RECALL THE T-TEST SETTINGS

t-test comes in three types:

1. A sample mean against a hypothesis (unknown population variance, small sample, normal population).

2.Two sample means compared to each other (two sample t-test).

3.Two means within the same sample (pairwise t-test).

TWO POPULATIONS → MULTIPLE?

We tested whether the means of two populations are equal.

Extend this to the comparison of more than two means.

Then, t-test is no longer useful here.

We will discuss two designs:

- Completely Randomized Design (CRD)
- Randomized Complete Block Design (RCBD)

COMPLETELY RANDOMIZED DESIGN

Example 1: There are *c* advertising messages (one factor, *c* levels, *c* treatments) in Internet; differ with respect to text, background color, font size, etc. [Leodolter and Swersey (2007), Chapter3]

Offer these to *distinct Internet users at random*; each user (experimental unit) responses one and only one advertising message.

Response is the sales volume generated from each advertising message, or the "hit ratio" (the proportion of those who access a particular Web site in response to the message).

COMPLETELY RANDOMIZED DESIGN

Example 2: A firm wants to test three different in-store promotions for a major product (one factor, *three* levels, *three* treatments) and identifies a group of 15 stores of similar size (experimental units) to participate in the experiment.

Each store will test one and only one of the promotions for a certain period of time (say three weeks).

The promotions are randomly assigned to the stores, with five different stores per promotion.

Since the treatments are assigned to the experimental units at random, we call this a completely randomized experiment.

RANDOMIZED COMPLETE BLOCK DESIGN

Example 2 (modified): Suppose the firm believes the 15 stores are not homogeneous, namely, there possibly are store effects (e.g, one particular store makes its sales particularly high under all three promotions).

Then, it would be difficult to recognize differences among the treatments, due to the additional noise from store effects.

It may be better to observe each store under all three in-store promotions (e.g., assign each promotion randomly to three one-week periods).

In this design, each of the 15 stores acts as a *block*.

Within each block, treatments are assigned at random.

It is a 'Complete' Block Design, because each of the 3 promotions is assigned to every block.

NOW, CRD

Example 1: Suppose there are three (c=3) different promotions.

Each promotion was used in five different stores. Promotions are stores were randomly assigned.

Sales volume for the week was measured and compared to the base sales of that store.

Percentage changes were calculated, given in Table next slide.

NOW, CRD

Example 1: Suppose there are three (c=3) different promotions.

Promotions			
	1	2	3
Sales vol. diff. %	9.5	8.5	7.7
	3.2	9.0	11.3
	4.7	7.9	9.7
	7.5	5.0	11.5
	8.3	3.2	12.4
Sample size	5	5	5
Sample mean	6.64	6.72	10.52
Sample variance	6.82	6.28	3.43

MORE GENERALLY, WE GET

	Groups		
	1	2	c
Response			
Sample size			
Sample mean			
Sample variance			

ONE-WAY ANALYSIS OF VARIANCE

Evaluate the difference among the means of three or more groups

Examples: Accident rates for 1st, 2nd, and 3rd shift
The productivity of three groups of workers

Assumptions

- Populations are normally distributed
- Populations have equal variances
- Samples are randomly and independently drawn

HYPOTHESES OF ONE-WAY ANOVA

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$$

- All population means are equal
- i.e., no factor effect (no variation in means among groups)

$$H_1 : \text{Not all of the population means are the same}$$

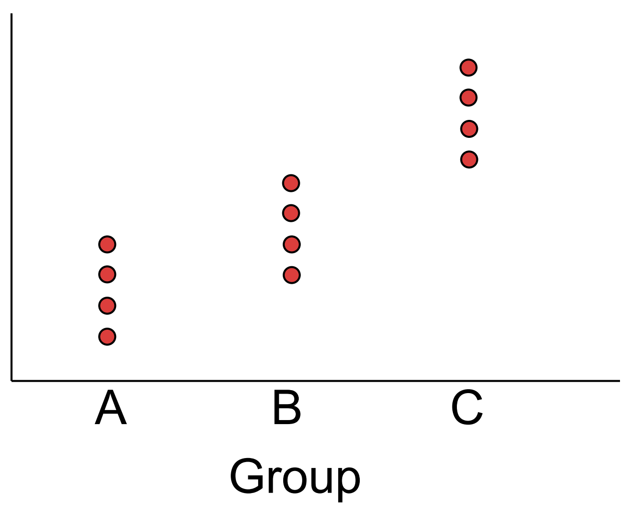
- At least one population mean is different
- i.e., there is a factor effect
- Does not mean that all population means are different (some pairs may be the same)

VARIABILITY

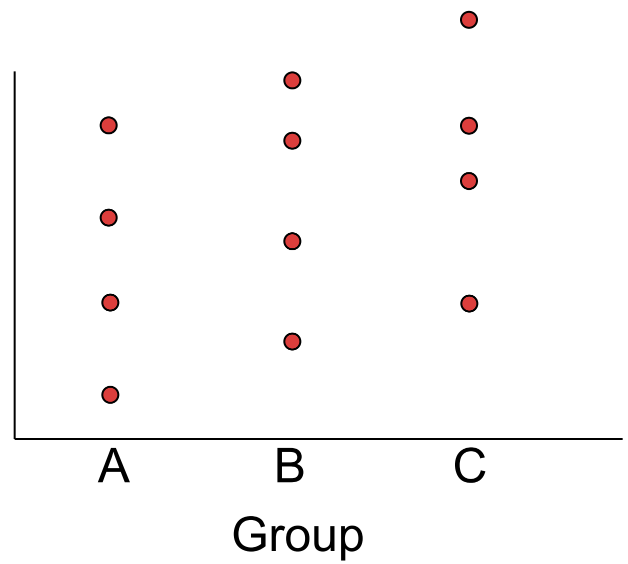
The variability of the data is key factor to test the equality of means

In each case below, the means may look different,

but a large variation makes the evidence that the means are different weak: Again, one should look at 'signal to noise ratio.'



Small variation within groups



Large variation within groups

ONE-WAY ANOVA

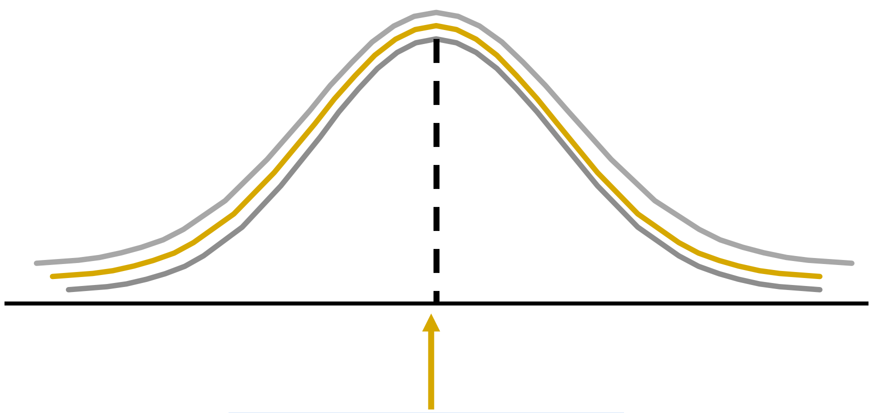
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$$

$$H_1 : \text{Not all } \mu_j \text{ are the same}$$

The Null Hypothesis is True

All Means are the same:

(No Factor Effect)



$$\mu_1 = \mu_2 = \mu_3$$

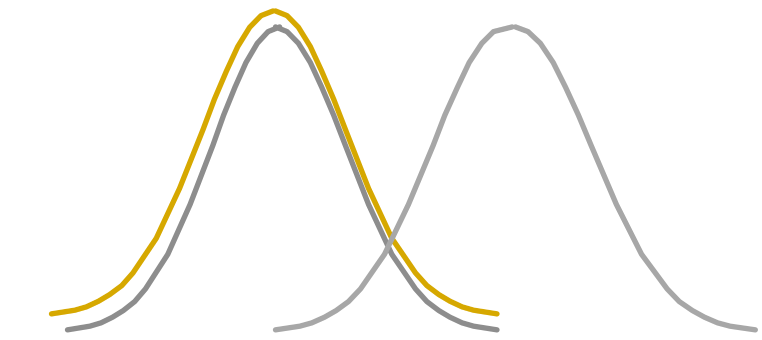
ONE-WAY ANOVA

(continued)

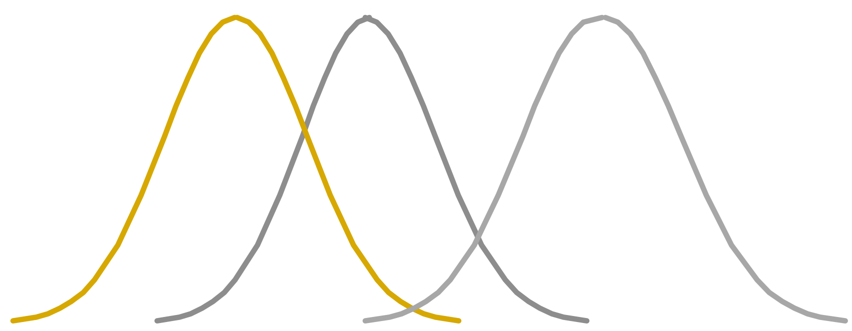
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$$

$$H_1 : \text{Not all } \mu_j \text{ are the same}$$

The Null Hypothesis is NOT true
At least one of the means is different
(Factor Effect is present)



or



$\mu_1 = \mu_2 \neq \mu_3$ $\quad$ $\mu_1 \neq \mu_2 \neq \mu_3$

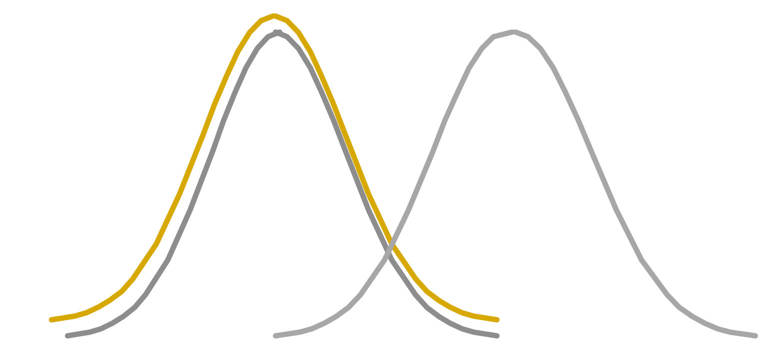
TEST IDEA: PARTITIONING THE VARIATION!

(continued)

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$$

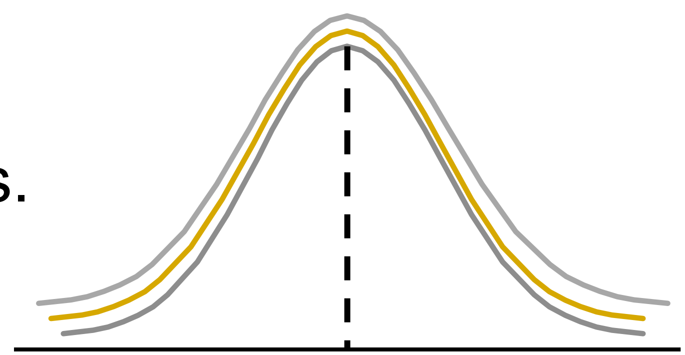
$$H_1 : \text{Not all } \mu_j \text{ are the same}$$

Under the null, we will examine the distribution of signal (between-group variation) to noise (within group-variation) ratio.



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

vs.



$$\mu_1 = \mu_2 = \mu_3$$

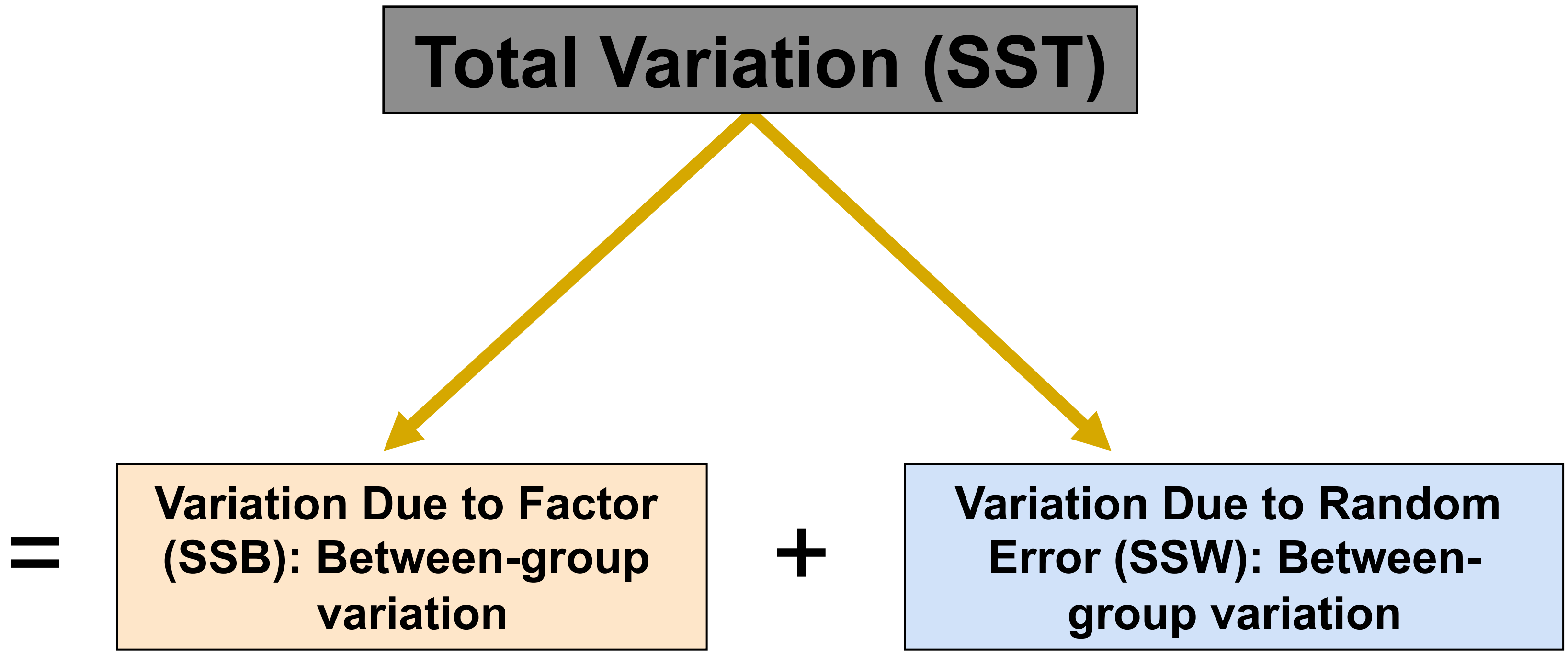
TEST IDEA: PARTITIONING THE VARIATION (HENCE, THE TERM *ANOVA*)

Total variation can be split into two parts:

$$\boxed{SST = SSB + SSW}$$

- SST = Total sum of [the observations minus the overall mean squared]
  (Total variation)
- SSB = Sum of Squares Between Groups
  (Between-group variation)
- SSW = Sum of Squares Within Groups
  (Within-group variation)

PARTITION OF TOTAL VARIATION



- The treatment (factor) component is the difference between the treatment mean and the overall mean. →*If treatment matters, this should be big? How big is big?*
- The error component is the difference between the observations and the treatment mean, i.e., the variation not explained by the treatments.

TOTAL SUM OF SQUARES

$$\boxed{SST} = SSB + SSW$$

$$\boxed{SST = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (X_{ij} - \bar{\bar{X}})^2}$$

Where:

SST = Total sum of squares

c = number of groups or levels

n_j = number of observations in group j

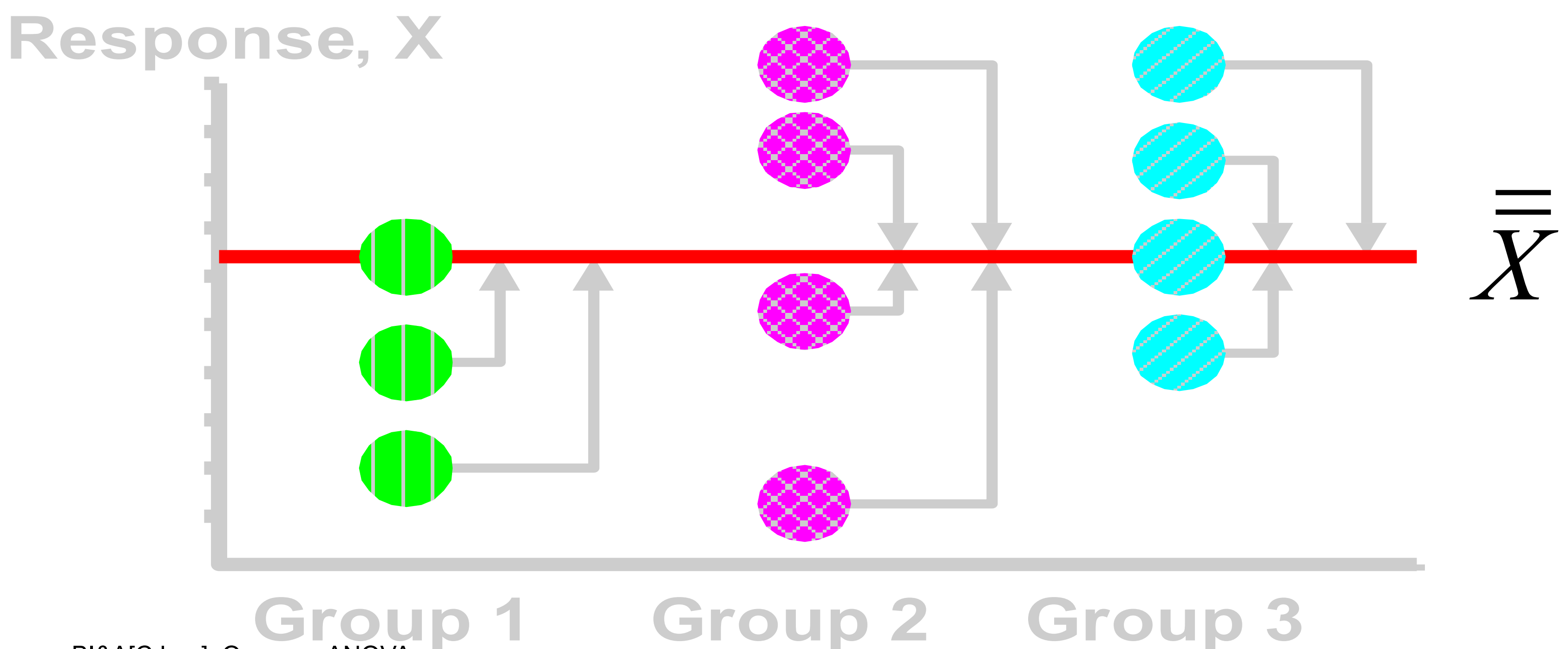
X_{ij} = i^{th} observation from group j

$\bar{\bar{X}}$ = overall mean (mean of all data values)

TOTAL VARIATION

(continued)

$$SST = (X_{11} - \bar{\bar{X}})^2 + (X_{12} - \bar{\bar{X}})^2 + \cdots + (X_{cn_c} - \bar{\bar{X}})^2$$



BETWEEN-GROUP VARIATION

$$SST = SSB + SSW$$

$$SSB = \sum_{j=1}^{c} n_j (\bar{X}_j - \bar{\bar{X}})^2$$

Where:

SSB = Sum of squares between groups

c = number of groups

n_j = sample size from group j

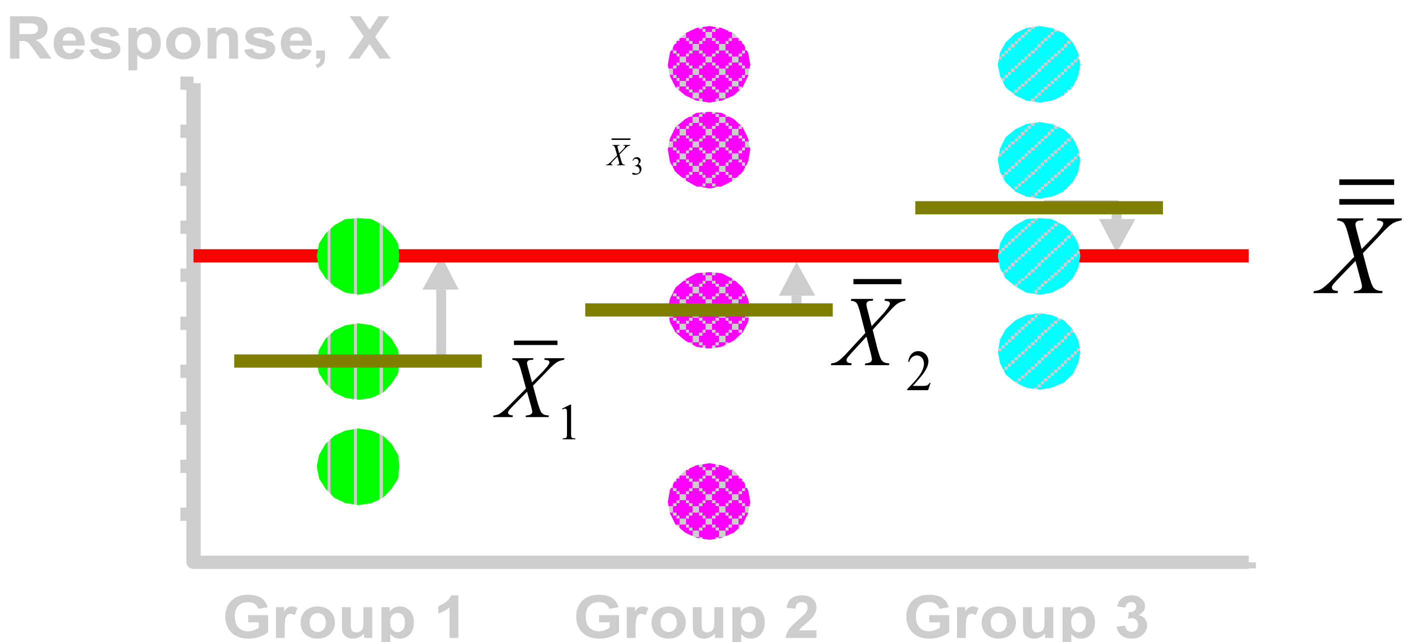
$\bar{X}_j$ = sample mean from group j

$\bar{\bar{X}}$ = grand mean (mean of all data values)

BETWEEN-GROUP VARIATION

(continued)

$$SSB = n_1(\bar{X}_1 - \bar{\bar{X}})^2 + n_2(\bar{X}_2 - \bar{\bar{X}})^2 + \cdots + n_c(\bar{X}_c - \bar{\bar{X}})^2$$

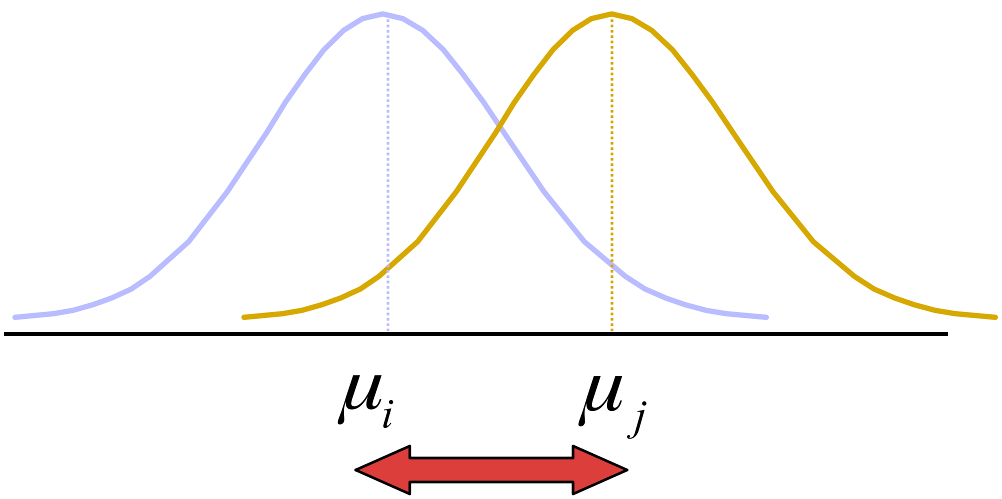


BETWEEN-GROUP VARIATION

(continued)

$$SSB = \sum_{j=1}^{c} n_j (\overline{X}_j - \overline{\overline{X}})^2$$

Variation Due to Differences Among Groups



$$\mathrm{MSB} = \frac{\mathrm{SSB}}{\mathrm{c} - 1}$$

Mean Square Between = SSB/degrees of freedom

WITHIN-GROUP VARIATION

$$SST = SSB + SSW$$

$$SSW = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$

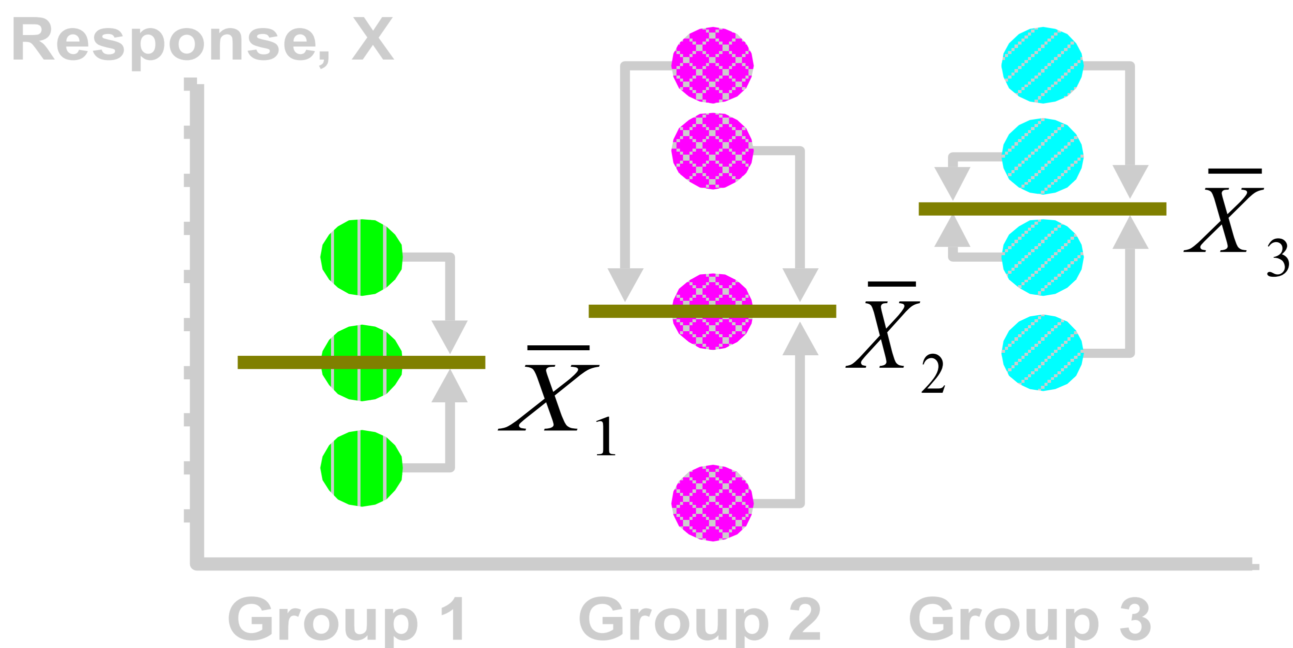
Where:

- SSW = Sum of squares within groups
- c = number of groups
- n_j = sample size from group j
- $\bar{X}_j$ = sample mean from group j
- X_{ij} = i^{th} observation in group j

WITHIN-GROUP VARIATION

(continued)

$$SSW = (X_{11} - \bar{X}_1)^2 + (X_{12} - \bar{X}_2)^2 + \cdots + (X_{cn_c} - \bar{X}_c)^2$$

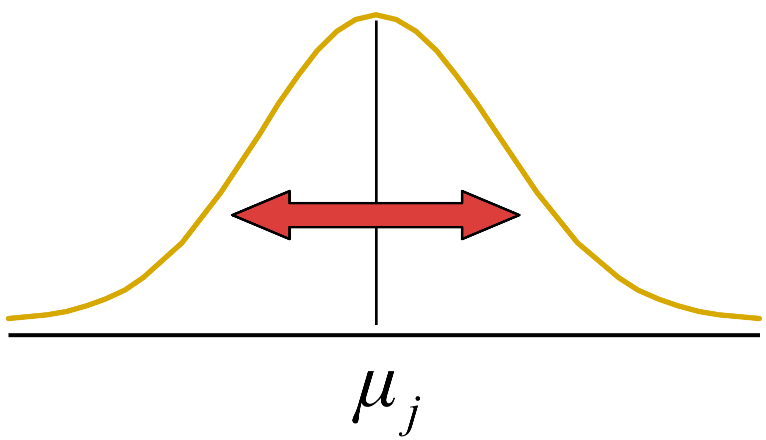


WITHIN-GROUP VARIATION

(continued)

$$SSW = \sum_{j=1}^{c} \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2$$

Summing the variation within each group and then adding over all groups



$$MSW = \frac{SSW}{n - c}$$

Mean Square Within = SSW/degrees of freedom

OBTAINING THE *MEAN* SQUARES

The Mean Squares are obtained by dividing the various sum of squares by their associated degrees of freedom

$$MSB = \frac{SSB}{c-1}$$

Mean Square Among
(d.f. = c-1)

$$\mathrm{MSW} = \frac{\mathrm{SSW}}{\mathrm{n-c}}$$

Mean Square Within
(d.f. = n-c)

$$\mathrm{MST} = \frac{\mathrm{SST}}{\mathrm{n-1}}$$

Mean Square Total
(d.f. = n-1)

One-Way ANOVA Table

Source of Variation	Degrees of Freedom	Sum Of Squares	Mean Square (Variance)	F
Between Groups	c - 1	SSB	$MSB = \frac{SSB}{c - 1}$	$F_{STAT} = \frac{MSB}{MSW}$
Within Groups	n - c	SSW	$MSW = \frac{SSW}{n - c}$	
Total	n – 1	SST		

c = number of groups
n = sum of the sample sizes from all groups

ONE-WAY ANOVA
F TEST STATISTIC

H_0: $\mu_1 = \mu_2 = \cdots = \mu_c$

H_1: At least two population means are different

Test statistic

$$F_{STAT} = \frac{MSB}{MSW}$$

MSB is mean squares between groups
MSW is mean squares within groups

Degrees of freedom

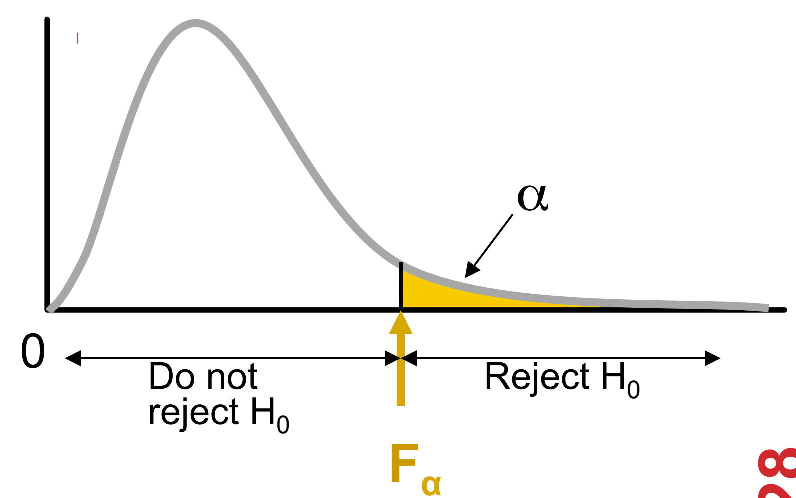
- $df_1 = c - 1$ (c = number of groups)
- $df_2 = n - c$ (n = sum of sample sizes from all populations)

INTERPRETING ONE-WAY ANOVA F STATISTIC

The F statistic is the ratio of the between estimate of variance and the within estimate of variance

Decision Rule:

- Reject H_0 if $F_{STAT} > F_\alpha$, otherwise do not reject H_0



ONE-WAY ANOVA F TEST EXAMPLE

You want to see if three different golf clubs yield different distances. You obtain five independent measurements from trials on an automated driving machine for each club. At the 0.05 significance level, *is there a difference in mean distance*?

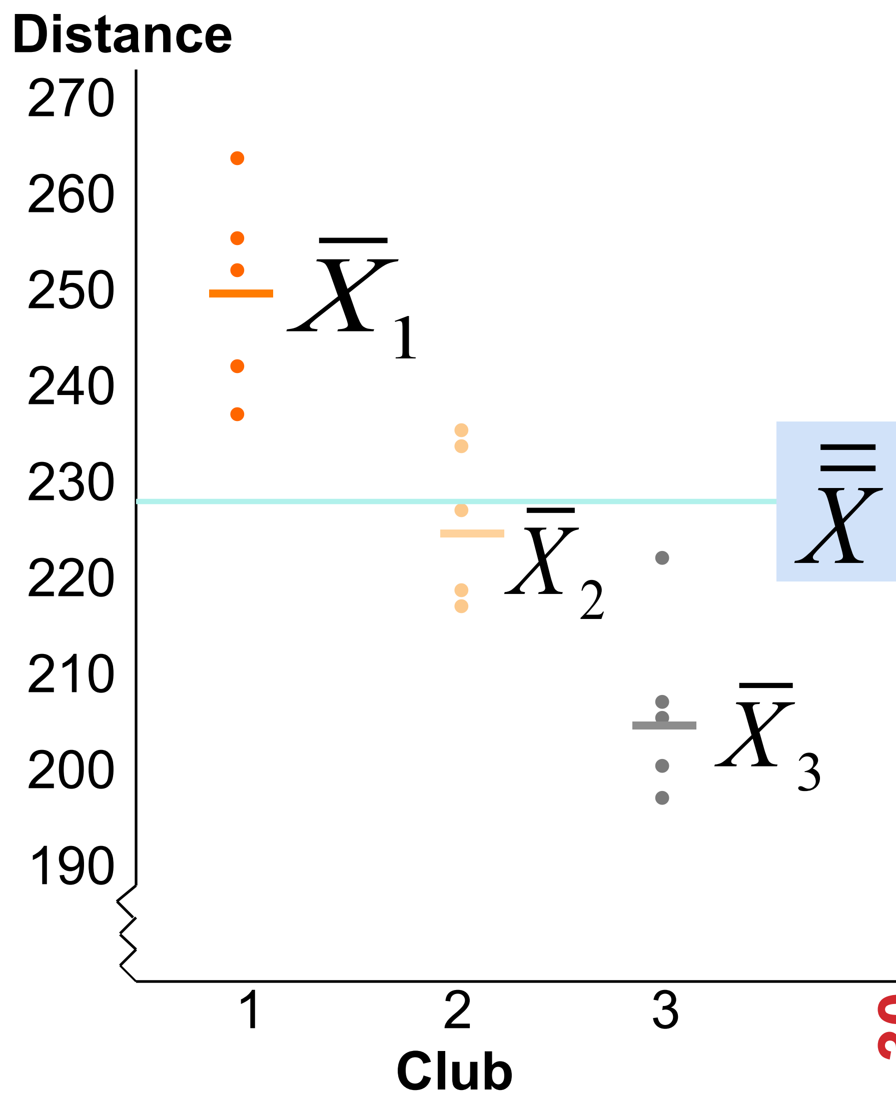
Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

ONE-WAY ANOVA EXAMPLE: SCATTER PLOT

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2$	$\bar{x}_2 = 226.0$	$\bar{x}_3 = 205.8$

$\bar{\bar{x}} = 227.0$



ONE-WAY ANOVA EXAMPLE COMPUTATIONS

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{X}_1 = 249.2$ $n_1 = 5$

$\bar{X}_2 = 226.0$ $n_2 = 5$

$\bar{X}_3 = 205.8$ $n_3 = 5$

$\bar{\bar{X}} = 227.0$ $n = 15$

$c = 3$

$$SSB = 5\,(249.2 - 227)^2 + 5\,(226 - 227)^2 + 5\,(205.8 - 227)^2 = 4716.4$$

$$SSW = (254 - 249.2)^2 + (263 - 249.2)^2 + \cdots + (204 - 205.8)^2 = 1119.6$$

$$MSB = 4716.4 \,/\, (3-1) = 2358.2$$

$$MSW = 1119.6 \,/\, (15-3) = 93.3$$

$$F_{STAT} = \frac{2358.2}{93.3} = 25.275$$

ONE-WAY ANOVA EXAMPLE SOLUTION

H_0: $\mu_1 = \mu_2 = \mu_3$

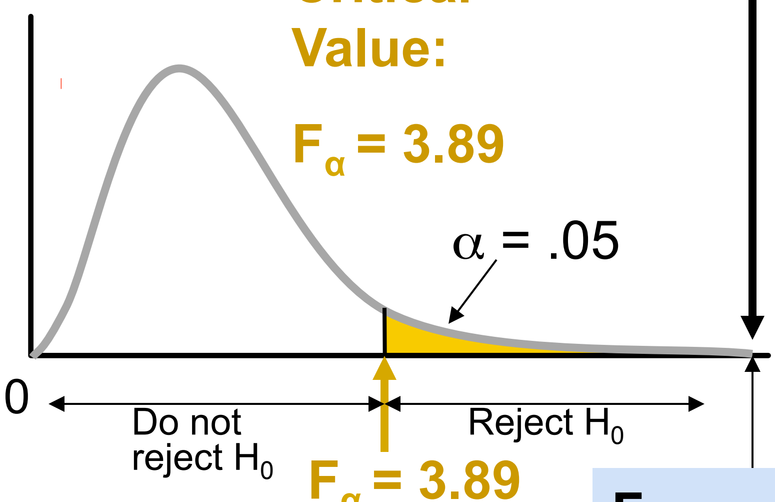
H_1: μ_j not all equal

$\alpha = 0.05$

$df_1 = 2$ $\quad$ $df_2 = 12$

Critical Value:

$F_\alpha = 3.89$



$F_{STAT} = 25.275$

Test Statistic:

$$F_{STAT} = \frac{MSA}{MSW} = \frac{2358.2}{93.3} = 25.275$$

Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

There is evidence that at least one μ_j differs from the rest

RECALL THIS EXAMPLE

Example 1: Suppose there are three (c=3) different promotions in department stores.

Promotions			
	1	2	3
Sales vol. diff. %	9.5	8.5	7.7
	3.2	9.0	11.3
	4.7	7.9	9.7
	7.5	5.0	11.5
	8.3	3.2	12.4
Sample size	5	5	5
Sample mean	6.64	6.72	10.52
Sample variance	6.82	6.28	3.43

EXERCISE

- Carry out by hand calculation.
- Carry out by software.

JMP EXAMPLE DEMO HERE

Companies.jmp

- **Companies' Profit ($M) vs. their Size (big, medium, small)**
- **Since the Prob > F is less than than 0.05, reject the null hypothesis. Conclude that there are differences between at least two of the means.**
- **To determine which means are different, a post hoc multiple comparison technique can be used.**
- **Each pair, t test.**
- **the mean for big is significantly different from the mean for small, but is not significantly different from the mean for medium.**

JMP EXAMPLE DEMO HERE

Analgesics.jmp

- **Thirty-three subjects were administered three different types of analgesics (A, B, and C). [Painkiller works?]**
- **The subjects were asked to rate their pain levels on a sliding scale. You want to find out if the means for A, B, and C are significantly different.**

NOTE: RELATION BETWEEN *T* AND *F* DISTRIBUTIONS

- **Recall the two sample t-test with two-sided alternative.**
- **If there are only two groups (c=2), then F-test and the above two sample t-test result should coincide.**
- **Indeed, it is known that:**
  the distribution of Squared-T distribution is the same that of F.
- **T measures the distribution of 'distance (mean),' whereas**
- **F measures the 'squared distance (variance).'**

JMP EXAMPLE DEMO HERE

Recall Bigclass.jmp data

- **Weight vs. [Male, Female] (two levels, c=2)**
- **Perform t-test (with two-sided alternative) and F-test and compare their p-values.**

RECALL: ANOVA ASSUMPTIONS!

1. Samples are randomly and independently drawn

- Select random samples from the c groups (golf club example)
- Or, randomly assign the levels (department store example)

2. Samples are drawn from a normal population

- The sample values for each group are from a normal population (use normal probability plot (Q-Q) plot, or goodness-of-fit test)

3. Homogeneity of Variance

- All populations sampled from the populations with the same variance
- Can be tested with Levene's Test

ANOVA ASSUMPTIONS LEVENE'S TEST

- **Tests the assumption that the variances of each population are equal.**
- **First, define the null and alternative hypotheses:**
  - H_0: $\sigma^2_1 = \sigma^2_2 = \cdots = \sigma^2_c$
  - H_1: Not all σ^2_j are equal
- **Second, compute the absolute value of the difference between each value and the median (or mean) of each group. (spread-measure)**
- **Third, perform a one-way ANOVA on these absolute differences.**

LEVENE'S HOMOGENEITY OF VARIANCE TEST EXAMPLE

H0: $\sigma^2_1 = \sigma^2_2 = \sigma^2_3$
H1: Not all σ^2_j are equal

Calculate Medians

Club 1	**Club 2**	**Club 3**	
237	216	197	
241	218	200	
251	227	204	**Median**
254	234	206	
263	235	222	

Calculate Absolute Differences

Club 1	**Club 2**	**Club 3**
14	11	7
10	9	4
0	0	0
3	7	2
12	8	18

Levene's Homogeneity Of Variance Test Example (continued)

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Club 1	5	39	7.8	36.2
Club 2	5	35	7	17.5
Club 3	5	31	6.2	50.2

Source of Variation	SS	df	MS	F	P-value	F c
Between Groups	6.4	2	3.2	0.092	0.912	3.8
Within Groups	415.6	12	34.6			
Total	422	14				

Since the p-value is greater than 0.05 we fail to reject H_0 & conclude there is an insufficient data evidence that supports variances are not equal.

RECALL: MULTIPLE POPULATIONS TEST

Comparison of more than two means

One way ANOVA → One factor and c (>2) treatments

We discussed two designs:

- Completely Randomized Design (CRD)
- Randomized Complete Block Design (RCBD)