**Math 17, Section 2 – Spring 2011**

# Homework 3 Solutions

**<u>Assignment</u>**
**Chapter 11: 40, 42**
**Chapter 12: 24, 36**
**Chapter 13: 2, 36, 42, 50**
**Chapter 14: 20, 33**

## Chapter 11

**11.40] Cell Phones.** The legislator claims cell phone use is below 12%. You notice that 4 out of 10 people, or 40% are using their cell phones one morning. Does this cast doubt on the legislator's claim? Let's use a simulation to see.

We want to see if it is unlikely to see 4 out of 10 people on the phone, assuming the percentage is 12%. At the most basic level, we are going to look at a group of 10 people and count how many are using a cell phone. We can do the following:

- Generate 10 random numbers between 0 and 1.
- If a number is less than 0.12, flag that as a cell phone use.
- If a number is more than 0.12, flag that as a non-cell phone use.
- Count the number of users out of 10.
- Repeat this many times, say, 100 times.
- See what proportion of your simulations showed 4 or more users.
    - Alternatively, you could calculate the percentage of cell phone users for each of your 100 samples, and look at a histogram of the percentages to see what kind of use you'd expect to see.

Create a dataset of random numbers between 0 and 1, with 100 samples (rows) of 10 columns (observations). The following code pasted into the script window will summarize your simulations. Be sure to replace "UniformSamples" in my code with whatever name you gave to your dataset of random numbers.

```
# Creates empty matrix
flagobs <- matrix(NA,100,10)

# Goes through data matrix, assigning 0 or 1
# to each element based on cell phone probability
for (i in 1:100){
  for (j in 1:10) {
    if (UniformSamples[i,j] <= 0.12){flagobs[i,j]=1}
    else{flagobs[i,j]=0}
   }
}
```

```
# Creates a vector of the sum of each row
phonetotal <- apply(flagobs,1,sum)

# If you want, create a vector of the percentage
# of cell phone users for each row.
phonepercent <- phonetotal/10

# Summarize counts, both as numeric and categorical
summary(phonetotal)
summary(as.factor(phonetotal))

# Summarize percentages if desired
summary(phonepercent)
```

The R output for my simulation is given below.

```
> summary(phonetotal)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    1.00    1.15    2.00    4.00

> summary(as.factor(phonetotal))
 0  1  2  3  4
31 35 25  6  3
```

Answering the Question. We observed 4 out of 10, or 40%, of people using cell phones. Based on the legislator's statement that the percentage is 12%, we'd expect that 0, 1, or 2 would be using cell phones.

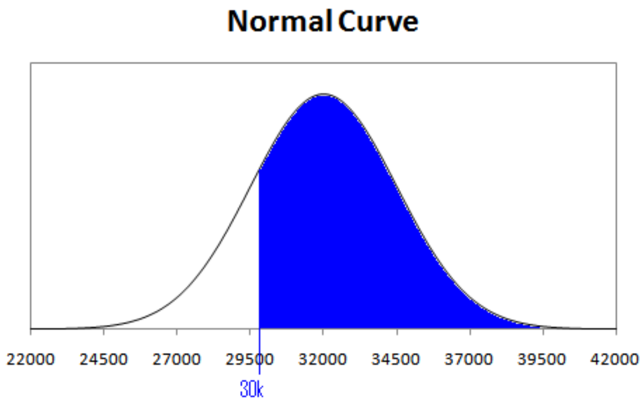
Our result of 4 users is possible but is highly unlikely. You could look at this a couple of ways:

1. Because our simulation showed that 4 callers is possible, this doesn't necessarily refute the claim. It is highly unlikely, but it could be that we were just "lucky" to have observed it.

2. We aren't very likely to observe 4 cell phone users. It could be that the true probability is really larger than 12%, if we think that taking another sample at the bus stop would result in another count of 4 users. In this case, our observation might refute the claim.

**11.42] Tires.** We are interested in finding the probability that four tires last 30,000 miles or longer. First, we need to find the probability that one tire lasts that long. We can use the normal distribution for that.

We are interested in the blue shaded proportion below:




Using R, we can compute this proportion:

**Distributions → Continuous Distributions → Normal Distribution → Normal Probabilities**

The R output is:

```
> pnorm(c(30000), mean=32000, sd=2500, lower.tail=FALSE)
[1] 0.7881446
```

So, the probability that a single tire lasts 30,000 miles or more is 0.7881.

Simulation. We can do the simulation in the same way as for problem 11.40. We'll simulate 4 observations, and repeat that many times. I'll try 1,000 this time. If the generated random number is less than or equal to 0.7881, we'll call that a tire that lasted. Otherwise, it will be a failed tire. You can adapt the code I provided above to summarize the results.

```
> summary(tiretotal)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   3.000   3.137   4.000   4.000

> summary(as.factor(tiretotal))
  0   1   2   3   4
  1  28 181 413 377
```

Answer. For our simulation, we found that all four tires lasted over 30,000 miles for 377/1000 = 0.377 of the simulations. I'd estimate that all four tires will last over 30,000 miles 37.7% of the time.

SIDE NOTE: We can also compute this theoretically. Assuming the tires are independent, the probability of all four lasting is $0.7881 \times 0.7881 \times 0.7881 \times 0.7881 = 0.7881^4 = 0.3858$. That's pretty close to my simulated value of 0.377. If we did more simulations we'd probably get closer to 0.3858.

# Chapter 12

**12.24] Playground.**
The managers will only get responses from people who come to the park to use the playground. Parents who are dissatisfied with the playground may not come.

**12.36] Happy workers?**

(a) A small sample will probably consist of mostly laborers, with few foremen, and maybe no project managers. Also, there is a potential for response bias based on the interviewer if a member of management asks directly about discontent. Workers who want to keep their jobs will likely tell the management that everything is fine!

(b) Assign a number from 001 to 439 to each employee. Use a random number table or software to select the sample.

(c) The simple random sample might not give a good cross section of the different types of employees. There are relatively few foremen and project managers, and we want to make sure their opinions are noted, as well as the opinions of the laborers.

(d) A better strategy would be to stratify the sample by job type. Sample a certain percentage of each job type.

(e) Answers will vary. Assign each person a number from 01-14, and generate 2 usable random numbers from a random number table or software.

# Chapter 13

**13.2] Heart attacks and height.**

(a) No, this is not an experiment. There are no imposed treatments. This is a retrospective observational study.

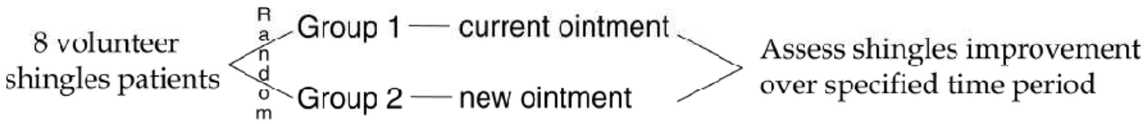
(b) We cannot conclude that shorter men are at higher risk of heart attack. There may be lurking variables that are associated with both height and risk of heart attack.

**13.36] Swimsuits.**
The "control" in this experiment is not the same for all swimmers. We don't know what "their old swim suit" means. They should compare their new swim suit to the same suit design. The order in which the swims are performed should be randomized. There may be a systematic difference form one swim to the next. For instance, swimmers may be tired after the first swim (or more warmed up). Finally, there is no way to blind this test. The swimmer will know which kind of suit they have on, and this may bias their performance.

**13.42] Full moon.**
Answers may vary. There's really no way to randomize people to a moon phase, and we would have to sample *a lot* of people to get people who have a police or emergency room incident. It would be better to use a retrospective observational study. For example, collect records from a random selection of police and emergency room logs for the past 3 years. Find the number of cases for the days when there is a full moon, when there is a waxing moon, a waning moon, and when the moon is nearly dark. Then, compare the numbers for each group.

**13.50] Shingles.**

(a) Answers may vary. This experiment has 1 factor (ointment), at 2 levels (current and new), resulting in 2 treatments. The response variables are the improvements in severity of the case of shingles and the improvements in the pain levels of the patients. Randomly assign the eight patients to either the current ointment or to the new ointment. Before beginning treatment, have doctors assess the severity of the case of shingles for each patient, and ask patients to rate their pain levels. Administer the ointments for a prescribed time, and then have doctors reassess the severity of the case of shingles, and ask patients to once again rate their pain levels. If the neither the patients nor the doctors are told which treatment is given to each patient, the experiment will be double-blind. Compare the improvement levels for each group. A picture of the design is given below:




(b) Answers may vary. Let numbers 1 through 8 correspond to letter A through H, respectively. Ignore digits 0 and 9, and ignore repeats. The first row contains the random digits, the second row shows the corresponding patient (X indicates an ignored or repeated digit), and the third row shows the resulting group assignment, alternating between Group 1 and Group 2.

```
41098 18329 78458 31685 55259
DAXXH XXCBX GXXEX XXF
11  1   12  2  2    2
```

Group 1 (current ointment): D, A, H, C
Group 2 (new ointment): B, G, E, F

(c) Assuming that the ointments looked alike, it would be possible to blind the experiment for the patient and the evaluating doctor. If both the subject and the evaluator are blinded, the experiment is double-blind.

(d) Before randomly assigning patients to treatments, identify them as male or female. Having blocks for males and females will eliminate variation in improvement due to gender.

# Chapter 14

**14.20] Stats projects.**
Since all of the events listed are disjoint, the addition rule can be used.

(a) *P*(two or more semesters of Calculus) = 1 – (0.55 + 0.32) = 0.13
(b) *P*(some Calculus) = *P*(one semester or two or more semesters) = 0.32 + 0.13 = 0.45
(c) *P*(no more than one semester) = *P*(no Calculus or one semester) = 0.55 + 0.32 = 0.87

**14.33] Disjoint or independent?**

(a) For one draw, the events of getting a red M&M and getting an orange M&M are disjoint events. Your single draw cannot be both red and orange.

(b) Answers here can vary based on with and without replacement. Based on the Question 14.33, the company is giving you the overall probabilities. This assumes that you are drawing from a very large population. For two draws, the events of getting a red M&M on the first draw and a red M&M on the second draw are independent events. Knowing that the first draw is red does not influence the probability of getting a red M&M on the second draw. However, if you just have a handful of M&M's and are drawing without replacement, then the draws are *dependent*. If you have 10 Red M&M's out of 100, the probability of a red on the second draw is 9/99 if you drew a Red first.

(c) Disjoint events can never be independent. Once you know that one of a pair of disjoint events has occurred, the other one cannot occur, so its probability has become zero. For example, consider drawing one M&M. If it is red, it cannot possibly be orange. Knowing that the M&M is red influences the probability that the M&M is orange. It's zero. The events are not independent.