LSU E. J. OURSO COLLEGE OF BUSINESS

Dr. Chun's Numb3rs & Løgic

Capture-Recapture Models

Young H. Chun, Ph.D.

Professor of Decision Science &
Cherie H. Flores Endowed Chair in MBA Studies

* 2x2 Contingency Table

		Event B		
		Yes	No	
Event A	Yes	a	b	$a+b$
	No	c	d	$c+d$
		$a+c$	$b+d$	$a+b+c+d$

* Probability Table

		Event B		
		Yes	No	
Event A	Yes	$a/(a+b+c+d)$	$b/(a+b+c+d)$	$(a+b)/(a+b+c+d)$
	No	$c/(a+b+c+d)$	$d/(a+b+c+d)$	$(c+d)/(a+b+c+d)$
		$(a+c)/(a+b+c+d)$	$(b+d)/(a+b+c+d)$	1

* Statistically independent: P[A] x P[B] = P[AB]

$$\frac{(a+b)}{(a+b+c+d)} \times \frac{(a+c)}{(a+b+c+d)} = \frac{a}{(a+b+c+d)}$$

or

$$ad = bc$$

1. Capture-Recapture Model

"Consider the problem of estimating the number of fish in a lake. A commonly used approach is to first catch a number n_1 of fish, tag or mark them, and release them again into the lake.

Sometime later, after giving the tagged fish an opportunity to mix well with the remaining fish in the lake, a second catch of n_2 fish is made, and the number m of tagged fish is observed:

		Second		Total
		Caught	Not	
First	Caught	$\boldsymbol{m}$	n_1-m	$\boldsymbol{n_1}$
	Not	n_2-m	N+m-n_1-n_2	N-n_1
Total		$\boldsymbol{n_2}$	N-n_2	$\boldsymbol{N}$?

Estimate the total number of fish N in the lake.

1. Capture-Recapture Model

For example, suppose that n_1 = 80, n_2 = 100, and m = 5.

		Second		Total
		Caught	Not	
First	Caught	**5**	**75**	**80**
	Not	**95**	***N*-175**	***N*-80**
Total		**100**	***N*-100**	***N***

Then, 5 x (*N*-175) = 75 x 95

The total number of fish *N* in the lake is

N = (75 x 95) / 5 + 175 = 1,600

2. Number of Remaining Errors

Two proof readers are checking two copies of the same manuscript. The first finds 30 errors, and the second finds only 24. When their completed proofs are compared, it turns out that only 20 errors have been spotted by both of them.

How many errors would you suspect remain, not detected by either of them?

- Martin Gardner, *Mathematical Circus*.

		B		Total
		Yes	No	
A	Yes	**20**		**30**
	No		***x***	
Total		**24**		

2. Number of Remaining Errors

* Number of Undetected Errors

		B		Total
		Yes	No	
A	Yes	20	10	30
	No	4	x	4+x
Total		24	10+x	34+x

Then, 20 x = 10 x 4. Thus, x = 2 errors

- Young H. Chun (2006), "Estimating the Number of Undetected Software Errors via the Correlated Capture-Recapture Model," *European Journal of Operational Research*, 175, pp. 1180~1192.

3. Cafeteria Problem

Ask Marilyn, *Parade Magazine* (February 19, 1995)

"If Andy eats at a cafeteria twice a week, and Andy sees Brad there about 75% of the time, Can Andy assume that Brad goes there more often than Andy does?"

		Brad		Total
		Yes	No	
Andy	Yes	1.5	0.5	2
	No	x-1.5	4.5-x	3
Total		**3.75**	5-x	5

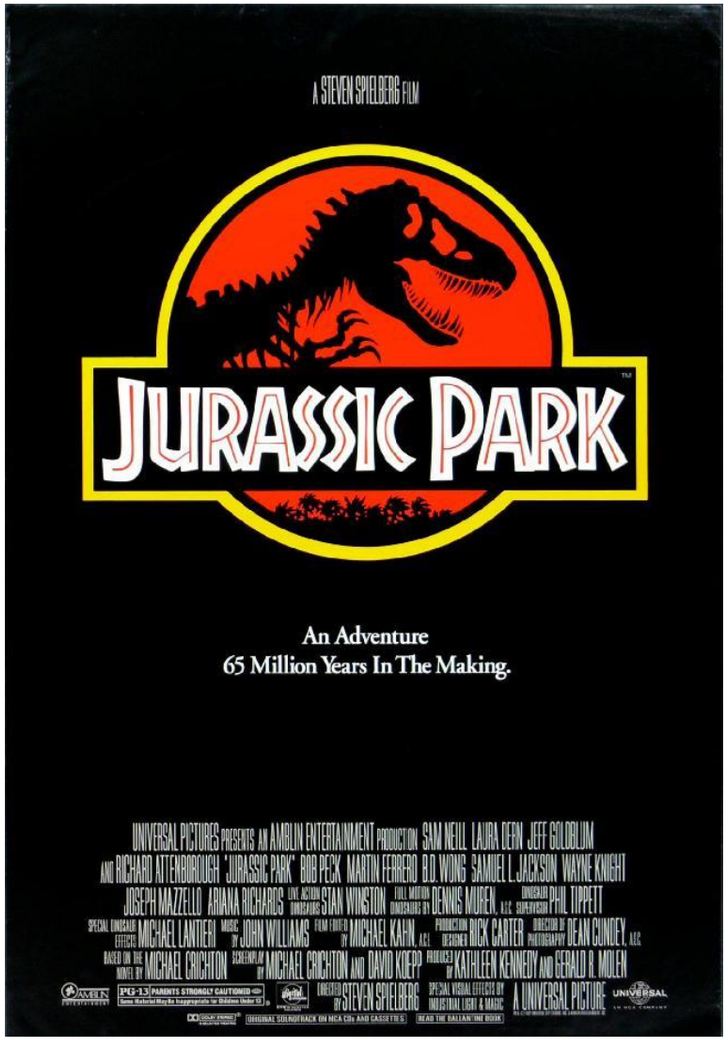
Then, $1.5 * (4.5-x) = 0.5 * (x-1.5)$

Thus, $x = 3.75$ **times a week !**

Movie Trivia

During a preview tour, a theme park suffers a major power breakdown that allows its cloned dinosaur exhibits to run amok.

Jurassic Park (1993)



During a preview tour, a theme park suffers a major power breakdown that allows its cloned dinosaur exhibits to run amok.