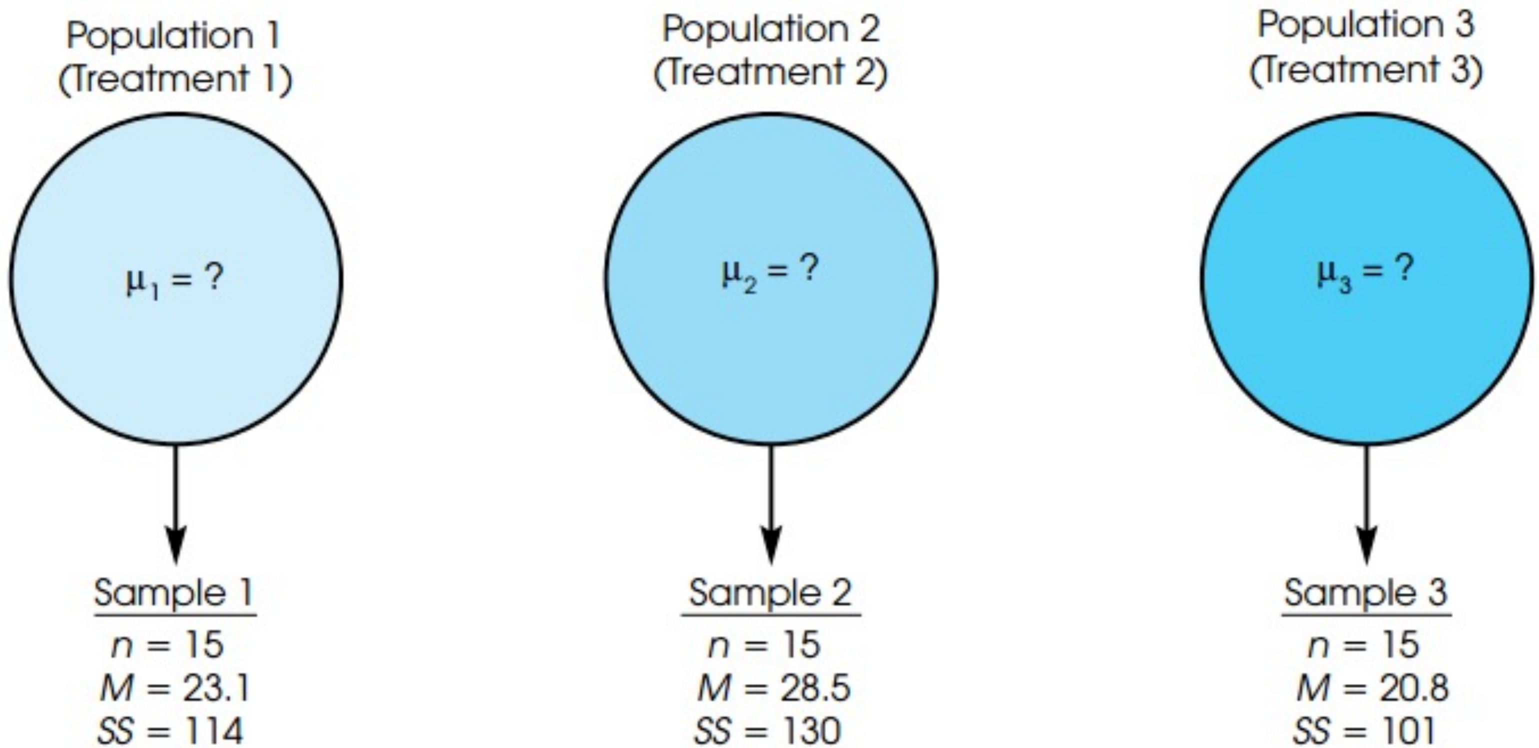
Lecture 10: One-way and Repeated Measure ANOVAs

Outline

- One-way ANOVA for testing the equivalence of the mean of multiple groups
- Tukey's HSD test and Scheffe Test for comparing specific pairs of groups
- Repeated-measures ANOVA for testing within-subject design

复习：The question for one-way ANOVA



- Compare the means from multiple conditions **(>2).** Similar to t tests, which only deals with two conditions. Here we have three **levels**.

- **Factor**: the treatment for different groups.
- **Independent factor**. We manipulate these independent groups by giving them different treatments. e.g., medical treatment conditions.
- **Quasi-independent factor**. The independent groups are formed naturally, e.g., gender or country.

Independent-measures v.s. Repeated-measures ANOVA

- Separate groups of subjects are tested, or the same subject group is tested for different treatments
- Just like t-tests have independent one-sample t-test vs. two-sample paired t-test

(b) Data from a nonexperimental design evaluating the effectiveness of a clinical therapy for treating depression.

	Depression Scores		
Participant	*Before Therapy*	*After Therapy*	*6-Month Follow-Up*
A	71	53	55
B	62	45	44
C	82	56	61
D	77	50	46
E	81	54	55

More complex designs, e.g., mixed ANOVA

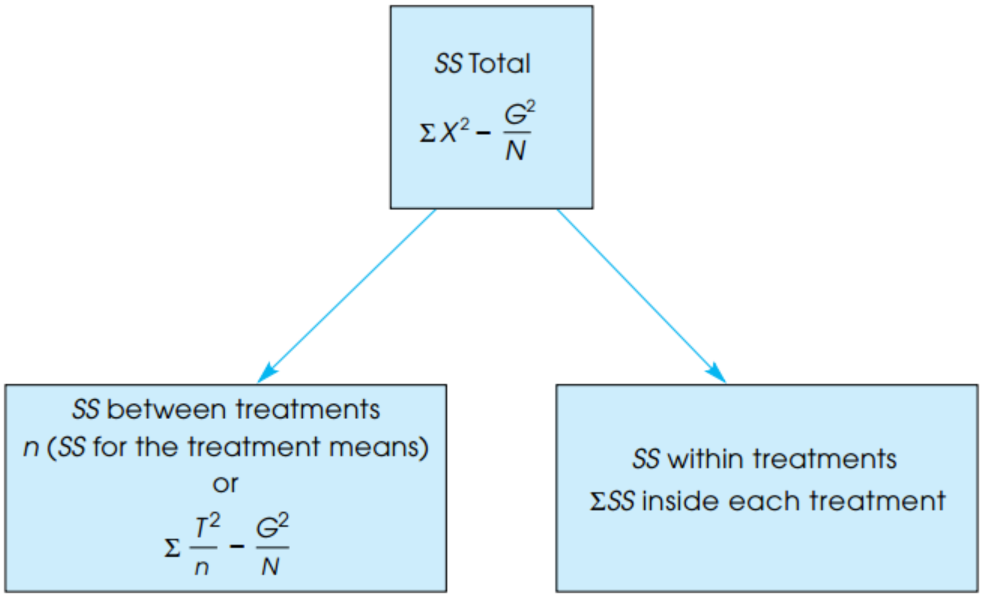
		TIME		
		Before Therapy	After Therapy	6 Months After Therapy
THERAPY TECHNIQUE	Therapy I (Group 1)	Scores for group 1 measured before Therapy I	Scores for group 1 measured after Therapy I	Scores for group 1 measured 6 months after Therapy I
	Therapy II (Group 2)	Scores for group 2 measured before Therapy II	Scores for group 2 measured after Therapy II	Scores for group 2 measured 6 months after Therapy II

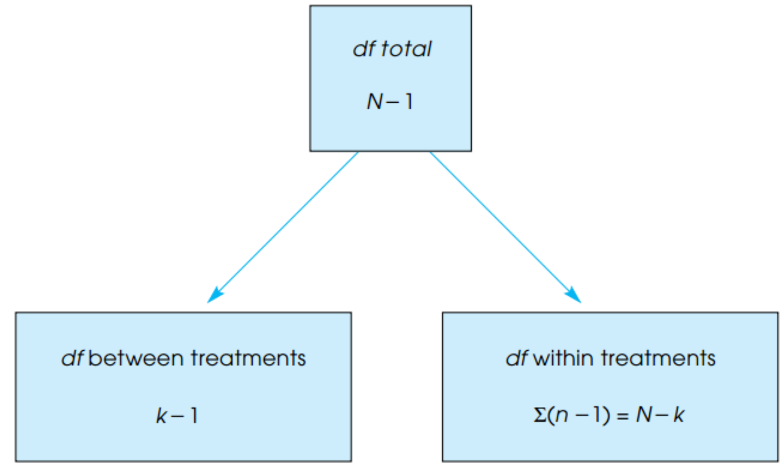
Driving performance example

We **randomly** select **three groups of subjects**, and compare the driving performance under three experimental conditions:

1. No phone
2. A hands-free phone
3. A hand-held phone

Question: Whether the phone use affects driven performance X





$$MS\,(\text{variance}) = s^2 = \frac{SS}{df}$$

$$MS_{\text{between}} = s^2_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{30}{2} = 15$$

$$MS_{\text{within}} = s^2_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{16}{12} = 1.33$$

$$F = \frac{s^2_{\text{between}}}{s^2_{\text{within}}} = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

=15/1.33 = 11.288

ANOVA table

Source	SS	df	MS	
Between treatments	30	2	15	$F = 11.28$
Within treatments	16	12	1.33	
Total	46	14		

$$F = \frac{0 + \text{ random, unsystematic differences}}{\text{random, unsystematic differences}}$$

Error term

Here, the error term is the within-treatment MS.

Effect size for ANOVA

- The percentage of variance accounted for by the treatment conditions, so called eta squared (η^2):

$$\text{The percentage of variance accounted for} = \frac{SS_{\text{between treatments}}}{SS_{\text{total}}}$$

= 60/122 = 0.492.

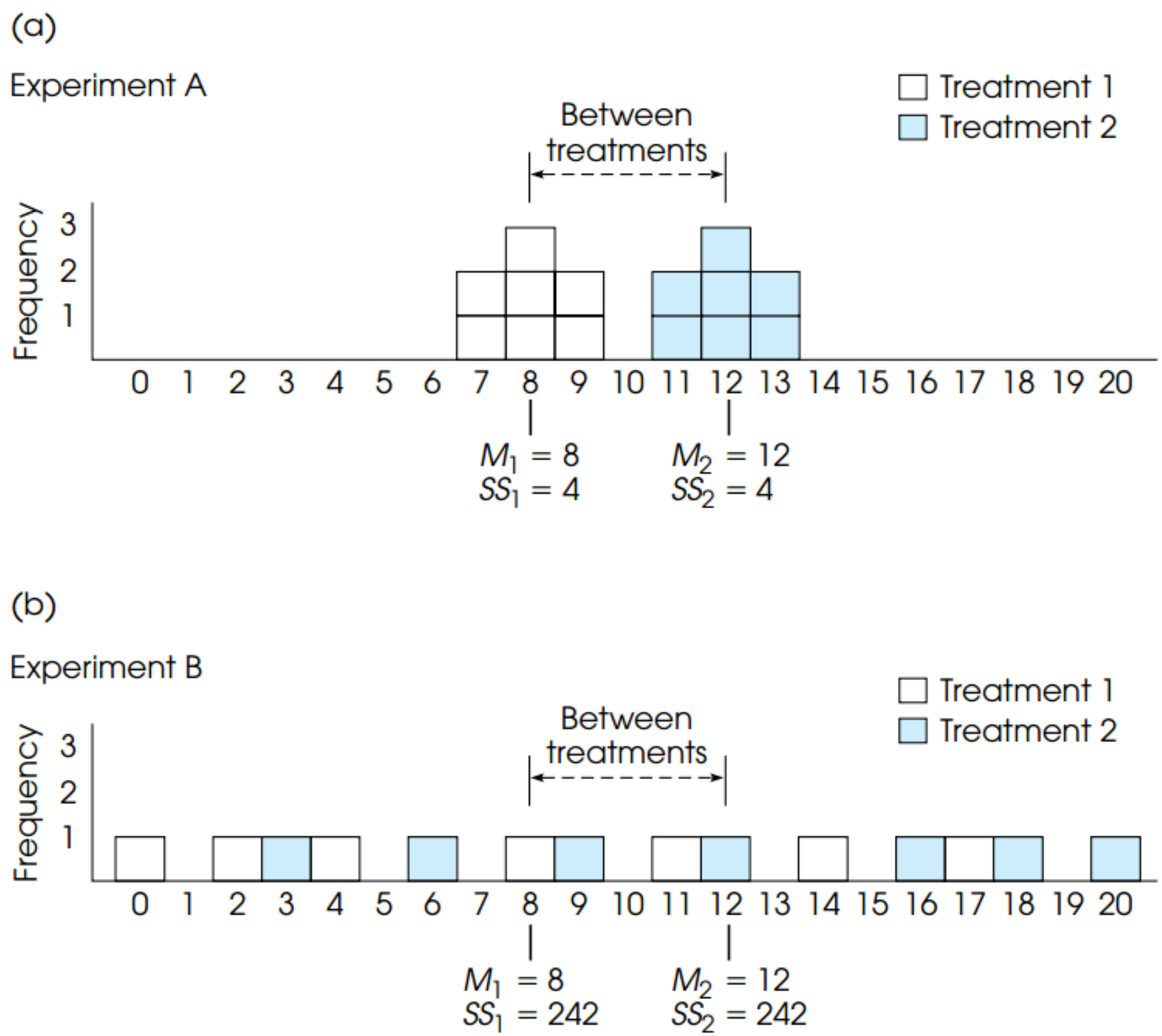
Analysis of Variance Table (ANOVA table)

ANOVA table					
Source	SS	df	MS	F statistic	p-value
Between	$SS_{between}$	k-1	$MS_{between}$	$\frac{MS_{between}}{MS_{within}}$	$Pr(F_{k-1,n-k}>F)$
Within	SS_{within}	n-k	MS_{within}		
Total	SS_{total}	n-1			

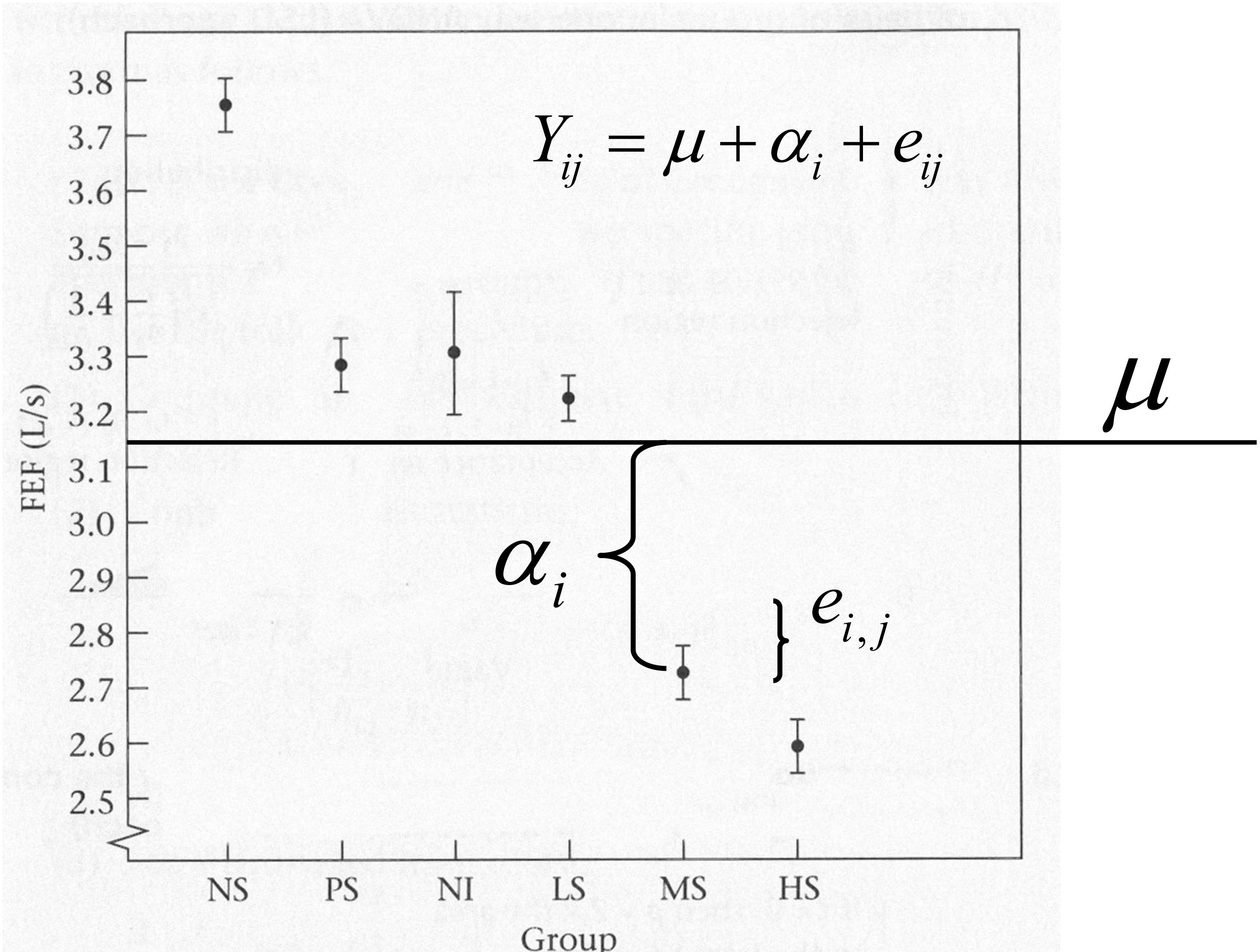
How to report results of ANOVA

- The mean and standard deviation of each phone condition is xxx. One-way ANOVA indicates that there are significant differences among the three phone conditions. $F(2,14) = 11.28$, $p<.01$, $\eta^2= 0.492$.

A visual representation



Another visual representation



Post Hoc test: pairwise comparisons between specific groups

- Conducted after ANOVA:

When the overall F test indicates that at least two means are different. The next step (usually) is to compare specific group means.

- Which means are significantly different from others?

u1, u2, u3 ➔ u1 v.s. u2, u2 v.s. u3 & u1 v.s. u3

Tukey's HSD test

Honestly Significant Difference

- Tukey's **HSD** = $q\sqrt{MS_{within}/n}$

- The value of q is found on Table B.5: using K (#group) and within-group df (or, more appropriate it should be df for error variance) and α level.

- Note n is the sample size of every group. HSD test requires that the sample size be the same for all treatments

- If the mean difference between any two treatment groups exceeds the number you get from Tukey's test, you can conclude that there is a significant difference between the treatments

Example for Tukey HSD

Treatment A	Treatment B	Treatment C
$n = 9$	$n = 9$	$n = 9$
$T = 27$	$T = 49$	$T = 63$
$M = 3.00$	$M = 5.44$	$M = 7.00$

Source	SS	df	MS
Between	73.19	2	36.60
Within	96.00	24	4.00
Total	169.19	26	
Overall $F(2, 24) = 9.15$			

$$HSD = q\sqrt{\frac{MS_{\text{within}}}{n}} = 3.53\sqrt{\frac{4.00}{9}} = 2.36$$

1. Treatment A is significantly different from treatment B (MA – MB = 2.44).
2. Treatment A is significantly different from treatment C (MA – MC = 4.00).
3. Treatment B is not significantly different from treatment C (MB – MC = 1.56).

Scheffe test

$$F = MS_{between} / MS_{within}$$

Original ANOVA:

$$SS_{between} = \sum_{i=1}^{k} n_i (M_i - \bar{G})^2 = \sum_{i=1}^{k} (T_i^2 / n_i) - G^2 / N$$

$$SS'_{between} = \sum_{i=1}^{2} n_i (M_i - Average\_of\_Two\_Means)^2$$

$$MS'_{between} = SS'_{between} / df_{between}$$

Scheffe Test:

What are changed: $SS_{between}$ is from the sum of the two compared groups; BUT, $df_{between}$ does not change.

SS_{within} and $F_{critical}$ remains the same!
$SS_{between}$ is *reduced*.

- ***This is a really conservative test!***
- ***Do not require equal number of samples for all groups.***

Summary

Total

$$SS = \Sigma X^2 - \frac{G^2}{N}$$

$$df = N - 1$$

Between treatments

$SS = n(SS \text{ for the means})$

$$\text{or } SS = \Sigma \frac{T^2}{n} - \frac{G^2}{N}$$

$$df = k - 1$$

$$MS = \frac{SS}{df}$$

Within treatments

$$SS = \Sigma SS_{\text{each treatment}}$$

$$df = N - k$$

$$MS = \frac{SS}{df}$$

$$F\text{-ratio} = \frac{MS \text{ between treatments}}{MS \text{ within treatments}}$$

Equal Variances Assumption

- Under H_0 in ANOVA it is assumed that the error variance is the same in each group.

- However, in cases where the underlying assumption of equal variances for each group is not believed to hold, <u>then</u> one-way ANOVA should not be used. t-test, which are based <u>only</u> on pairs of groups, are appropriate.

- OR, MORE ADVANCED METHODS CAN BE USED (e.g., **Welch correction)**. They are typically some transformation of original distribution.

Repeated-measure ANOVA

- Separate groups of subjects are tested vs. the same subject group is tested for different treatments
- Similar to t-tests: two-sampled independent t-test vs. paired t-test

(b) Data from a nonexperimental design evaluating the effectiveness of a clinical therapy for treating depression.

	Depression Scores	
Participant	*Before Therapy*	*After Therapy*
A	71	53
B	62	45
C	82	56
D	77	50
E	81	54

Repeated-measures ANOVA examples

(a) Data from an experimental study evaluating the effects of different types of distraction on the performance of a visual detection task.

Participant	Visual detection scores: No distraction	Visual distraction	Auditory distraction
A	47	22	41
B	57	31	52
C	38	18	40
D	45	32	43

(b) Data from a nonexperimental design evaluating the effectiveness of a clinical therapy for treating depression.

Participant	Depression scores: Before therapy	After therapy	6-month follow up
A	71	53	55
B	62	45	44
C	82	56	61
D	77	50	46
E	81	54	55

- Experimental study: manipulate an independent variable (here different sensory conditions) to test the same group repetitively.
- **Non-experimental study: the same group was simply observed repetitively.**

The setup of repeated-measures ANOVA

- Hypotheses are the same as for the independent-measures ANOVA
  - H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots$
  - H_1: at least one comparison is different

- The test statistic for the repeated-measures ANOVA is similar to the test statistic for the independent-measures ANOVA, i.e, using F-ratio.
- However, the F-ratio is differently defined.

The error variance caused by individual differences is removed!

The logic behind repeated-measures ANOVA

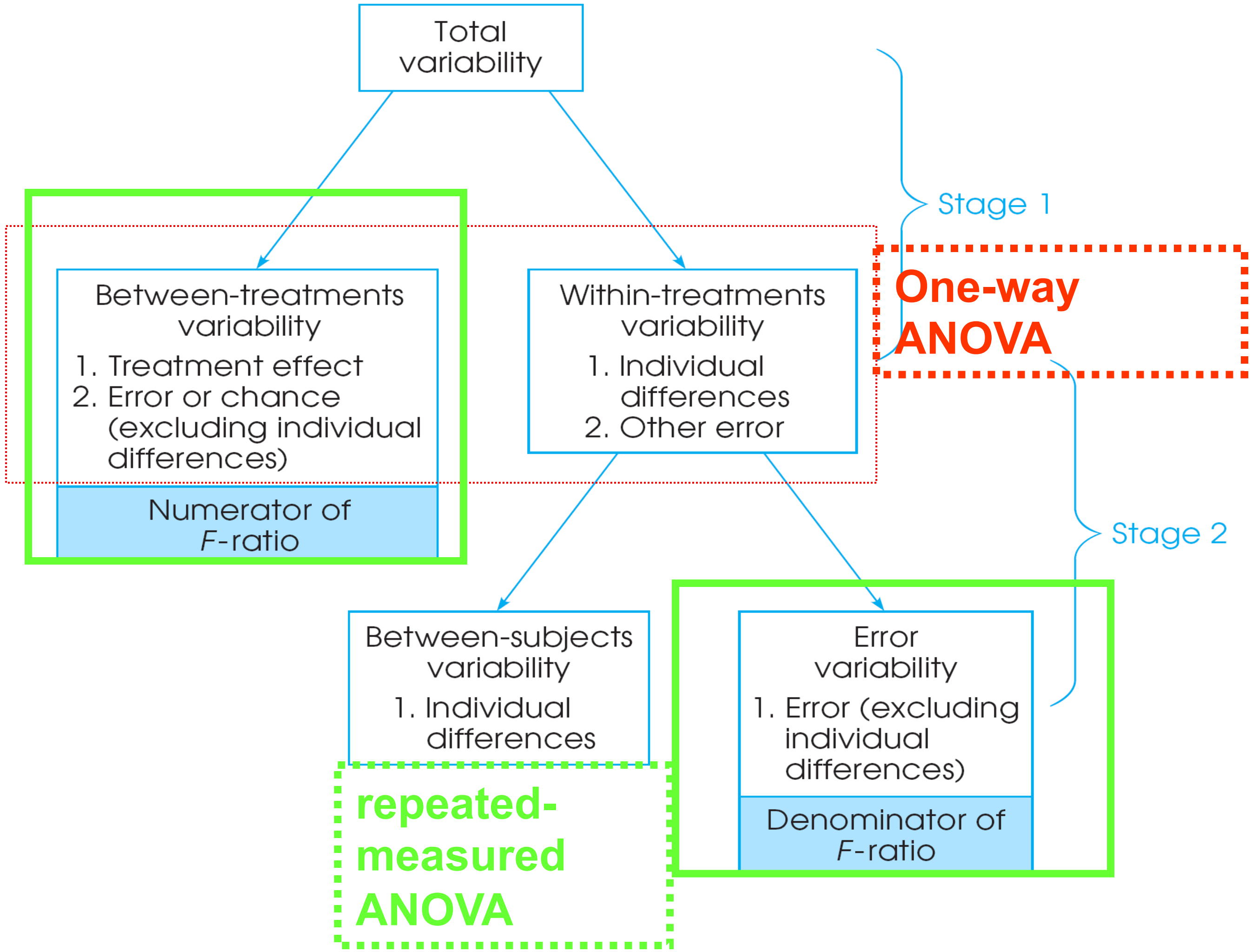
$$F = \frac{\substack{\text{variance between treatments} \\ \text{(without individual differences)}}}{\substack{\text{variance with no treatment effect} \\ \text{(individual differences removed)}}}$$

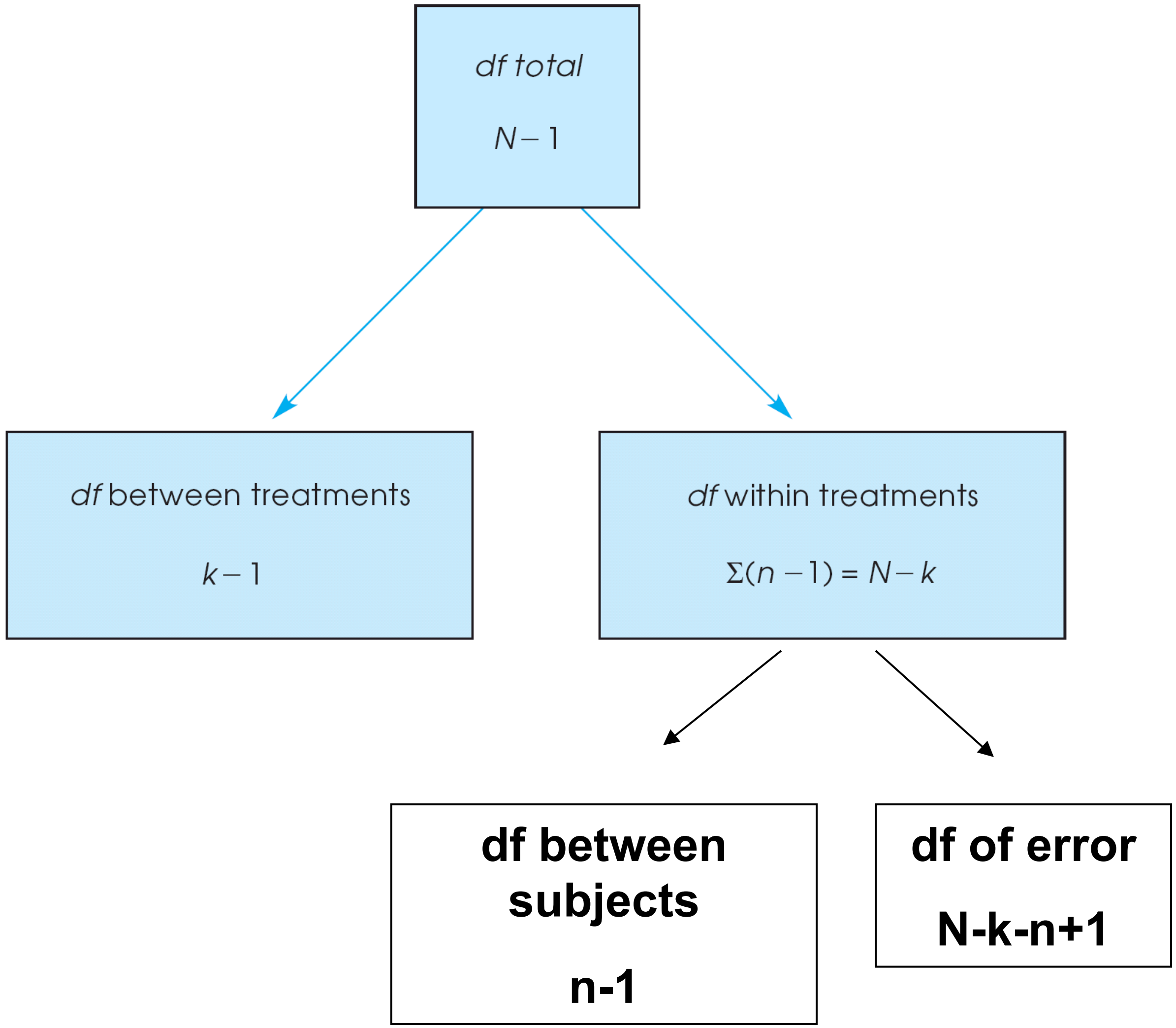
$$F = \frac{\text{between-treatments variance}}{\text{error variance}}$$

$$= \frac{\text{treatment effects} + \text{random, unsystematic differences}}{\text{random, unsystematic differences}}$$

- The numerator of the F-ratio can be explained by two factors:
  - – Treatment effect
  - – Error or chance
- The denominator of the F-ratio measures the differences that would occur by chance:
  - — individual differences
  - — random error

The partitioning of variability




df total
N − 1
df between treatments
k − 1
df within treatments
Σ(n − 1) = N − k
df between subjects
n-1
df of error
N-k-n+1

Repeated-measure ANOVA example

Can four learning strategies lead to any difference in quiz scores?

Strategies for Studying Text Passages

Student	Read Once	Read and Reread	Answer Prepared Questions	Create and Answer Questions	Person Totals	
A	3	5	8	8	$P = 24$	$n = 6$
B	3	3	5	9	$P = 20$	$k = 4$
C	4	5	8	7	$P = 24$	$N = 24$
D	6	7	9	10	$P = 32$	$G = 168$
E	6	8	8	10	$P = 32$	$\Sigma X^2 = 1298$
F	8	8	10	10	$P = 36$	
	$T = 30$ $M = 5$ $SS = 20$	$T = 36$ $M = 6$ $SS = 20$	$T = 48$ $M = 8$ $SS = 14$	$T = 54$ $M = 9$ $SS = 8$		

Similar to previous one-way ANOVA

Total variance

$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{N} = 1298 - \frac{168^2}{24} = 1298 - 1176 = 122$$

$$df_{\text{total}} = N - 1 = 23$$

n: the number of subjects;

k: the number of treatments;

G is the grand total of all measures.

Within-treatments variance

$$SS_{\text{within treatments}} = \Sigma SS_{\text{inside each treatment}} = 20 + 20 + 14 + 8 = 62$$

$$df_{\text{within treatments}} = \Sigma df_{\text{inside each treatment}} = 5 + 5 + 5 + 5 = 20$$

Between-treatments variance

$$SS_{\text{between treatments}} = SS_{\text{total}} - SS_{\text{within treatments}}$$

$$= 122 - 62 = 60$$

$$df_{\text{between treatments}} = k - 1 = 3$$

New: the variance caused by individual differences

$$SS_{\text{between subjects}} = \Sigma \frac{P^2}{k} - \frac{G^2}{N}$$

$$df_{\text{between subjects}} = n - 1$$

$$df_{\text{between subjects}} = 6 - 1 = 5$$

$$\begin{aligned} SS_{\text{between subjects}} &= \frac{24^2}{4} + \frac{20^2}{4} + \frac{24^2}{4} + \frac{32^2}{4} + \frac{32^2}{4} + \frac{36^2}{4} - \frac{168^2}{24} \\ &= 144 + 100 + 144 + 256 + 256 + 324 - 1176 \\ &= 48 \end{aligned}$$

The formula is exactly the same for between-treatment SS and for between-person SS!

Strategies for Studying Text Passages

Student	Read Once	Read and Reread	Answer Prepared Questions	Create and Answer Questions	Person Totals	
A	3	5	8	8	$P = 24$	$n = 6$
B	3	3	5	9	$P = 20$	$k = 4$
C	4	5	8	7	$P = 24$	$N = 24$
D	6	7	9	10	$P = 32$	$G = 168$
E	6	8	8	10	$P = 32$	$\Sigma X^2 = 1298$
F	8	8	10	10	$P = 36$	
	$T = 30$	$T = 36$	$T = 48$	$T = 54$		
	$M = 5$	$M = 6$	$M = 8$	$M = 9$		
	$SS = 20$	$SS = 20$	$SS = 14$	$SS = 8$		

F ratio of the Repeated-measures ANOVA

$$SS_{\text{error}} = SS_{\text{within treatments}} - SS_{\text{between subjects}}$$

$$SS_{\text{error}} = 62 - 48 = 14$$

$$df_{\text{error}} = df_{\text{within treatments}} - df_{\text{between subjects}}$$

$$df_{\text{error}} = (k - 1)(n - 1)$$ **=15**

$$F = \frac{MS_{\text{between treatments}}}{MS_{\text{error}}} = \frac{20}{0.933} = 21.43$$

$$df = df_{\text{between treatments}},\ df_{\text{error}}$$ **=3, 15**

Source	SS	df	MS	F
Between treatments	60	3	20.00	$F(3,15) = 21.43$
Within treatments	62	20		
Between subjects	48	5		
Error	14	15	0.933	
Total	122	23		

Effect size for the repeated-measures ANOVA

Percentage of variance that is explained by the treatment differences

$$\eta^2 = \frac{SS_{\text{between treatments}}}{SS_{\text{total}} - SS_{\text{between subjects}}}$$

$$\eta^2 = \frac{SS_{\text{between treatments}}}{SS_{\text{between treatments}} + SS_{\text{error}}}$$

$$\eta^2 = \frac{60}{74} = 0.811 \quad (\text{or } 81.1\%)$$

Compared to that of one-way ANOVA:

$$\eta^2 = \frac{SS_{\text{between treatments}}}{SS_{\text{between treatments}} + SS_{\text{within treatments}}} = \frac{SS_{\text{between treatments}}}{SS_{\text{total}}}$$

Thus, 81.1% of the total variance in the data (except for the individual differences) is accounted for by the differences between treatments.

How to report the results

The means and variances of the quiz scores for the four strategies are shown in Table 1. A repeated-measures analysis of variance indicated significant mean differences in the four methods for studying text passages, $F(3, 15) = 21.43$, $p < .01$, $\eta^2 = 0.811$.

TABLE 1

Quiz scores for students using four different study strategies

	Read Once	Read and Reread	Answer Prepared Questions	Create and Answer Questions
M	5.00	6.00	8.00	9.00
SD	2.00	2.00	1.67	1.26

Post hoc tests with repeated-measures ANOVA

- **For Tukey's HSD, we can use the same formula while substituting the new error variance:**

- **Instead of using MS_{within}, you will use MS_{error}**
- **Instead of using df_{within} when looking up the value for q, you use df_{error}**

$$HSD = q\sqrt{\frac{MS_{error}}{n}} = 4.08\sqrt{\frac{0.933}{6}} = 4.08(0.394) = 1.61$$

	Read Once	Read and Reread	Answer Prepared Questions	Create and Answer Questions
M	5.00	6.00	8.00	9.00

Which treatments differ?

Assumptions of the repeated-measures ANOVA

The first three assumptions are identical to those for independent-measures ANOVA:

1. The observations within each treatment condition must be independent.
2. The population distribution within each treatment must be normal (As before,
the assumption of normality is important only with small samples).
3. The variances of the population distributions for each treatment should be equivalent.
4. For the repeated-measures ANOVA, there is an additional assumption called sphericity assumption (Mauchly’s test). If violated, then the Greenhouse-Geisser correction or the Huynh-Feldt correction should be used.

Advantages and disadvantages of the repeated-measures ANOVA

Good: few subjects, individual differences removed, F-ratio increased thus more likely to have significant results compared to independent-measures. Thus, it has more power (more likely to detect a real difference).

Bad: Order effect as for paired t-test. Some differences between treatments are caused by factors other than treatment.
E.g., 1) meta learning (learning over different learning strategies) instead of learning strategy itself causes the score difference;
2) “Natural” cognitive development instead of schooling improves cognitive performance.

Thus, we should have a better design (e.g., counter-balanced between participants to have consistent individual differences) and cautious interpretation of the data (keeping the order effect in mind).