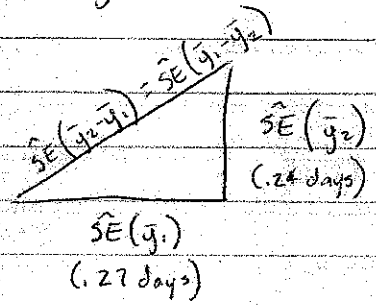
Feb. 24, 2009

- Analysis of 2 independent samples (con't)

Daphnia Ex.

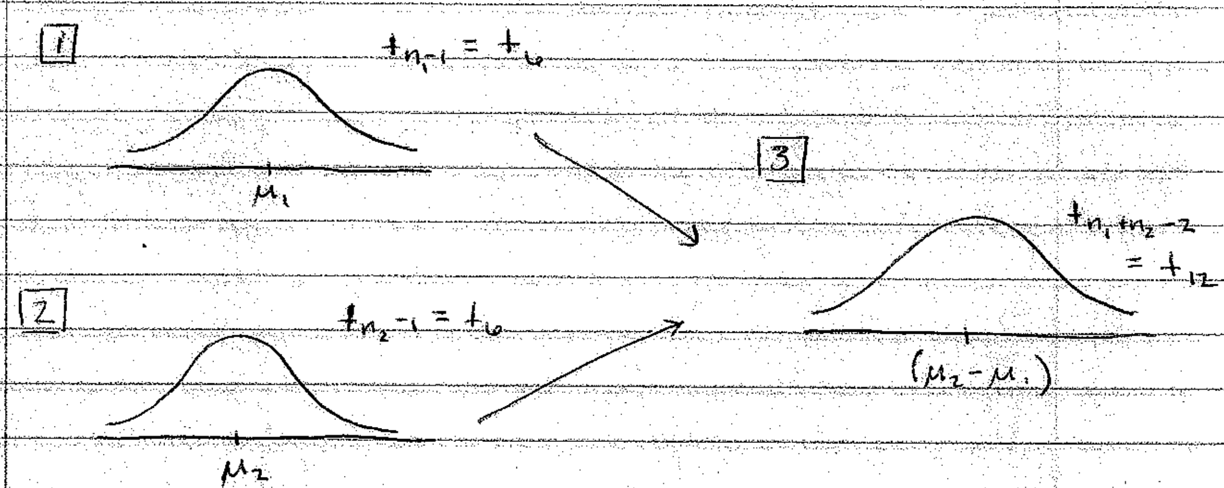
SD = ? $\bar{y}_2 - \bar{y}_1 = \bar{y}_2 + (-\bar{y}_1)$

* math fact: $\hat{SE}(\bar{y}_1)$ and $\hat{SE}(\bar{y}_2)$ combine in calculating $\hat{SE}(\bar{y}_2 - \bar{y}_1)$ like the edges of a right triangle.



$$SE_{diff.} = \sqrt{\left[\hat{SE}(\bar{y}_1)\right]^2 + \left(\hat{SE}(\bar{y}_2)\right)^2} = .3614 \text{ day}$$

$$= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = SE_{difference}$$



- 2 Independent Samples: Dichotomous Outcomes

Ex: Parasite that causes sudden oak death also affects redwoods

Feb. 24, 2009

California: 265 trees, 9 infected

Oregon: 281 trees, 20 infected

- Is the parasite affecting trees more in Oregon than CA?
  → need a data set for each population

CA:

Population — all redwood trees in CA grove

Sample — observed redwood trees

Imaginary — possible $\hat{p}$'s

Infected? 1 = yes, 0 = no

Population: 1's or 0's

mean p_1 = ?

SD σ_1 = ?
$= \sqrt{p_1(1-p_1)}$

Sample: Infected? 1's or 0's, $n_1 = 265$

mean $\hat{p} = \frac{9}{265} = 3.4\%$

[ ] $n_1 = 265$

mean $\hat{p} = 3.7\%$

Imaginary: 3.4%, 3.7%, … $M_1 = \infty$

- long run mean: $E_{IID}(\hat{p}) = p_1$
- long run SD: $\hat{SE}(\hat{p}) = 1.1\%$

OR:

Population — all trees in OR grove

Sample — observed trees

Imaginary — possible $\hat{p}$'s

Population: [ ]

mean p_2 = ?

SD $\sigma_2 = \sqrt{p_2(1-p_2)}$

Sample: 1's or 0's, $n_2 = 281$

mean $\hat{p} = \frac{20}{281} = 7.1\%$

Imaginary: 7.1%, … $M_2 = \infty$

- long run mean: ~~EED~~ $E_{IID} = p_2$
- long run SD: $\hat{SE}_{IID}(\hat{p}_2) = 1.5\%$

AR

3/5

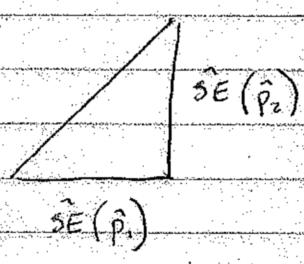
Feb. 24, 2009

Inferential Summary

Unknown Quantity of Interest	$(p_1 - p_2)$ = pop. mean difference in infection rates in OR & CA
Estimate	$(\hat{p}_1 - \hat{p}_2) = 3.4\% - 7.1\% = -3.7\%$
Give or take	$\hat{SE}(\hat{p}_1 - \hat{p}_2) = 1.9\%$
95% CI for $(p_1 - p_2)$	$(\hat{p}_1 - \hat{p}_2) \pm 2\,\hat{SE}(\hat{p}_1 - \hat{p}_2)$

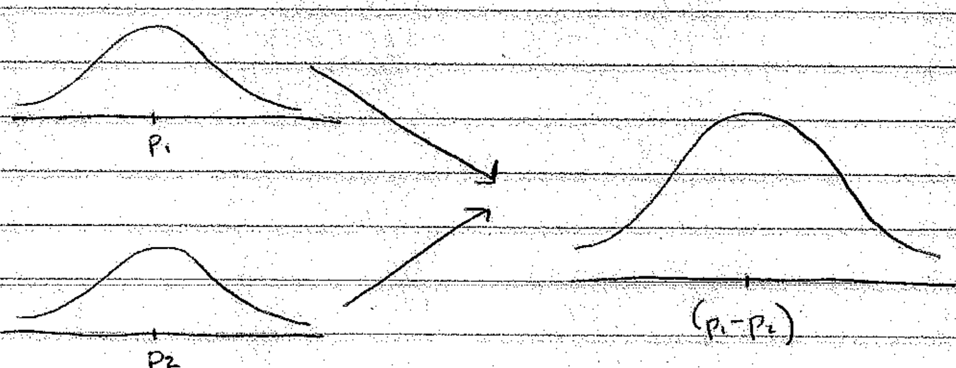
$$\hat{SE}_{IID}(\hat{p}_1) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}} = 1.1\%$$

$$\hat{SE}_{IID}(\hat{p}_2) = \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n}} = 1.5\%$$



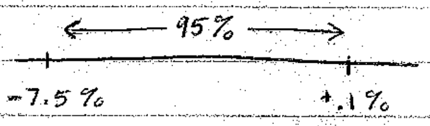
$$\hat{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{(.011)^2 + (.015)^2} = 1.9\%$$

long run histogram



Feb. 24, 2009

$$95\%\ CI = (\hat{p}_1 - \hat{p}_2) \pm 2\,\hat{SE}(p_1 - p_2)$$
$$= (-3.7\%) \pm 2(1.9\%)$$
$$= -3.7\% \pm 3.8\%$$

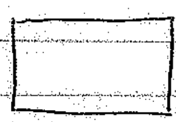


So, 0 is in interval and difference is not statistically significant at 95% level (need more data)

Simple Correlation and Regression

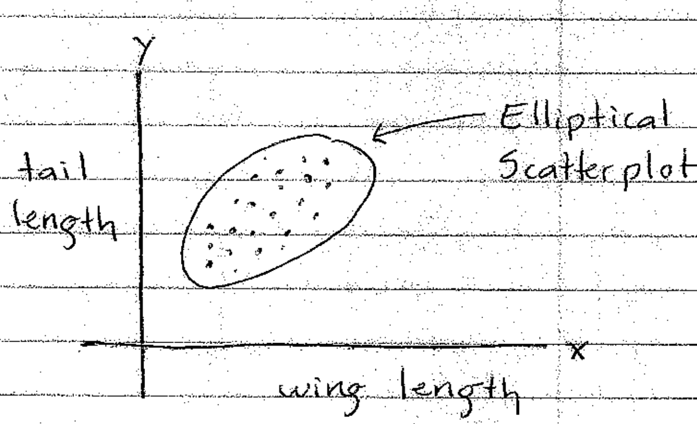
Sparrows: y = tail length (cm)
x = wing length (cm)

Sample

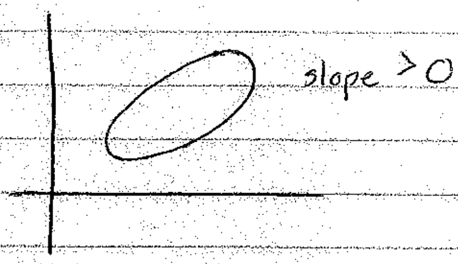


	Y	X
	7.4	10.4
	7.6	10.8
	⋮	⋮
	8.3	11.4
mean	$\bar{y} = 7.6$	$\bar{x} = 10.7$
	$s_y = .35$	$s_x = .40$

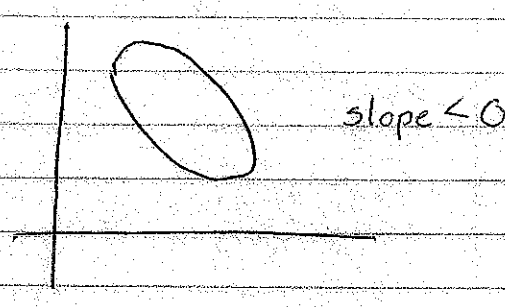
$n = 12$



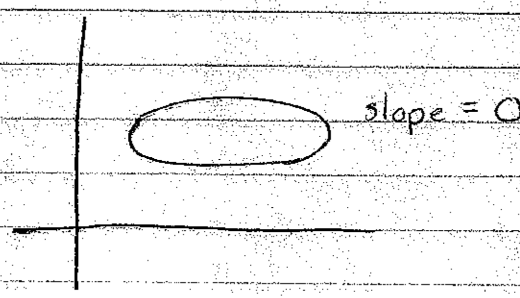
Feb. 24, 2009



positive association

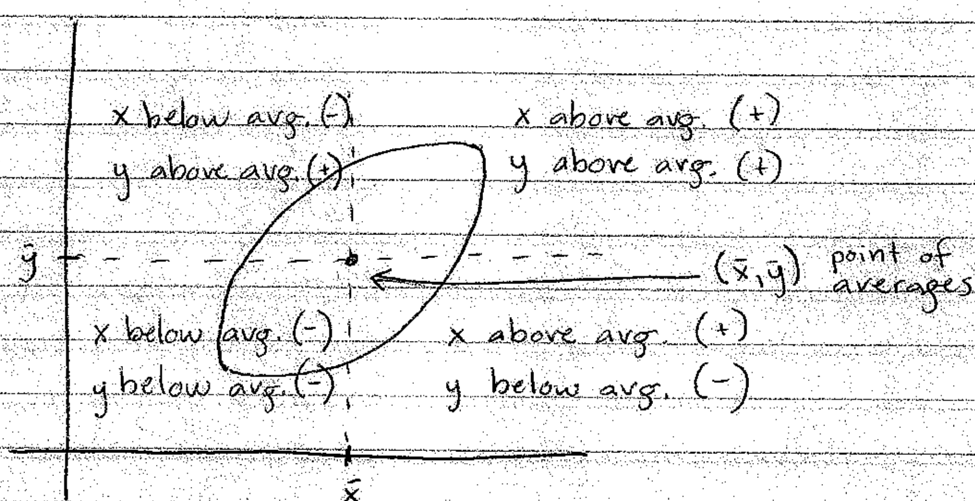


negative association



no association

Strength of Association?



Convert to standard values & multiply: $\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$

take mean: r (product moment correlation coefficient)

$$= \frac{1}{n} \sum^{n} \left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$$

where: $s_x^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_2 - x_1)^2}$ $\quad s_y^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_2 - y_1)^2}$