parametric tests

everything we've done so far except for Spearman correlation

1. hypothesizing about POPN PARAMETERS given sample statistics
2. assumptions:

   popn is normal
   popns have equal variances
   samples are indep.

   } some common ones

   each test has its own assumptions

3. interval/ratio data

Non-parametric tests

1. no assumptions about popn distribution
2. often not trying to estimate a popn parameter like μ or σ; instead measuring something more sketchy
3. any type of data

χ^2 - goodness of fit

H_0: popn preferences are

Tool A	Tool B	Tool C
a%	b%	c%

a common H_0 is a% = b% = c%
(popn has no preference)

but it is possible to have other values for
a%, b%, and c%

H_1: population has different preferences
than what is stated in H_0

example:

Tool A	Tool B	Tool C
15	19	6

$n = 40$

H_0:

A	B	C
25%	50%	25%

H_1: $\neq$

A	B	C
25%	50%	25%

① Compute expected frequencies if H_0 true.

A: 25% of 40 = 10
B: 50% of 40 = 20
C: 25% of 40 = 10

$freq_{exp.} =$

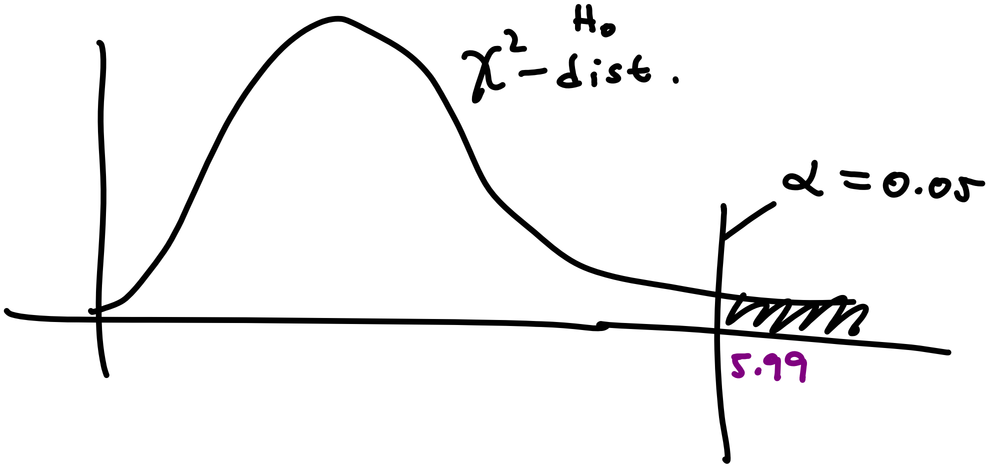
A	B	C
10	20	10

② Compute χ^2-statistic.

$$\chi^2 = \sum \frac{(freq_{observed} - freq_{exp})^2}{freq_{expected}}$$

$$\chi^2 = \frac{(15-10)^2}{10} + \frac{(19-20)^2}{20} + \frac{(6-10)^2}{10} = 4.15$$

③ Figure out how likely this statistic is from the H_0 χ^2 - distribution.



$\alpha = 0.05$ for $\chi^2(C-1)$ is at critical value is 5.99

\# categories

$4.15 < 5.99$

Cannot reject H_0.

χ^2 test for independence

relationship between two variables

Favorite Class

	631	660	711	734	TOT
M.S. students	10	3	15	22	50
PhD students	90	17	25	18	150
TOT:	100	20	40	40	

TOTAL: 200

popn MS

popn PhD

H_0: no difference in prefs. among the 2 popns

H_0: no difference b/w distribution of favorite class preferences between M.S. students and PhD students.

H_1: there is a difference

Favorite Class

	631	660	711	734	TOT
M.S. students	10	3	15	22	50.
PhD students	90	17	25	18	150.
TOT:	100	20	40	40	

$N = 200$

① Compute H_0 percentages.

631	660	711	734
$\frac{100}{200}$	$\frac{20}{200}$	$\frac{40}{200}$	$\frac{40}{200}$

② Compute Expected Frequencies (if H_0 true)

Expected Proportions

631	660	711	734
50%	10%	20%	20%

expected frequencies

MS	25	5	10	10
PhD	75	15	30	30

③ Compare expected frequencies to observed frequencies.

expected frequencies

	631	660	711	734
MS	25	5	10	10
PhD	75	15	30	30

observed frequencies

					TOT
M.S. students	10	3	15	22	50
PhD students	90	17	25	18	150
TOT:	100	20	40	40	

$$\chi^2 = \sum \frac{(\text{freq}_{\text{observed}} - \text{freq}_{\text{expected}})^2}{\text{freq expected}}.$$

$$= \frac{(10-25)^2}{25} + \frac{(3-5)^2}{5} + \frac{(15-10)^2}{10} + \frac{(22-10)^2}{10}$$

$$+ \frac{(90-75)^2}{75} + \frac{(17-15)^2}{15} + \frac{(25-30)^2}{30} + \frac{(18-30)^2}{30}$$

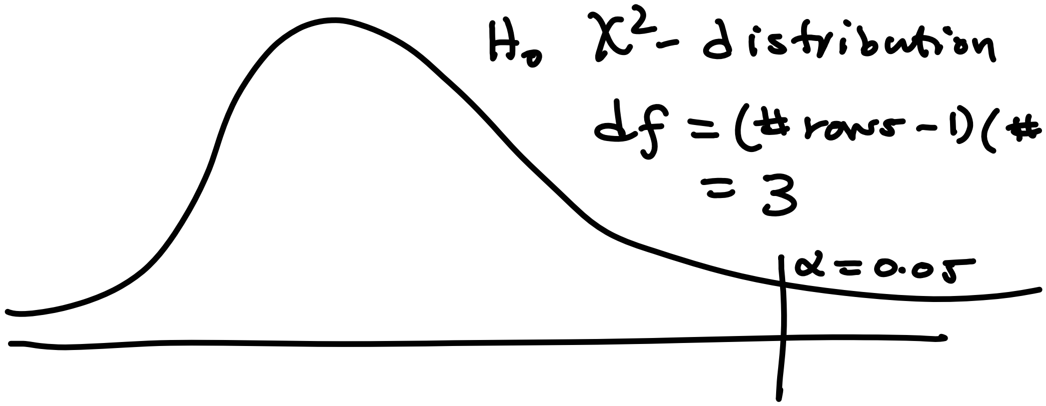
$$= 9 + 4/5 + 2.5 + 14.4 + 8\tfrac{1}{3} + \frac{4}{15} + \frac{25}{30} + 4.8$$

$$= 40.93$$

H_0 χ^2- distribution

$df = (\#\text{ rows} - 1)(\#\text{ cols} - 1)$

$= 3$



$\alpha = 0.05$ for critical value

value $\chi^2(3) = 7.81$

$40.93 \gg 7.81$,

so H_0 disproved.

Assumptions for χ^2-tests

① observations are independent

② exp. frequencies ≥ 5 (b/c $freq_{exp}$ in denominator)

③ each subject counted in ONLY ONE CATEGORY!

Sign Test

For repeated measures, we want to say something about the <u>trend</u> of scores.

ex: time to recognize handwriting on day 1
vs. time to recognize handwriting on day 10

	A	B	C	D	E	F	G	H
Direction of Change Day 1 $\longrightarrow$ Day 10	−	−	−	+	−	−	−	−

The sign test has a formula and z-statistic, but we can conduct an equivalent test using the χ^2- statistic goodness-of-fit.

	A	B	C	D	E	F	G	H
Direction of Change Day 1 ⟶ Day 10	−	−	−	+	−	−	−	−

$n = 8$

H_0 : proportion (plus) = proportion(minus) = $\frac{1}{2}$

Observed Data

Plus	Minus
1	7

Expected Data if H_0 true

Plus	Minus
4	4

$$\chi^2 = \frac{(1-4)^2}{4} + \frac{(7-4)^2}{4} = \frac{18}{4} = 4.5$$

At $\alpha = 0.05$, critical value of $\chi^2(1)$ is 3.84.

$4.5 > 3.84$, so H_0 disproved.

sign-test reported with z-score:

$$z^2 = \chi^2$$

$$\chi^2 = 4.5 \longrightarrow z = \sqrt{4.5} = 2.12$$

$$2.12 > \underbrace{1.96}_{\alpha = 0.05,\ \text{unit normal table}} .\ H_0 \text{ disproved.}$$

Note on sign test

Great for when people don't finish.

Time to Complete:	Day 1	Day 2	Diff.	
A	20	18	2	+
B	40	45	-5	−
C	60	30	-30	−
D	— >60	55	?	−

More tests for Ordinal Data

1. Mann-whitney U test

   alternative to indep. measures t-test

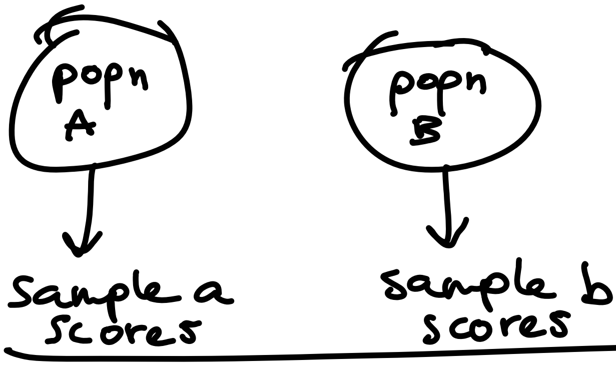
2. Wilcoxon test

   alternative to repeated measures t-test.

3. Kruskal-Wallis test

   alternative to single-factor indep. measures ANOVA

4. Friedman test : alternative to 1-factor repeated meas. ANOVA



Do $\bar{a}$, $\bar{b}$ tell us anything about μ_A and μ_B?

Assumptions of indep. measures t-test

1. indep. samples
2. A, B normal
3. $\sigma_A^2 = \sigma_B^2$

Assumptions of Mann-Whitney U test

1. indep. samples
2. A, B continuous

Mann-Whitney U test

	English texts	Arabic texts
#words with multiple meanings	27	71
	2	63
	9	18
	48	68
	6	94
	15	8

① Combine the data in a single list by rank.

Rank	Score	
1	2	ENG
2	6	ENG
3	8	ARA
4	9	ENG
5	15	ENG
6	18	ARA
7	27	ENG
8	48	ENG
9	63	ARA
10	68	ARA
11	71	ARA
12	94	ARA

$U_{ENG} = 6+6+5+5+4+4$
$= 30$

Let
Eng = Category A
Ara = Category B

note: ties in rank handled as they are for Spearman

$U_{ARA} = 4+2 = 6$

H_0: no difference between ranks of ENG and ARA.

number of b_i's below a_i

$$U_A = \sum_{i=1}^{n_A} \text{below}(a_i)$$

$$= 6 + 6 + 5 + 5 + 4 + 4 = 30$$

$$U_B = \sum_{i=1}^{n_B} \text{below}(b_i)$$

$$= 4 + 2 = 6$$

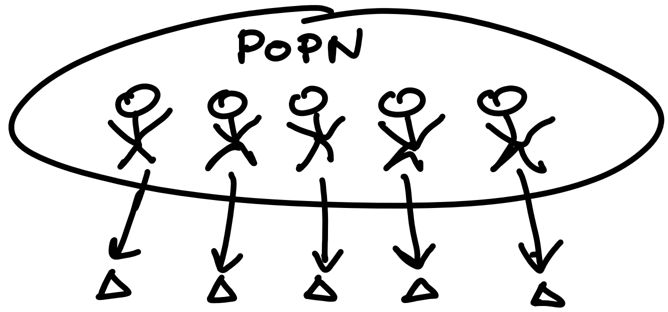
$$U_{final} = \min(U_A, U_B) = 6.$$

Compare U_{Final}

to critical value of $U(n_A, n_B)$

$= U(6,6) = 5$ at $\alpha = 0.05$

$6 \nleq 5$, so H_0 NOT DISPROVEN



Δ b/w Treatment A, Treatment B.

use sample data to infer something about μ_Δ

Assumptions of repeated measures t-test

1. sample scores within a treatment are indep.
2. distribution of differences for the popn is normal

Assumptions of Wilcoxon Signed Ranks

1. sample scores within a treatment are indep.
2. distribution of differences for the popn is CONTINUOUS

	pre test	post test	diff	rank of |diff|
A	18	18	∅	1.5
B	24	24	∅	1.5
C	31	30	−1	3
D	28	24	−4	4
E	17	24	+7	5
F	16	24	+8	6
G	15	26	+11	7.5
H	18	29	+11	7.5
I	20	36	+16	9
J	9	28	+19	10

H_0: no tendency for difference scores to be systematically positive or negative

---

① Divide ranks into positive or negative.

Positive		Negative	
Diff	Rank	Diff	Rank
$\emptyset$	1.5	$\emptyset$	1.5
7	5	−1	3
8	6	−4	4
11	7.5		
11	7.5		
16	9		
19	10		

split $\emptyset$ diffs equally

Positive		Negative	
Diff	Rank	Diff	Rank
∅	1.5	∅	1.5
7	5	−1	3
8	6	−4	4
11	7.5		
11	7.5		
16	9		
19	10		
TOTAL: 46.5		TOTAL: 8.5	

Wilcoxon $T = \min\left(\sum \text{pos ranks}, \sum \text{neg ranks}\right) = 8.5$

Critical value for $T(n=10) = 8$, $\alpha = 0.05$

$$\boxed{\begin{array}{c} 8.5 \nleq 8 \\ H_0 \text{ holds} \end{array}}$$