**Problem Set 3Answers**

**Problem 1.** A politician is concerned that her support on a current issue may have dropped below 50%. To determine her support, she commissioned a random survey of 186 of her constituents. This survey found 88 people supporting her position.

a. plot the binomial for n = 186, p = 0.5

b. What are the mean, variance, and standard deviation for this problem?

c. What is the probability that X, the number of people supporting her position, will be 88 or less based on this distribution?

d. Should the politician conclude that her support is less than 50%?

**See Excel Spreadsheet**

**Problem 2**. Clerical workers process checks in batches of 1000. This process results in a 4% average error rate.

a. For a batch of 1000 checks, what is the expected number of errors?

b. Provide a 3-std. deviation (3-sigma) lower and upper bound for the number of errors in a batch of 1000.

c. One worker had 60 errors in a batch of 1000. He claimed that 60 errors is not an unusual occurrence. Is this claim reasonable?

**See Excel Spreadsheet**

**Problem 3.** In an automobile engine plant, cylinders are bored and fitted with machined pistons. For the engine to work correctly, the clearance between each piston cylinder pair (defined as the difference between the inner diameter of the cylinder and the outer diameter of the piston) must be within certain design specifications. Suppose that the probability of any given piston/cylinder pair meeting design specifications is 95%. Furthermore, assume that the clearances of different piston/cylinder pairs are independent random variables. What is the probability that an entire engine will be functional (that is, all piston/cylinder pairs will meet design specifications)? Assume the engine has 6 cylinders.

**Let N = # functional piston/cylinder pairs**

**N is a binomial random variable with parameters n = 6 and p = 0.95. We want P{N = 6} = binomdist(6, 6, 0.95, 0) = 0.735 (approximately).**

**Problem 4.** You company has 400 employees. Suppose there is a 20% chance that each employee will develop cancer.

(a) What is the probability that, out of 20 randomly chosen employees exactly 5 will develop cancer?

(b) What is the probability none will develop cancer?

(c) What is the probability that they will all develop cancer?

**Let N = # employees that develop cancer**

**N is a binomial random variable with parameters n = 20, p = 0.2.**

**P{N = 5} = binomdist(5, 20, 0.2, 0) = 0.17456**

**P{N = 0} = binomdist(0, 20, 0.2, 0) = 0.011529**

**P{N = 20} = binomdist(20, 20, 0.2, 0) = 1.05 E-14**

**Problem 5.** Consider a population consisting of 1000 married couples, all of whom carry the sickle cell trait (but do not develop the disease)---that is, they are coded SA. Suppose each married couple produces exactly two children.

**Note that there will be 2000 children in total. Each will have a probability of ¼ of being coded SS, a probability of ½ of being coded SA (remember, either S from father and A from mother OR A from father & S from mother codes as SA), and a probability of ¼ of being coded AA.**

1. What is the expected number of children that will develop sickle cell disease?

**Let N = # children coded SS. N is a binomial random variable with parameters n = 2000 and p = ¼. The expected value = n x p = 2000 x 0.25 = 500.**

1. What is the expected number of children that will carry the sickle cell trait but not develop the disease?

**Let N = # children coded SA or AS. N is a binomial random variable with parameters n = 2000 and p = 1/2. The expected value = n x p = 2000 x 0.5 = 1000.**

1. What is the expected number of children that will NOT carry the sickle cell trait at all?

**Let N = # children coded AA. N is a binomial random variable with parameters n = 2000 and p = ¼. The expected value = n x p = 2000 x 0.25 = 500.**

1. What is the probability that the number of children who develop sickle cell disease is greater than or equal to 500?

**Let N = # children coded SS. N is a binomial random variable with parameters n = 2000 and p = ¼. We seek P{N > 500} = 1 – binomdist(499, 2000, 0.25,1) = 0.508583**

**Problem 6.** Suppose that each couple in the initial population from problem 6 continues to have children until each couple produces exactly two children who do not develop sickle cell disease.

1. What is the expected number of children who do NOT carry the sickle cell trait?

**There will be 2000 children that code either SA, AS, or AA since these are exactly those who do not develop sickle cell disease. For a given child from this population of 2000, each of these three possibilities is equally likely, so the probability that a child chosen randomly from this population does NOT carry the sickle cell train is 1/3.**

**Let N = # children who do not carry the sickle cell trait. N is a binomial random variable with parameters n = 2000 and p = 1/3. The expected value = n x p = 2000 x 1/3 = 667.**

1. What is the probability that the number of children who do NOT carry the sickle cell trait is less than or equal to 500?

**We seek P{N < 500} = binomdist(500, 2000, 1/3, 1) = 0 (approximately).**

**Problem 7.** What is the probability that the World Series will last 4 games? 5 games? Assume that the teams are evenly matched---that is, each team has an equal 50-50 chance of winning each game. Also, assume that each game is independent of any other game---that is, the probability of any team winning any game is 50%, independent of the outcome of previously played games. In the World Series, there are two baseball teams, one from the National League and one from the American League. The series ends when the first (and winning) team wins 4 games. Thus, the series cannot end in less than four games. Since each game is played until there is a winner (no ties allowed), the series cannot extend past 7 games.

**This is a nontrivial application of the binomial distribution. In the world series, there are two baseball teams. The series ends when the winning team wins 4 games. Therefore, we define a success as a win by the team that ultimately becomes the world series champion.**

**Let's look first at the simplest case. What is the probability that the series lasts only 4 games. This can occur if one team wins the first 4 games. Let N = the number of wins by the National League in the first four games. N is a binomial random variable with parameters n = 4, and p = 0.5. The probability of the National League team winning 4 games in a row is, therefore, the probability that N = 4 which is given by:**

**binomdist(4, 4, 0.5, 0) = 0.0625**

**Similarly, when we compute the probability of the American League team winning 4 games in a row, we find that it is also 0.0625. Therefore, probability that the series ends in four games would be 0.0625 + 0.0625 = 0.125; since the series would end if either the American or National League team won 4 games in a row.**

**Now let's tackle the question of finding the probability that the world series ends in 5 games. The subtlety in finding this solution is to recognize that the series can only end in 5 games if one team has won 3 out of the first 4 games. So let's first find the probability that the American League team wins exactly 3 of the first 4 games. Let N be defined as before. The desired probability is the probability that N = 3, which is given by:**

**binomdist(3, 4, 0.5, 0) = 0.25**

**Given that the American League team has won 3 of the first 4 games, the American League team has a 50/50 chance of winning the fifth game to end the series. Therefore, the probability of the American League team winning the series in 5 games is 0.25 \* 0.50 = 0.125. Since the National League team could also win the series in 5 games, the probability that the series ends in 5 games would be 0.125 + 0.125 = 0.25.**

**If you continue on in this way, you can figure out the probabilities of the series ending in 6 or 7 games. You should find that the probability of the series ending in 6 games is 0.3125; and the probability of the series ending in 7 games is also 0.3125.**

**While this is statistically correct in theory, (given the assumptions) over the years the actual world series has turned out differently, with more series than expected (given the evenly-matched assumption) lasting 7 games. Note also, that if the teams are not evenly matched, a 7-game series is even less likely. For an interesting discussion of why world series reality differs from theory, see Ben Stein's explanation of why in the following NY Times article:** [**http://www.nytimes.com/2003/10/22/sports/baseball-a-7-game-world-series-is-unusually-common.html**](http://www.nytimes.com/2003/10/22/sports/baseball-a-7-game-world-series-is-unusually-common.html)**.**