AMS 7
Notes
1
11/10
Julia Larson

11/10 2 independent sample dichotomous outcomes

ex. Sudden Oak death parasite in Redwood trees comparison in CA and OR. CA 9 out of 265 randomly choosen trees in a grove had it
OR 20 out of 281 " " " " " " " "

(CA)

Pop.	Sample	Imag. data
All Redwood trees in chosen grove	The observed Redwood trees	Possible $\hat{p}$

Infected?

N_1 = Thousands

1 = Yes
0 = no

$\begin{bmatrix} 1\text{'s} \\ + \\ 0\text{'s} \end{bmatrix}$ SRS = IID, Hypo. IID

Infected?

$\begin{bmatrix} 1\text{'s} \\ + \\ 0\text{'s} \end{bmatrix}$ $n_1 = 265$

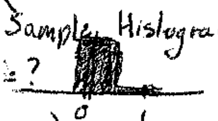
mean $\hat{p}_1 = \frac{9}{265} = 3.4\%$

$\begin{bmatrix} 3.4\% \\ 3.6\% \\ \vdots \end{bmatrix}$ $M = \infty$

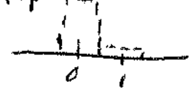
mean p_1 = ?

SD $\sigma_1 = \sqrt{p_1(1-p_1)}$ = ?

Sample Histogram



Pop. Hist.



$\begin{bmatrix} 1\text{'s} \\ + \\ 0\text{'s} \end{bmatrix}$ $n_1 = 265$

$\hat{p} = 3.6\%$

Long run mean $\boxed{1}$ EV of $\hat{p}_1 = p_1$

(est.) long run SD $\boxed{2}$ $\hat{SE}$ of $\hat{p}_1 = 1.1\%$

long run histogram $\boxed{3}$

A similar diagram exists for OR grove, with some differences mean $\hat{p} = \frac{20}{281} = .7\%$, with long run mean $\hat{p}_2 = p_2$, & est. long run SD = 1.5%

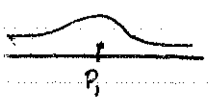
Inferential Summary

Unknown pop. quantity of interest	$p_1 - p_2$ = Pop. difference in % infection of (CA-OR)
Estimate of $(p_1 - p_2)$	$\hat{p}_1 - \hat{p}_2 = 3.4\% - 7.1\% = -3.7\%$
Give or take for $(\hat{p}_1 - \hat{p}_2)$ as estimated from $(p_1 - p_2)$	$\hat{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{(1.1\%)^2 + (1.5\%)^2} = 1.9\%$
95% interval for $(p_1 - p_2)$	$-3.7\% \pm (1.96)(1.9\%) = 3.8\%$

$\boxed{1}$ EV of $\hat{p}_1 = E_{IID}(\hat{p}_1) = p_1$

$\boxed{2}$ $SE_{IID}(\hat{p}_1) = \sqrt{\frac{p_1(1-p_1)}{n_1}} = \sqrt{\frac{(.034)(.966)}{265}} = 0.011 = 1.1\%$

$\boxed{3}$

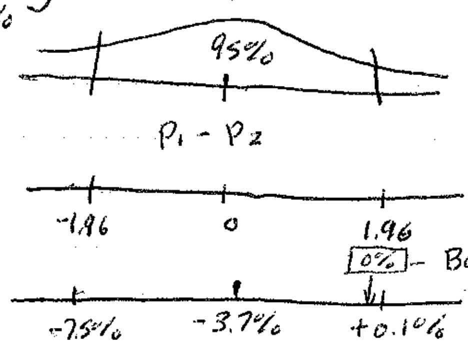


AMS 7
Notes
2
11/10
Julia Larson

Math Fact: $\hat{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Long run Histogram $\hat{p}_1 - \hat{p}_2$

$\hat{SE}$ 1.9%



- Barely not statistically significant, but since most of the data is ⊖ it seems it should be significant, to fix it we need more tree samples

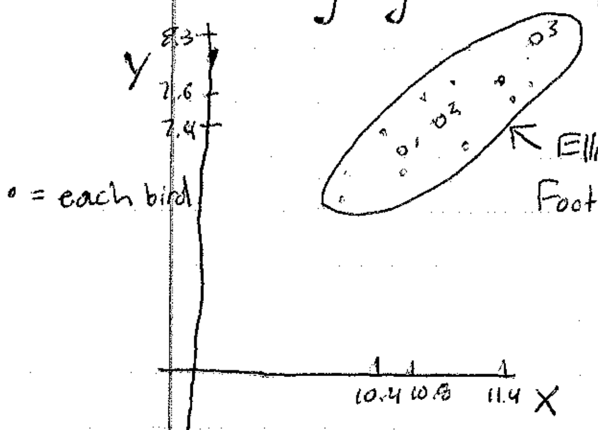
Simple Correlation and Linear Regression

Case Study

comparing tail and wing length correlation of birds

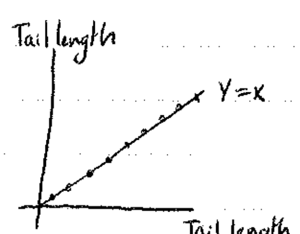
Y = Tail length (cm) (outcome)
X = Wing length (cm) (predictor)

Y	X
7.4	10.4
7.6	10.8
⋮	⋮
8.3	11.4
mean $\bar{Y}$ = 7.6cm	mean $\bar{X}$ = 10.7cm
SD S_Y = 0.35cm	SD S_X = 0.4cm
hist.	

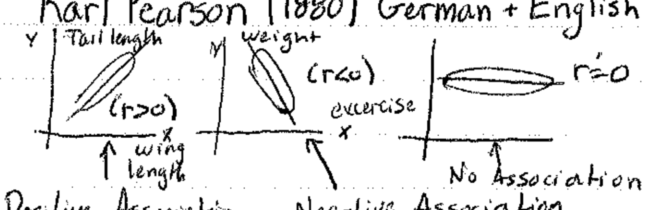


Scatter Plot or
Scatter Diagram

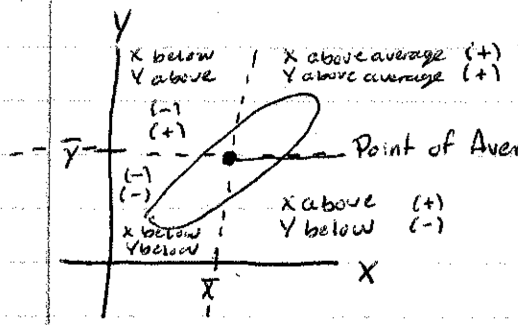
Bivariate, normal distribution (elliptical scatterplot)



Karl Pearson (1880) German + English Scientists (Eugenist)



AMS 7
Notes
3
11/10
Julia Larson



Point of Average (moved X and Y axis to this point

Positive Association

$$\begin{bmatrix} Y_1 & X_1 \\ Y_2 & X_2 \\ Y_3 & X_3 \\ \vdots & \vdots \\ Y_n & X_n \end{bmatrix} \; n$$

$$r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_2 - \bar{X}}{S_x^*}\right) \cdot \left(\frac{Y_2 - \bar{Y}}{S_Y^*}\right)$$ = Pearson's product moment correlation coefficient

$$S_x^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_2 - \bar{X})^2}$$

$$S_Y^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_2 - \bar{Y})^2}$$