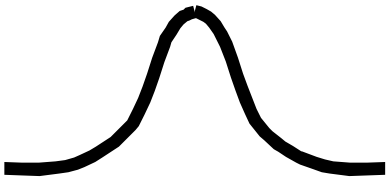
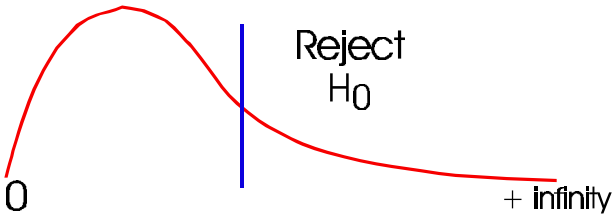
ONE FACTOR COMPLETELY RANDOMIZED ANOVA

Sampling Distribution of F

F is a test statistic





[ ][ ][ ][ ]

Test Statistic: **F =**

ONE WAY ANOVA FOR INDEPENDENT GROUPS

STATISTICS-SCIENCE INTERFACE

ANOVA is short for ...

RESEARCH CONTEXT

Several independent groups

Each group is treated differently

Examples:

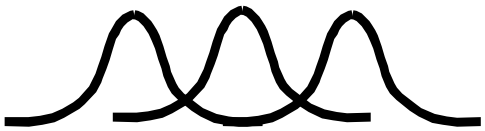
TREATMENT LEVELS

Example: Low Fat, Normal, and High Fat diet

MODEL USED BY SCEPTIC TO FORM H_0

Lo	Norm	Hi
137	145	159
146	166	182
:	:	:
Mean =135	148	169

MODEL USED BY SCIENTIST TO FORM H_1



	Lo	Norm	Hi
	137	145	159
	146	166	182
	:	:	:
Mean =	135	148	169

IMPORTANT ASSUMPTIONS IN ANOVA

NORMAL POPULATIONS

HOMOGENEOUS VARIANCES

H_0 IS TRUE

TREATMENT EFFECTS IN ANOVA

If Scientific Hypothesis is true...

Then different levels of ....

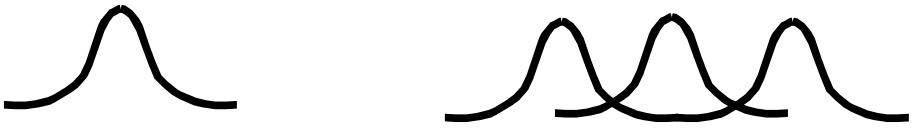
High, Normal and Low Fat Diet example

Low Fat treatment effect...

Normal treatment effect...

High Fat treatment effect...

Population of
Congolese Male
Weights



Low Fat treatment effect =

Normal treatment effect =

High Fat treatment effect =

TASK MEANING EXAMPLE

EXAMPLE:

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 subjects. 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 subjects 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 intellectual problem. (I'm not sure how she got this last group through the Human Subjects Review Committee.) Her performance measure is number of errors made in solving the puzzles.

Scientific hypothesis #1

Scientific hypothesis #2

Skeptical hypothesis

PCH OF CHANCE

DATA MATRIX:

Number Errors during Problem Solving		
Neutral	***Positive***	***Negative***
3	2	5
4	3	7
5	4	6
6	2	4
4	0	5
3	2	6
2	3	5
5	2	4
	1	3
	1	
M_1=4	**M_2=2**	**M_3=5**

Does the data pattern fit the scientific hypothesis or the anti-scientific hypothesis?

BUT! is this data pattern due to chance?

ANOVA can deal with this PCH OF CHANCE.

BACK TO THE GENERALITIES OF ANOVA

DATA MATRIX

What does Xij mean?

Subject #	Group 1	Group 2	Group 3
	X11=	X12=	X13=
	X21=	X22=	X23=
	X31=	X32=	X33=
	X41=	X42=	X43=
	M1	M2	M3

Grand Mean

Some vocabulary:

Treatment means (M1, M2, M3)

Grand mean

x_{ij}

From the example, what is the value of $x_{7\,2}$?

Some sums

means ...

means ....

means ....

LINEAR MODEL FOR AN INDIVIDUAL'S SCORE (x_{ij})

x_{ij} =

means...

means...

means...

DEMONSTRATE THE USE OF MODEL TO GENERATE DATA MATRIX

Diet Example

In reality we don't know values of parameters...

But... Let

$\mu = 150$

$\alpha_1 = -20$

$\alpha_2 = 0$

$\alpha_3 = +20$

Table of errors (εij)			
Subject #	**Group 1**	**Group 2**	**Group 3**
1	+3	-3	-3
2	-4	+5	+5
3	+1		
4			

If an individual's score were totally determined by baseline

x_{ij} =

Table of DATA			
Subject #	**Group 1**	**Group 2**	**Group 3**
1			
2			
3			
4			

If an individual's score were determined by baseline plus treatment effect of the individual's group.

x_{ij} =

Table of DATA			
Subject #	**Group 1**	**Group 2**	**Group 3**
1			
2			
3			
4			

If a score is determined by baseline, treatment, and random error

x_{ij} =

Table of DATA			
Subject #	**Group 1**	**Group 2**	**Group 3**
1			
2			
3			
4			

STATISTICAL HYPOTHESES

COMPUTATIONAL FORMULAS

SUMS OF SQUARES

Between groups

Within groups

Total

$SS_{TOT} = SS_{BET} + SS_{WITHIN}$

DEGREES OF FREEDOM

MEAN SQUARES

F RATIO

F =

F IS THE TEST STATISTIC

EXAMPLE COMPUTATIONS

SS_{TOT} =

SS_{BET} =

SS_{WITHIN} =

df

MEAN SQUARES

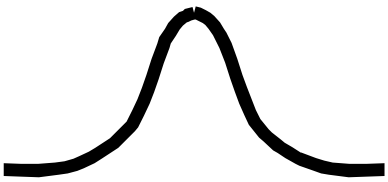
F Ratio

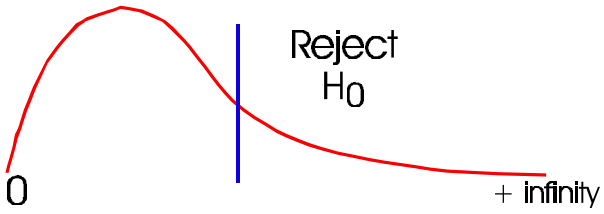
F =

ANOVA SUMMARY TABLE

ANOVA Summary Table					
Source of Variance	**df**	**SS**	**MS**	**F**	***p***
Total					
Between					
Within (error)					

STATISTICAL CONCLUSION VALIDITY





[ ] [ ] [ ] [ ] **F =**

USE F TABLES TO FIND CRITICAL F values

Need to know **df**(between), **df**(within) and α level to use tables

BUT WHAT ABOUT SPECIFIC DIFFERENCE BETWEEN TREATMENT MEANS?

MEANINGS & INTUITIONS

MEANING OF MEAN SQUARE **WITHIN** GROUPS

Variances measure...

Little **s** =

Top of formula is called...

Bottom of formula is called...

3-group example

With s we can measure the variability within each group

MS within is...

FYI (Don’t need to know this except for general idea)

Both the Null and Alternative hypotheses say...

So the meaning of MS within is...

MEANING OF MEAN SQUARE **BETWEEN** GROUPS

3 group example

SS between finds the variability between...

The Null Hypothesis says...

So...

But if the Alternative Hypothesis is true, then...

MS between means...

Comparison of meaning of MS between for H_0 and H_1	
H_0 says	H_1 says