http://www.stat.ucla.edu/~dinov/courses_students.dir/10/Winter/STAT13.1.dir/

STAT 13 Homework 2 Solutions

Question 1: Suppose that in a certain population of married couples 33% of the husbands smoke, 17% of the wives smoke and in 9% of the couples both the husband and wife smoke. Is the smoking status of the husband independent of that of the wife? Why or why not?

Answer:

No; smoking status of husband is not independent of that of the wife.

We will define events H and W as follows: H: Husband smokes; W: Wife smokes

The following is given in the problem:

P(H) = 0.33; P(W) = 0.17; P(H & W) = 0.09

If H and W are independent, then P(H & W) = P(H)*P(W)

0.33 * 0.17 = 0.0561 which is not equal to 0.09. Therefore, we know that the smoking status of husbands and wives are NOT independent.

Question 2: A certain drug treatment cures 88% of cases of hookworm in children. Suppose that 20 children suffering from hookworm are to be treated, and that the children can be regarded as a random sample from the population. Find the probability that:

Answer:

For every child, there are two outcomes: the child is cured of hookworm, or the child is not cured. Therefore, this problem uses the binomial distribution.

p = 0.88 (this is the probability of being cured – "success")
1-p = 0.12 (this is the probability of not being cured – "failure")
n = 20 (the number of children)
k = the number of "successes" we are interested in. (success = children being cured)

The formula for the probability of having k successes in n trials is:

$$\Pr(K = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

For simplicity, this will be written as: (n choose k) * p ^ k * (1-p)^(n-k)

a) *All 20 will be cured.*
   - k = 20
   - Pr(k = 20) = (20 choose 20) * .88^20 * .12^0 = **0.07756**
b) *All but 1 will be cured.*
   - k = 19
   - Pr(k = 19) = (20 choose 19) * .88^19 * .12^1 = **0.2115**
c) *Exactly 17 will be cured.*

- k = 17
- Pr(k = 17) = (20 choose 17) * .88^17 * .12^3 = **0.2242**

d) *Exactly 80% will be cured.*
- k = 16
- Pr(k = 16) = (20 choose 16) * .88^16 * .12^4 = **0.1299**

Question 3: Childhood lead poisoning is a public health concern in the US. In a certain population, one child in seven has a high blood lead level (>30 µg/dLi). Compute the following probabilities for a randomly chosen group of 16 children from this population:

Answer:

Again, this is binomial. The labels of "success" and "failure" are arbitrary.

p = 1/7 = 0.14286 (this is the prob of having high lead – "success")
1-p = 6/7 = 0.85714 (this is the prob of not having high lead – "failure")
n = 16 (the number of children)
k = the number of "successes" we are interested in. (success = child has high lead)

The formula is: (n choose k) * p ^ k * (1-p)^(n-k)

a) *P(none have high blood lead)*
- k = 0
- P(k=0) = (16 choose 0) * (1/7)^0 * (6/7)^16 = **0.08489**

b) *P(one has high blood lead)*
- k = 1
- P(k=1) = (16 choose 1) * (1/7)^1 * (6/7)^15 = **0.22637**

c) *P(two have high blood lead)*
- k =2
- P(k=2) = (16 choose 2) * (1/7)^2 * (6/7)^14 = **0.28296**

d) *P(three or more have high blood lead)*
- k = 3, or 4, or 5, … , or 19, or 20
- This is equivalent to: 1-P(0 or 1 or 2 have high blood lead)
- 1 - (P(k=0) + P(k=1) + P(k=2))
- Those answers were found in parts a through c
- 1 - (0.08489 + 0.22637 + 0.28296) = **0.40578**

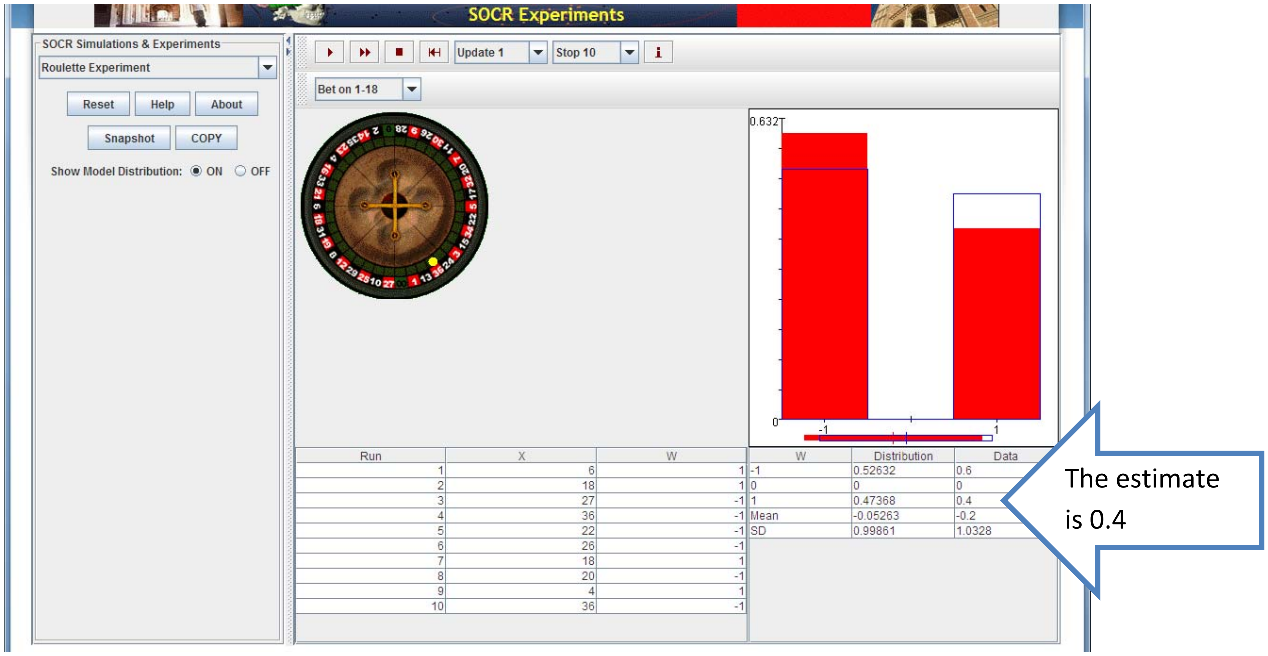
Question 4: Use the SOCR Roulette Experiment to design and run a simulation estimating the probability that a number <= 18 turns up is we spin the Roulette Wheel. Compute the exact probability of this event and list the sample-driven estimates of this event for samples of size 10, 100 and 1,000. What is your observation about these probability estimates?

Answer:

There are 38 possible outcomes in the game of Roulette. We are interested in 18 of them.

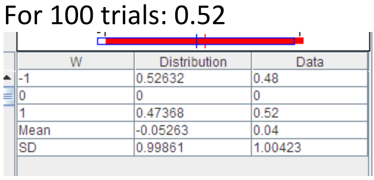
The exact probability is 18/38 = **0.47368**

The empirical estimates for every student will vary based on their results from SOCR.



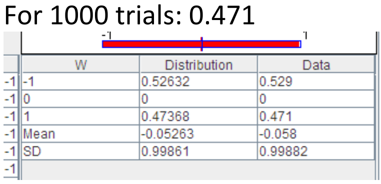
The estimates will come from the Data column for the row that says "1"

My results:
For 10 trials: 0.4
For 100 trials: 0.52



W	Distribution	Data
-1	0.52632	0.48
0	0	0
1	0.47368	0.52
Mean	-0.05263	0.04
SD	0.99861	1.00423

For 1000 trials: 0.471



W	Distribution	Data
-1	0.52632	0.529
0	0	0
1	0.47368	0.471
Mean	-0.05263	-0.058
SD	0.99861	0.99882

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

Question 5: Suppose that a long stretch of DNA has only Adenine (A), Thiamine (T), Cytosine (C) and Guanine (G), which occur with the following probabilities 0.25, 0.3, 0.25, 0.2, respectively. The A, T, C and G nucleotides make up the core of the genetic code for any species. What is the probability that

Answer:

a) A random drawing of 10 A's in a row in a sample of 11 randomly chosen nucleotides?
   - We must have 10 A's in a row. There are only two possible ways to achieve this: AAAAAAAAAAX or XAAAAAAAAAA (where X is a T, C or G)
   - P(A) = 0.25; P(Not A) = 0.75

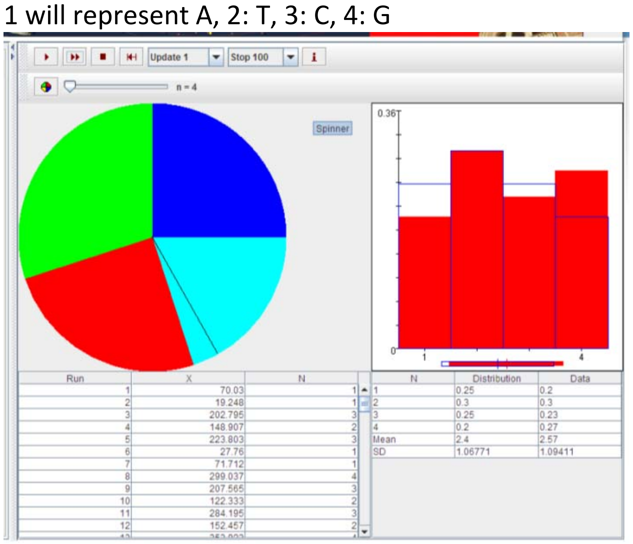
- The first chain's probability is:
  = .25*.25*.25*.25*.25*.25*.25*.25*.25*.25*.75 = $0.25^{10} * 0.75$
- In the same way the second chain's probability is = $0.75 * 0.25^{10}$
- Add those two probabilities and we get **$1.43051147 \times 10^{-6}$**
- We might also say that 11As in a row count as having 10As in a row.
  If we do, we must add 0.25^{11} to get **$2.38418579 \times 10^{-6}$**

b) A random sample of 5 nucleotides has equal number of A's and T's?
- We are looking at only one side of the DNA strand. This problem uses the multinomial distribution.
- Here, there are three ways for us to have an equal number of A's and T's.
  0A and 0T and 5 of something else
  1A and 1T and 3 of something else
  2A and 2T and 1 of something else
- The probabilities of A, T and "something else" are:
  P(A) = 0.25; P(T) = 0.3; P(X)= 0.45
- The formula for the multinomial distribution is:

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

  For simplicity, it will be written as: n!/(x1! * … xk!) * p1^x1 * … * pk^xk
- P(0,0,5) = 5!/(0!*0!*5!) * 0.25^0 * 0.3^0 * 0.45^5 = 0.0184528125
- P(1,1,3) = 5!/(1!*1!*3!) * 0.25^1 * 0.3^1 * 0.45^3 = 0.1366875
- P(2,2,1) = 5!/(2!*2!*1!) * 0.25^2 * 0.3^2 * 0.45^1 = 0.0759375
- Total probability of having the same numbers of A and T = **0.231077813**

c) Use the SOCR Spinner experiment
- 1 will represent A, 2: T, 3: C, 4: G



- For the first part, I drew 100 samples of the spinner. I looked at 9 groups of 11, and searched for 10 1's in a row. None found. Empirical rate = 0
- For part 2, I looked at 20 groups of 5 and counted how many had the same numbers of 1's and 2's.
  I found 6. Empirical results = 6/20 = 0.30