INTRODUCTION TO ANOVA

Examples

Example

Caffeine (Michael T. Brannick http://faculty.cas.usf.edu/mbrannick/)

Computational Example: Caffeine

G1: Control	G2: Mild	G3: Jolt
	Test Scores	
75	80	70
77	82	72
79	84	74
81	86	76
83	88	78

Computational Example: Caffeine

G1: Control	G2: Mild	G3: Jolt
	Test Scores	
75=79-4	80=84-4	70=74-4
77=79-2	82=84-2	72=74-2
79=79+0	84=84+0	74=74+0
81=79+2	86=84+2	76=74+2
83=79+4	88=84+4	78=74+4
	Means	
79	84	74
Mean of means	79	

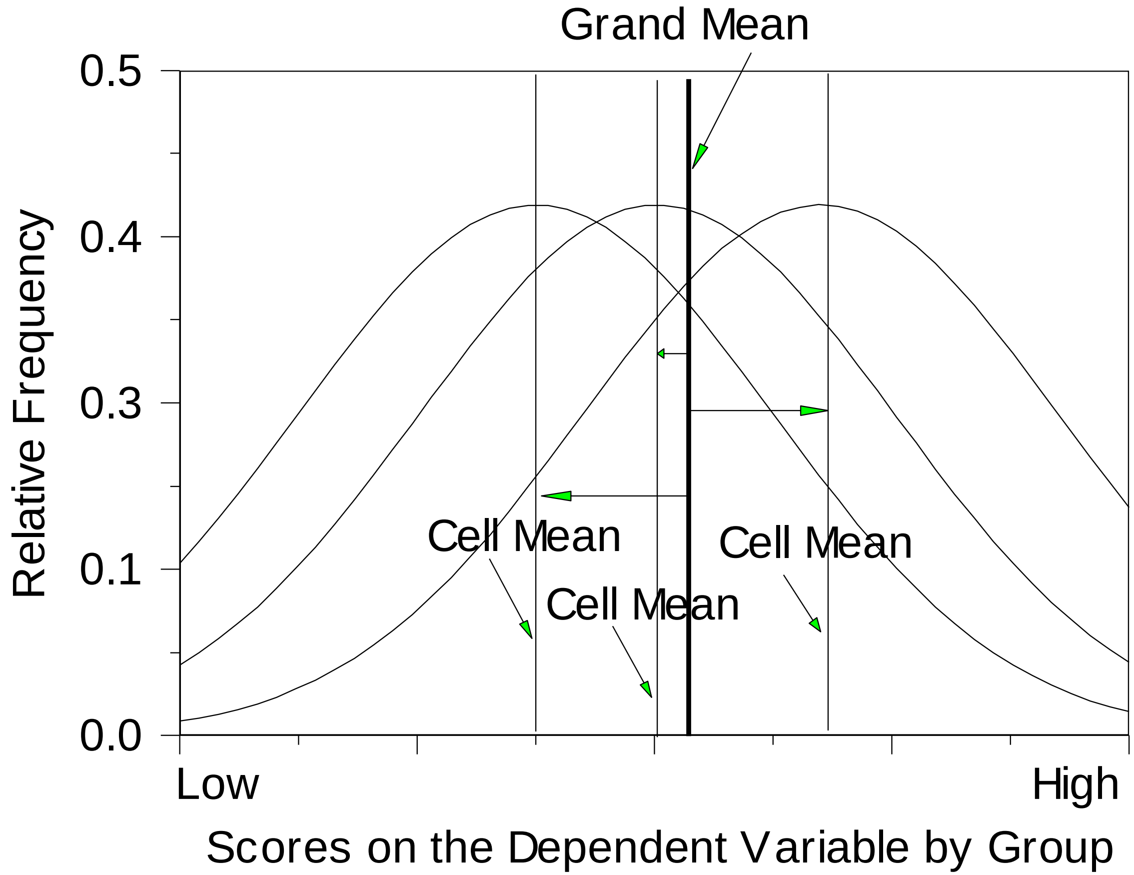
Between Sum of Squares

$$SS_B = \sum N_A(\bar{X}_A - \bar{X}_G)^2$$

	$\bar{X}_A$	$\bar{X}_G$	$(\bar{X}_A - \bar{X}_G)^2$
G1	79	79	0
Control	79	79	0
M=79	79	79	0
SD=3.16	79	79	0
	79	79	0
G2	84	79	25
M=84	84	79	25
SD=3.16	84	79	25
	84	79	25
	84	79	25
G3	74	79	25
M=74	74	79	25
SD=3.16	74	79	25
	74	79	25
	74	79	25
Sum			250

The between sum of squares relates the Cell Means to the Grand Mean. This is related to the variance of the means.

$$SS_B = \sum N_A(\bar{X}_A - \bar{X}_G)^2$$



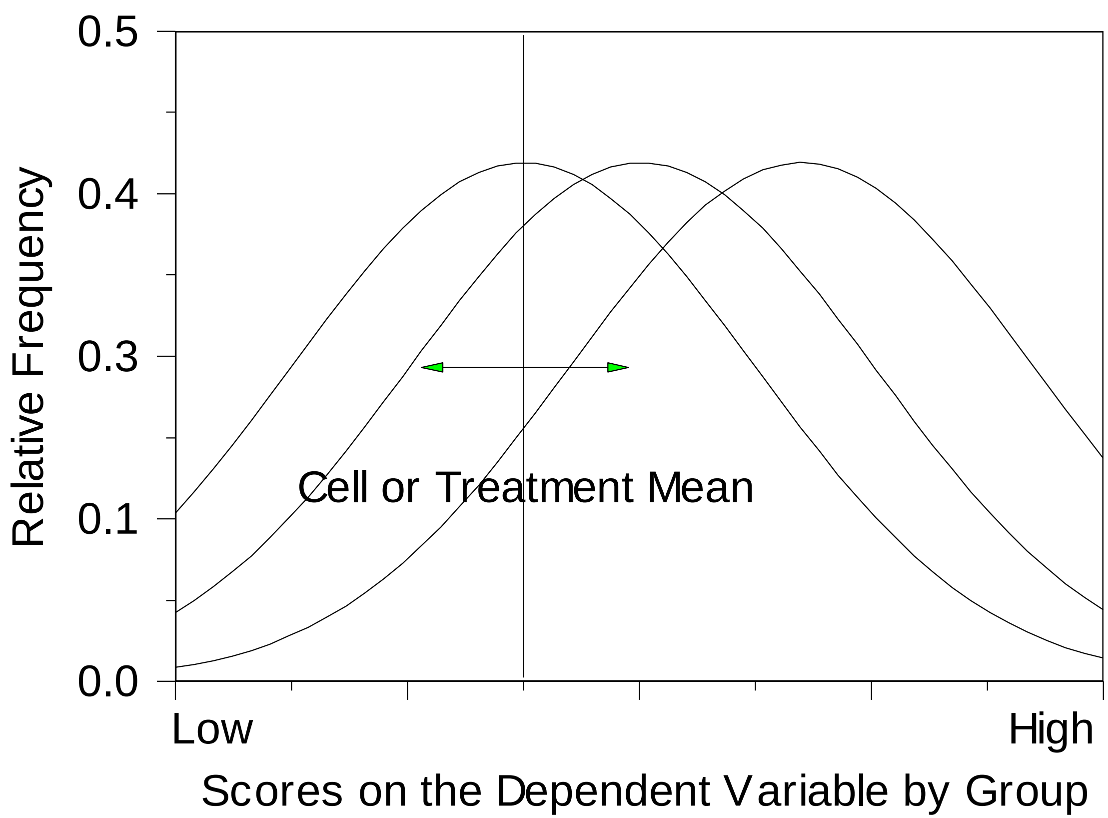
Within Sum of Squares

$$SS_W = \sum (X_i - \bar{X}_A)^2$$

	X_i	$\bar{X}_A$	$(X_i - \bar{X}_A)^2$
G1	75	79	16
Control	77	79	4
M=79	79	79	0
SD=3.16	81	79	4
	83	79	16
G2	80	84	16
M=84	82	84	4
SD=3.16	84	84	0
	86	84	4
	88	84	16
G3	70	74	16
M=74	72	74	4
SD=3.16	74	74	0
	76	74	4
	78	74	16
Sum			120

Within sum of squares refers to the variance within cells. That is, the difference between scores and their cell means. SS_W estimates error.

$$SS_W = \sum (X_i - \bar{X}_A)^2$$



- Calculate the MSE

$$MSE = \frac{1}{N-K}\sum_{j}\sum_{i}\left(x_{ij} - \overline{X}_{j}\right)^{2} = SS_{w} / N - K$$

- 120/(15-3)=10

$$F = \frac{SSB/(K-1)}{MSE}$$

- F=250/(3-1)/10=12.5

ANOVA Source Table (1)

Source	SS	df	MS	F
Between Groups	250	k-1=2	SS/df 250/2= 125 $=MS_B$	F = MS_B/MS_W = 125/10 =12.5
Within Groups	120	N-k= 15-3=12	120/12 = 10 = MS_W	
Total	370	N-1=14		

ANOVA Source Table (2)

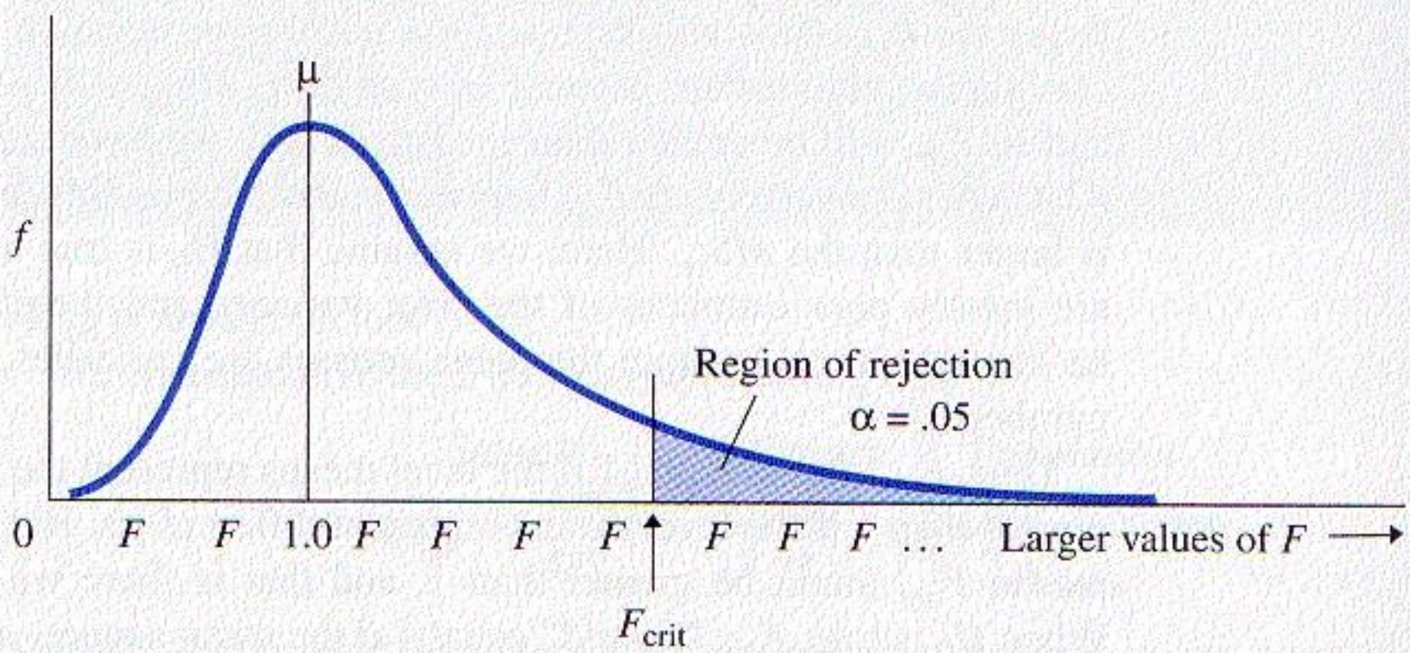
- *df* – Degrees of freedom. Divide the sum of squares by degrees of freedom to get
- *MS*, Mean Squares, which are population variance estimates.
- *F* is the ratio of two mean squares. *F* is another distribution like *z* and *t*. There are tables of *F* used for significance testing.

The F Distribution

- F" distribution, with K-1 and N-K degrees of freedom (where K is the number of groups and N is the total number of observations)
- K=3, N=15 → F(2,12) (working at 0.05)

The F Distribution

FIGURE 17.2 Sampling Distribution of F When H_0 Is True



F Table – Critical Values

	Numerator df: df_B				
df_W	1	2	3	4	5
5 5% 1%	6.61 16.3	5.79 13.3	5.41 12.1	5.19 11.4	5.05 11.0
10 5% 1%	4.96 10.0	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64
12 5% 1%	4.75 9.33	3.89 6.94	3.49 5.95	3.26 5.41	3.11 5.06
14 5% 1%	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.04	2.96 4.70

Review 6 Steps

- Set alpha (.05).
- State Null & Alternative
- H0: $\mu_1 = \mu_2 = \mu_3$
- H1: not all μ are =.
- Calculate test statistic: F=12.5
- Determine critical value F.05(2,12) = 3.89
- Decision rule: If test statistic > critical value, reject H0.
- Decision: Test is significant (12.5>3.89). **population are different.**