Note Title 9/9/2009

STT 200 9-9-09

WWW.STT.MSU.EDU/~LEPAGE

REC. ONE ASSIGNMENT (OPTIONALLY) DUE NEXT TUESDAY.

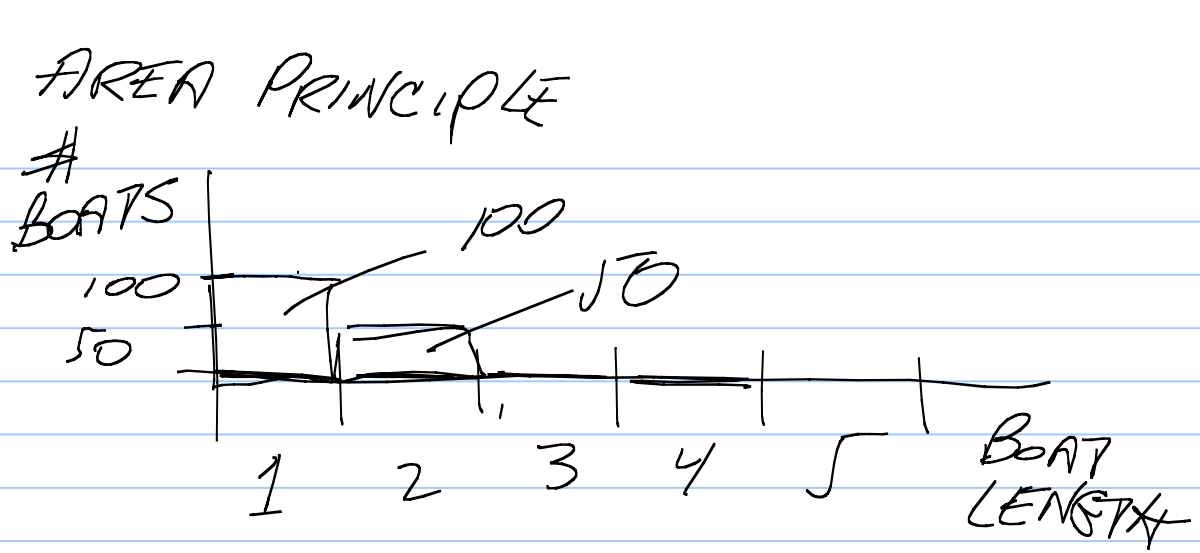
EXTRA CREDIT FOR THOSE WHO DID IT RIGHT FIRST TIME.

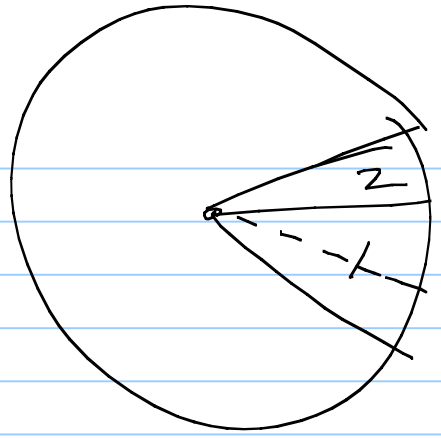
REC. 2 SUBMIT (FROM CH 3 - Ex. 13, 19, 22, 29, 35, 38.)
PLUS T.B.A FROM CH 4.

TODAY AREA PRICIPLE
CATEGORICAL DATA
BAR CHART
PIE CHART

CONTINGENCY TABLES
JOINT DISTRIBUTION
MARGINAL DISTRIBUTION(S)
CONDITIONAL DISTRIBUTION(S)
INDEPENDENCE (PROPORTIONALITY)
SIMPSON'S "PARADOX" (CONFOUNDING)

AREA PRINCIPLE





CONTINGENCY TABLES.

eg TITANIC

	TICKET CLASS 1	2	3	CREW	TOTAL
ALIVE	203	118	178	212	711
DEAD	122	167	528	673	1490
	325	285	706	885	2201

ROW TOTAL

COL TOTALS.

RELATIVE FREQUENCIES

	1
ALIVE	
DEAD	122/2201

TOTAL IS $\frac{2201}{2201} = 1$

CAN INTERPRET $\frac{122}{2201}$ AS CHANCE (PROB) OF GETTING A PERSON IN CATEGORY 1-DEAD IF SELECT AT RANDOM FROM 2201.

(4) JOINT DISTRIBUTION IS COMPLETE TABLE OF THESE "PROBABILITIES."

MARGINAL DISTRIBUTION OF LIVE-DIE

	1	2	3	CREW	
ALIVE	203	118	178	212	711/2201
					1490/2201

MARGINAL PROB'Y DIST^N LIFE-DEATH

ALIVE	711/2201
DEAD	1490/2201
	1

ALSO

MARGINAL FOR (BOOKING)

	1	2	3	4
ALIVE	203			
DEAD	122			

$$P(\text{TICKET } 1) = \frac{203 + 122}{2201} = 325/2201$$

JOINT DIST (ON ROW-COL)

$\Rightarrow$ MARGINAL FOR "MORTALITY"

ALSO FOR TICKET CLASS

FOR PERSON WHO LIVED (REFER ROW 1)
WHAT IS THEIR CHANCE OF HOLDING TICKET 1?

	1	2	3	CREW	TOTAL
ALIVE	203	118	178	212	711
	122	167	528	673	1490

⇒ CONDITIONAL DIST^N OF TICKET TYPE CONDITIONAL ON HAVING LIVED.

	1	2	3	CREW
ALIVE	203/711	118/711	178/711	212/711
LIKEWISE FOR DEAD	122/1490	167/1490	528/1490	673/1490

← TOTAL OF THESE IS 1

RATHER THAN GO INTO THE COMPARISON OF THE ABOVE TWO CONDITIONAL DISTRIBUTIONS - LET'S LOOK AT CONDITIONAL PROBABILITY DIST^N OF MORTALITY GIVEN TICKET TYPE.

FIX. TICKET TYPE 1 (FIRST CLASS) INSTEAD DO IT FOR TICKET 2

	1	PROBS.
ALIVE	203	203/325
DEAD	122	122/325
	325	1 = 325/325

	2	PROBS
ALIVE	118	118/285
DEAD	167	167/285
	285	1

WRITE PR ALIVE GIVEN 1 > PR ALIVE GIVEN 2

INDEPENDENCE: WHERE CHANCE OF DEATH FOR TICKET TYPE i (i.e. CONDITIONAL PR(Y DEATH GIVEN TICKET i)) DOES NOT DEPEND UPON i.

	MED	PLACEBO
HELPS	140	80
NOT	70	(40)

CONSTANT PROPORTION IN ROWS & COLUMNS IS THE APPEARANCE OF "INDEPENDENCE."

SO TOO

	18	6
	12	(4)

SIMPSON'S "PARADOX"

ILLUSTRATE W/R ADMISSIONS TO BERKELEY (1973?)

FOUND ADMISSIONS RATE FOR MEN WAS HIGHER THAN FOR WOMEN.

RAISED QUESTION OF BIAS.

YOU APPLY TO A PROGRAM / MERGE OF PROGRAMS MIGHT BE APPLES/ORANGES.

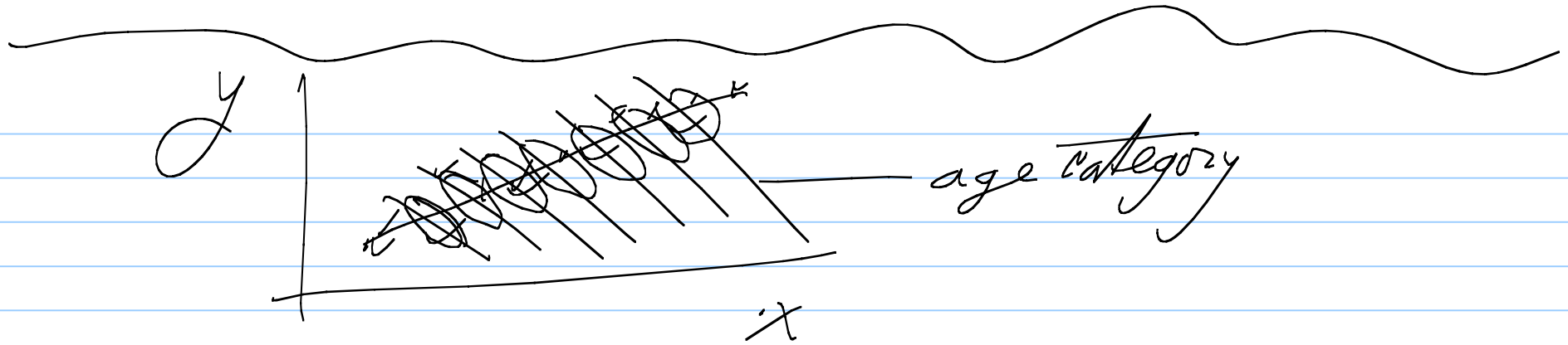
PROGRAM	ADM MEN	ADM. WOMEN	DEC ENROLL
A			
B			
C			
D			
E			
F			

- MATH POSSIBLE (ARTIFICIAL DATA)

  CHANCE MALE ADMITTED > CHANCE FEMALE ADMITTED

  BUT FOR EVERY PROGRAM OF STUDY WOMEN ARE ADMITTED AT A HIGHER RATE THAN MEN

  MECHANISM: WOMEN APPLY IN RELATIVELY GREATER NUMBERS (THAN MEN) TO PROGRAMS THAT ARE HARDER FOR BOTH SEXES TO GET INTO.


y
x
age category