Two-Factor ANOVA

two independent variables

example: independent measures, equal n designs

"cell"

	JVM Heap size 128 MB	256 MB	512MB
sun java	n = 15 java programs	n = 15	n = 15
jikes	n = 15	n = 15	n = 15

scores are times to completion (performance).

Is performance affected by choice of compiler.

or heap size?

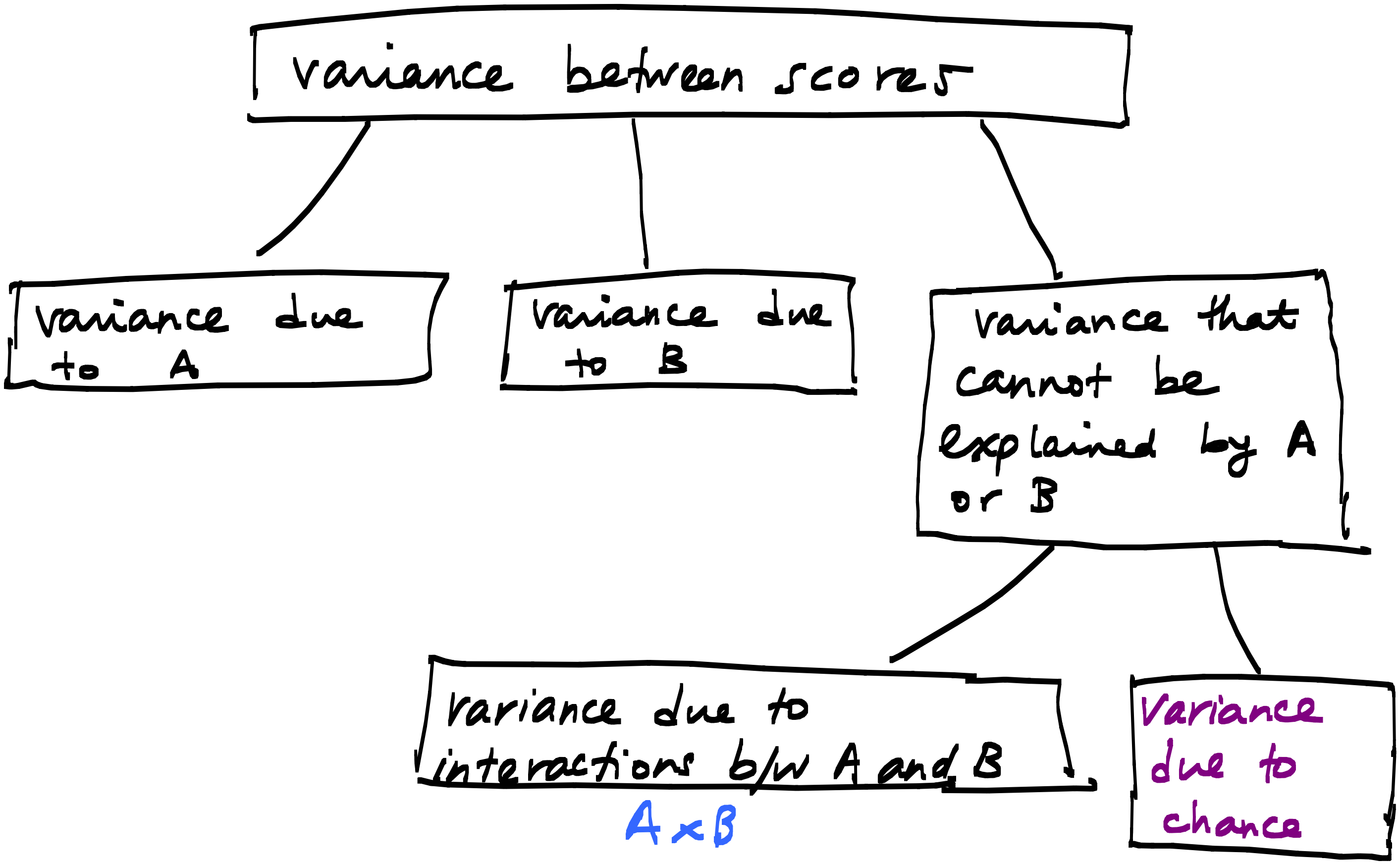
JVM Heap size

	B_1 128 MB	B_2 256 MB	B_3 512 MB
A_1 sun java	n = 15 java programs	n = 15	n = 15
A_2 jikes	n = 15	n = 15	n = 15

3 Hypothesis Tests

1. $H_0: \mu_{java} = \mu_{jikes}$
2. $H_0: \mu_{128} = \mu_{256} = \mu_{512}$
3. H_0: no interaction b/w heap size, compiler

Hypothesis 3 searches for differences between scores that cannot be explained by Hypothesis 1 or Hypothesis 2.


variance between scores
variance due to A
variance due to B
variance that cannot be explained by A or B
variance due to interactions b/w A and B
A×B
Variance due to chance

3 Hypothesis Tests

① $H_0: \mu_{A1} = \mu_{A2}$

② $H_0: \mu_{B1} = \mu_{B2} = \mu_{B3}$

} main effects

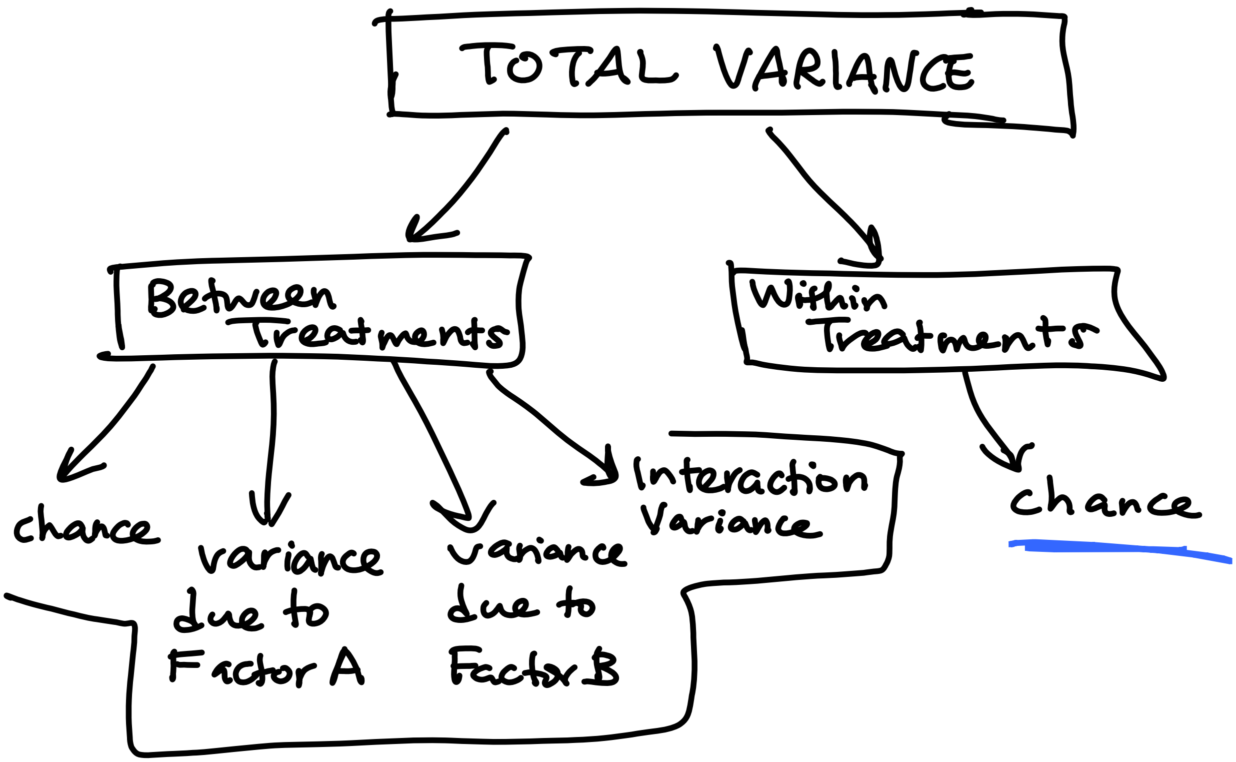
③ H_0: no interaction b/w A and B

(A×B interaction: effect of A depends on level of B
and/or vice versa)

For each test, compute corresponding F-statistic.

Main Effects: $F_A = \dfrac{\text{variance between means for factor A}}{\text{variance expected by chance}}$

→ var. within subjects


TOTAL VARIANCE
Between Treatments
Within Treatments
chance
variance due to Factor A
variance due to Factor B
Interaction Variance
chance

Example

	B_1	B_2	B_3	ROW
A_1	3, 1, 1, 6, 4 $\bar{X}=3$, $SS=18$	2, 5, 9, 7, 7 $\bar{X}=6$, $SS=28$	9, 9, 13, 6, 8 $\bar{X}=9$, $SS=26$	$n=15$ TOTAL = 90 MEAN = 6
A_2	0, 2, 0, 0, 3 $\bar{X}=1$, $SS=8$	3, 8, 3, 3, 3 $\bar{X}=4$, $SS=20$	0, 0, 0, 5, 0 $\bar{X}=1$, $SS=20$	$n=15$ TOTAL = 30 MEAN = 2
COL	$n=10$ TOTAL = 20 MEAN = 2	$n=10$ TOTAL = 50 MEAN = 5	$n=10$ TOTAL = 50 MEAN = 5	

$N = 5 \times 6 = 30$

TOTAL = 120

OVERALL MEAN = 4

Source	SS	df	var	F
b/w treatments	240	5		
Factor A	120	1	120	$F_A = 24$
Factor B	60	2	30	$F_B = 6$
A×B	60	2	30	$F_{A\times B} = 6$
Within Treatments	120	24	$120/24 = 5$	
Total	360	29		

At $\alpha = 0.05$, critical values

for $F(1,24) = 4.26$

$F(2,24) = 3.40$

$F(2, 24) = 3.40$

df of numerator

df of denominator

$F_A = 24 >> 4.26$

$F_B = 6 > 3.40$

$F_{A\times B} = 6 > 3.40$

w/ 95% confidence, we can reject all 3 null hypotheses!

remember: The F-distribution is a <u>family</u> of distributions. Each curve in the family defined by df_{num} and df_{denom}.

$\underline{df}$

$N - 1 = 29.$

$6 - 1 = 5$

$2 - 1 = 1$

$3 - 1 = 2$

$A \times B = 5 - 1 - 2 = 2$

Within = Total − b/w = 29 − 5 = 24

SS

overall mean

TOTAL: $(3-4)^2 + (1-4)^2 + (1-4)^2$

$+ \ldots + (5-4)^2 + (0-4)^2 = 360$

SS within treatments

$= 18 + 28 + 26 + 8 + 20 + 20$

$= 120$

				ROW
A_1	3, 1, 1, 6, 4	2, 5, 9, 7, 7	9, 9, 3, 6, 8	n = 15 TOTAL = 90 MEAN = 6
A_2	0, 2, 0, 0, 3	3, 8, 3, 3, 3	0, 0, 0, 5, 0	n = 15 TOTAL = 30 MEAN = 2

OVERALL MEAN = 4

$$SS_A = n_A \left[\sum_{i=1}^{2} \left(\bar{X}_{A_i} - \text{OVERALL MEAN} \right)^2 \right]$$

$$= 15\left(2^2 + 2^2\right) = 120$$

	B_1	B_2	B_3
	3, 1, 1, 6, 4	2, 5, 9, 7, 7	9, 9, 13, 6, 8
	0, 2, 0, 0, 3	3, 8, 3, 3, 3	0, 0, 0, 5, 0
COL	$n = 10$	$n = 10$	$n = 10$
	TOTAL = 20	TOTAL = 50	TOTAL = 50
	MEAN = 2	MEAN = 5	MEAN = 5

OVERALL MEAN = 4

$$SS_B = n_B \left(\sum \left(\bar{x}_{B_i} - \text{OVERALL MEAN} \right)^2 \right)$$

$$= 10 \cdot \left(2^2 + 1 + 1 \right) = 60$$

Assumptions of Two-Factor ANOVA (indep. measures)

1. observations within each sample are indep.
2. popns are normal
3. popns have homogeneity of variance

Three Factor Anova

SS_A $\quad$ $SS_{A \times B}$ $\quad$ $\underbrace{SS_{A \times B \times C}}$

SS_B $\quad$ $SS_{A \times C}$ $\quad$ means that

SS_C $\quad$ $SS_{B \times C}$ $\quad$ $A \times B$ interactions depend on C

7 hypotheses!

Beyond 3 factors, the interaction differences are hard to explain.

So do 4+ way ANOVAs at your own risk!