Inferential Statistics

Analysis of Variance
(ANOVA)

What We Will Cover in This Section

- Introduction.
- Overview.
- Simple ANOVA.
  - Theory
  - Practice

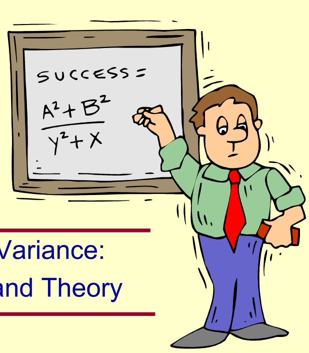
Situation

The management of Saul T. Dogg's restaurant wanted to see if the saltiness of appetizers would influence the number of drinks people purchased. Three sections of the club are targeted to receive appetizers that have either low, medium, or high saltiness. The dependent variable is the number of drinks ordered.

Appetizer saltiness and number or drinks ordered.		
Group 1 Low Salt	Group 2 Medium Salt	Group 3 High Salt
2	3	3
3	4	1
1	4	2
1	5	2
2	6	3
2	4	1
1	3	1
2	2	2
2	4	1
4	4	2
M = 2.00	M = 3.90	M = 1.80

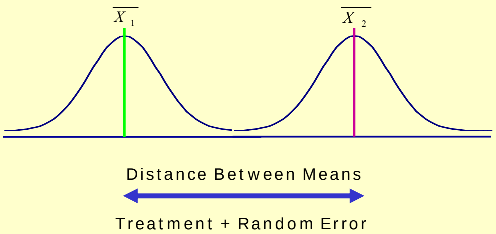
Issue

How to determine if one mean is significantly different from the other means while minimizing the probability of committing a Type I error.

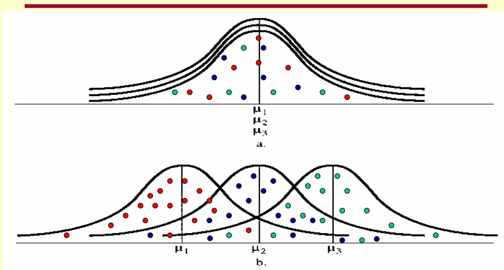


Analysis of Variance: Background and Theory

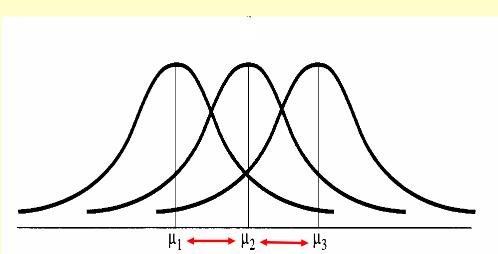
t-test Logic



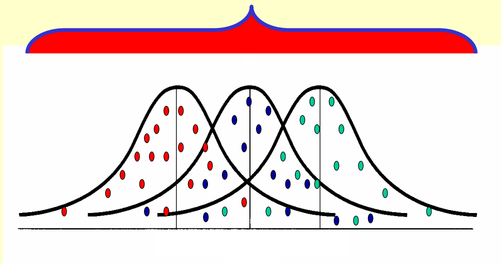
ANOVA Logic

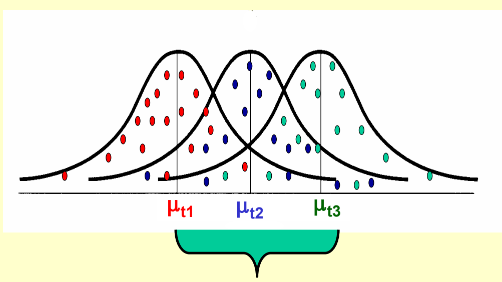


Treatment Effects



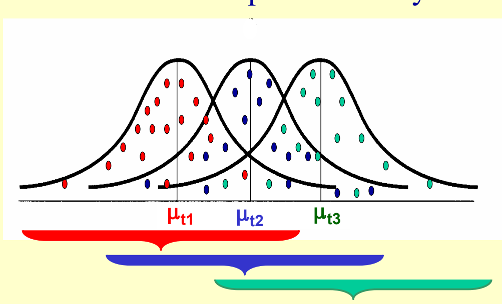
Total Variability



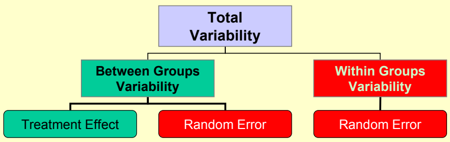


Between Groups Variability

Within Groups Variability



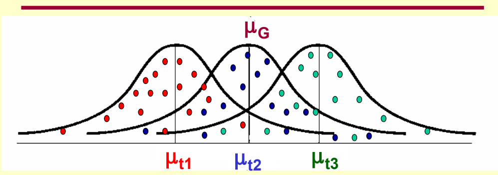
Partitioning the Variance



Where We Are Going

$$\frac{(\text{Treatment}) + (\text{Random Error})}{\text{Random Error}}$$

ANOVA Model



Total Effect = Treatment Effect + Random Error

Total Variability = Between Groups Variability + Within Groups Variability

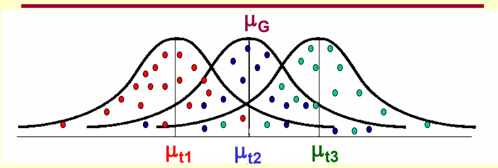
Partitioning the Variance

Total Variability = Between Groups Variability + Within Groups Variability

$$\frac{SS_{Total}}{df_{Total}} = \frac{SS_{Between\,Groups}}{df_{Between}} + \frac{SS_{Within\,Groups}}{df_{Within}}$$

(N - 1) (k – 1) (N – k)

The F ratio



$$F_{(k-1,N-k)} = \frac{MS_B}{MS_W}$$

F-test Compared to t-test

$$t_{(df)} = \frac{X_1 - X_2}{\sigma_{\bar{X}}} \qquad F_{(k-1,N-k)} = \frac{MS_B}{MS_W}$$

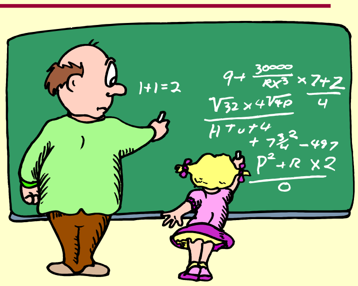
What Does the F Mean?

$$F = \frac{\text{Treatment Variability} + \text{Random Error}}{\text{Random Error}}$$

The Critical Value of F

- See page 485 in text book.
- Notice
  - Need *df* $_{between}$ (numerator) for columns.
  - Need *df* $_{within}$ (denominator) for rows.
  - As *df* increases the critical values get smaller.

The Problem Returns



The Situation (in case you forgot)

The management of Sal T. Dogg's restaurant wanted to see if the saltiness of appetizers would influence the number of drinks people purchased. Three sections of the club are targeted to receive appetizers that have either low, medium, or high saltiness. The dependent variable is the number of drinks ordered.

Hypotheses

Research Hypothesis.

Saltiness of the appetizers will influence the number of drinks that people buy.

Null Hypothesis.

Saltiness will not influence the number of drinks that people buy.

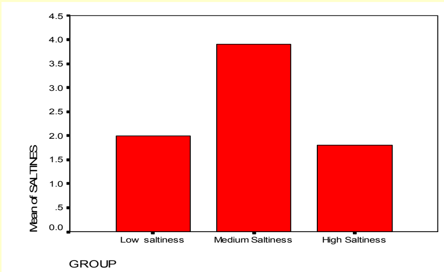
Statistical Hypothesis.

$\mu_1 \neq \mu_2 \neq \mu_3$

Appetizer saltiness and number or drinks ordered.

Group 1 Low Salt	Group 2 Medium Salt	Group 3 High Salt
2	3	3
3	4	1
1	4	2
1	5	2
2	6	3
2	4	1
1	3	1
2	2	2
2	4	1
4	4	2
$\bar{X}$ = 2.00	$\bar{X}$ = 3.90	$\bar{X}$ = 1.80

Graph of Saltiness Ratings



ANOVA Summary Table

Source	SS	df	MS	$F_{(crit=3.35)}$
Between Groups	26.87	2	13.435	14.77
Within Groups	24.50	27	.91	
Total	51.37	29		

How to Express F

$F_{(2,27)} = 14.77, \; p<.05$

- Degrees of freedom (between, within)
- Calculated F
- Alpha

Post Hoc Tests

When F is significant, how do you determine which of the means differs from the others?

Tukey Honestly Significant Difference Test (HSD)

$$HSD = q_{(\alpha, df_{within}, k)}\sqrt{\frac{MS_{within}}{n}}$$

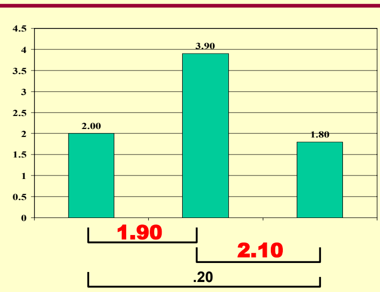
$$HSD = q_{(.05,27,3)}\sqrt{\frac{.91}{10}}$$

$$HSD = 3.53 x \sqrt{.091}$$

$$HSD = 1.06$$

q = Value from table, page 488
α = desired significance level
df_{within} = within groups df
k = Number of groups.

How To Use Tukey ($HSD = 1.06$)



Effect Size: Eta Squared (η^2)

$$\eta^2 = \frac{SS_{Between}}{SS_{Total}}$$

$$\eta^2 = \frac{26.87}{51.37}$$

$$\eta^2 = .523$$

Assumptions

1. The observations within each sample are independent.
2. The population from which the samples are selected is normally distributed.
3. The population from which the samples are selected have equal variances (homogeneity of variance)

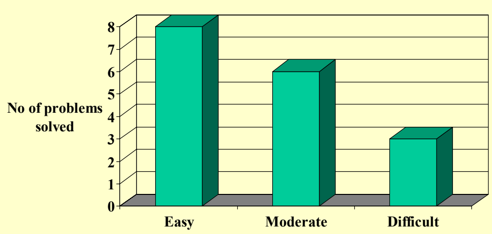
Another ANOVA Example

Sal O. Gysm felt that the perceived difficulty of logic problems would influence performance on these problems. Sal developed a set of problems and gave them to three groups. One group was told that the problems was easy, another was told that they were moderately difficult, and the third was told that they were difficult. The dependent variable was the number of problems solved.

ANOVA: Example 2

Easy	Moderate	Difficult
9	4	1
12	6	3
4	3	4
8	2	5
7	10	2
M = 8.0	M = 6.0	M = 3.0

Logic Problem Results



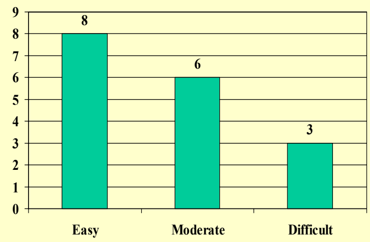
ANOVA: Summary Table

Source	SS	df	MS	F
Between	63.33	2	31.67	4.52
Within	84.00	12	7.00	
Total	147.33	14		

Is this value significant?

Post hoc Analysis: Tukey HSD

$$HSD = 4.46$$



Effect Size: Eta² (η^2)

$$\eta^2 = \frac{SS_{between}}{SS_{total}}$$

$$\eta^2 = \frac{63.33}{147.33}$$

$$\eta^2 = .428$$

Key Learning Points

1. The F test is used for comparing three or more means.
2. The Total Variance is broken down into two components
   - Between Group Variance (Treatment plus error)
   - Within Group Variance (Error)
3. The F ratio is computed by dividing ***Mean Square Between*** by the ***Mean Square Within.***

Key Learning Points

4. The F ratio tells you ONLY if all of the means together are different.
5. The post hoc tests tell you which pairs of means differ.
6. Eta^2 tells you the amount of variability accounted for in the treatment.

Another Practice Problem

Tess Tosterone is studying aggression among adolescent girls. She believes that there is a relationship between the level of interaction a girl has with her mother and the girl's level of aggression. She has identified fifteen girls who fall into one of three maternal interaction levels (low, medium, and high) and has measured their aggression scores.

The scores are shown on the next slide.

Data Summary Table

Low interaction	Moderate Interaction	High Interaction
6	6	0
5	8	4
9	5	0
4	4	1
6	2	0
M = 6.00	M = 5.00	M = 1.00

ANOVA Summary Table

	Sum of Squares	df	Mean Square	F
Between	70.00	2		
Within				
Total	116.00	14		

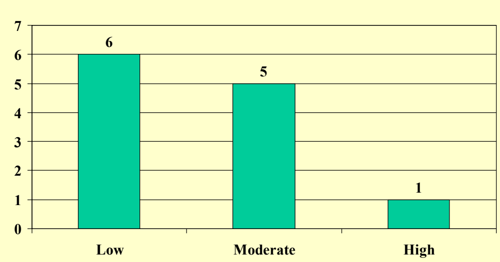
Tukey HSD

$$HSD = q_{(\alpha, df_{within}, k)} \sqrt{\frac{MS_{within}}{n}}$$

$$3.77\sqrt{\frac{3.833}{5}}$$

HSD = 3.30

Which Means are Different?



Eta^2

$$\eta^2 = \frac{SS_{between}}{SS_{total}} \qquad \frac{70.00}{116.00}$$

.60