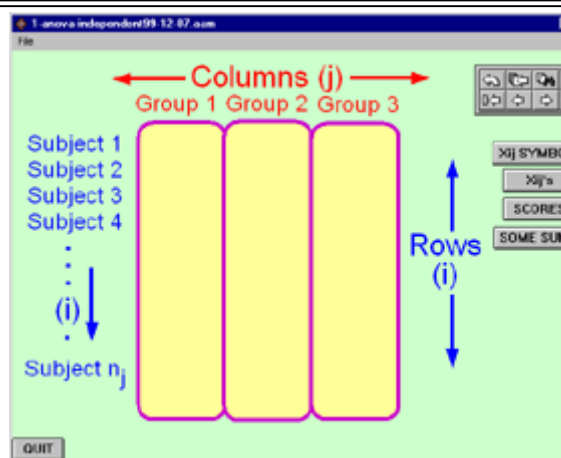
A matrix is a highly abstract mathematical object which might represent many things. For our use, just think of the matrix as a bunch of symbols (often numbers) organized into rows and columns.

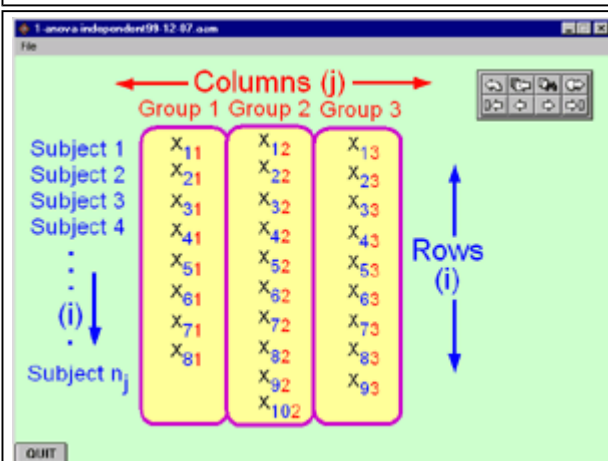
**COLUMNS.** What goes across the top of the matrix is usually the IV and its levels (i.e., the names of the groups in the study). In other words, the columns of the matrix are the groups in the study.

**ROWS.** What goes down the rows within each column are the measurements made on each participant in a group.

In our specific example the names of the columns are "Neutral," "Positive," and "Negative." In the general case, shown on the screen, we refer to our groups (columns) as "one," "two," and "three." The index for columns is "*j*." That means that when we want to refer to a specific group but we don't know which one, we'll call it "group *j*". In our example, *j* can vary between 1 and 3, since we have three treatment groups.

Participants within a specific group are indexed by "*i*." That means if I want to talk about a specific participant in, say, group 1 but I don't know which one, I can refer to "**participant *i*** in group 1."





Look at the symbolic data matrix. Notice that in the one-way ANOVA we have several Participants in each of several groups. So we want to know the score of any individual participant we need an address for the score that specifies both the group and the participant within that group.

We need a group identifier (like a family or last name) and then we need an individual identifier (like a first name). To locate a person in a town, we first need to know a last (family) name; is this person a Jensen

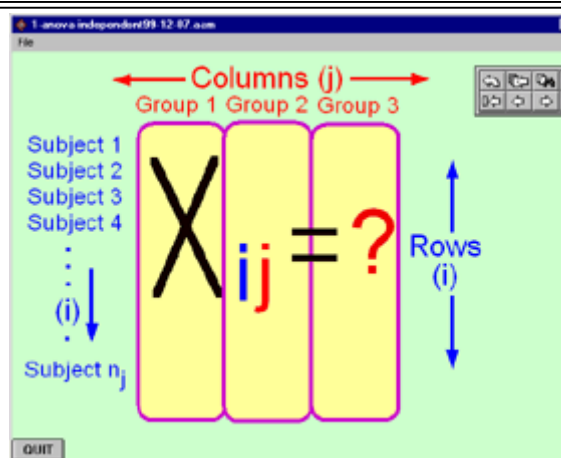
or a Sanchez or a Chang? Once we know the person's family we need to know specifically which family member. If the last name is Sanchez, we need to know if the person we are seeking is Maria or Veto or Jorge or Lisa. To identify a particular person we need both a first and last name, for example, Lisa Sanchez

On the graphic, each Participant's score is indicated by an X. But there are many Participants in many groups. So each X has to have a first name (i) and a last name (j). So on the graphic you see that the X's are all doubly indexed (X11, X21, X31, etc.) The "i" or row index (first name) is shown in blue. The "j" or column index (last name) is shown in red.

In this kind of ANOVA it's okay for the groups to have a different number of Participants. Here for instance, there's eight in the first group, ten in the second group, and nine in the third group.

The same is true in the one-way ANOVA. We need a first and last name. We will use a symbol like  $X_{ij}$ . In this case "i" is the place marker for "first name" and "j" is the place marker for "last name." If I say "Go find (Person first name last name)", you will reply that you can't because you need to know what specifically is this person's first and last names. If I say "Go find (Person Gary Jensen)" then you will be able to do it.

"First Name" and "Last Name" are general markers found on nearly every form you fill out. However, first name and last name need specification before they can point to a particular person. Wherever you see "First Name" and "Last Name" on a form, you put in your particular name. In the same way  $X_{ij}$  is a general way of





referring to any participant in the study without saying which one. If you want to find a specific participant then you have to specify  $i$  and  $j$ . You have to say  $i = 3$  and  $j = 2$ ; then you know you want the score from the third participant in the second group. It is equivalent to changing the phrase "Some person who has a first name and a last name" to "Michael Chang."

**THE  $X_{ij}$  SYMBOL.** We don't always want to talk about a specific person. We don't always want to talk about Betty Goldstein. We often want to talk about a person as a general category, or just people in general. In a data matrix, when we want to talk about some score in general, we say " $X_{ij}$ ." It's the score of  $i$ th participant in the  $j$ th group.

	Group 1	Group 2	Group 3
Subject 1	3	2	5
Subject 2	4	3	7
Subject 3	5	4	6
Subject 4	6	2	4

## THE DATA

The actual data shown here is in your course notes so you don't have to rewrite it. The scores in the data matrix are the actual number of errors that were generated by each participant in the various groups, where group one is neutral instructions, group two is positive instructions, and group three is negative instructions.

**TREATMENT MEANS.** The mean of each of the groups is often referred to as a **treatment mean or a group mean**. For example, it makes sense to call the mean of group 1 the neutral instruction treatment mean because everyone in this group was given the same treatment that is, they were all given neutral instructions. Referring to the graph we saw way back in this lecture, the group mean for the first group is 4.

Referring back to the same graph, the treatment mean for group 2 is 2, and the treatment mean for group 3 is 5.

In general, we can symbolize the mean of any treatment group as  $M_j$ .

**GRAND MEAN.** The **grand mean** is the mean of all of the scores combined without regard for their treatment group. If you wanted a grand mean, you would just add up all these scores in the entire data matrix, and divide by the total number of scores that are in the data matrix. The Grand Mean is often symbolized as an  $M$  with a subscripted  $G$  (which is hard to write in html at this time). The Grand Mean is also often symbolized by " $M_{..}$ " (that is, an  $M$  with two dots after it).

Now we have more than one kind of mean that we're interested in. We need the mean for each treatment group (treatment means), and we will need to have the mean of all of the data for the entire data matrix (grand mean).

These aren't hard concepts, they're just ways of distinguishing and pointing at



different aspects of the data.

**SPECIFIC SCORES.** It should be clear what  $X_{ij}$  means; it's the  $i$ th participant of the  $j$ th group. We can specify any score we want. Using this example, what is the value of the **seventh** score of the **second** group?

To answer the question, you should be able to go back to your data matrix to find the correct group (second) and then look down that column to the right row (seventh) and finally notice that the correct score is 3.

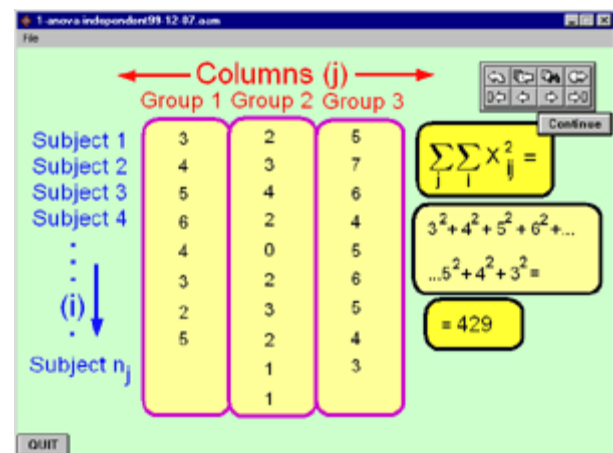
You might want to practice some more. If I asked for the first participant in the third group, that person's score would be 5. That is,  $X(1,3) = 5$ .

What is  $X(2,3)$ ? What is  $X(3,2)$ ?

**ANSWERS.**  $X(2,3) = 7$ .  $X(3,2) =$

## SOME SUMS

We will now go through the details of 3 summation formulas that are the basis of the calculations in ANOVA. These three summations are actually the only calculations that you have to do. Later in the lecture you will see various complex formulas; but each of those formulas only combines these three sums in some combination or another.



**1. SQUARE FIRST, THEN SUM.** (Refer to the graphic.) This first summation formula means that you square each piece of raw data first and then add all these squared numbers up across all the treatment groups.

Look at the example. We square the score for the first participant in the first group (which is  $3 \times 3 = 9$ ), then square the score of second participant in the first group ( $4 \times 4 = 16$ ) and just keep going until we get to the very last participant in the third group ( $3 \times 3 = 9$ ). Then we add up all those squared scores ( $9 + 16 + \dots + 9$ ).

There's nothing very complicated about that. It's just a matter of knowing what the summation symbols mean. This first formula simply asks you to "Square things first, then add them up." If you do that then in this example you'll get 429.



**2. SUM FIRST, THEN SQUARE.** (Refer to

	Group 1	Group 2	Group 3
Subject 1	3	2	5
Subject 2	4	3	7
Subject 3	5	4	6
Subject 4	6	2	4
⋮	4	0	5
⋮	3	2	6
⋮	2	3	5
⋮	5	2	4
Subject $n_j$	1	1	3

the graphic.) What this second summation formula means is that you sum all the numbers first and then (once you have a total for all the numbers) square the total and divide by N. N in this case is the total number of scores in the whole data matrix.

In our example, if you were to add up all the data in the data matrix the sum total would be 97. Next the formula asks you to square the total.  $97 \times 97$  is 9409.

Finally you are asked to divide 9409 (the squared total) by 27 (N). This results in a value of 348.481.

### 3. SQUARE THE COLUMN TOTALS.

(Refer to graphic.) What the last summation formula means is bit more complicated.

**First** you add up each column to get a column total. In this example the first column totals to 32. The second column totals to 20, and the third column totals to 45.

**Second**, the formula asks that you take those column totals and square each one which gives you three squared column totals (302, 202, 452 in the example).

**Third**, the formula asks you to divide each squared column total by the number of scores in that column (or treatment group). In the example there are 8 Participants in group one ( $32 \times 32 / 8$ ), 10 Participants in group two ( $20 \times 20 / 10$ ), and 9 in group 3 ( $45 \times 45 / 9$ ). So we divide 32 squared by 8, then we divide 20 squared by 20 and finally we divide 45 squared by 9. In short, the formula asks you to divide the squared total for each group by the number of Participants in each group. Upon doing the arithmetic, we get three numbers (one for each group): 128, 40, and 225.

**Fourth**, the formula asks you to add these three numbers up. When we do that in this particular example we get  $128 + 40 + 225 = 393$ .

So the third summation formula gives you instructions to add up each column, square the total in the column, divide by the number of scores the

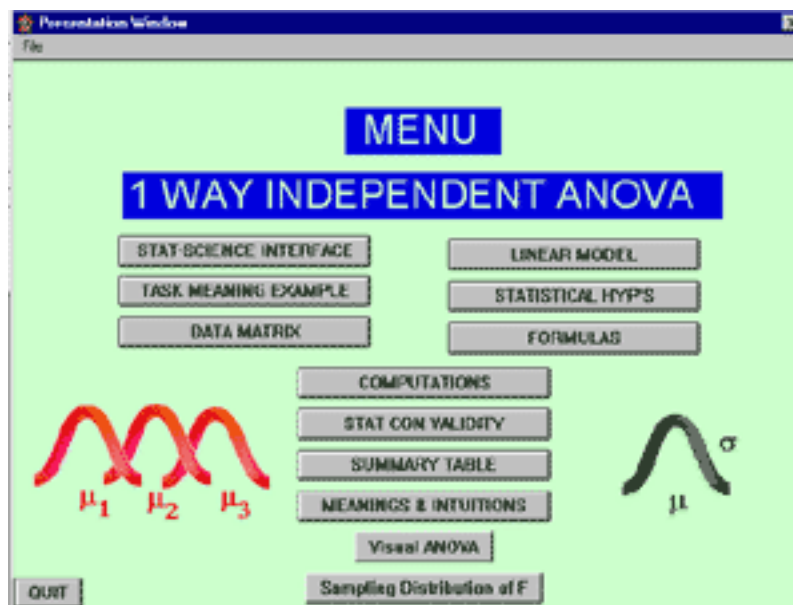
	Group 1	Group 2	Group 3
Subject 1	3	2	5
Subject 2	4	3	7
Subject 3	5	4	6
Subject 4	6	2	4
⋮	4	0	5
⋮	3	2	6
⋮	2	3	5
⋮	5	2	4
Subject $n_j$	1	1	3

$\sum x_{ij} = 32 \quad 20 \quad 45$

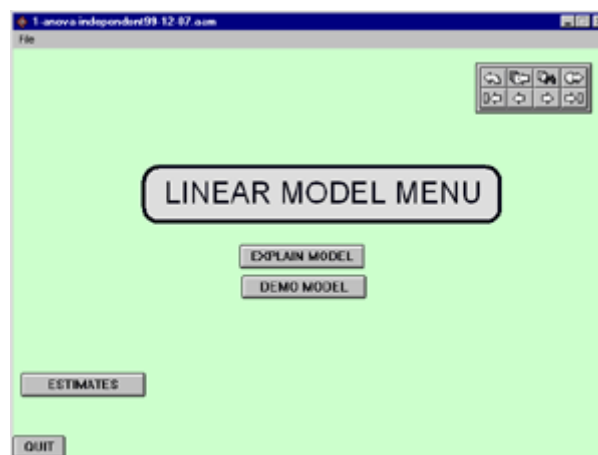
column, and then add all the resulting numbers across all columns.

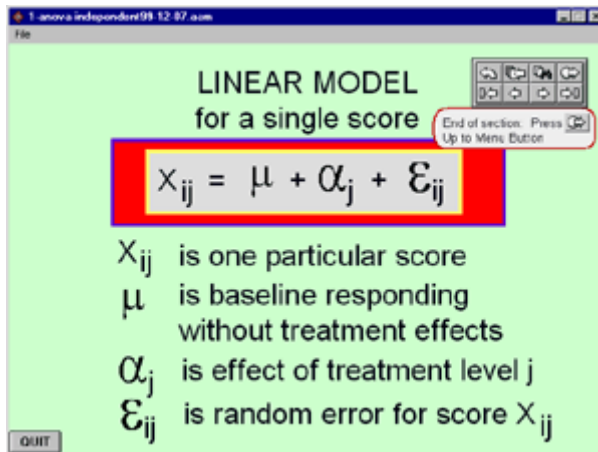
**MAIN POINT FOR 3 SUMMATION FORMULAS.** The main thing is that you should be able to look at these three summation formulas and know what you're supposed to do. Those are the only three sums that we're going to have to calculate in order to perform a simple one-way ANOVA for independent groups.

Now we will let all these details of formulas and number crunching settle and return to the theory of the our probability model.



## The Linear Model





## SCIENTIFIC MODELING

$X_{ij}$  is our symbol for any score in the data matrix.

In terms of the probability model, how does any score,  $X_{ij}$ , take on a specific value?

One of the most creative and interesting aspects of science is building models of natural processes. Many of these models are mathematical equations. The

simplest form of equation is the linear equation. (An equation is linear if the important terms, or variables, are only raised to the first power. Notice on the graphic that none of the terms in the equation are squared or raised to higher powers.) Because of the simplicity of linear equations, many scientific models are linear. [We have already learned about simple linear equations when we studied regression.]

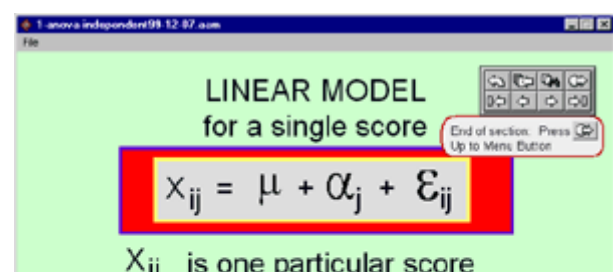
The linear model we will learn introduces you to a whole class of linear models in science and statistics. Statistically it is an important concept. The ideas you will learn about this particular linear model are fundamental to many statistical procedures beyond those covered in this course. But even more than its relevance to statistics, what you learn about linear modeling will introduce you to a very general form of scientific thinking.

[NOTE: Linear models are certainly not the only useful types of models in science. For example nonlinear models are fundamental to the emerging disciplines of chaos, dynamic systems, and complexity theory. In fact, one of the major meta-questions about the kinds of statistics we are learning in this course is how well they can be applied to dynamic systems.]

So in this section we are studying a specific case of very general issues in science. Understanding how this linear model underlies statistical procedures helps you understand important issues in science and modeling. This is worthwhile in itself. And it will be essential to a real understanding ANOVA.

## AN EQUATION FOR ASSIGNING A VALUE TO ANY SINGLE SCORE, $X_{ij}$ .

Recall that a single data point is represented by the symbol  $X_{ij}$ . The linear model (see graphic) is an equation which specifies one theory of how a



specific score takes on a specific value.

Write the linear equation down in your course notes. In case you want to say it in words, you can say, " **$X_{ij}$**  is equal to  **$\mu$** , plus  **$\alpha_j$**  plus  **$\epsilon_{ij}$** ." In this web text I will generally use the small English "e" for epsilon.

$\mu$	is baseline responding without treatment effects
$\alpha_j$	is effect of treatment level j
$\epsilon_{ij}$	is random error for score $X_{ij}$

In this model,  $X_{ij}$  is one particular score, and what we're going to do is build a model of what the score is conceptually in terms of probability theory.

**MU.** The  $\mu$  term refers to the  $\mu$  of the (green) baseline population. This is the starting point of all scores. Remember that  $X_{ij}$  is the score that we get when we take DV measurements on the "i-th" participant in the "j-th" group. And recall that through out this course we have typically modeled our DV as a normal distribution. So if our DV is IQ, we model that as a normal distribution with  $\mu = 100$ . In the Normal lecture when the DV was the height of northern European males, we modeled our DV as a normal distribution with  $\mu = 150$  cm. In the case of the weight of our Congolese male research volunteers, we model the DV as a normal distribution with  $\mu = 150$  lb.

In the linear model, the  $\mu$  term is always the  $\mu$  of the baseline, untreated distribution.

**ALPHA j.** The next term in the equation is  $\alpha_j$ . It represents the effect of the j-th level of the IV. Or, "How is the DV affected by Treatment Level j?"  $\alpha_j$  is the effect of the j-th Treatment Level.

Remember the study examining the effects of instructions on puzzle solving performance. We have three groups in which people are given either neutral (group 1), playful (group 2), or negative (group 3) instructions. So the IV is Type of Instruction and the three levels are neutral, positive or negative. Group 2 ( $j = 2$ ) has playful, positive instructions. So  $\alpha_2$  is the effect on puzzle solving of having positive instructions.  $\alpha_3$  is the effect of having negative instructions. And so on.

Everyone in the same group receives the same treatment. So the 3 treatment effects create differences BETWEEN the 3 groups.

**ERROR:  $\epsilon_{ij}$ .** The final term is  **$\epsilon_{ij}$** , or  **$\epsilon_{ij}$** .  $\epsilon_{ij}$  is a random error added to every DV score. Scientists admit that measurements are always imprecise to some degree. There will always be some measurement error, no matter what we are measuring. To model this imprecision, the linear model assumes that there will be a random error added to every DV measurement.

**HOW DOES THE MODEL ASSIGN AN ERROR?** The model assumes that errors are randomly sampled from a normal distribution with  $\mu = 0$ . Some errors

will be positive and make DV measurement larger; others will be negative and make the DV measurement smaller. But over many, many measurements, the errors should average out to 0 because the distribution from which they are being sampled has  $\mu = 0$ .

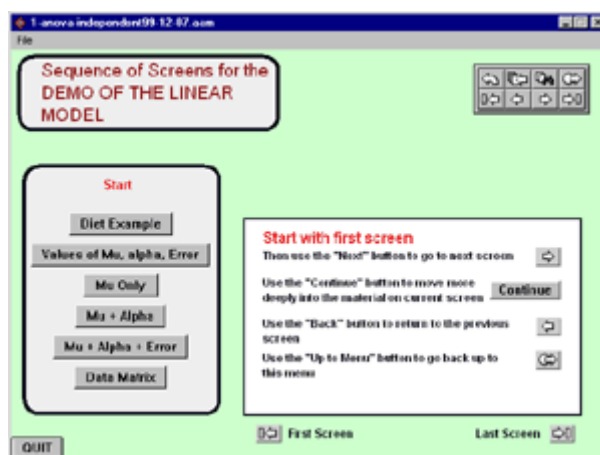
So the linear model asks us to visualize yet another normal distribution. The error distribution is normal with a mean of 0 and a variance called, sensibly enough, error variance.

Error creates differences **WITHIN** groups. Every person in Group 1 is treated identically. But random error makes their DV scores different even though they are treated identically.

Error also creates differences **BETWEEN** groups. All the groups have errors in their measurements and this will make the groups different from each other to some degree.

**SUMMARY.** In the linear model, a score is broken into three parts: 1) the baseline value for our DV measures, 2) an effect resulting from a certain treatment, and 3) a measurement error. This is a pretty sensible way to think a score. A person's score will be a basic value plus some value depending on how they have been treated plus some unexplainable error in the measurement.

## *Demonstrating the Model*

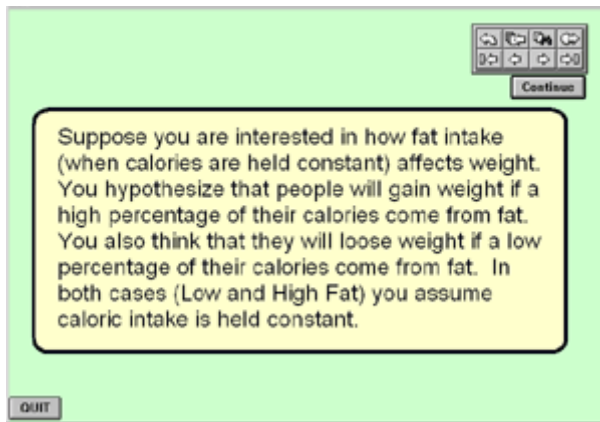


The linear model is an equation that simulates data. We put in values of  $\mu$  and  $\alpha_j$  and  $e_{ij}$  and the equation calculates a score for us. Let's demonstrate how the model works to simulate data that appears realistic.



**DIETARY FAT HYPOTHESIS.** Imagine a





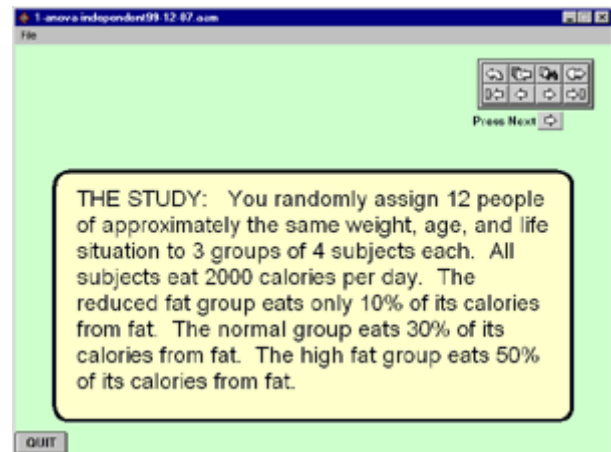
research team is interested in how the percentage of fat in a person's diet affects weight when calories are held constant. A common hypothesis in nutrition right now is that people gain weight if a high percentage of their calories come from fat. Conversely, they might lose weight if a low percentage of their calories came from fat.

## RESEARCH DESIGN

A pool of 12 volunteer Participants is randomly assigned to three groups of four Participants each. Assume that all of the 12 Participants are approximately the same weight, age, and have a similar life style.

All Participants eat 2000 calories per day during the study. The reduced fat group gets only 10 % of its calories from fat, the normal group gets 30 % of its calories from fat, and the high fat group gets 50% of its calories from fat.

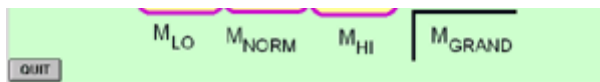
All the groups take in the same number of calories, but the calories are coming from different sources.



	Lo Fat	Normal	High Fat
Subject 1	$X_{11}$	$X_{12}$	$X_{13}$
Subject 2	$X_{21}$	$X_{22}$	$X_{23}$
Subject 3	$X_{31}$	$X_{32}$	$X_{33}$
Subject 4	$X_{41}$	$X_{42}$	$X_{43}$

## DATA MATRIX

So our data matrix might look something like the graphic on the left. We have three groups, low fat, normal, and high fat with four Participants in each group. We will calculate a treatment mean for low fat, and a treatment mean for normal fat, and a treatment mean for high fat, and a grand

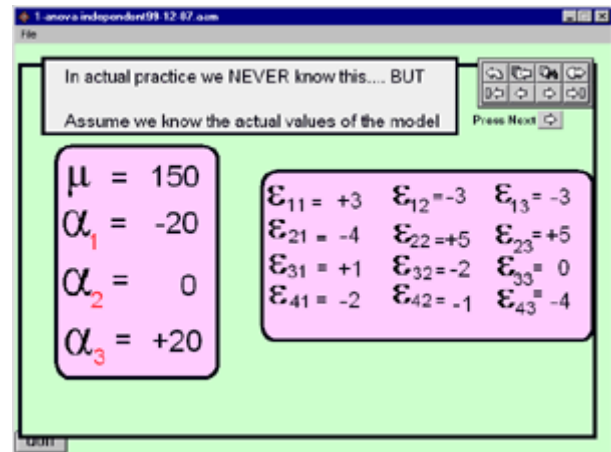


mean which is the average of all of the three means.

## SPECIFY PARAMETERS OF THE MODEL.

In mathematical jargon,  $\mu$ ,  $\alpha_j$ , and  $\epsilon_{ij}$  are called the parameters of the model.

Suppose the research team actually knew the truth about the values of the parameters. Suppose they knew that, for this group of people, the baseline weight was 150 pounds. Suppose also that the research team knew the effect of being on the low fat diet would be a 20 pound reduction in weight, the effect of eating normally would be no change in weight, and the effect of eating a high fat diet would be a 20 pound gain in weight.



The treatment effects (symbolized as  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ ) are then known to be -20, 0, and +20 lb.

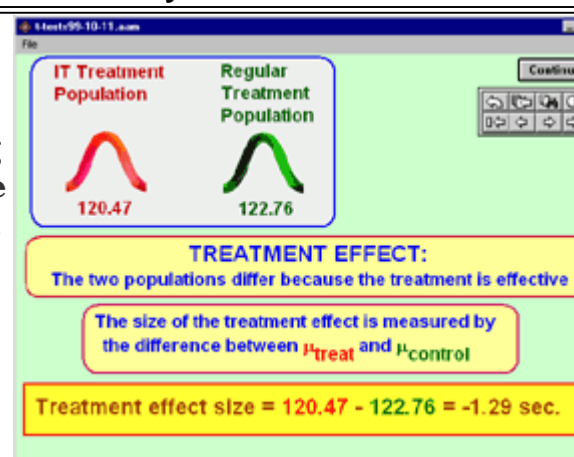
## ANOTHER WAY TO THINK ABOUT ALPHA-j

In this table we'll describe another way to think about treatment effects ( $\alpha_j$ 's). We think of them as differences among treatment  $\mu$ 's. In the **DOUBLE SAMPLE** lecture at the **INDEPENDENT-t** lecture we talked about treatment effects as the differences between treatment  $\mu$ 's. Because you are familiar with it from those lectures, this way thinking about treatment effects may seem more natural to you.

### REMEMBER t-TEST TREATMENTS

Here's another way to think about what each treatment effect ( $\alpha_j$ ) means. Using the current graphic as a reminder, recall the elite skiing example from the independent-t lecture. The treatment effect was defined as the difference between the treatment population and the untreated population.

So a treatment effect is the difference between the  $\mu$  of a treatment population



and the  $\mu$  of a baseline, untreated population.

## MULTIPLE TREATMENTS IN ANOVA

In ANOVA the scientist thinks there are multiple treatment populations. The skeptic thinks there is only one untreated, baseline population. In the percent fat in diet example, suppose we specify the  $\mu$ 's of the scientist's three treatment populations as  $\mu_1 = 130$ ,  $\mu_2 = 150$  and  $\mu_3 = 170$ . The  $\mu$  of the skeptic's baseline population 150.

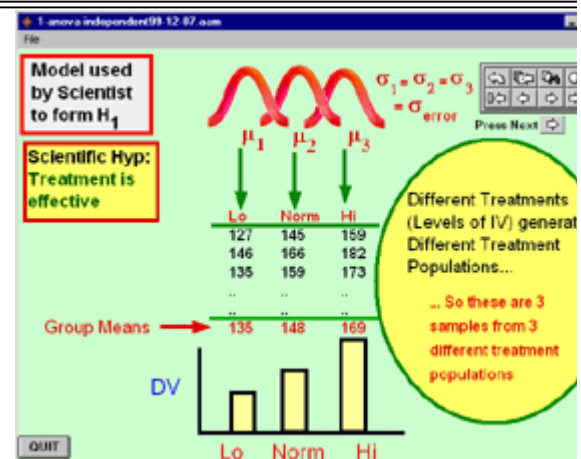
Then:

$$\text{Alpha-1} = 120 - 150 = -20.$$

$$\text{Alpha-2} = 150 - 150 = 0.$$

$$\text{Alpha-3} = 170 - 150 = +20.$$

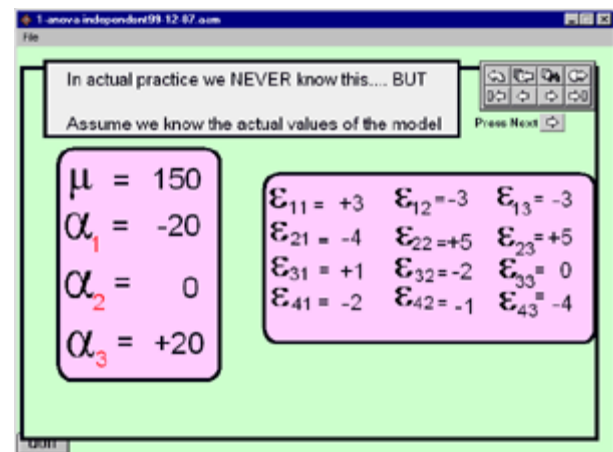
So specifying the  $\mu$ 's of the treatment populations is perfectly equivalent to specifying the  $\alpha$ -j's.



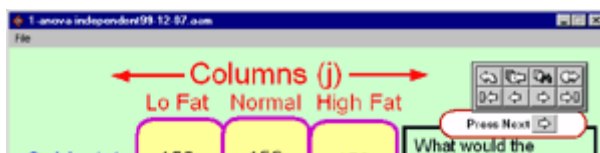
## BACK TO THE LINEAR MODEL: ERROR

Suppose finally that the measurement errors (symbolized as  $\epsilon_{11}$  to  $\epsilon_{43}$ ) are known for every participant in every group. The graphic shows the random error for every participant in every group. They are +3, -4, +1, -2, -3, +5, -2, -1, -3, +5, 0 and 4, respectively.

Remember we are assuming that the researchers actually know the truth; they know  $\mu$  (150), they know the  $\alpha$ 's (-20, 0, and +20), and they know the measurement error for every score. This is, of course, unrealistic; scientists never know the truth of models nor their parameters. But to learn about the model we will suspend belief and pretend that we know these things.



Let's see how the model performs when we add one parameter to the equation at a time. We'll start with a very simple model where  $X_{ij}$  only equals  $\mu$ .



$X_{ij} = \mu$  ONLY

What would the data look like if the model had only one parameter,  $\mu$ ?

Subject 1	150	150	150
Subject 2	150	150	150
Subject 3	150	150	150
Subject 4	150	150	150
	$M_{LO}$	$M_{NORM}$	$M_{HI}$
	$M_{GRAND}$		

data look like if the linear model were simply:

$$X_{ij} = \mu$$

$$X_{ij} = 150$$

Look at the equation on right side of the graphic. We have reduced the model so that  $X_{ij} = \mu$ . If the model was really that simple, and every score was equal to  $\mu$  (150), then every score in the data matrix would have to be 150. So all the groups means would be 150. The grand mean would be 150.

Therefore the data matrix would look like what is shown the the graphic. If the truth of the situation was that every score had to be the baseline, and nothing else mattered then everybody would weigh 150 pounds. All people in all groups would weigh the same. Okay, so obviously that model doesn't simulate data very well. Data doesn't look like that.

$$X_{ij} = \mu + \alpha_j$$

Look at the graphic. Notice that now we've made the equation more complete.  $X_{ij} = \mu + \alpha_j$ . What would the data look like if the model functioned with  $\mu$  and  $\alpha_j$  as parameters?

Here everybody's weight is equal to  $\mu$  again, but we now have included in the model the effect of each person's diet. We have added the treatment effect to a person's baseline weight. Remember, we previously said that the treatment effects are - 20, 0, and + 20.

	Columns (j)		
	Lo Fat	Normal	High Fat
Subject 1	130	150	170
Subject 2	130	150	170
Subject 3	130	150	170
Subject 4	130	150	170
	$M_{LO}$	$M_{NORM}$	$M_{HI}$

What would data look like if the model were:

$$X_{ij} = \mu + \alpha_j$$

$$\alpha_1 = -20$$

$$\alpha_2 = 0$$

$$\alpha_3 = +20$$

Look at the data matrix again. The people with the low fat treatment effect, that is those people in group one, would all have a weight of  $\mu$  minus  $\alpha_1$ . Each person in group 1 would weigh  $150 - 20 = 130$ . All of them would have the same weight.

People in the normal group would have a weight of 150 minus 0, which would keep their weight at 150.

People in the third group on the high fat diet would weigh 170 pounds because a treatment effect of +20 lb. would be added to all their weights.

**GROUP MEANS.** The group means would be  $M_1 = 130$ ,  $M_2 = 150$ , and  $M_3 = 170$ .

Let's look at two important conceptual points. Notice two things. Every group has different weights. And every person in the same group has the same weight.

**THE IV CREATES DIFFERENCES BETWEEN GROUPS.** An important conceptual point is that the levels of the IV (or treatment levels) create differences **BETWEEN** groups. The alpha-j's make the groups different.

**THE IV DOES NOT CREATE DIFFERENCES WITHIN GROUPS.** A second important conceptual point is that the alpha-j's (treatment levels) do **NOT** create any differences **WITHIN** the groups. The variance within each group is 0, because all the scores within a group are the same.

Back to the data. Given that the model is  $X_{ij} = \mu + \alpha_j$ , this simulation of data by the model is still unrealistic. What is portrayed is that everyone who is treated the same, or eats the same, should weigh the same, assuming that they started off with the same constitutions and the same  $\mu$  or base line weight. But we know that this would never be true, data just does not look like that.

So let's see what happens if we just throw some random error into the linear model.

	Lo Fat	Normal	High Fat
Subject 1	133	147	167
Subject 2	126	155	175
Subject 3	131	148	170
Subject 4	128	149	166

what if the model were:  
 $X_{ij} = \mu + \alpha_j + \epsilon_{ij}$

$\epsilon_{11} = +3$	$\epsilon_{12} = -3$	$\epsilon_{13} = -3$
$\epsilon_{21} = -4$	$\epsilon_{22} = +5$	$\epsilon_{23} = +5$
$\epsilon_{31} = +1$	$\epsilon_{32} = -2$	$\epsilon_{33} = 0$
$\epsilon_{41} = -2$	$\epsilon_{42} = -1$	$\epsilon_{43} = -4$

$$X_{ij} = \mu + \alpha_j + \epsilon_{ij}$$

Now we have the whole linear model displayed on the graphic. Each person's weight would be equal to the base line plus a treatment effect plus some random error. The full model would give us the data shown in the data matrix.

On the previous graphic, everybody in the low fat group weighed 130; but now, with an error term added each score, everyone in group 1 has a different weight. The first person is 3 pounds bigger than 130, the second person is 4 pounds below 130, person #3 is 1 pound above, and so forth.

Including random error in the model allows the model to simulate data that appears realistic. If you calculate the means of the three groups now, you will get  $M1 = 129.5$ ,  $M2 = 149.75$  and  $M3 = 169.5$ .

Let's look at some important implications of the random error component of the linear model.

**ERROR CREATES DIFFERENCES WITHIN GROUPS.** Notice that adding a



random value to every score makes all the scores **WITHIN** a group different. On the previous graphic all scores in group 1 were 130 lb. Now the scores within group 1 differ from each other. Now the variance within group 1 is greater than 0. The same is true of all groups. Now the scores differ within groups and there is a non-zero variance within each group.

**ERROR ALSO CREATES DIFFERENCES BETWEEN GROUPS.** Notice also that adding a random value to every score makes scores different **BETWEEN** groups. Even if the treatment effects were all zero, the random error would make the scores (and their means) in different groups different from each other.

### **MORE ABOUT THE ERROR TERM**

So far we've described the error assigned to each person's score as "measurement error." That's correct, but let's look deeper into what we mean by that. The first and most literal thing it means is that measurement is never completely precise in science. No matter how careful you are, you will make errors in your measurements. If nothing else, someone will invent a better instrument in a few years.

But error can be conceptually more than that. It also is an admission that the independent variable is not the only variable affecting a person's score. In the diet example, the IV is percent of calories from fat. But many other variables affect weight that are not included in the research. Exercise routine is an obvious and important variable. Even though we treat everyone in the low fat group identically in terms of fat intake, they probably all differ in their exercise; therefore their weights will vary accordingly. So error doesn't mean just measurement noise. It also includes systematic variables that we are not paying attention to in a particular research project.

Other important variables that certainly affect weight are metabolic rate, water retention, and sugar intake. You can easily think of more. So even though we treat everyone the same in the same treatment group, all these other variables will, willy nilly, make each person's weight different at the end of the study.

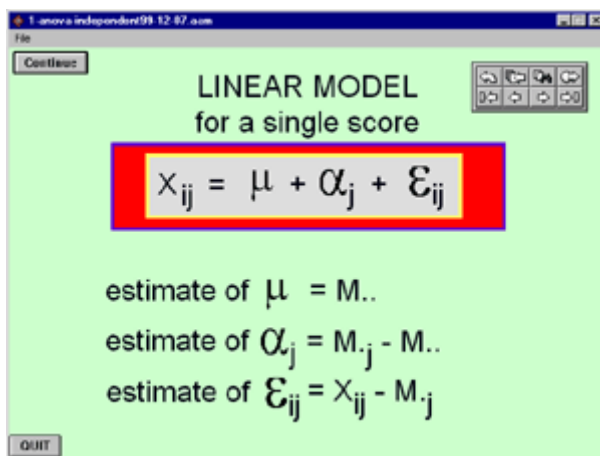
**ERROR VARIANCE.** Think about the model and think about variance. Clearly, if the baseline,  $\mu$ , were the only factor, there would be no variance whatsoever; all scores would be the same and there would be no variance within the groups. There would be only variance between the groups. Every one sampled from the same treatment population would have the same score. Let's repeat that, **"Everyone sampled from the same treatment population would have the same score. There would be no variance in the scores within groups"** Conceptually, that means there must not be any variance in the treatment populations. If there were variance in the treatment populations, then there would be variance in the samples.



**THE POINT.**  $\mu$  is a baseline; it does not create variance.  $\alpha_j$  is a treatment effect; it creates variance between groups; it does not create variance within groups. The only real source of variance within the treatments is error. So the variance of the treatment populations will be called **ERROR VARIANCE**. Error is what creates variance in treatment populations. Treatment effects create variance between treatment populations. Just let that settle; we'll have more to say about it.

**SUMMARY.** The main point is that the linear model assumes that a score is made up of 3 parts, a baseline, treatment effects, and a random error specific to the individual score. The variances of any population, whether baseline or treatment population will be called "Error Variance."

**CONCEPTUAL NOTE.** Notice that we made the treatment effects in this example much larger than the errors. The treatment effects ranged from -20 to 20 lb. The errors ranged from -4 to +5 lb. For now, just note this. Later in the lecture we will make clear that when the treatment effects are larger and errors are small, the value of the test statistic,  $F$ , will be very large.



## ESTIMATING PARAMETERS

In the Estimating Parameters lecture we discussed the procedures for estimating various parameters in various populations and sampling distributions.

In the case of ANOVA, we won't have any explicit need to estimate parameters (on homeworks or exams, for example).

But we will lay out the parameter estimates for the linear model because

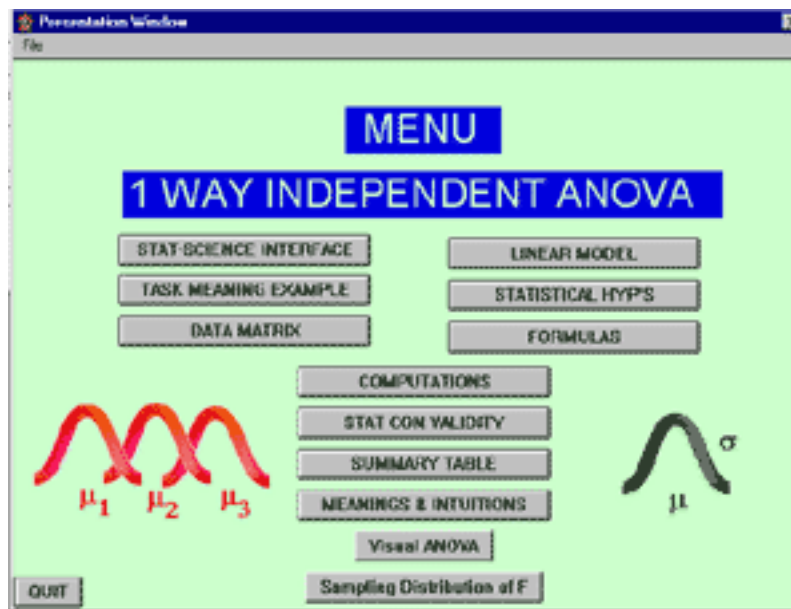
they are simple and do provide some insight into data analysis we will get to further along in this lecture. The estimates are shown in the graphic.

The estimate of  $\mu$  is the grand mean ( $M_{\text{sub } G}$  or  $M_{\text{dot dot}}$ ).

The estimate of the treatment effect size is the group mean minus the grand mean.

The estimate of a score's error is the score minus the group mean. Notice that conceptual consequences of this error estimate is that all differences between the individual scores and their group mean are due to error.

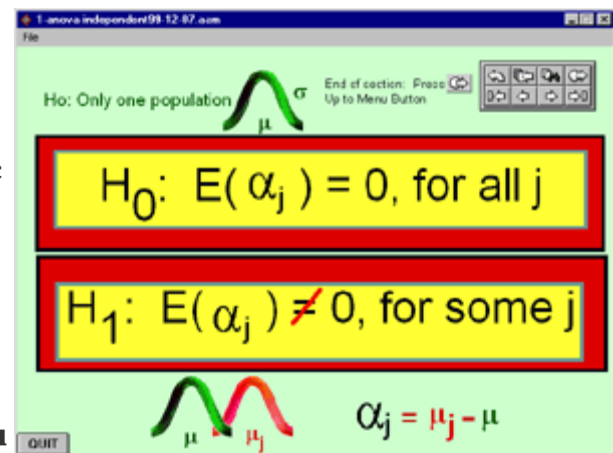
Understanding the linear model allows you to understand the statistical hypotheses. Let's now look at the null and alternative hypotheses.



## Null and Alternative Hypotheses

Back in the realm of science, the skeptic does not believe the IV affects the DV.

S/he thinks that percent fat in diet (when calories are held constant) has no effect on weight. Or, in the Task Meaning study, s/he thinks that the meaning you give a task does not affect the number of errors you make when you do the task. In other words, skeptic believes there are no treatment effects. Any differences between group means will be due to chance. In fact, any differences anywhere will be due to chance.



As you can see on the graphic, The null hypothesis is that the expected value of alpha j is equal to zero for all j. That's just a formal way of saying that the treatment effect is 0 or all groups.

Notice what the linear model reduces to if Ho is true. Every alpha-j will be 0. So

$$X_{ij} = \mu + e_{ij}$$

is the the linear model if Ho is true. No treatment effects. Just a baseline

plus random error.

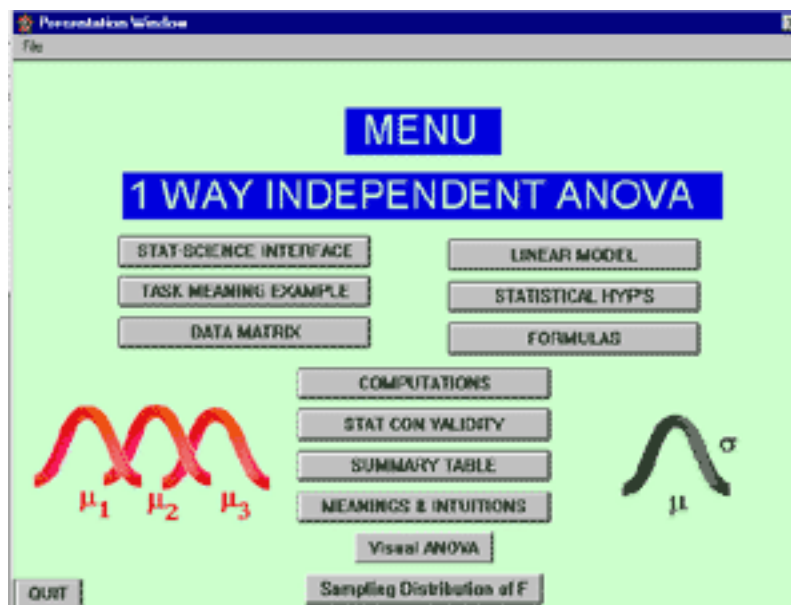
**ALTERNATIVE HYPOTHESIS.** Back in the realm of science, the scientist thinks that the IV does affect the DV. Percent fat does affect weight even when calories are held constant. Or, the meaning of a task does affect task performance. There ARE treatment effects.

As you can see on the graphic, the alternative hypothesis says that we expect  $\alpha_j$  not to be equal to 0 for at some  $j$ . This hypothesis holds that at least one, maybe all, maybe some, of the  $\alpha_j$ 's will be different from zero. Some treatment effect somewhere exists. This is the converse of the  $H_0$ .

The alternative hypothesis is not specific about which group will be different from zero. It doesn't say that they **all** have an effect, but it says that at least one of these  $\alpha_j$ 's, one of these treatments, is going to be non-zero.

The scientist might like to believe all levels of the IV will have a treatment effect. But this particular statistical alternative hypothesis only says there's an effect somewhere in the study.

So now you have a fair amount of theory to digest. Let's return to the details of number crunching. We'll start with some (fairly complex) formulas.



## Formulas



The statistical formulas in ANOVA are extensive enough that we have to divide them up into pieces. You can't just write out the formula for an analysis of variance; it wouldn't fit whatever you're writing on. The calculations are quite extensive and for this reason we're going to break them up into four chunks: 1) SUMS OF SQUARES; DEGREES OF FREEDOM; MEAN SQUARES; and the F RATIO.

**SUMS OF SQUARES.** The sums of squares formulas are the complex ones. There are three sums of squares we have to calculate: a) SUMS OF SQUARES BETWEEN GROUPS; b) SUMS OF SQUARES WITHIN GROUPS, and c) SUMS OF SQUARES TOTAL.

**REMEMBER:** We have already gone through the data matrix calculating pieces of formulas. Get ready to refer back to your notes on that section. In fact, when you did those calculations you did all the really hard work in calculating an ANOVA. In this section we are just going to organize the calculations you have already done in various different ways.

$$SS_{BG} = \sum_j \left[ \frac{(\sum_i x_{ij})^2}{n_j} \right] - \frac{(\sum_j \sum_i x_{ij})^2}{N}$$

Sum of Squares Between Groups

## Sum of Squares Between Groups

This is the formula for the sum of squares between groups, or **SSBG**. SS stands for Sum of Squares and the small BG, stands for *between groups*. We have already talked about how to calculate these different summations. The first term is the sum of the sum of the columns squared divided by the number of scores in the column. The second

term is the sum of all the scores squared first and then divided by the total number of scores.

## Sum of Squares Within Groups

The next formula is called the sum of squares within groups (**SSWG**), and it is shown in the graphic on the right. This formula uses terms that you already are familiar with. The first part says to square every single score and add them up. The second term says to sum of each of the columns, square the column sum and then divide that by the number of scores in the column.

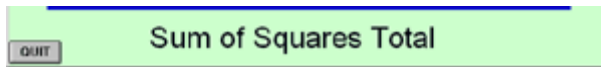
$$SS_{WG} = \sum_j \sum_i x_{ij}^2 - \sum_j \left[ \frac{(\sum_i x_{ij})^2}{n_j} \right]$$

Sum of Squares Within Groups

$$SS_{TOT} = \sum_j \sum_i x_{ij}^2 - \frac{(\sum_j \sum_i x_{ij})^2}{N}$$

## Sum of Squares Total

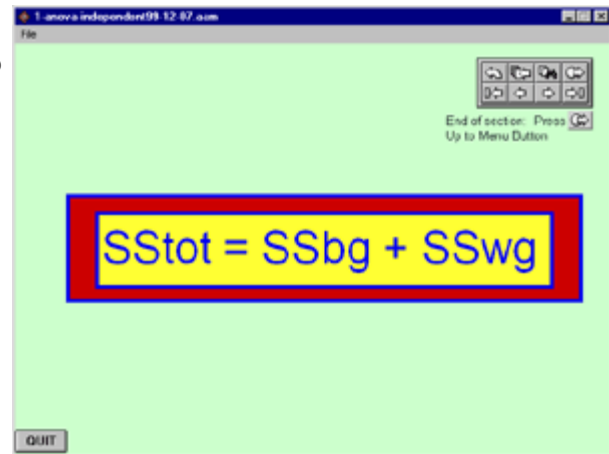
The sum of squares total is calculated using the formula on the left. You should notice that you have already written down parts of these formulas. So if you calculated the two previous sums of squares, you already have the values necessary for both parts of this formula. In essence to calculate an *F*, you have to



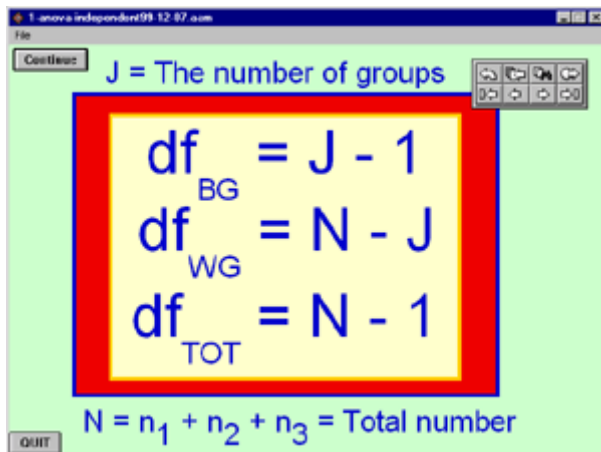
other terms you need to calculate F.

The statement at the right is always true. The sum of squares total is equal to sum of squares between groups and sum of squares within groups. I recommend that you use this formula to check your answers so you can make sure you didn't make an arithmetic error.

calculate three different values. Then you use those three values to find the



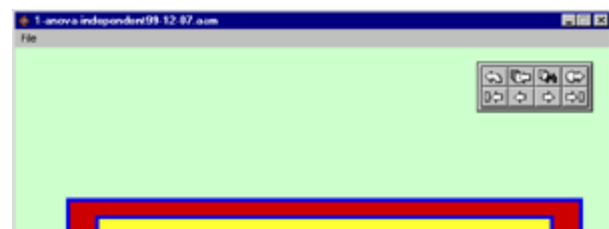
## Degrees of Freedom



Degrees of freedom are easier to calculate than the sums of squares terms. To calculate degrees of freedom (*df*), we let big J equal the total number of groups, and big N equal the total number of Participants Big N is equal to *n*<sub>1</sub> (number of scores in group one) plus *n*<sub>2</sub> (number of scores in group two) plus *n*<sub>3</sub> (number of scores in group three). In other words by adding up the total number of scores in each group you can determine the value of big N.

The degrees of freedom between groups is just J minus 1; the number of groups you have minus 1. The degrees of freedom within groups is total n minus j, or big N minus the number of groups. Finally, the degrees of freedom total, is big N minus one, the total number of observations that you have minus one.

Here is another check you can do, although usually people don't make mistakes on degrees of freedom. Degrees of freedom total is equal to degrees of freedom between groups plus degrees of freedom within groups.





A screenshot of a software window with a light green background. At the top, a yellow rectangular box with a red border contains the formula  $df_{tot} = df_{bg} + df_{wg}$  in blue text. Below the box, there is a small "QUIT" button in the bottom left corner.

## Mean Square Between Groups

Mean square between groups is easy to calculate at this point because you have already figured out the values of its component parts. You divide the value you calculated for Sum of Squares between groups by the number of degrees of freedom between groups.

A screenshot of a software window titled "1 anova independent09 12 07.asm". It has a light green background. A yellow rectangular box with a red border in the center contains the formula  $MS_{BG} = SS_{BG} / df_{BG}$  in blue text. There are "Continue" and "QUIT" buttons visible.

## Mean Square Within Groups

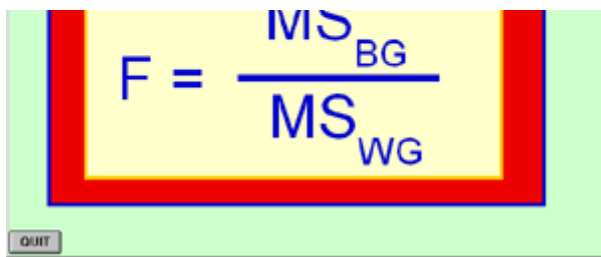
Mean square within groups is easy to calculate too. You divide the value you calculated for Sum of Squares within groups by the number of degrees of freedom within groups.

A screenshot of a software window titled "1 anova independent09 12 07.asm". It has a light green background. A yellow rectangular box with a red border in the center contains the formula  $MS_{BG} = SS_{BG} / df_{BG}$  in blue text. There are "Continue" and "QUIT" buttons visible.

A screenshot of a software window titled "1 anova independent09 12 07.asm". The main text "The F Ratio" is displayed in a large font. Below it, there is a yellow rectangular box with a red border. At the bottom of the window, there is a small text box that says "End of section. Press Up to Menu Button".

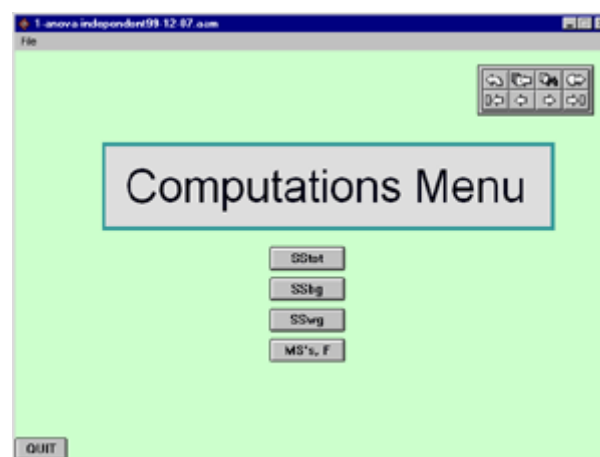
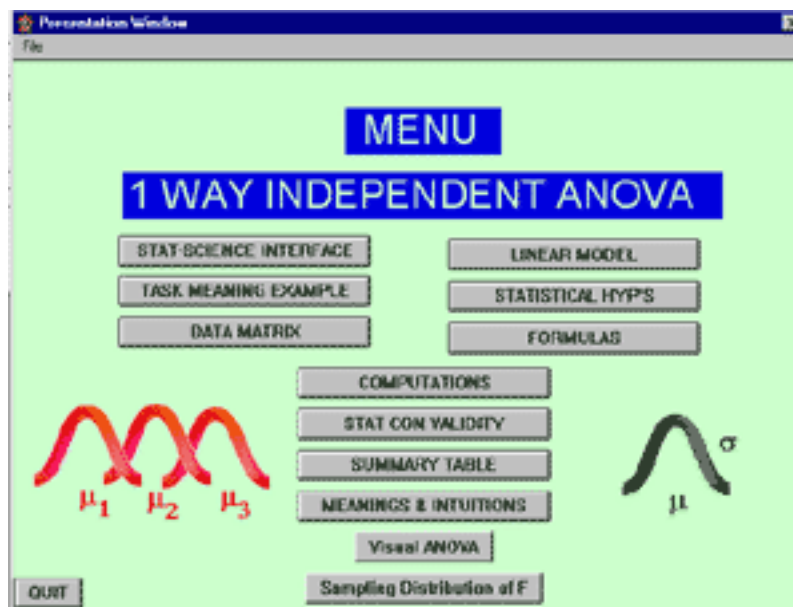
## F Ratio

You are finally ready to calculate the actual *F*. It is mean square between groups divided by mean square within



$$F = \frac{MS_{BG}}{MS_{WG}}$$

groups.



Let's go ahead and do the computations, and apply these formulas to the example about the effect of instructions on performance.



*Computation of Sum of*

$$SS_{TOT} = 429 - \frac{(97)^2}{27}$$

$$\sum_j \sum_i X_{ij}^2 - \frac{(\sum_j \sum_i X_{ij})^2}{N}$$

## Squares Total

The first thing we'll compute is the sum of squares total. We are going to go through and actually do an example of each of these. Here's the formula again for your review. This this case, sum of squares total is equal to 429 minus 97 squared over 27. In your notes from our earlier work, you should have that 429 was the sum of all the scores squared. 97

was the total of all scores, and so squaring and dividing by the n gives us the value we need to calculate SS tot.

These are only pieces that we need to get the SS tot. Sum of squares total is equal to 80.519.

$$SS_{TOT} = 429 - \frac{(97)^2}{27}$$

$$SS_{TOT} = 429 - \frac{9409}{27}$$

$$= 429 - 348.481 = 80.519$$

$$SS_{BG} = 393 - \frac{(97)^2}{27}$$

$$\sum_j \left[ \frac{(\sum_i X_{ij})^2}{n_j} \right] - \frac{(\sum_j \sum_i X_{ij})^2}{N}$$

## Cor of Sum of Squares Between

At the bottom of this graphic is the sum of squares between groups formula for your reference. If you plug in the values that we calculated earlier, SS bg is 393 minus the term we just used in SS tot. You'll notice that some of these terms are repeating themselves so there isn't

all that much to calculate. You just have to calculate the three basic values and then substitute them in the right places in the formulas.

Sum of squares between groups is equal to 393 minus 97 squared over 27, and so that gives us 393 minus 348.2481 or 44.519.

$$SS_{BG} = 393 - \frac{(97)^2}{27}$$

$$SS_{BG} = 393 - 348.481$$

$$= 44.519$$

$$SS_{WG} = 429 - 393$$

$$\sum_j \sum_i x_{ij}^2 - \sum_j \left[ \frac{(\sum_i x_{ij})^2}{n_j} \right]$$

## Computation of Sum of Squares Within

The final piece of the formula is the sum of squares within groups. From our previous calculations, we know that sum of squares within is equal to 429 minus 393. So the sum of squares within groups is 36. And then finally you can double check yourself--sum of squares total is 80.519 and it's equal to 44.519 plus 36. This means that we have

correctly computed our sums of squares.

Now you know how to compute the three sums of squares, so we're going to calculate the other things that are important, the degrees of freedom, the mean squares, and, ultimately, the F statistic.

$$SS_{WG} = 429 - 393$$

$$SS_{WG} = 36$$

$$SStot = SSbg + SSwg$$

$$80.519 = 44.519 + 36$$

## Degrees of Freedom

$$df_{tot} = N - 1 = 27 - 1 = 26$$

$$df_{bg} = J - 1 = 3 - 1 = 2$$

$$df_{wg} = N - J = 27 - 3 = 24$$

$$df_{tot} = df_{bg} + df_{wg}$$

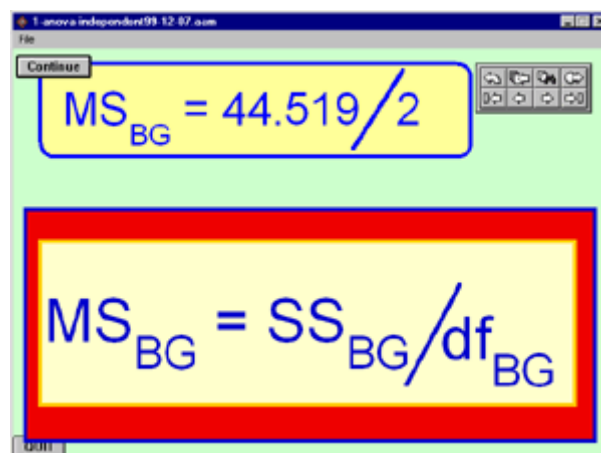
$$26 = 2 + 24$$

The **degrees of freedom total** ( $df_{tot}$ ) formula is quite simple: it's  $N$  minus 1, so for this example that's 27 minus 1, which is equal to 26. **Degrees of freedom between groups** ( $df_{bg}$ ) is equal to  $j$  minus 1, where  $j$  is the number of groups. In our example, three groups minus 1 equals 2. **Degrees of freedom within groups** ( $df_{wg}$ ), is the total number of observations minus the number of groups so that's  $N$  minus  $j$ . For our study this is equal to 27 minus 3 or 24.

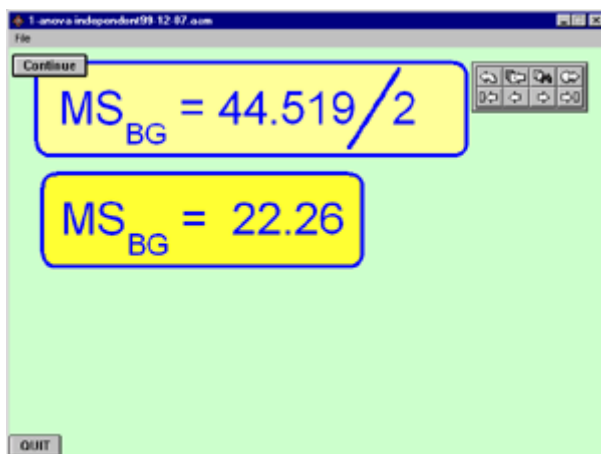
**Notice** that the degrees of freedom add up. The degrees of freedom total is equal to degrees of freedom between groups plus the degrees of freedom within groups.

## Mean Squares

Let's define the mean squares. These are just the sum of squares divided by their corresponding degrees of freedom. So, if you look in the red box you can see that the mean square for between groups is equal to sum of squares between groups divided by degrees of freedom between groups. So in our case, the mean square between groups is 44.519 divided by 2, or 22.26.



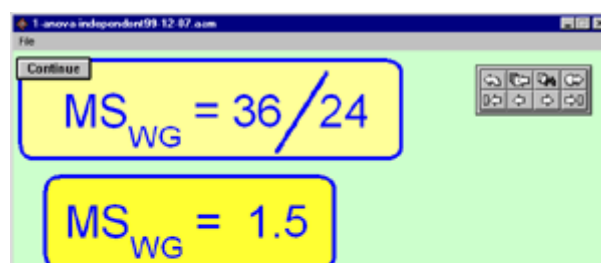
A screenshot of a software window titled "1 anova independent99.12.07.asm". It shows a yellow box with the formula  $MS_{BG} = 44.519 / 2$ . Below it, a red-bordered box contains the formula  $MS_{BG} = SS_{BG} / df_{BG}$ . The window has a "Continue" button at the top left and a "QUIT" button at the bottom left.



A screenshot of a software window titled "1 anova independent99.12.07.asm". It shows two yellow boxes. The top box contains the formula  $MS_{BG} = 44.519 / 2$ . The bottom box contains the result  $MS_{BG} = 22.26$ . The window has a "Continue" button at the top left and a "QUIT" button at the bottom left.

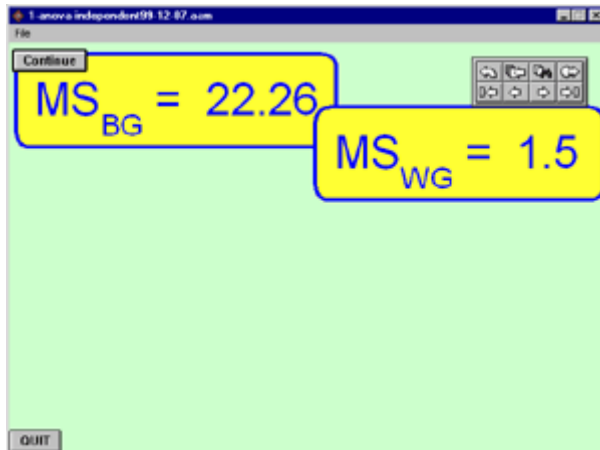
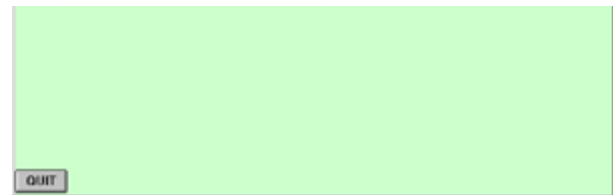
The 2 in the denominator is the degrees of freedom between groups. The mean square between groups is defined as the sum of squares between groups over degrees of freedom between groups. The degrees of freedom between groups is  $j$  minus 1, three groups minus one, which is two. The mean square between groups is 22.26.

The mean square within groups is just the sum of squares within groups over the degrees of freedom within groups. In this case, 36 was the sum of squares within groups divided by 24 which was degrees of freedom within groups. Mean square within groups is 1.5.



A screenshot of a software window titled "1 anova independent99.12.07.asm". It shows two yellow boxes. The top box contains the formula  $MS_{WG} = 36 / 24$ . The bottom box contains the result  $MS_{WG} = 1.5$ . The window has a "Continue" button at the top left and a "QUIT" button at the bottom left.

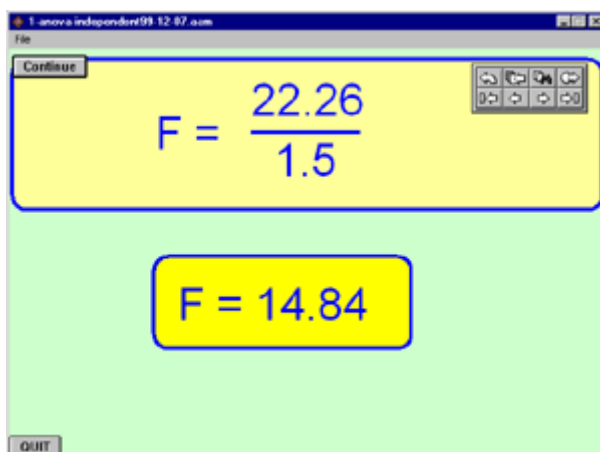
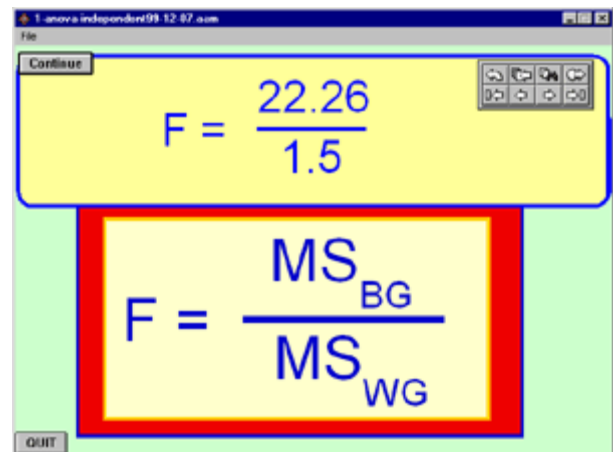




Now you can calculate the mean squares between groups and mean squares within groups. There is not a calculation of mean squares total, no one ever calculates that.

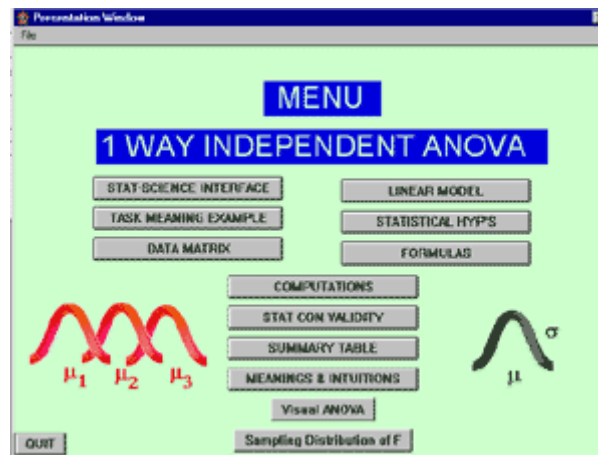
### TEST STATISTIC: F

Finally we get to our test statistic,  $F$ . The formula for  $F$  is mean square between groups, divided by mean square within groups. This is called the  $F$  ratio. For our study the  $F$  is 22.26 divided by 1.5. So  $F$  in this case is 14.84.

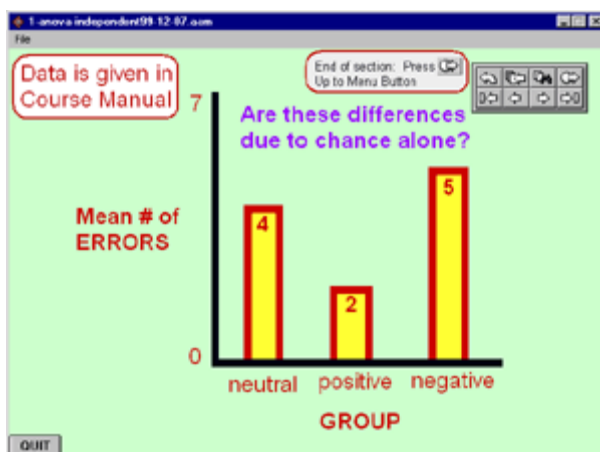
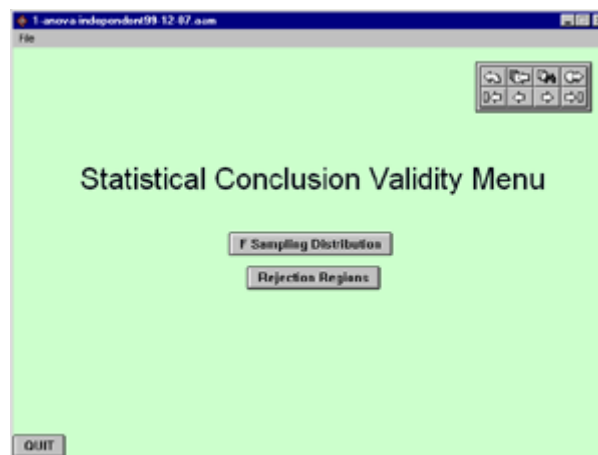


Finally, after all these computation we have calculated the value of  $F$ .

Now that we have calculated F, let's turn to the standard issues of statistical conclusion validity. Is this F value significant?



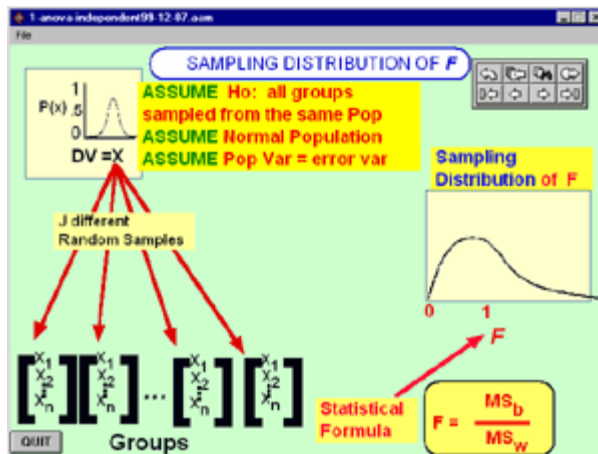
### Statistical Conclusion Validity



### TASK MEANING DATA

As we've seen, the group means came out just as the scientist predicted. The data pattern fits with the scientific hypothesis.

PCH OF CHANCE. But the question all skeptics will want answered right away is, "Do these means differ by chance alone?"



## OVERVIEW OF MODEL.

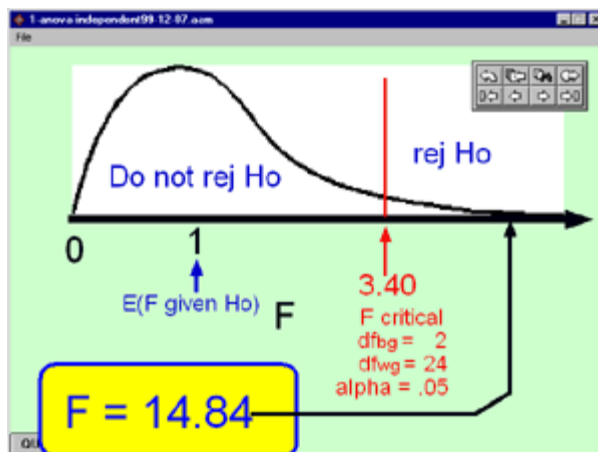
Let's go back to the big picture. It is a schema we have used many times for summarizing in a holistic way a complicated process.

We have modeled our DV (# of errors) as a Normal Population. We assume that  $H_0$  is true and that all groups in the study are random samples from the same untreated, baseline population. Then we calculated a test statistic, the F ratio in

this case, on the data from our groups.  $F = MS_b / MS_w$ .

Finally, given that we assume  $H_0$  is true, we want to know what is the probability of getting an F value this high by chance alone. We use the sampling distribution of F to answer that question.

Let's look at the sampling distribution of F more closely.



## $H_0$ : TO REJECT OR NOT?

The range of  $F$  goes from 0 to positive infinity. Unlike the  $t$  statistic which has a range from negative to positive infinity, the  $F$  statistic starts from 0. The reason that there are no negative  $F$  values is that every quantity which goes into the  $F$  formula is a squared quantity. Since every number squared becomes a positive value, the whole range of the  $F$  statistic starts at 0 and goes to positive infinity.

Later, in the Meanings and Intuitions section of this lecture, it will become apparent why the expected value of  $F$  is 1, given that  $H_0$  is correct. For now, just accept that we expect the  $F$  ratio to be around 1 if  $H_0$  is true.

Determining the critical region requires that you know the alpha level and two different degrees of freedom (between groups and within groups degrees of freedom). You will find an F table for  $\alpha = .05$  on the website with the other statistical tables. We won't use any alpha other than .05

From this table for  $\alpha = .05$  with  $df$  between = 2 and  $df$  within = 24, you can find that the critical value of  $F$  is 3.40.

The graphic shows a red line marking the value of 3.40 on the sampling distribution of  $F$ . There is only a 1 in 20 (.05) chance of getting an  $F$  value beyond that line if  $H_0$  is true. So using our usual logic we reject  $H_0$  for calculated values of  $F$  beyond 3.40. We do not reject for values below 3.40.

With our Task Meaning example, the calculated  $F$  value does fall out there in the rejection region, so we would reject  $H_0$ .



## Analysis of Variance Summary Table

Source of Var	df	SS	MS	F	p
Total					
Between					
Within					

Between corresponds to treatment (IV)  
Within corresponds to error

### Summary Table

An outline of an ANOVA summary table is shown on the current graphic. When we do our calculations we will put the results in this table.

**SOURCE OF VARIANCE.** The first column in the summary table is the "source of variance." After going through the calculations we have just completed, it should be clear that the total variance in the table is broken down into the variance that occurs between groups and

the variance that occurs within groups.

**DEGREES OF FREEDOM.** The next column lists the degrees of freedom. As you know, you have to calculate three degrees of freedom, total, between and within.

**SUMS OF SQUARES.** The next column is where you put the results of the calculations for the three sums of squares.

**MEAN SQUARES.** In the next column you list your calculated mean squares. There is no total mean square. So you only list the mean square between and within.

**F RATIO.** The F ratio goes in the fifth column. Of course there is only one F ratio in this analysis so you only list the one F you calculated. Put it in the "Between" row.

**p VALUE.** We selected an alpha value of .05. So in this final column we would indicate that the probability of an F value of this magnitude being due to chance alone is less than .05.

## SUMMARY TABLE FILLED OUT

The current graphic shows an ANOVA summary table filled out for the Task Meaning example that we just computed.

The purpose of having an analysis of variance summary table is to organize your results and to allow you to see an overview all of the ANOVA calculations. You can use this table to put all the pieces you calculated into a standard format. Just by looking you can see that the between and within degrees of freedom add up to the total degrees of freedom. The same is true for the Sum of Squares column. The summary table also helps you to spot errors.

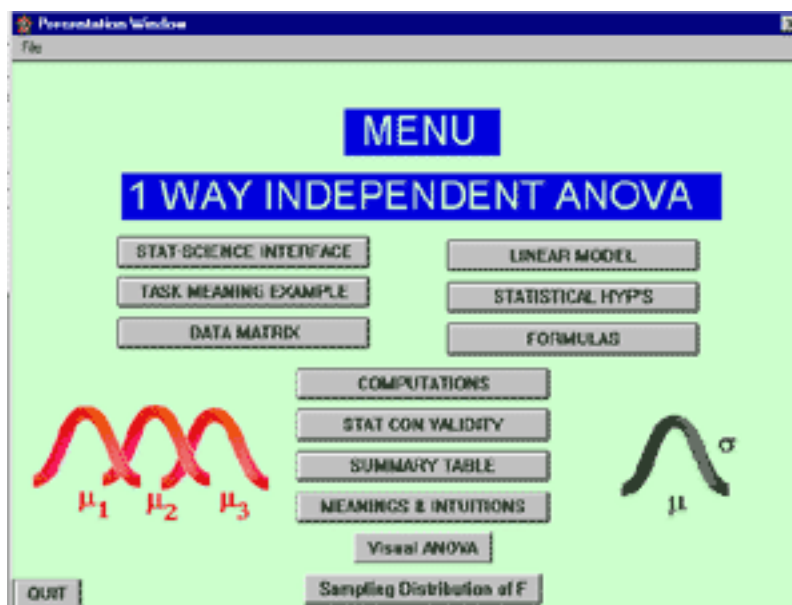
You'll be asked to put your results in a summary table in homeworks and on exams. And you'll also need to be able to read one for the exam. There's nothing new in the analysis of variance summary table, it's just a way to organize a fairly complicated process.

Source of Var	df	SS	MS	F	p
Total	26	80.52			
Between	2	44.52	22.26	14.84	<.05
Within	24	36	1.5		

Between corresponds to treatment (IV)  
Within corresponds to error

Notice that for df and SS, Between and Within add up to Total

QUIT

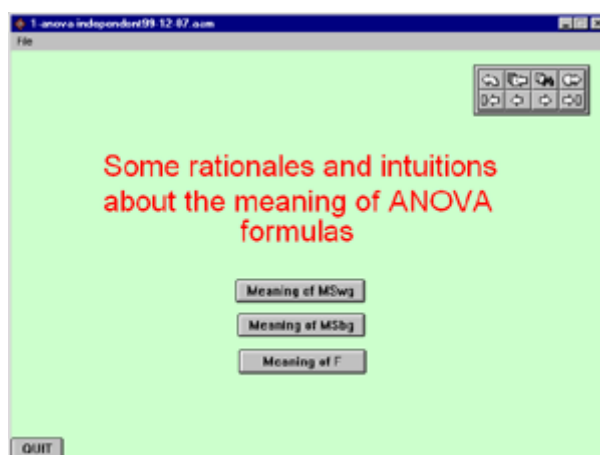


## 1-ANOVA Meanings & Intuitions

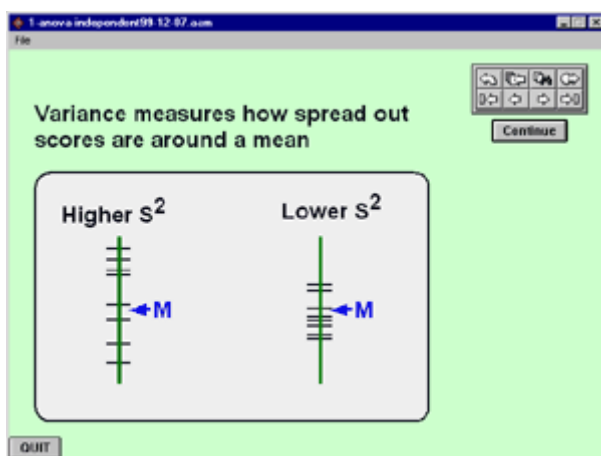
In this section we are going to develop some intuitions about why the whole system of ANOVA works the way it does. Up to this point we've developed a lot of theory and a lot of computational details. Now we are going to bring that all together into a conceptual whole. And we any luck we will satisfy some of your intellectual curiosity about ANOVA.

We will address the meaning of the *mean square within groups*. Next we will discuss what the *mean square between groups* means. Finally, we will look at what the *F test* mean and why this particular ratio (MSbg over MSwg) makes sense?

Our focus is on the main theoretical ideas not any numbers or details. You will get to play with a simple program called "Visual ANOVA" after you read this section. It will allow you "see" the ideas we are talking about here in a visual, holistic form.







## Variance

To start out let's review the meaning of variance. Two of the fundamental concepts in this course are variance and central tendency. Variance measures how spread out scores are around the mean.

This slide shows examples of two data sets shown graphically, where the hash marks are scores. Notice on the left side

that the scores are more spread out from the mean, consequently the value for variance or "big S squared" will be greater than on the right side of the screen where the scores are more compact. On the right the scores are closer to the mean, which will result in a smaller variance. So variance measures how spread out things are.

### LITTLE s: AN ESTIMATE OF POPULATION VARIANCE

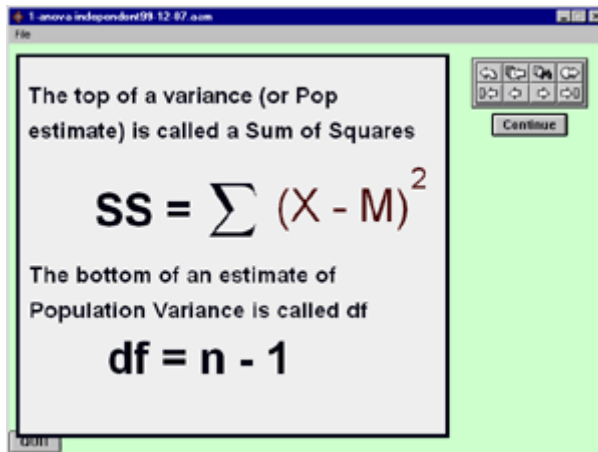
ANOVA uses "little s" rather than "big S." That is, it uses estimates of population variance rather than descriptions of sample variance. Little s uses an  $n - 1$  as a denominator, while big S uses an  $n$  as a denominator. So big **S** and little **s** have different terms on the bottom of the formula.

As a review, here is the conceptual formula for little **s**, which is the sample estimate of the population variance. (You have also learned computational formulas for little s.) This conceptual formula is the sum of the squared deviations around the mean divided by  $n$  minus one. The top of the formula or numerator is where the sum of the squared deviation can be found. So if you want SS, you just multiply little **s** by  $(n - 1)$ .

(Alternatively, you could multiply big **S** by  $n$  to get SS.)

In analysis of variance we use population estimates, so there will also be **n minus one** in the denominators of formulas. That's all a review something that you are already familiar with, but it never hurts to have all the details refreshed.

**CONCEPTUAL VERSUS COMPUTATIONAL FORMULAS.** In this section on Intuitions and Meanings, we will use conceptual formulas for the various parts of ANOVA. This is, of course, because they are conceptual and lend themselves to understanding. But for calculating they are a bother. So we recommend that you use the computational formulas you learned above to do calculations. We will not require you to use the conceptual formulas for calculating ANOVA's in this course.



## Sum of Squares in Variance

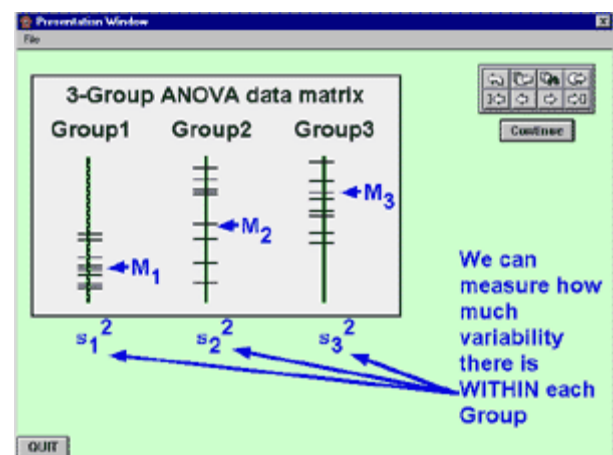
The top of either variance formula (big S squared) or the population estimate formula (little s squared), is called the sum of squares. The top part is where we have the **sum of (XI minus the mean) squared**. Obviously, this is the sum of squared deviations around the mean, or for short the sum of squares. The bottom part of an estimate of population

variance is called its degrees of freedom; and the degrees of freedom here on this graphic is **n minus one**.

## A POPULATION ESTIMATE FOR EACH GROUP

Imagine that we have a three-group analysis of variance data matrix. (Of course we don't have to be limited to three groups, but we can make all the conceptual points that we need to make using three groups.)

This graphic shows the data points in each of the groups as hash marks. For our purposes here, we don't need numbers; we're trying to develop intuitions and visualizations rather than calculations or proofs.



The three means are shown in blue with arrows. Mean 1 in Group 1 is shown

on the left. In the center we have Mean 2 with all its scores spread around it, and over on the right we have Mean 3 with all of its scores spread around it. Notice that three means differ from one another and that the scores within the groups differ from their means.

**THREE POPULATION ESTIMATES.** Along the bottom of the grey box on the graphic are the symbols for three population estimates, one for each group. Using the little  $s$  formula we just reviewed, if I DID give you numbers rather than hash marks you COULD find little  $s^2$  for each of the groups. It would not be particularly difficult for you to calculate those three variance estimates.

### WHAT ARE THE THREE ESTIMATES ESTIMATING?

Let's examine what these three estimates are estimating in terms of the skeptic's model first and then in terms of the scientist's model.

**SKEPTIC SAYS "ONE POPULATION VARIANCE."** If the skeptic's model (there is only one baseline population) is right, then we have taken three samples from the same population; so these are three different estimates of the same population variance.

**SCIENTIST SAYS "THREE POPULATION VARIANCES."** If the scientist's model (there are three treatment populations) is right, then the three groups of score are samples from the three treatment populations. So each "little  $s^2$ " estimates the variance of one of the three treatment population variances.

**BUT** remember one of our assumptions.

**HOMOGENEITY OF VARIANCE (AGAIN).** Recall that even if you do have three treatment populations as the scientist claims, homogeneity of variance assumption says that all THREE have the same variance. By a formal assumption of the model, all population variances are equal.

So it doesn't really matter which model, the skeptic's or the scientist's, we use. There is only one variance to estimate since all populations have the same variance. So the three "little  $s^2$ 's" in the three groups all are an estimate of the same population variance.

That's a fair amount of theory. But it comes down to one simple point. All three variance estimates estimate the same population variance.

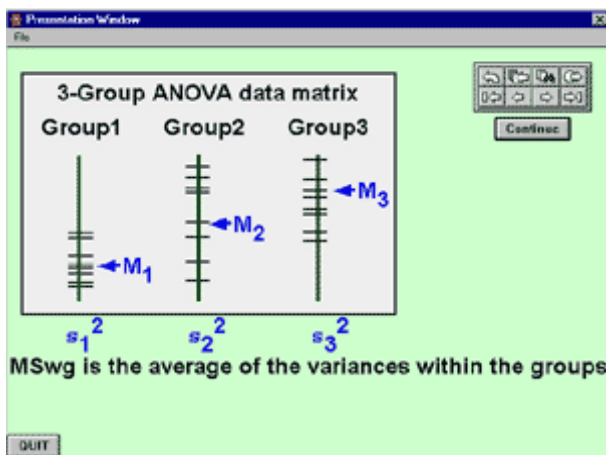
How do we interpret this one population variance we are estimating?

**ERROR VARIANCE (AGAIN).** All the Participants within a particular group are

treated identically so there is no reason for their scores to differ except for error (measurement imprecision, uncontrolled and unknown variables, etc.). All variance within a treatment is error. So, as we've said before, all the variance within a treatment population must be error variance. Everyone given the same treatment should generate the exact SAME score (except that everyone is genetically different, has had a different life history, is affected by different contexts, and so on, all of which we call ERROR). Everyone from the same treatment population is the same except for the variance generated by error.

**POPULATION VARIANCES ARE ERROR VARIANCES.** In the formal model then, all population variances are considered to be error variances.

**ALL THREE LITTLE s's ARE EACH ESTIMATING ERROR VARIANCE.**



## What is Mean Square Within?

**COMPUTATIONALLY.** Computationally, the mean square within is just an average of the little s's from these three groups.

**ERROR VARIANCE.** As we just discussed, the model interprets all variance within a treatment population as error variance.

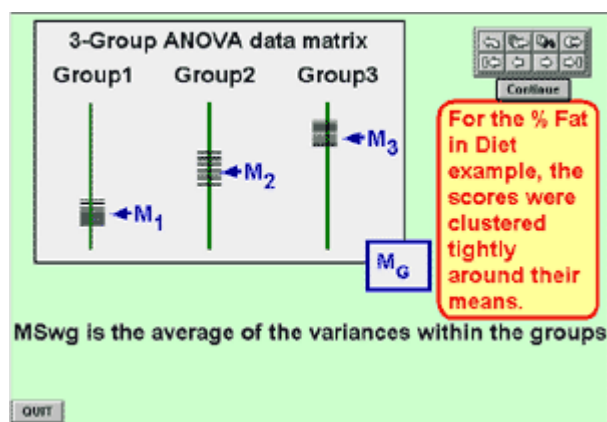
So each little s is an estimate of population error variance. Since we have three data samples, we've got three different estimates of error variance-- we've got an estimate from Group 1, an estimate from Group 2, and an estimate from Group 3. The mean square within simply is the average of these three population variance estimates.

Let's make this a little more concrete by using the percent of calories from dietary fat example. We've had a lot of experience with developing the linear model using that example so these theoretical intuitions can be grounded in that experience.

Remember, we are building intuitions and visualizations here. Some numbers and formulas will be used but you should be attending to the big picture. The numbers and formulas are there to help tie down the meaning of the ideas.



## FAT DATA REPRESENTATION



For the Percent Fat in Diet example, the treatment effects were large (+ or - 20 lb.) and the random errors were small, only a few pounds. Therefore, the scores were clustered very tightly around their group means, as shown by hash marks on the current graphic.

In more realistic data sets the picture is not usually so clear cut. Many times the errors are very large relative to the treatment effects.

## CONCEPTUAL FORMULAS FOR SSwg & MSwg

The current graphic looks like a lot of formulas and symbols, but conceptually it is pretty simple. This slide is just a repetition of what was presented earlier but is a little more explicit. AND you don't need to learn to use these formulas.

The sum of squares within Group 1, or SS-1, is equal to the sum of the squared deviations of the individual data points around the mean of Group 1. That should make good conceptual sense to you at this point in this class.

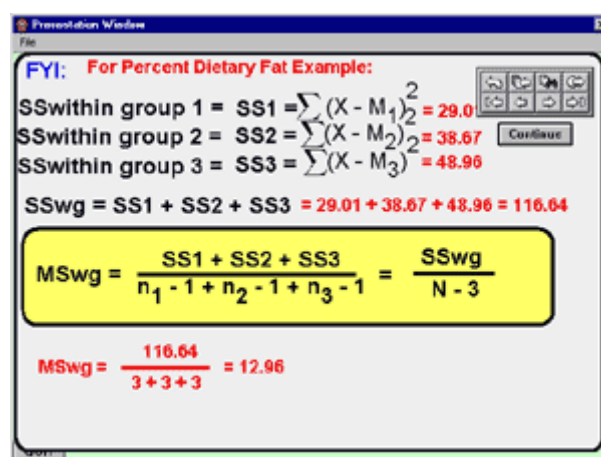
The sum of squares within Group 2, or SS-2, is the sum of squared deviations around the mean of Group 2. And the same idea is true for the sum of squares for Group 3.

There is no need for you to calculate the following numbers, just get the general idea. (But if you want, you can go back to the data for the Dietary Fat example and confirm the numbers.)

SS-1 = 29.01; SS-2 = 38.67; and SS-3 = 48.96.

**SUM OF SQUARES WITHIN GROUPS.** To get SSwg all you do is add up the SS's from the individual groups. So SS wg = 116.64.

[NOTE: To get this number we have used conceptual formulas rather than computational formulas. But if you used the computational formulas you learned above, you would get the same result, within rounding error. As we said before, the course will not require you to use the conceptual formulas in





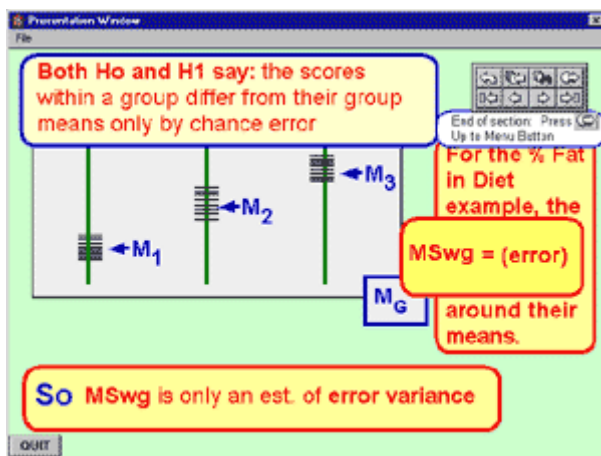
this section for calculations. So just follow the train of thought rather than worrying about learning new ways to calculate an ANOVA.]

**MEAN SQUARE WITHIN GROUPS.** To get MS<sub>wg</sub> you merely divide the SS<sub>wg</sub> by the degrees of freedom within groups.  $df_{wg} = N - J$ , where  $N$  is the total number of scores and  $J$  is the number of groups.  $N$  in this example = 12 (3 groups times 4 scores each = 12). So  $df_{wg} = 12 - 3 = 9$ .

But now there is a chance you can begin understanding the degrees of freedom. Look at the yellow box where MS<sub>wg</sub> is defined. The top is the sum of the individual SS's from the three groups; the bottom is the sum of the divisors from the three little s's. Each little s has as a divisor of  $(n - 1)$ . Adding them up gives us  $N$  (the total number of scores) - 3.

So MS<sub>wg</sub> = 116.64 divided by 9 = 12.96.

This is a very different way to calculate MS<sub>wg</sub> than by using the computational formulas you learned previously. It is more oriented toward understand how the calculations go with the theory.



**HO AND H1 AGREE FOR ONCE**

**A SINGLE VARIANCE ESTIMATE.** Mean Square within groups is just a way pooling together the three population variance estimates (little s's) from the three groups. MS<sub>wg</sub> puts the data for the three groups together to get a single population variance estimate.

MS<sub>wg</sub> is just an estimate of population variance.

**ERROR VARIANCE.** I review one more time why this variance being estimated is error variance. If this seems obvious to you skip down to the next section; otherwise read on.

What variance does MS<sub>wg</sub> estimate? For once the skeptic (Ho) and the scientist (H1) agree about something, they both think that MS<sub>wg</sub> estimates error variance.

The skeptic thinks that nothing is going on in the study except chance error. Ho asserts that all treatment effects are 0. So every score is just the baseline  $\mu$  plus some error. So all variance is error variance. MS<sub>wg</sub> therefore is estimating error variance.

The scientist thinks that every person within a group is treated identically. So the only reason scores differ within groups is due to error. So all the

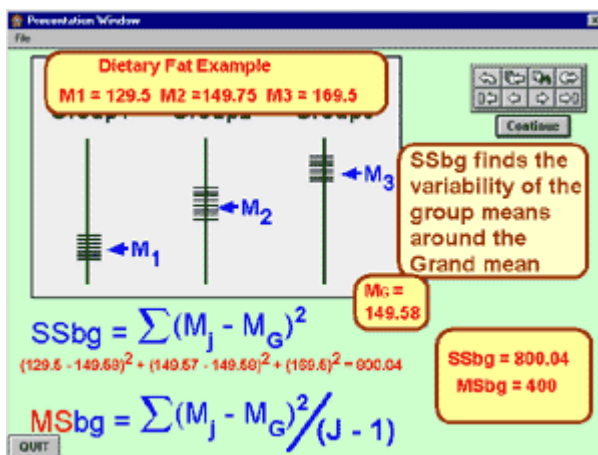
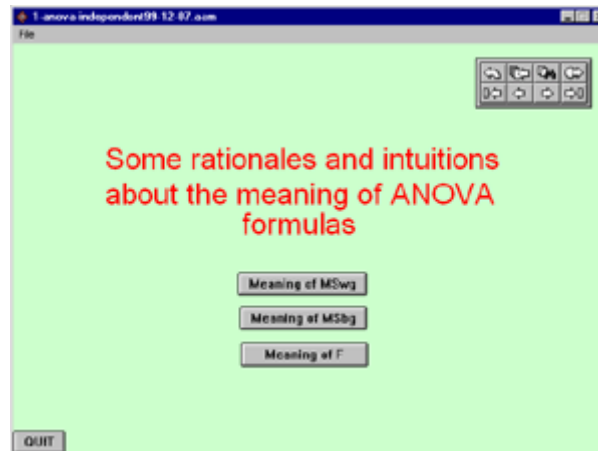


variance within groups is due to error. Therefore MSwg estimates pure error variance.

Let's restate the scientist's point of view. As experimenters, we randomly assign Participants to each group and then within each group we treat every participant identically. That is, we give them each the same reinforcement schedule, or the same diet, or the same psychotherapy, or the same leadership training. It doesn't matter what the IV (treatment) is, the important point is that everybody in a particular group gets treated identically. So we have no trouble modeling within group variance as error.

**WITH THE MODEL WE HAVE BUILT, MSwg ESTIMATES ERROR VARIANCE.**

## MEANING OF MEAN SQUARE BETWEEN GROUPS



## Conceptual Formulas for SSbg & MSbg

The grand mean is the mean of the group means. So we can see how much variance the group means have around their mean.

**SUM OF SQUARES BETWEEN.** Let's look at the computational formula for the SSbg. As you can see on the graphic, SSbg is the sum of the squared deviations of the group means around the grand

mean.

**MEAN SQUARE BETWEEN GROUPS.** We're talking about the variance of the group means around the grand mean. That's a little bit like word salad, but the means are just numbers and they vary, and if they vary then they have variance. So we are going to apply our ideas from variance to the group means.

**VARIANCE IS CRAZY.** One of the confusing and annoying things about variance is that it has a split personality: There are two of it. There is (are) big S squared and little s squared. But by now we're used to variance's personality disorder. Big S is a descriptive statistic. Little s is a population estimate. This personality disorder of variance is one reason why people like sums of squares; SS's get you to the heart of the variation--and then you can divide by whatever you like later (either divide by  $n$  or  $n - 1$ ).

Back to MSbg. We've got a formula for SSbg. Now we have to divide it by something. Now we have to deal with the split personality ( $n$  versus  $n-1$ ). In our example we have 3 means. What should we divide by? Should we divide by 3 (i.e., the number of scores) or by 2 (i.e., the number of scores minus 1)? [In general case of  $J$  groups, the number of means is  $J$ .]

The answer to the split personality question is that MSbg is SSbg divided by  $J - 1$ . We use  $J - 1$  because, as you will see, we want an estimate of population variance not a description of sample variance.

### **PERCENT FAT IN DIET EXAMPLE**

If you go back to the dietary fat example you'll find that  $M1 = 129.5$ ,  $M2 = 149.75$ , and  $M3 = 169.5$ . The grand mean, which is the mean of these means, is 148.58.

**SSbg.** As the graphic shows we can find the SSbg by finding the sum of square deviations if these three means around their mean.  $SSbg = 380.25 + 0.0289 + 396.81 = 800.04$ .

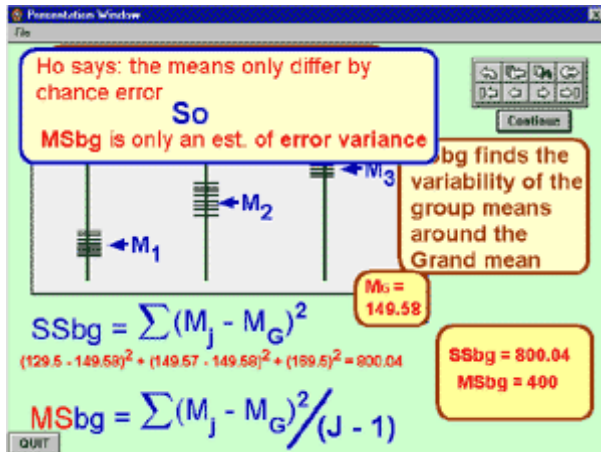
**MSbg.** The graphic also shows that  $MSbg = SSbg$  divided by  $J - 1 = 800.04$  divided by 2 = 400.

So we have gone back to basic variance formulas to understand what the calculations for MSbg give you.

But what is the theoretical meaning of MSbg in terms of the probability model? We want meaning in this section, not another way to calculate.

In the next two graphics we will find a disagreement in interpretation between  $H0$  and  $H1$ .

### **$H0$ AND $H1$ DISAGREE ABOUT MEANING OF MSbg**



## HO SAYS "ERROR"

The skeptic thinks the IV does not work and has no effect on the DV. The PCH of Chance and H0 come from the skeptic who believes that our three diets don't work. Therefore, these are just three random samples from the same probability distribution. By chance alone we'd expect the three means to vary around the grand mean.

**LINEAR MODEL IF HO IS TRUE.** Ho says that the alpha-j term of the linear model is 0 for all groups. So assuming Ho is true reduces the linear model to  $X_{ij} = \mu + e_{ij}$ .

So if Ho is true, the only possible source of variance is error variance. MSbg must estimate error variance because there is no other variance to estimate.

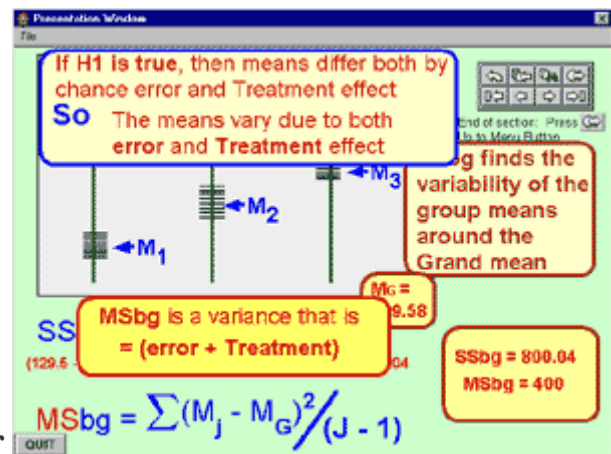
For the skeptic and Ho the only reason that there was variability in the means is error. Therefore MSbg is an estimate of population error variance.

## MSbg ESTIMATES ERROR VARIANCE IF HO IS TRUE.

How do we interpret MSbg if H1 is true?

## H1 SAYS "ERROR PLUS TREATMENT"

H1 follows from the scientific hypothesis. If H1 is true, if in fact there are a treatment effects, these treatment effects are making the scores in the various groups different from one another. In the diet example, the 10% Fat treatment is making the weights in Group 1 lower; and the 50% Fat treatment is making the weights in Group 3 higher. If the scores are lower or higher then, of course, the means will be lower or higher. Therefore the treatment effects are making the means vary. The treatment effects are creating variability in the means.



Treatment effects, if they exist, will make the means vary.

Error also will make the means vary.

So the Scientist and H1 assert that MSbg is an estimate of variance due to

both error and treatment.

Let's repeat that idea more formally.

**LINEAR MODEL IF H1 IS TRUE.** If H1 is true then some of the  $\alpha_j$ 's are not equal to 0. Some may be positive, others negative, but some of them are not equal to 0. So the linear model is  $X_{ij} = \mu + \alpha_j + \text{error}$ . So some of the groups will have their means affected by treatment effects. So treatment effects (where they exist) make the means vary.

In other words from the point of view of the scientist and H1, the variability that we're measuring between the group means and the grand means is due to both error and treatment effects.

**MS<sub>bg</sub> ESTIMATES ERROR PLUS TREATMENT VARIANCE IF H1 IS TRUE.**

### SUMMARY OF MEANING OF MS<sub>bg</sub>

H0 and H1 represent two rather different interpretations of what the mean square between groups is indicating.

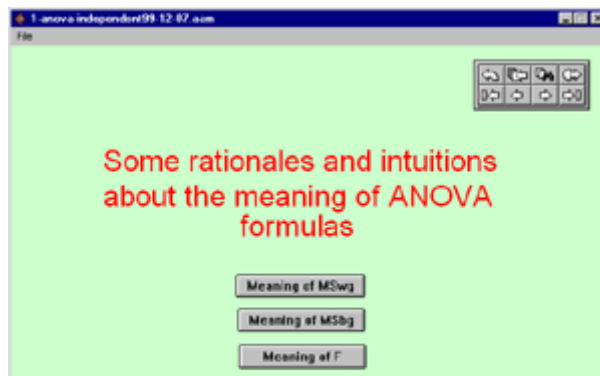
H0, the null hypothesis, says that the treatment effects are all zero and so the mean square between groups is simply due to error. MS<sub>bg</sub> is an estimate of error variance.

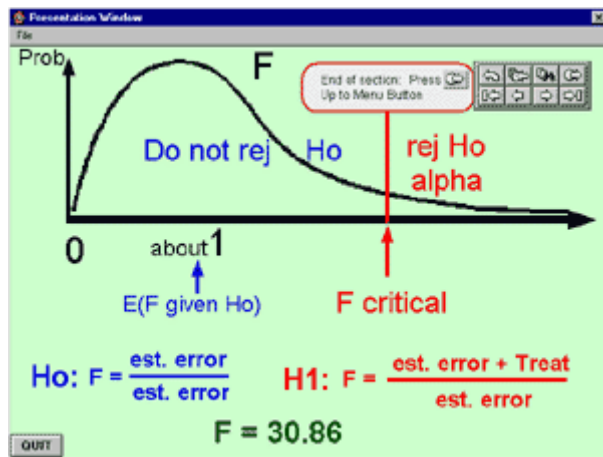
On the contrary, H1 says there ARE treatment effects and those treatment effects are forcing the means to be different from each other. Therefore, the mean square between groups is partly due to error (no one can contest that error makes the means different from each other) but its also due to treatment.

Now let's go on an look at the rationale for the the F ratio.

---

### The Meaning of the F Ratio





## Summary

**MEAN SQUARE WITHIN GROUPS.** The null and alternative hypotheses agree on the meaning of MS<sub>wg</sub>. They both contend that the mean square within groups is an estimate of population error variance.

**MEAN SQUARE BETWEEN GROUPS.** The null and alternative hypotheses disagree on the meaning of MS<sub>bg</sub>. If H<sub>0</sub> is true all alpha-j's are 0. So group means can only

differ by chance. Therefore MS<sub>bg</sub> only estimates error variance.

If H<sub>1</sub> is true, the means will vary both due to error and due to the effects of treatments. Therefore MS<sub>bg</sub> is affected by more than error; it is also affected by treatment.

## THE F RATIO

The F ratio is MS<sub>bg</sub> divided by MS<sub>wg</sub>

**H<sub>0</sub>.** If H<sub>0</sub> is true, both MS<sub>bg</sub> and MS<sub>wg</sub> are estimates of error variance. Therefore, if H<sub>0</sub> is true the F ratio is one estimate of error variance divided by another estimate of error variance. As a consequence we expect the calculated F ratio to be in the neighborhood of 1. These two estimates of error variance (MS<sub>bg</sub> and MS<sub>wg</sub>) won't be exactly the same due to chance variation, so we don't expect F to be exactly equal to one. But H<sub>0</sub> is expecting F to be in the neighborhood of 1.

[TECHNICAL NOTE. Mathematically we expect F to be "df-wg divided by (df-wg -2)" which is slightly less than 1. For our purposes in this introductory course we lose no intuitive insight by saying that we expect F to be around 1.]

### HO EXPECTS F TO BE NEAR 1.

On the other hand, the scientist is proposing the alternative hypothesis. This hypothesis says there are treatment effects that are making the means different. H<sub>1</sub> is making the claim that the top of the F ratio is an estimate of error plus treatment effects and the bottom of the ratio is just an estimate of error.

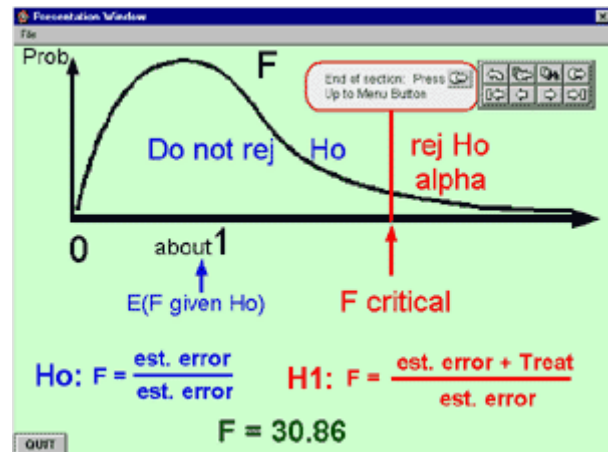
So the scientist thinks that the F test should be larger than one.

## H1 EXPECTS F TO BE LARGER THAN 1.

### REJECTING H0

#### STATISTICAL CONCLUSION VALIDITY

Let's review this slide again. Here is the sampling distribution of F. Notice that it starts at zero and goes to positive infinity. It is skewed out to the right, meaning it comes to a point (like a skewer) on the right side of the distribution. The F critical cuts off a region beyond which only .05 of the probability under the F curve lies.



The null hypothesis is expecting values toward the left side of the critical value.  $H_0$  thinks that the F ratio is just an estimate of error divided by another estimate of error and therefore ought to be close to a value of 1. Notice that the big wave of probability in this distribution is where the value of 1 is.

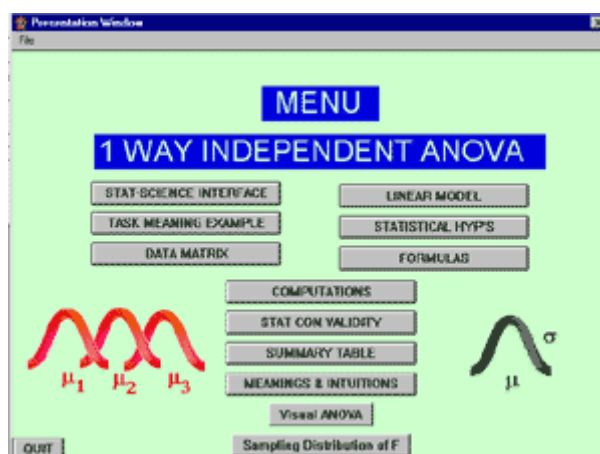
On the other hand the alternative is expecting F to be an estimate of error plus treatment effect divided by an estimate of error. The top of the equation should be larger than the bottom. Therefore the calculated value of F will be larger than one. The calculated value of F has reason to be out in the rejection region.

At this point the use of the F ratio to reject or not reject  $H_0$  should make sense. If  $H_0$  is true the overwhelming probability is that the calculated F should be near 1 well away from the rejection region. There only a small probability (alpha) that F would fall in the rejection region.

In contrast, if  $H_1$  is true, then F should be higher than 1 and falling in the rejection region should make sense.

In terms of the percent fat in diet example, our calculated  $F = 30.86$ . This means that the top of the F ratio ( $MS_{bg}$ ) is 30.86 times as large as the bottom ( $MS_{wg}$ ). Such a value makes you suspect that something is going on (like treatment) that is making the top systematically larger than the bottom. Therefore rejecting  $H_0$  makes sense. Of course the data we made up for the percent fat in diet example was extreme because we were making teaching points. Generally, you won't see F values that large.



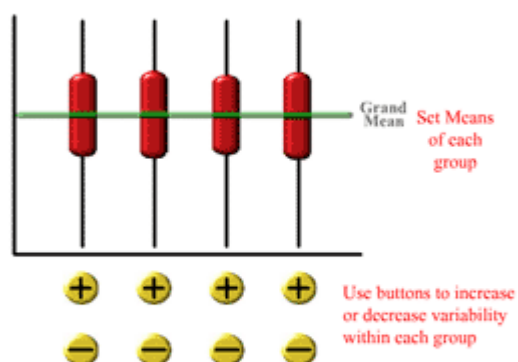


### Visual ANOVA tool

Visual ANOVA is a simple little program that lets you put all this theory we've been describing into a simple visual whole. It assumes that you've read the Meanings and Intuitions section and have have understood the the general ideas at least. Even if your understanding of the previous section is incomplete at this time, it is worth playing with Visual ANOVA since that may clear up the big picture for you. You can go back and forth between the Meanings and Intuitions section and Visual ANOVA tool.

### Understanding ANOVA Visually

MSBetween MSWithin Instructions



$$F = \frac{\text{Var Between Means}}{\text{Var Within Groups}} = \frac{MS_{\text{Bet}}}{MS_{\text{Within}}} = \frac{\text{ }}{\text{ }}$$

F =

### INSTRUCTIONS

Right below the "Understanding ANOVA Visually" title are three little buttons labeled MS between, MSwithin, and Instructions. Running your mouse over each of these button will bring up brief text to remind you of various concepts or to tell you the point of the Visual ANOVA tool.

The tool interface is a graph representing a four group study. The length of the red jelly bean icons represents how much variability there is within each of the four groups.

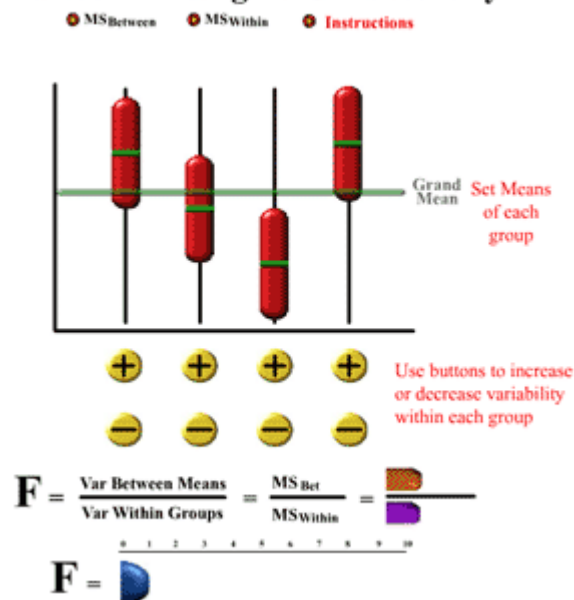
**DRAG THE RED JELLY BEANS.** You can click and drag the red jelly bean icons on the graph. Doing so will allow you to move each group mean up or down. That way you can increase the variability between the four group means.

**CLICK ON THE YELLOW BUTTONS.** Click on the + and - buttons for each

group. Doing so will increase or decrease the variability within each group.

The conceptual formula for F is shown below the graph. We'll talk about it in the next graphics.

### Understanding ANOVA Visually



#### HI BETWEEN AND HI WITHIN

The current graphic shows a case where the Visual ANOVA tool has been set so that the differences between the means are large. The variability within the groups is also set to be large.

**GROUP MEANS.** Notice that now you can see a green line in the middle of the red group icons. The green line represents the group mean.

**GRAND MEAN.** You also can see a long green line across the whole graph. It represents the mean of the groups means (the Grand Mean).

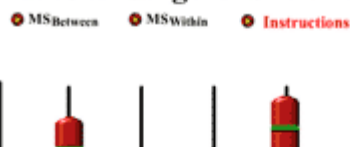
**MSbg DIVIDED BY MSwg.** Just below the yellow within group variability buttons, you can see a conceptual formula for F. Conceptually, the F ratio is variability between groups divided by variability within groups or MSbg divided by MSwg. **This F ratio is represented visually as length of a gold bar divided by the length of a purple bar.**

**F RATIO.** At the very bottom of the tool, **the value of F is represented by a large blue bar.** There is a scale from 0 to 10 above the blue bar so you can have some sense of how large the F value is.

**NO NUMBERS.** Other than the scale above the blue bar there are no numbers. The purpose of this tool is to get away from all the convoluted words and complex calculations and get you some experience playing visually with the holistic ideas which give all these numbers and words meaning.

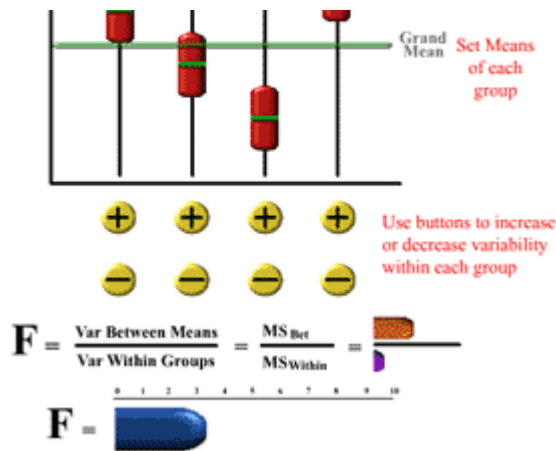
Notice that for the way the Visual ANOVA tool is set in this graphic, the gold MSbg bar is about the same length as the purple MSwg bar. So the blue F bar extends out to about 1 on the scale.

### Understanding ANOVA Visually



#### HI BETWEEN AND LO WITHIN

The current graphic is pretty much the same as the previous one, except that the variability within the groups has been



decreased.

Now you'll notice that the gold bar representing  $MS_{\text{bg}}$  is longer by about 3 or 4 times as the purple bar representing  $MS_{\text{wg}}$ . Consequently, blue bar is now out to about 3 on the scale.

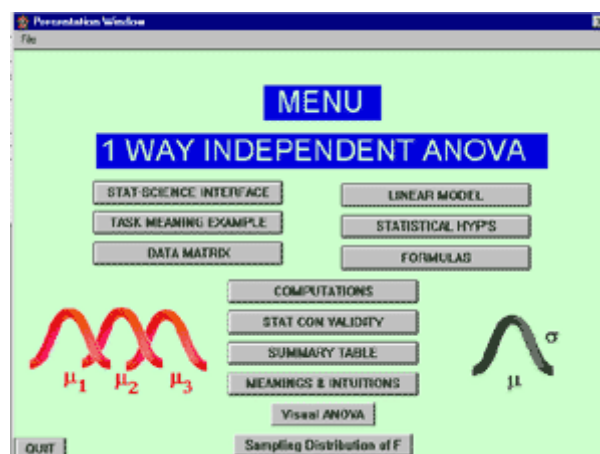
These lecture graphics are just static snapshots. Play with the Visual ANOVA tool to get a feel for how variability between groups and variability within groups interact

to change the value of the F ratio.

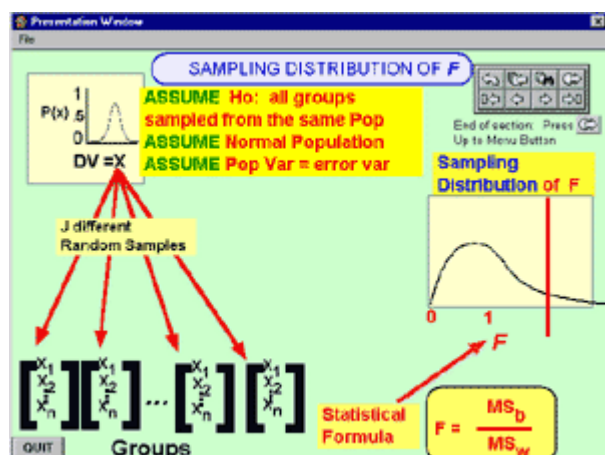
Have fun.

**DISCLAIMERS AND COMMENTS.** As we said, this tool is meant to direct your attention to relationships among the components of ANOVA by representing them visually. It is not meant to be a calculation device. In the programming, we have scaled various values so that they can be presented on the screen in a way that looks good rather than in a way that is highly accurate computationally. For example, F can actually vary from 0 to infinity. But on the tool F can only vary from 0 to 10. We placed similar restrictions on  $MS_{\text{bg}}$  and  $MS_{\text{wg}}$ .

Also, the red icons represent VARIABILITY as a concept. Their lengths are a transformation of actual variance values. These transformations are simply to make the graph work as a visual whole. Variance is a squared value and its length is very long compared the distance between means. The standard deviation was visually unappealing because it was too short. So the length of the red bars while an accurate representation of variability in general is not specifically the range nor the variance nor the standard deviation.



This section is an optional overview of the entire 1-ANOVA lecture



## OVERVIEW OF SAMPLING DISTRIBUTION LOGIC

Let's discuss the general 4-step sampling distribution schema which we have used throughout the course as it applies to ANOVA. We will do this without an example, just to get the big picture.

**NORMAL POPULATION.** We assume, as a start, that the dependent variable can be modeled by a normal distribution. That

means we think of our measurements as samples from a normal distribution. We call the population variance **ERROR VARIANCE**.

**MULTI-GROUP RESEARCH.** We set up our research study with 3 or more independent treatment groups. Scientists treat each group differently so we expect that the groups will perform differently. (You can, of course, apply ANOVA to a two-group study, but generally we use a t-test for that case.)

**NULL HYPOTHESIS.** If  $H_0$  is true, there is just a single population, and our study is sampling several random samples from this one population. [Remember that in statistics we assume that  $H_0$  is true so that we can test it.]

**HOMOGENEITY OF VARIANCE.** Even if the scientist is right and there are treatment effects and therefore treatment populations, all the groups have the same variance. We call that variance error variance.

**WITHIN GROUPS:** From the point of view of our study, everybody ought to perform or score exactly the same in the individual groups, because we are treating them all identically. Of course, they all end up with different scores, even within a single group with identical treatment, and so any kind of variance within the members of a single group is perceived of as error.

**BETWEEN GROUPS.** If the scientist is right, the groups ought to differ due to treatment. The skeptic says that the groups will differ only due error.

The F test for ANOVA is based on a very complicated statistical formula that can be summarized as **mean square between** divided by **mean square within**. The mean square between is based on the variance between the group means. The mean square within is based on the variance within the individual groups.

**If  $H_0$  is true, both the  $MS_{bg}$  and  $MS_{wg}$  are estimates of error variance. So the  $F$  ratio is one estimate of error divided by another estimate of error. Other than chance variation, these two estimates should be the same. So  $H_0$  expects  $F$  to be around 1.**

**This  $F$  statistic has a sampling distribution. Based on this  $F$  distribution, we can set up a critical region to evaluate just how far the calculated  $F$  is from 1, just how far it is from what  $H_0$  predicted. If it is too far from  $H_0$ 's prediction, we reject  $H_0$ .**

**End of one-way ANOVA independent lecture.**