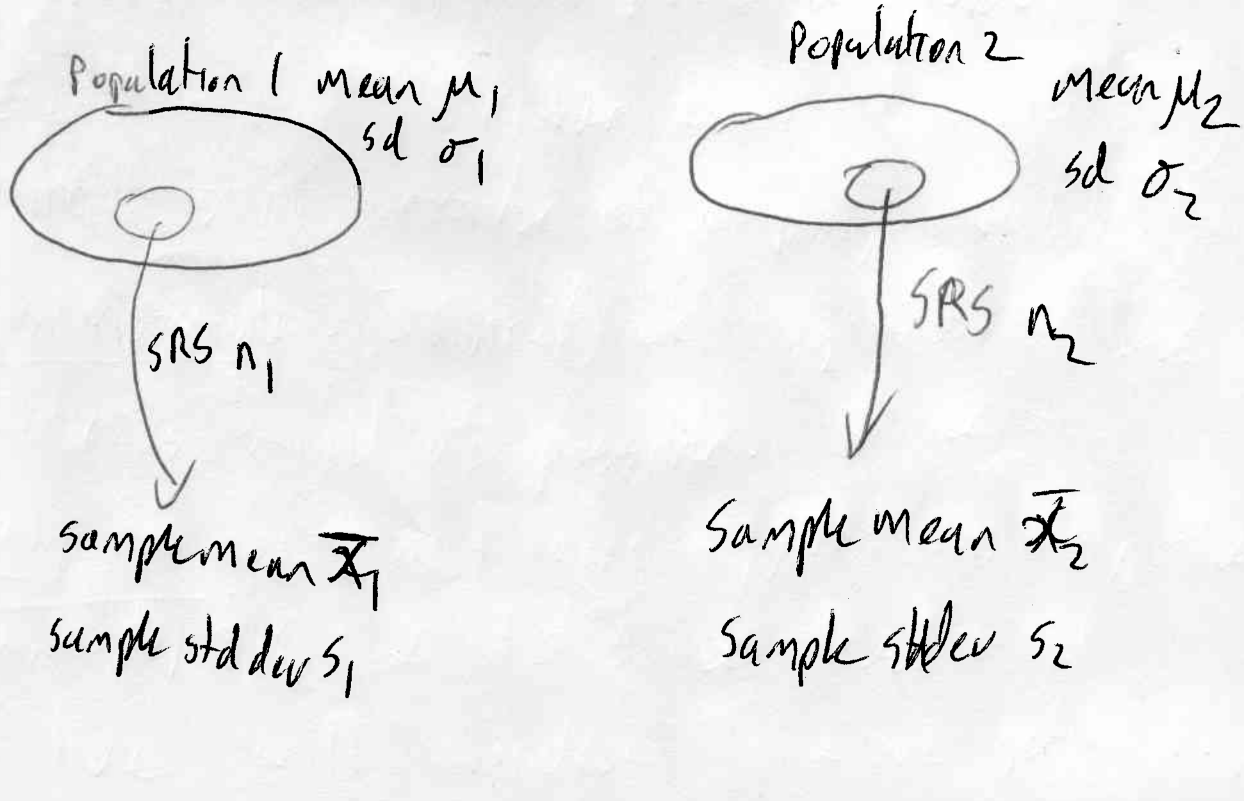
Lecture 28

Today we begin our discussion of two sample problems in statistical inference.

Two-Sample problem

- Goal is to compare responses in two groups
- Each group considered to be a sample from distinct population
- Responses in each group are independent of those in the other group.

ie



Goal is to use $\bar{x}_1, \bar{x}_2, s_1, s_2$ to make inferences about differences between μ_1 and μ_2.

When would we use two sample methods?

Comparative experiments between two groups.

eg randomly assign patients to placebo or drug group. Then later measure their response/value of some variable to decide whether the drug was helpful or not.

Note we use

$$\mu_1 - \mu_2 \quad (\text{or } \mu_2 - \mu_1)$$

to compare population means. If ~~three~~ two (population) means are equal this implies

$$\mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$$

Note we estimate $\mu_1 - \mu_2$ with $\bar{x}_1 - \bar{x}_2$ and it can be shown that $SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Two sample confidence interval for $\mu_1 - \mu_2$

Suppose we take a SRS of size n_1 from a normal population with unknown mean μ_1 and an independent SRS of size n_2 is taken from a normal population with unknown mean μ_2.

Then a level C Confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $P(-t^* < T < t^*) = C$ and T has t distribution with $\min(n_1 - 1, n_2 - 1)$ df.

Two sample hypothesis test

Again assume conditions above.

or rewriting clearer

$H_0: \mu_1 - \mu_2 = 0$

$H_A: \mu_1 - \mu_2 \neq 0$

$$\left.\begin{aligned} H_0&: \mu_1 = \mu_2 \\ H_A&: \mu_1 \neq \mu_2 \end{aligned}\right\} \text{typical null and alternative}$$

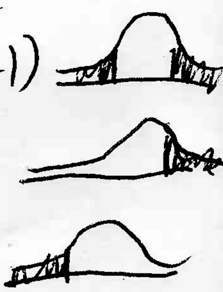
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

← test statistic

has t distribution with $\min(n_1 - 1, n_2 - 1)$ df.

P-values come from t distribution, as before the tail that we look at comes from the form of the alternative

Null	Alternative	P-value
$H_0: \mu_1 - \mu_2 = 0$	$H_A: \mu_1 - \mu_2 \neq 0$	$2P(T > \lvert t \rvert)$
$H_0: \mu_1 - \mu_2 \leq 0$	$H_A: \mu_1 - \mu_2 > 0$	$P(T > t)$
$H_0: \mu_1 - \mu_2 \geq 0$	$H_A: \mu_1 - \mu_2 < 0$	$P(T < t)$



Example

Researchers studying the learning of speech often compare measurements made on the recorded speech of adults and children. One variable is called the voice onset time.

We have results for six year old children and adults asked to pronounce the word "bees". The VOT measured in milliseconds may be positive or negative.

Results

Group	n	$\bar{x}$	s
Children	10	-3.67	33.89
Adults	20	-23.17	50.74

Test whether there is a difference in VOT for the word "bees" between adults and children.

First let
μ_1 = mean of children
μ_2 = mean of adults

$H_0: \mu_1 - \mu_2 = 0$

$H_A: \mu_1 - \mu_2 \neq 0$

$$t = \frac{-3.67 - -23.17}{\sqrt{\frac{33.89^2}{10} + \frac{50.74^2}{20}}} = 1.25 \text{ (2dp)}$$

df = min(10−1, 20−1) = 9

1.125 is between 1.100 and 1.383

so pvalue is between 2*.15 and 2*.1

is

$.2 < \text{pvalue} < .3$

so cannot reject H_0. No evidence to show there is a difference.