read:
LN pp. L-180+
L-213
today:
LN pp. L-188→

this time: 2-sample paired comparisons

next time: 2 independent samples

Ashley Tai
AMS 7
17 May 18

2-Sample Problem

1. paired comparisons

CORTEX STUDY

pair #	T	C	D = T − C (difference)
1	689	657	+32
2	656	623	+33
59	649	602	+47
mean	$\bar{y}_T = 683$	$\bar{y}_c = 647$	$\bar{d} = +36.2$
SD	$S_T = 32$	$S_c = 30$	$S_d = 31.5$

* here, there are 2(59) = 118 individuals grouped into n = 59 pairs by holding PCF (genetics) constant
* 2-sample problem converts into a 1-sample problem when attention in inference is focused on the difference column

SYSTOLIC BP STUDY

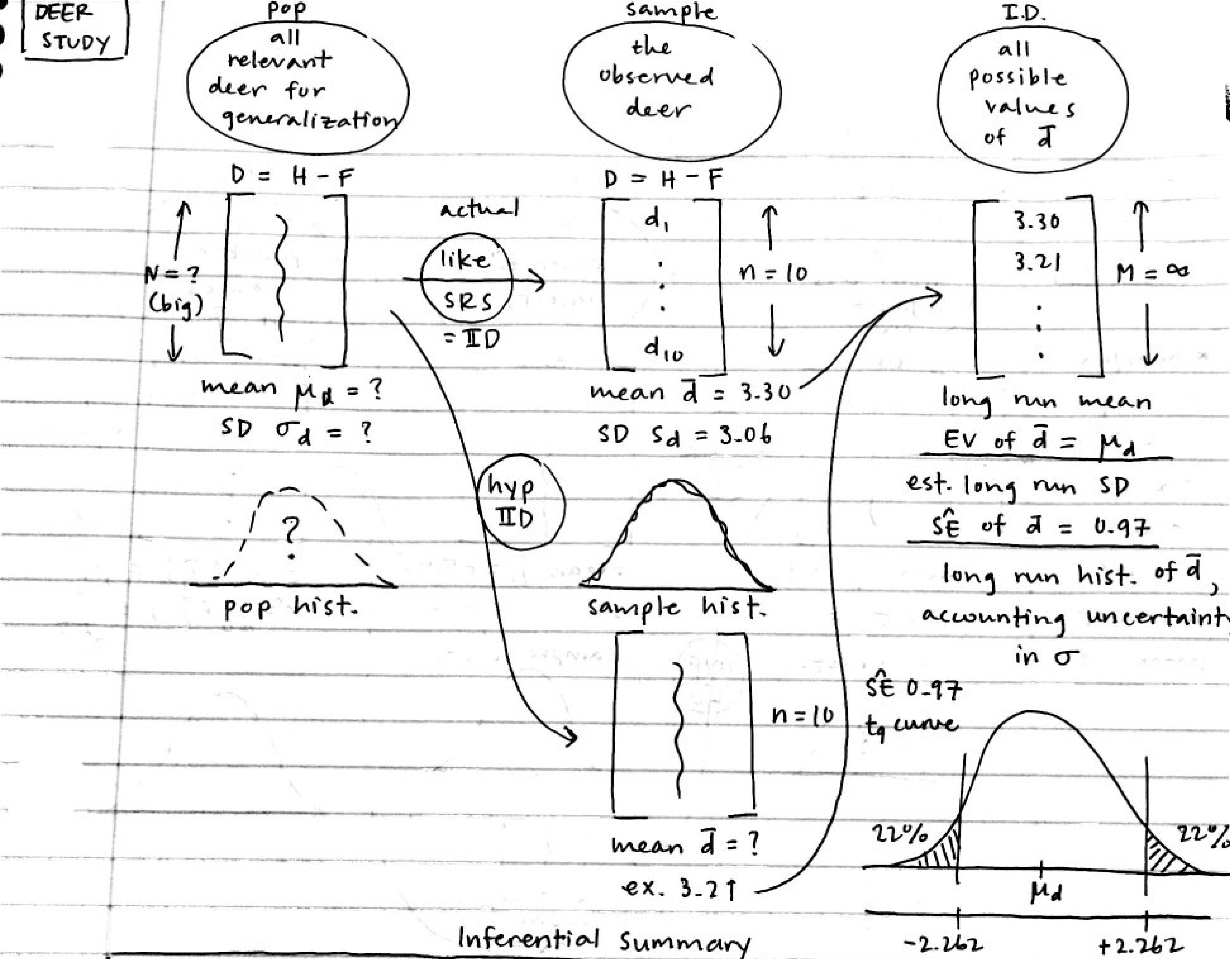
← blood pressure →

person	before	after	diff. (A − B)
1			
2			
⋮			
n			
mean	$\bar{y}_B = 174$	$\bar{y}_A = 160$	$\bar{d} = -14$
SD	$S_B = 20$	$S_A = 21$	$S_d = 11$

statsig

* here, there are only n individuals, not 2n as in matched pairs
* likely to be accurate because 2 measurements per person, which holds entire person constant

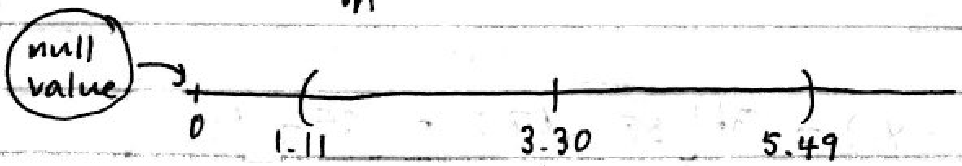
$$\frac{\bar{y}_A - \bar{y}_B}{\bar{y}_B} = \frac{160 - 174}{174} = -0.08 = -8\% \text{ (pract sig)}$$



Inferential Summary

P	unknown pop quantity of main interest	μ_d = pop. mean diff. in H vs. F length
S	estimate	$\bar{d}$ = 3.30 cm
ID	give or take for $\bar{d}$ as est. of μ_d	$\hat{SE}(\bar{d})$ = 0.97 cm
	95% CI for μ_d	$\bar{d} \pm 2.262\,\hat{SE}(\bar{d})$ = (1.11, 5.49)

$$\bar{d} \pm \left(t_{n-1}^{0.95}\right)\frac{s}{\sqrt{n}} = 3.30 \pm (2.262)(0.97) = 3.30 \pm 2.19$$



* 0 is not in 95% CI, so diff <u>is</u> statsig.

Sokal & Rohlf (1995)

* analysis of sample I & II are the same, but w/ different notations

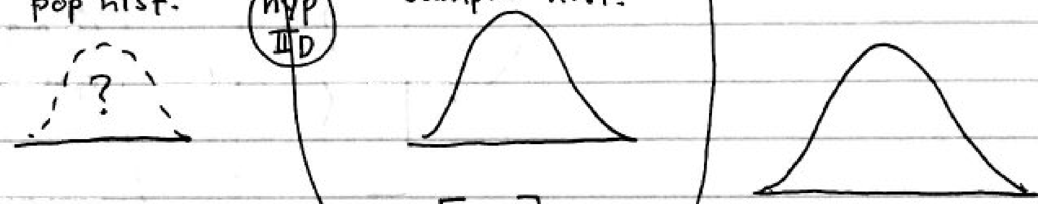
2 independent samples

I

pop	sample	I.D.
all female insects from this genetic cross	the observed insects	all possible values of $\bar{Y}_1$
time	time	7.5143, 7.58, ⋮
$N = ?$		$M_1 = \infty$
mean $\mu_1 = ?$	mean $\bar{Y}_1 = 7.5143$	$E_{IID}(\bar{Y}_1) = \mu_1$
SD $\sigma_1 = ?$	SD $s_1 = 0.7104$	$\hat{SE}_{IID}(\bar{Y}_1) = \frac{s_1}{\sqrt{n_1}}$
pop hist.	sample hist.	

like SRS

hyp IID

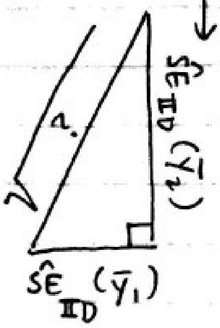


mean $\bar{Y}_1$
ex 7.58

II repeat (ditto for genetic cross II)

Inferential Summary

P	unknown pop. quantity of interest	$(\mu_2 - \mu_1)$ = pop mean diff in time to reproduce b/t gen cross I & II
S	estimate	$(\bar{Y}_2 - \bar{Y}_1) = 7.5571 - 7.5143 = +0.0428$ days
ID	give or take $(\bar{Y}_2 - \bar{Y}_1)$ as est of $(\mu_2 - \mu_1)$	$\hat{SE}(\bar{Y}_2 - \bar{Y}_1) = 0.3614$ days
	95 % CI $(\mu_2 - \mu_1)$	$(\bar{Y}_2 - \bar{Y}_1) \pm (2.179)\,\hat{SE}(\bar{Y}_2 - \bar{Y}_1)$



$$\hat{SE}_{IID}(\bar{Y}_2 - \bar{Y}_1) = \sqrt{[\hat{SE}_{IID}(\bar{Y}_1) + \hat{SE}_{IID}(\bar{Y}_2)]}$$

indep

$$= \sqrt{\left(\frac{s_1}{\sqrt{n_1}}\right)^2 + \left(\frac{s_2}{\sqrt{n_2}}\right)^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\hat{SE}_{IID}(\bar{y}_1) = \frac{s_1}{\sqrt{n_1}} = \frac{0.7104 \text{ days}}{\sqrt{7}} = 0.2685 \text{ days}$$

$$\hat{SE}_{IID}(\bar{y}_2) = \frac{s_2}{\sqrt{n_2}} = \frac{0.6399 \text{ days}}{\sqrt{7}} = 0.2419 \text{ days}$$

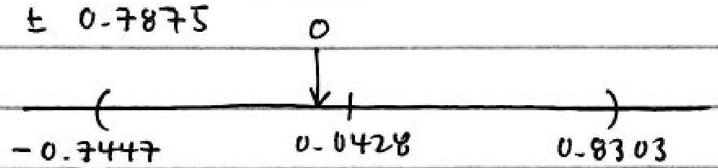
* uncertainty (SE) in $(\bar{y}_2 - \bar{y}_1)$ is related to uncertainty in $\bar{y}_1$ (SE) & $\bar{y}_2$ (SE)

$$\hat{SE}_{IID}(\bar{y}_2 - \bar{y}_1) = \sqrt{(0.2685)^2 + (0.2419)^2} = \boxed{0.3614 \text{ days}}$$

$\boxed{\text{OR}}$

$$\hat{SE}_{IID}(\bar{y}_2 - \bar{y}_1) = \sqrt{\frac{0.7104^2}{7} + \frac{0.6399^2}{7}} \doteq \boxed{0.3614 \text{ days}}$$

$$(\bar{y}_2 - \bar{y}_1) \pm 2.179\, \hat{SE}(\bar{y}_2 - \bar{y}_1)$$
$$= 0.0428 \pm 2.179\,(0.3614)$$
$$= 0.0428 \pm 0.7875$$



★ not statsig ★