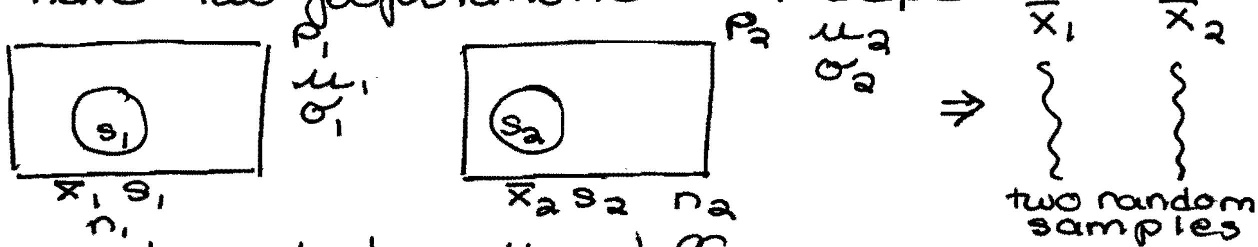
Recall: Question → Hypotheses → Data Collection → Data Analysis → Inference about the larger population

For one population, z test for σ known
t — σ unknown

I. If we have two populations: & independent samples



We are interested in the difference in means $(\mu_1 - \mu_2)$ for which the best estimator is $(\bar{x}_1 - \bar{x}_2)$

$H_0: (\mu_1 - \mu_2) = 0$

For one pop., focus was only σ known vs unknown. Here need to understand if $\sigma_1^2 = \sigma_2^2$ or not.

Two approaches:

1. Test assuming $=$ & then $\neq$ & compare.
2. Rule of thumb: If one of the s values is four or more times the other, assume unequal variance.

→ If $s_1 \geq 4 s_2$ → unequal and $\sigma_1^2 = \sigma_2^2$

For two pop. we need to compute a pooled variance = the weighted average of the two sample variances: s_p^2

$$s_p^2 = \frac{[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]}{(n_1 + n_2 - 2)}$$

We then find our test statistic and df

$$df = (n_1 + n_2 - 2)$$

value from H_0

$$t = \frac{\left[(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right]}{\left[s_p^2 \left(1/n_1 + 1/n_2\right)\right]}$$

If. $\sigma_1^2 \neq \sigma_2^2$,

$$df_{\neq} = \frac{\left[(s_1^2/n_1) + (s_2^2/n_2)\right]^2}{\left[\frac{(s_1^2/n_1)^2}{(n_1-1)} + \frac{(s_2^2/n_2)^2}{(n_2-1)}\right]}$$

and $$t_{\neq} = \frac{\left[(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right]}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

larger df, larger n have lower prob. of Type II error.

$df_{=} > df_{\neq}$ → if in doubt, choose = variance test

* Remember, z- & t-tests assume a normal distribution!
→ plot the data first!

II. If we have two samples that are somehow linked (before / after, old part/new part, in the wild / in captivity) with two measures on each subject:

	A		B	Diff
1	•	→	•	x
2	•	→	•	x
3	•	→	•	x
4	•	→	•	x
5	•	→	•	x

our statistic of interest is not two two separate measures, but the diff. btwn them X_D

$X_D = X_A - X_B$

Our null hypothesis then takes the form:

$H_0: \underline{\underline{\underline{\mu_D}}} = 0$ & $H_a: \mu_D > < \neq 0$

↑ "Matched pairs"
"Paired sample"

The test statistic is

$$t_0 = \frac{\bar{x}_D - \mu_0}{s_D / \sqrt{n_D}} \quad \text{with } df = n_D - 1$$

(looks like a one population test!)

III. If you data are very non-normal, you can't use tests involving t statistics

→ need a non-parametric test

Wilcoxon Signed Rank Sum Test

for two populations

nonnormal data

matched pairs data

Process: Estimates the differences x_D

Take absolute value of x_D

Eliminate all $x_D = 0$

Rank the remainder from 1 = smallest to n = highest value, but separately for + diff. & − diff.

Sum the ranks to get totals T_+ & T_-

If T_+ & T_- differ significantly, we can infer H_0 of no difference is false

Ex:

x_1	x_2	D	|D|	Rank +	Rank −
5	6	−1	1		1
7	7	0	0		
9	4	5	5	4	
11	14	−3	3		3
13	15	−2	2		2
				4	6
				T_+	T_-

$T_+ > T_-$

for samples of $n > 30$ the totals are approximately normally distributed. with

$$E(T) = \frac{n(n+1)}{4} \qquad \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

and $z = \dfrac{T_+ - E(T)}{\sigma_T}$

T_+ is the test statistic in this case.