Business Analytics

Prof. Phil Jones

Fall 2016

**Exam #1**

**Instructions:**

1. Please write the name of everyone in your group

at the top of each page.

* The exam is open book, open note, and open computer.
* You are expected to do your work within your group alone.

Note: Problems are worth points as indicated.

Providing short explanations may allow awarding partial credit in some cases.

The exam is due **September 21**. Please return your exam (hardcopy) at the beginning of class. If you are unable to return your exam in class, you can submit via email, but please do so only if absolutely necessary.

*Good Luck!*

Names:

1. (10 points) **A height problem:** Suppose that the height of women in the United States is normally distributed with a mean of 65 inches and std. deviation of 2.5 inches. Suppose you select a random sample of 200 women within the United States. Let N = the number of those women whose height is less than or equal to 62 inches.

* 1. (5 points) What kind of probability distribution does N follow? Specify the name and parameters of the distribution.

**Binomial with n = 200 and p = normdist(62, 65, 2.5, 1) = 0.11507**

* 1. (5 points) What is the probability that N is less than or equal to 25?

**Binomdist(25, 200, 0.11507, 1) = 0.715849**

1. (20 points) **An urn problem:**  An opaque urn contains 30 red balls, 40 blue balls, and 30 green balls. Suppose you draw exactly twice from the urn (with replacement) so that the odds of a given color are the same on the second draw as they were on the first draw.

With no other information, we have the following 9 possibilities for our two draws:

RR

RB

RG

GR

GG

GB

BR

BG

BB

But urn # 1 has 30 R, 40 B, and 30 G, so these possibilities are not all equally likely:

The corresponding probabilities are:

RR = 0.3 x 0.3 = 0.09

RB = 0.3 x 0.4 = 0.12

RG = 0.3 x 0.3 = 0.09

GR = 0.3 x 0.3 = 0.12

GG = 0.3 x 0.3 = 0.09

GB = 0.3 x 0.4 = 0.12

BR = 0.4 x 0.3 = 0.12

BG = 0.4 x 0.3 = 0.12

BB = 0.4 x 0.4 = 0.16

The sum of the probabilities = 4 x 0.9 + 4 x 0.12 + 0.16 = 0.36 + 0.48 + 0.16 = 1

We can also simplify this by using N to denote a non-blue ball and then listing the possibilities and their corresponding probabilities as:

NN = 0.6 x 0.6 = 0.36

NB = 0.6 x 0.4 = 0.24

BN = 0.4 x 0.6 = 0.24

BB = 0.4 x 0.4 = 0.16

1. (5 points) What is the probability that at least one ball is blue given that at least one ball is either green or red?

Let us denote by B1 the event that at least one ball is blue

Let us denote by N1 the event that at least one ball is non-blue.

Let us denote by B1 and N1 the joint event.

We seek P{B1 | N1}. But P{B1 | N1} = P{B1 and N1}/P{N1}

P{N1} = 1 – P{two blue balls} = 1 – 0.16 = 0.84

The event B1 and N1 occurs if and only if we have one of the two outcomes: a blue ball followed by a non-blue ball (BN) or a non-blue ball followed by a blue ball (NB). The probability of the event B1 and N1 is therefore given by 0.24 + 0.24 = 0.48.

So P{B1 |1N1} = 0.48/0.84 = 0.57 (approximately)

1. (5 points) What is the probability that at least one ball is blue given that the second ball drawn is either green or red?

Let us denote by N2 the event that the second ball drawn is non-blue.

As before, let B1 denote the event that at least one ball is blue.

Let B1 and N2 denote the joint event.

We seek P{B1 | N2}.

P{B1 | N2} = P{B1 and N2}/P{N2}

The event N2 is comprised of the outcomes NN and BN.

So P{N2} = 0.36 + 0.24 = 0.6.

The event B1 and N2 is comprised of the single outcome BN which has probability 0.24.

P{B1 | N2} = P{B1 and N2}/P{N2} = 0.24/0.6 = 0.4

c. (5 points) What is the probability that at least one ball is blue given that at least one ball is either green or red?

Let us denote by B1 the event that at least one ball is blue

Let us denote by N1 the event that at least one ball is non-blue.

Let us denote by B1 and N1 the joint event.

We seek P{B1 | N1}. But P{B1 | N1} = P{B1 and N1}/P{N1}

P{N1} = 1 – P{two blue balls} = 1 – 0.16 = 0.84

The event B1 and N1 occurs if and only if we have one of the two outcomes: a blue ball followed by a non-blue ball (BN) or a non-blue ball followed by a blue ball (NB). The probability of the event B1 and N1 is therefore given by 0.24 + 0.24 = 0.48.

So P{B1 |1N1} = 0.48/0.84 = 0.57 (approximately)

1. (5 points) What is the probability that at least one ball is blue given that the second ball drawn is either green or red?

Let us denote by N2 the event that the second ball drawn is non-blue.

As before, let B1 denote the event that at least one ball is blue.

Let B1 and N2 denote the joint event.

We seek P{B1 | N2}.

P{B1 | N2} = P{B1 and N2}/P{N2}

The event N2 is comprised of the outcomes NN and BN.

So P{N2} = 0.36 + 0.24 = 0.6.

The event B1 and N2 is comprised of the single outcome BN which has probability 0.24.

P{B1 | N2} = P{B1 and N2}/P{N2} = 0.24/0.6 = 0.4

3. (10 points) **A Card Problem**: Suppose you draw from a randomly shuffled deck of 52 cards (4 suits, each with 13 cards numbered 2 through Ace) 5 times **without** replacement. Without replacement means that after each draw, the drawn card is removed from the deck and the remaining cards in the deck are re-shuffled before drawing again. Hint: how is this similar to the “multiple birthdate” problem gone over in class?

* 1. (3 points) What is the probability that you get a royal flush in spades? A royal flush in spades means that after drawing the five cards, you have the ace, king, queen, jack, and 10 of spades.

**Probability of getting one of the 5 cards on the first draw is 5/52.**

**On the second draw, there are 51 cards left and only 4 possibilities, so the probability of getting one of the 4 remaining cards on the second draw is 4/51.**

**The probability of getting one of the 3 remaining cards on the third draw is 3/50, and so on.**

**The probability of getting the royal flush in spades is the product of**

**5/52 x 4/51 x 3/50 x 2/49 x 1/48 = 120/[52 x 51 x 50 x 49 x 48] = 120/311,875,200**

**= 3.85 x 10 -7, approximately.**

* 1. (2 points) What is the probability you get a royal flush (in any of the four suits ---clubs, diamonds, hearts, or spades) ?

**This is just 4 x probability computed in part a, so 480/311,875,200 = 1.54 x 10 -6, approximately.**

* 1. (3 points) If one million people repeat this experiment using different but otherwise identical decks of 52 cards, what is the expected number that get a royal flush?

**Let R = the number of people drawing a royal flush.**

**R is a binomial random variable with parameters n = 1,000,000 and p = 1.54 x 10 -6**

**The expected value of R is n x p = 1.54.**

* 1. (2 points) If one million people repeat this experiment using different but otherwise identical decks of 52 cards, what is probability that at least one person gets a royal flush?

**P{R > 1} = 1 – P{R < 0} = 1 – binomdist(0, 1000000, 1.54 x 10 -6, 1)**

**Working the formula out in Excel we get:**

**P{R > 1} = 1 – 0.214579 = 0.785421. So, although not absolutely certain, the odds are that at least one person will draw a royal flush.**

1. (25 points) **A World Series Problem:** Recall the world Series Problem from problem set #3. 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. 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.
2. (5 points) Assume that the teams are evenly matched---that is, each team has an equal 50-50 chance of winning each game. In that case, what is the overall probability that the American League team wins the series?

**50% or ½. Note that we worked out in problem set 3 the following probabilities:**

**P{AL wins series in 4 games} = 0.0625**

**P{AL wins series in 5 games} = 0.125**

**P{AL wins series in 6 games} = P{AL wins series in 7 games} = .15625**

**Summing them all up, we get P{AL wins series} = 0.0625+0.125 +2\* 0.15625 = 0.5.**

1. (15 points) Now assume that the odds are in favor of the American League team: for each individual game, the probability that the American League team will win is 60% (which means that the National League team has a 40% chance of winning each individual game). Now, what is the overall probability that the American League team wins the series?

**Now we need to work out the following:**

**P{AL wins in 4} = binomdist(4, 4, 0.6, 0) = 0.1296**

**P{AL wins in 5} = binomdist(3,4,0.6, 0) x 0.6 = 0.20736**

**P{AL wins in 6} = binomdist(3,5,0.6.0) x 0.6 = 0.20736**

**P{AL wins in 7} = binomdist(3,6,0.6,0) x 0.6 = .165888**

**Summing them up, we get**

**P{AL wins series} = 0.710208**

**Note that this is considerably larger than the 60% chance the AL team wins an individual game.**

1. (5 points) For the original problem from problem set 3 (equal probabilities of winning each game), let N = # games actually played in the series. What are the mean and standard deviation of N?

|  |  |
| --- | --- |
| **Length of Series** | **Probability** |
| **4** | **.125** |
| **5** | **.25** |
| **6** | **.3125** |
| **7** | **.3125** |

**Expected # games = 4 x .125 + 5 \* .125 + 6 \* .3125 + 7 \* .3125 = 5.8125**

**Variance of N = .125 x (4 – 5.8125)^2 +…..+ .3125x (7 – 5.8125)^2** **= 1.027344**

**Std. Dev. of N = sqrt(33.22314) = 1.01358**

5. (10 points) **Rattlesnakes and Prairie Dogs**: A youth baseball league in Wyoming plans to institute a playoff series between the championship teams from each of their two divisions: the prairie dog division and the rattlesnake division. Unlike professional baseball, games in the youth league are allowed to end in ties. Thus, there are three possible outcomes for any game: prairie dog team wins & rattlesnake team loses, prairie dog team loses and rattlesnake team wins, they tie. The youth league plans to play the series until three games with the same outcome have occurred. If there are three wins by one of the divisions, that division’s team will win the championship, but if there are three ties, the championship will be declared a tie. Assume that the outcomes of successive games are independent events, and that the probability of either team winning a specific game is 45% (the same for both teams) while the probability of a tie is 10%.

1. (2 points) Let N denote the number of games the series lasts. What is the maximum possible value for N? The minimum?

**Clearly the minimum is 2 games.**

**The only way the series can have gone 6 games without ending is for the PD (prairie dog) team to have two wins, the R (rattlesnake) team to have two wins, and for there to be exactly two ties---any other outcome for the first six games would have resulted in the series ending. The seventh game MUST produce one of these previous 3 results, so the maximum number of games is 7.**

1. (4 points) What is the probability that the series lasts exactly 3 games?

**The series can end in 3 with a tie, a rattlesnake win, or a prairie dog win.**

**To end in 3 with a tie, all of the 1st three games must have been ties. We can compute this probability as:**

**P{end in 3 with a tie} = Binomdist(3,3,0.1,0) x 0.1 = 0.001**

**Similarly, we can compute:**

**P{end in 3 with a R win } = P{end in 3 with a P win} = binomdist(3, 3, 0.45, 0) x 0.45 = 0.091125**

**P{end in 3} = 0.001 + 0.091125 + 0.091125 = 0.18325**

1. (4 points) What is the probability that the series lasts exactly 4 games?

**To end in 4 with a tie, exactly two of the 1st three must have been ties AND the fourth must be a tie:**

**P{end in 4 with a tie} = binomdist(2,3,0.1,0) x 0.1 = 0.0027**

**P{end in 4 with a R win} = P{end in 4 with a P win} = binomdist(2,3,0.45,0) x 0.45 = .150356**

**P{end in 4} = 0.0027 + 0.150356 + 0.150356 = .303412**

1. (15 points) Suppose the gender of a newborn child is equally likely to be male or female (a 50% chance either way). Suppose further, that you know a couple with exactly four children (no twins or triplets or quadruplets). You may assume that the gender of successive births are independent events.

**The (equally likely) possibilities (with no additional information) are:**

**BBBB BGGG GBBB GGGG**

**BBGG BGBB GBGG GGBB**

**BGBG BBGB GGBG GBGB**

**BGGB BBBG GGGB GBBG**

**Here I use the convention that the order is oldest to youngest.**

1. (5 points) What is the probability that at least two of the children are girls given that at least one is a boy?

**The GGGG possibility is deleted, so 15 remain. Of the 15, 10 have at least 2 G’s. Since each of the 15 are equally likely, the probability is 10/15 or 2/3.**

1. (5 points) What is the probability that at exactly two of the children are girls given that the youngest child is a girl?

**There are 8 remaining possibilities are**

~~BBBB~~  **BGGG** ~~GBBB~~  **GGGG**

**BBGG** ~~BGBB~~  **GBGG** ~~GGBB~~

**BGBG** ~~BBGB~~  **GGBG** ~~GBGB~~

~~BGGB~~  **BBBG** ~~GGGB~~  **GBBG**

**Of these equally likely possibilities, three have exactly 2 G’s. So the probability is 3/8.**

1. (5 points) What is the probability that at exactly two of the children are girls given that at least one of the children is a girl?

**Only BBBB is deleted. Of the 15 remaining possibilities, 6 have exactly two G’s. So the probability is 6/15 = 2/5.**

1. (10 points) **Sickle Cell Children.** Consider a population consisting of 1,000,000 adult females and 1,000,000 adult males. In this population, 10,000 females and 10,000 males code SS for the Sickle Cell gene, 100,000 of each gender code SA for the sickle cell gene, and the remainder code AA. Suppose that all of the people in this population eventually marry someone of the opposite gender and suppose that the pairing is done randomly (at least as far as the sickle cell gene goes). That is, let us consider a random male---since 1% of the females code SS, 10% code SA, and 89% code AA, then there is a 1% chance that this male will wind up marrying a female who codes SS, a 10% chance that the male will wind up marrying a female who codes SA, and an 89% chance that the male will wind up marrying a female who codes AA. Suppose that each of the 1,000,000 resulting couples has two children so that there will be 2,000,000 children born. In this new generation, what will be the expected numbers of offspring coding SA, AA, and SS? Would we expect the number fall into each category in the new generation to rise, lower, or stay the same?

**Let us consider the 10,000 females who code SS. Since 1% of the males code SS and the females are equally likely to marry any man, we would expect 1% or 100 of the 10,000 SS females to marry a male who codes SS. Likewise, we would expect 10% or 1,000 to marry a male who codes SA and 89% or 8900 to marry a male who codes AA. From these total 10,000 pairings (marriages) we would expect them to break out as follows:**

**Both of the pair code SS: 100**

**One of the pair codes SS and the other codes SA: 1000**

**One of the pair codes SS and the other codes AA: 8900**

**Now let us consider the 100,000 females who code SA. We would expect 1,000 to marry men who code SS, 10,000 to marry men who code SA, and 89,000 to marry men who code AA. From these total 100,000 pairings we would expect them to break down as follows:**

**One of the pair codes SS and the other codes SA: 1000**

**Both of the pair code SA: 10,000**

**One of the pair codes SA and the other codes AA: 89,000**

**Now let’s consider the 890,000 females who code AA. We would expect 1% or 8900 to marry men who code SS, 10% or 89,000 to marry men who code SA, and 89% or 792,100 to marry men who code AA. From these total 890,000 pairings we would expect them to break down as follows:**

**One of the pair codes AA and the other codes SS: 8900**

**One codes AA and one codes SA: 89,000**

**Both code AA : 792,100**

**In total, we would expect the pairings to break down as follows:**

**Both of the pair code SS: 100**

**One codes SS and one codes SA: 2000**

**One codes SS and the other codes AA: 17,800**

**One codes SA and the other codes AA: 178,000**

**Both code SA: 10,000**

**Both code AA: 792,100**

**Note that the total is 1,000,000 as it should be since there are 1,000,000 pairs. Now let us consider the children from each of the groups above:**

**Both parents code SS: We will observe 200 children, all of whom will code SS.**

**One parent codes SS and one codes SA: We will observe 4000 children. Each child must pick up an “S” gene from the “SS” parent and there is a 50-50 chance that the corresponding gene pair from the other parent will be either “S” or “A”. Thus, we expect 2000 SS children and 2000 SA children.**

**One parent codes SS and the other codes AA: We will observe 35,600 children, each of whom must code SA.**

**One parent codes SA and the other codes AA: We will observe 356,000 children. Half or 178,000 can be expected to code SA and the remaining 178,000 can be expected to code AA.**

**Both parents code SA: We will observe 20,000 children. We would expect 25% or 5000 to code AA, 5000 to code SS, and 10,000 to code SA.**

**Both parents code AA: We will observe 1,584,200 children, all of whom will code AA.**

**In total,**

**the expected number of children coding SS: = 200 + 2000 + 5000 = 7200**

**the expected number of children coding SA: = 2000 + 35,600 + 178,000 + 10,000 = 225600**

**the expected number of children coding AA: = 178,000 + 5000 + 1,584,200 = 1,767,200**

**Note that the new generation can be expected to have fewer people coding SS (7200 as compared to 10,000) and fewer people coding AA (1,767,200 as compared to 1,780,000). But the number coding SA can be expected to increase from 200,000 to 225,600.**