

**STAT 13, section 1, Winter 2011, UCLA Statistics
HW 2; Problem Solution**

HW 2.1

Since $\text{Prob}(\text{husband smokes}) \cdot \text{Prob}(\text{wife smokes}) = (.3) \cdot (.18) = .054$ is not equal to $\text{Prob}(\text{both smoke}) = .1$, the smoking habit for the husband is not independent of that of the wife.

HW 2.2

The number of cured children is a binomial random variable with $n=20$, and $p=.9$

- (a) $\text{Prob}(\text{all 20 cured}) = (.9)^{20} = .1216$
- (b) $\text{Prob}(\text{all but 1 cured}) = P(\text{exactly 1 not cured}) = 20 \cdot (.9)^{19} \cdot (.1) = .2702$
- (c) $\text{Prob}(\text{exactly 17 cured}) = (20C17) \cdot (.9)^{17} \cdot (.1)^3 = .1901$
- (d) $\text{Prob}(\text{exactly 80\% cured}) = P(\text{exactly 16 cured}) = (20C16) \cdot (.9)^{16} \cdot (.1)^4 = .0898$

HW 2.3

Similar to 2.2, the number of children who has a high blood lead level is $\text{Bin}(16, 1/7)$

- (a) $\text{Prob}(\text{none has high blood lead}) = (6/7)^{16} = .0849$
- (b) $\text{Prob}(\text{one has high blood lead}) = 16 \cdot (1/7) \cdot (6/7)^{15} = .2264$
- (c) $\text{Prob}(\text{two have high blood lead}) = (16C2) \cdot (1/7)^2 \cdot (6/7)^{14} = .2830$
- (d) $\text{Prob}(\text{three or more have high blood lead}) = 1 - .0849 - .2264 - .2830 = .4058$

HW 2.4

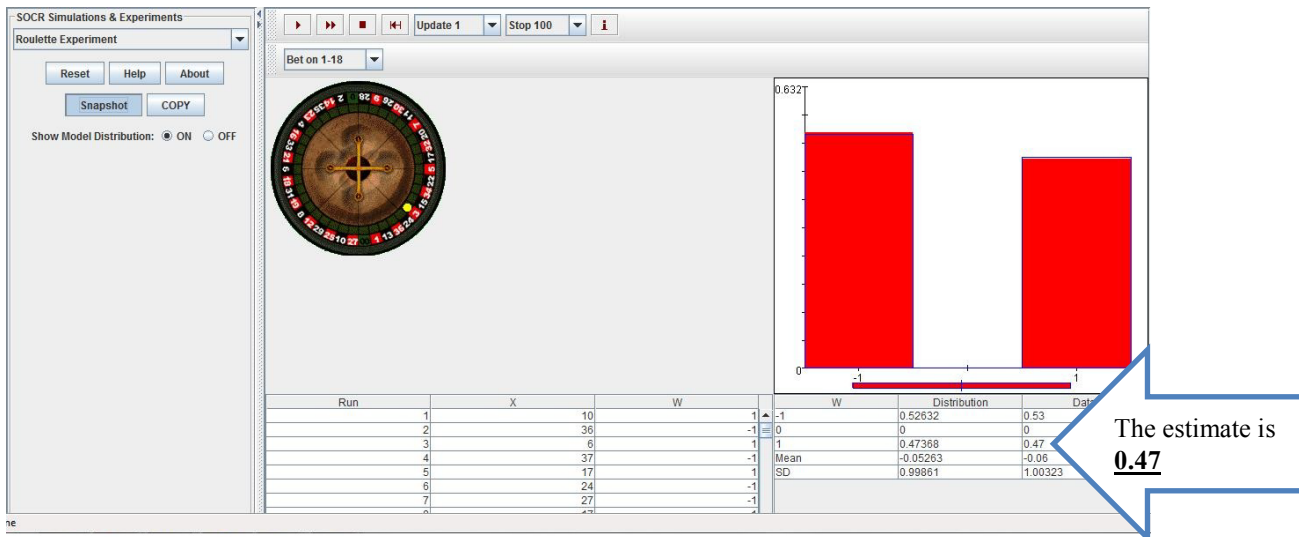
Note that betting numbers 1-18 has exactly the same probability as betting 19-36. Therefore, we can use this alternative in SOCR simulation. The probability of getting a number between 19-36 is exactly $18/38 = .474$. We only need to look at how many times we get 1's (wins) from the simulator.

For example, for 100 runs, I get 47 wins and 53 losses. Therefore, the empirical probability based on this simulation of 100 runs is .47.

Repeat the experiment and report honestly your own results for different number of experiments. Two examples are shown below, but the results of your experiments may vary.

For 100 trials: 0.52				For 1000 trials: 0.471			
W	Distribution	Data	W	Distribution	Data		
-1	0.52632	0.48	-1	0.52632	0.529		
0	0	0	-1	0	0		
1	0.47368	0.52	-1	0.47368	0.471		
Mean	-0.05263	0.04	-1	Mean	-0.05263	-0.058	
SD	0.99861	1.00423	-1	SD	0.99861	0.99882	

As the number of samples increases, the empirical results should come closer and closer to the theoretic results.



HW 2.5

(a) $\text{Prob}(5 \text{ A's in a row in a sample of } 6) = 2 * (.25)^5 * (.75) = .00146$

(b) Regarding C and G as a singleton with probability .45.

$\text{Prob}(\text{equal number of A's and T's}) = \text{Prob}(\text{no A, no T, all C/G}) + \text{Prob}(1A, 1T, 3C/G) + \text{Prob}(2A's, 2T's, 1C/G)$
 $= (.45)^5 + (5C1,1,3) * (.25) * (.3) * (.45)^3 + (5C2,2,1) * (.25)^2 * (.3)^2 * (.45) = .2311$

Here you need to use the multinomial formula

(http://wiki.stat.ucla.edu/socr/index.php/AP_Statistics_Curriculum_2007_Distrib_Multinomial):

$$\frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

(c) For part (a), simulation (with update 1, stop 100 for 6 times) shows that a 100 runs of a random sample of 6 contain no case of 5 A's in a row. Therefore, the empirical probability is 0 based on 100 runs.

For part (b), out of 100 runs (i.e. you need 500 spins), there are 2 cases of 0 A and T, 23 cases of 1A and 1T and 11 cases of 2A's and 2T's.

Therefore the empirical result is .36, see figure below:

