Anova (Analysis of Variance)

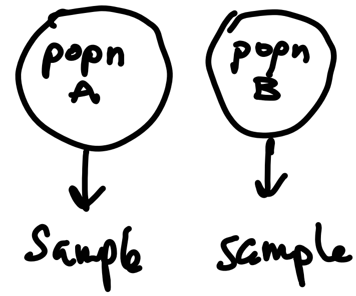
extend t-tests to situations
with more than one indep. var.
and/or more than two treatment levels
for any indep. var.

REVIEW!

Between-Subjects t-test

$H_0: \mu_A = \mu_B \quad (\mu_A - \mu_B = \emptyset)$

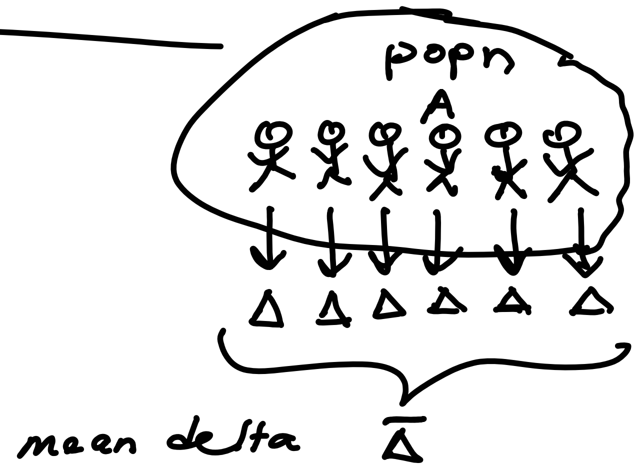
$H_1: \mu_A \neq \mu_B \quad (\mu_A - \mu_B \neq \emptyset)$



Within/Matched Subject t-test

$H_0: \mu_{\bar{\Delta}} = \emptyset$

$H_1: \mu_{\bar{\Delta}} \neq \emptyset$

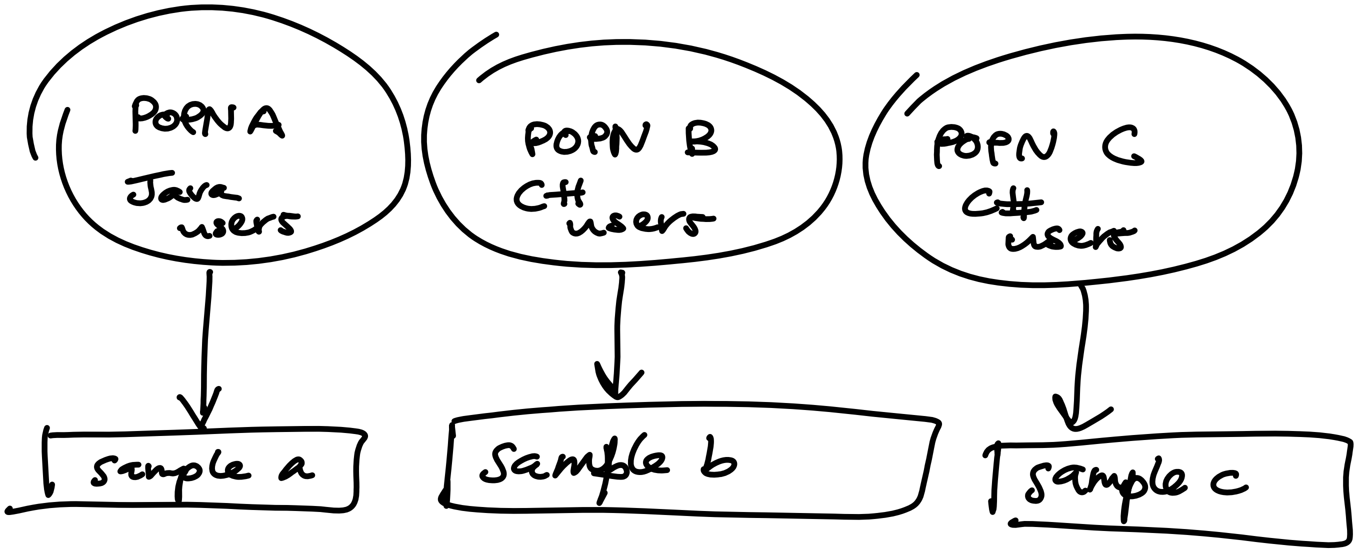


t-tests rely on measuring <u>differences</u> b/w sample scores or sample statistics.

ANOVA generalizes that approach by measuring <u>variances</u> between and within sample groups.

Independent Measures One-Way ANOVA

(between-subjects, only one indep. var.)



H_0 : $\mu_A = \mu_B = \mu_C$ is true

H_1 : $\mu_A = \mu_B = \mu_C$ is not true.

prove or disprove H_0 with

<u>F-statistic</u> : $\dfrac{\text{variance between sample means}}{\text{variance expected by chance}}$

ANOVA Vocab

SS = sum of squares $\sum(x_i - \bar{x})^2$
(numerator of variance)

df = degrees of freedom $n-1$
(denominator of variance)

MS = mean square (variance) s^2 or var

factor: indep. variable

	$n = 5$	$n = 5$	$n = 5$	$k = 3$
	Java	C++	C#	
# Errors Made	0	4	1	$N = 15$
	1	3	2	
	3	6	2	
	1	3	0	
	0	4	0	
	$\bar{X}_{java} = 1$	$\bar{X}_{C++} = 4$	$\bar{X}_{C\#} = 1$	

Variability everywhere!

2 sources of variability

variances between treatment levels

chance

true treatment difference

variances within treatment levels

chance

chance: natural variation b/w subjects, measurement error

F - statistic

$$F = \frac{\text{variation between treatments (chance + true treatment effect)}}{\text{variation within subjects (chance)}}$$

H_0: true treatment effect $= \emptyset$

probably true when $F \approx 1$

probably false when $F >> 1$

	n = 5 Java a	n = 5 C++ b	n = 5 C# c	k = 3
#Errors Made	0	4	1	
	1	3	2	N = 15
	3	6	2	
	1	3	0	
	0	4	0	
	$\Sigma a = 5$	$\Sigma b = 20$	$\Sigma c = 5$	TOTAL Σ: 30
	$\bar{a} = 1$	$\bar{b} = 4$	$\bar{c} = 1$	OVERALL MEAN = 2

F-statistic $= \dfrac{\text{variance between treatments}}{\text{variance within tr.}}$

	SS	df	$\text{var} = \frac{SS}{df}$	$F = \frac{\text{var}_{b/w}}{\text{var within}}$
b/w tr.	30	2	15	11.28
within tr	6+6+4 = 16	12	$1\frac{1}{3}$	$15/\frac{4}{3}$
TOTAL	46	N-1 = 14	$46/14 = 3.28$	

$\sum(x - \text{OVERALL MEAN})^2$

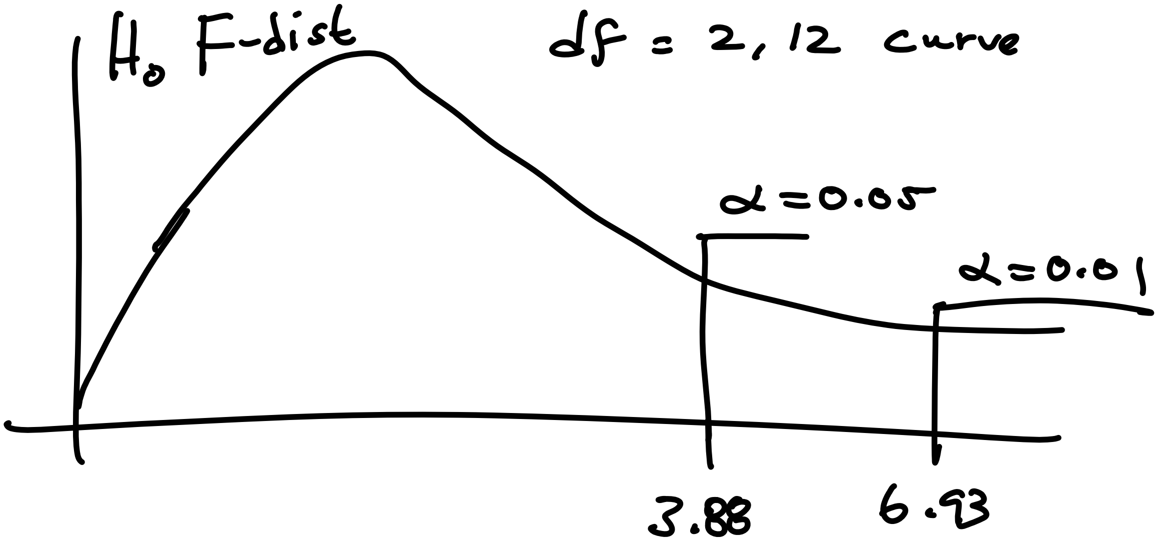
$$
\begin{aligned}
SS_{\text{TOTAL}} &= (-2)^2 + (-1)^2 + (1)^2 + (-1)^2 + (-2)^2 \\
&\quad + (2)^2 + (1)^2 + (4)^2 + (1)^2 + (2)^2 \\
&\quad + (-1)^2 + (0)^2 + (0)^2 + (-2)^2 + (-2)^2 \\
&= 46.
\end{aligned}
$$

$$SS_{within} = \sum SS_{each\ treatment\ level}$$

$$= 6 + 6 + 4 = 16$$

F-statistic : 11.28

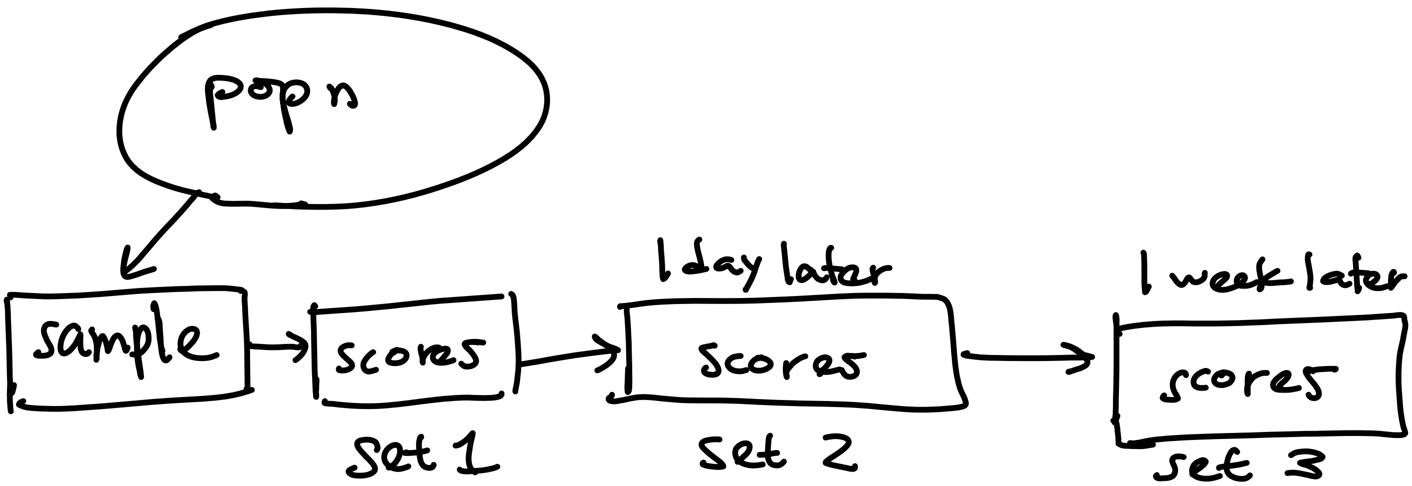
<u>Family of F-distributions</u>



11.28 > 6.93 H_0 disproven.

Repeated Measures Single Factor Anova

- Within-subjects / matched subjects
- one indep. var.



Within Subjects / Matched Subjects ANOVA

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

$H_1: \mu_1 = \mu_2 = \mu_3$ is not true

F-statistic: $\dfrac{\text{variance b/w treatments (no indiv. differences)}}{\text{variance expected by chance} - \text{individual differences}}$

	# Questions Answered Right				
	Google G	Ask.com A	Clusty C	Ice Rocket I	subj
Amir	3	4	6	7	5
Bela	0	3	3	6	3
Dina	2	1	4	5	3
Jayant	0	1	3	4	2
Lisa	0	1	4	3	2
	$\Sigma g=5$	$\Sigma a=10$	$\Sigma c=20$	$\Sigma i=25$	
	$\bar{g}=1$	$\bar{a}=2$	$\bar{c}=4$	$\bar{i}=5$	

$n=5$

$k=4$

$N=20$

TOTAL SUM $=60$

OVERALL MEAN $=3$

$$\text{F-statistic} = \frac{\text{variance between treatments} \\ \text{(true treatment effect + residual variance)}}{\text{residual variance} \\ \text{(chance – individual differences)}}$$

	SS	df	$var = \frac{SS}{df}$	$F = \frac{\text{var b/w tr.}}{\text{var resid.}}$
b/w tr.	50	3	$50/3 = 16.67$	25
within tr	32	16		$F(3, 12)$
b/w subj.	24	4		
residual	8	12	$8/12 = 0.67$	
Total	82	19		

$$SS_{\text{between subject}} \quad \left(\text{aka } SS_{\text{block}}\right)$$

$$= \quad k \sum \left(\overline{\text{subject}} - \begin{matrix}\text{OVERALL}\\ \text{MEAN}\end{matrix} \right)^2$$

$$= 4\left[\underset{\text{AMIR}}{(5-3)^2} + \underset{\text{BELA}}{0^2} + \underset{\text{DINA}}{0^2} + \underset{\text{JAY}}{(2-3)^2} + \underset{\text{LISA}}{(2-3)^2} \right]$$

$$= 24.$$

Assumptions of ANOVA

1. observations between each treatment condition must be indep.
2. within each treatment → normal
3. popn variances are equal.

Post - Hoc Tests

H_1 only says $\left\{ \begin{array}{l} \mu_A = \mu_B = \mu_C \ldots \\ \mu_1 = \mu_2 = \mu_3 \ldots \end{array} \right.$ is FALSE.

But → which means are significantly different?

Need to do pairwise comparisons of means

1. Pairwise t-tests
2. Tukey's HSD
3. Scheffé Test

use data from ANOVA computation