

this probability  
 time: models for  
 next weeks &  
 time: weeks

read: DD ch. AMS7  
 10 15 Oct 09  
 hwk 2 due  
 Tue 20 Oct in class ①

lab 2 due by 5pm Fri  
 23 Oct in class or  
 lab or disc. see or  
 at my office (BE  
 135)

MLP gender (F, M)

Y	F
Y	M
N	F
Y	F
⋮	⋮

$n = 106$

MLP = marijuana

legalization preference  
 (Y, N)

Y	F	4
⋮	⋮	29
Y	F	+
Y	M	+
⋮	⋮	52
Y	M	+
N	F	↑
⋮	⋮	20
N	F	↓
N	M	↑
⋮	⋮	5
N	M	↓
		106

gender

	MLP		
	Y	N	
F	29	20	49
M	52	5	57
	81	25	106

2x2 contingency  
 table (example of  
 categorical data  
 analysis)

Q: Are gender, MLP independent in this data set, or are they dependent (associated)? <sup>②</sup>

A:  $P(Y) = \frac{81}{106} \approx 76\%$

$P(Y|F) = \frac{29}{49} \approx 59\%$

$P(Y|M) = \frac{52}{57} \approx 91\%$

Choose 1 person  
at random from

106

ELM ✓

so gender,  
MLP are  
(highly)  
dependent

in this data set because

$76\% \neq 59\% \text{ or } 91\%$

finish  
T-S

$P(\text{1 or more T-S in 5}) =$

$1 - P(0 \text{ T-S in 5})$

$= 1 - P(\text{not T-S on 1st}) \textcircled{and} \text{not T-S on 2nd} \textcircled{and} \dots \textcircled{and} \text{not T-S on 5th}$

indep

$$= 1 - P(\text{not TP}) \cdot P(\text{not TP on 2nd}) \cdots P(\text{not TP on 5th})$$

identical distribution

$$= 1 - \left(1 - \frac{1}{4}\right) \cdot \left(1 - \frac{1}{4}\right) \cdots \left(1 - \frac{1}{4}\right)$$

$$= 1 - \left(1 - \frac{1}{4}\right)^5 = 76\%$$

each pen 1/4

$$P(DP) = \frac{36}{326} \approx 11\%$$

$$P(DP | DW) = \frac{19}{160} \approx 12\%$$

$$P(DP | DB) = \frac{17}{166} \approx 10\%$$

outcome:

DP vs. (Y) not

treatment (X)

race (W vs. B) dependent

basic design: obs. study

causing: bias from RCT

MLP: race of victim (W vs. B)

Z, Y assoc? ✓

Z, X assoc? ✓

how defect PCF at analysis time:

hold it constant

(4)

$$P(DP | VW) = \frac{30}{214} \approx 14\%$$

$$P(DP | DW \text{ and } VW) = \frac{19}{151} \approx 13\%$$

$$P(DP | DB \text{ and } VW) = \frac{11}{63} \approx 17\%$$

$$P(DP | VB) = \frac{6}{112} \approx 5\%$$

$$P(DP | DW \& VB) = \frac{0}{9} = 0\%$$

$$P(DP | DB \& VB) = \frac{6}{103} \approx 6\%$$

direction of relationship

between  $X$  &  $Y$  is total when

$Z$  was controlled for: Simpson's Paradox

11 (coming out ahead 1 spin, right?) (5)

$$= \frac{1}{38} (\text{ELM} \checkmark) P(\text{split}) = \frac{2}{38}$$

$\approx 2.5\%$

$\approx 5\%$

$$\mu = \frac{(-8) + (-1) + \dots + (-1) + (+35)}{38}$$

$$= \frac{-2}{38} = -0.0526 : \text{ each time I bet}$$

\$1 on a single # I expect to

lose about 0.05, give or take

about

$$\sigma = \$5.76$$

pop sd  $\sigma$  : with fact: if pop.

has only 2 values,

$$\sigma = \left[ \left( \frac{\text{larger}}{\#} \right) - \left( \frac{\text{smaller}}{\#} \right) \right]$$

$$\sqrt{\left( \frac{\text{proportion of larger}}{\#} \right) \left( \frac{\text{proportion of smaller}}{\#} \right)}$$

pop possible outcomes of a single spin your net gain

single #

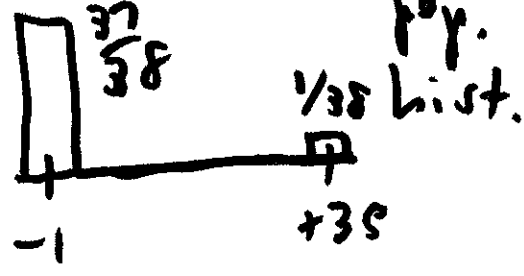
sample observed spins

imaginary data set all possible values of  $\bar{x}$

$N = 38$

-81	0
-1	00
...	...
+35	6
-1	?
...	...
-1	36

mean  $\mu = 0.05$   
 SD  $\sigma = 5.76$



IID

your net gain  
 $\begin{bmatrix} -1 \\ -1 \\ +35 \\ -1 \\ \vdots \end{bmatrix} n = 1000$

sum  $\bar{x} = ?$   
 (ex. -64)

IID

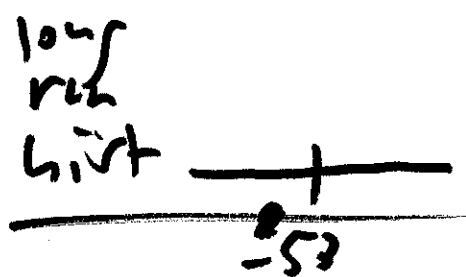
$\{ \} n = 1000$

sum  $\bar{x} = ?$   
 (ex. -28)

$\begin{bmatrix} -64 \\ -28 \\ \vdots \end{bmatrix}$  a set of  $M=100$

long run expected value of mean  $\bar{x} = 0.53$

long run  $\bar{x}$



here  $\sigma = \left[ \binom{1}{+35} - \binom{1}{-11} \right] \sqrt{\left( \frac{1}{38} \right) \left( \frac{37}{38} \right)}$  (7)

$\approx 5.76$

$P(\text{coming out ahead}) = P(S > 0) = ?$

1000 spins,  $\frac{1}{38}$  of  $\binom{1}{+35}$ ,  $\frac{37}{38}$  of  $\binom{1}{-1}$

I expect around 26  $\binom{1}{+35}$  & 974  $\binom{1}{-1}$ .

$\text{sum} = 26 \binom{1}{+35} + 974 \binom{1}{-1} = -64$

another possibility: 27  $\binom{1}{+35} + 973 \binom{1}{-1}$

$= -28$

$64 - 28 = 36$

← 36 →
$\frac{-1}{-1} \quad \frac{+1}{+35}$

math fact:

Expected value of  $S = EV = E_{IID}(S) = n\mu$

$= \left( \frac{\# \text{ spins}}{\text{mean}} \right) = (1000) (-0.05) = -50$

after 1000 \$1 play of single #  
I expect to be behind by  
about \$53 (EV) given take

