ASSIGNMENT 1

Submitted by

VINAY KUMAR RANGANATH BABU
(University ID: 0002897242)

As part of "Introduction to Statistics"

In
DATA SCIENCE
SCHOOL OF INFORMATICS AND COMPUTING

INDIANA UNIVERSITY

AUG 2016

1.For each of the following scientific questions, state whether the question is best answered using a randomized experiment or an observational study, and briefly explain why.

(a) Does the flu vaccine prevent the flu?

Answer: This experiment can be clearly categorized as randomized experiment because 2 groups of people can be randomly picked (all different age groups, gender) and one group gets a flu vaccination and the other does not get. Over the course of time, this experimentation shows how many from group 1 and 2 respectively are affected by flu. Hence by this way, randomized experiment would give an appropriate answer for this question.

(b) Has support for same-sex marriage increased over time?

Answer: The conclusion/answer for this statement can only be given by observational studies. Same-sex marriage cannot be forced on anyone and hence it cannot be random experiment. We are accumulating data in a way that it does not matter on how actually the information did arise. And we are collecting data from the past (like past few years) and then deciding on the outcome. Hence observational studies can be the best way to determine this.

(c) Does banning laptops in class improve exam scores?

Answer: The answer for this can be best given by randomized experiment because we can think of a finite group of population to perform the experiment. First group of population (let's say one class) are allowed to use laptops at class for a fixed period of time (say whole semester) and the other group of population (second class) are banned of bringing the laptops. At the end of semester, we could make out which class scores good marks and which other do not. Although there are many other factors like students do not study for exams, searching only useful information on laptops in class which might affect our result, randomized experiment gives a more promising result compared to other models.

(d) Are voters with college degrees more likely to support Hillary Clinton than voters without college degrees?

Answer: Voting is something confidential and people might not be open to disclose the fact on whom did they vote for. Hence choosing a randomized experiment might be a bit of challenge in evaluating this. Hence depending on the number of people who voted (both with college degrees and ones without), a close observation can be made using other factors like the place, gender etc. Hence I feel observational study would fit to get probable solution.

(e) Does bacon cause colorectal cancer?

Answer: This experiment can be done with randomized approach using cause-and-effect question. However, one cannot force a group of eating bacon because cancer is a life threatening disease and the people might deny to participate in the experiment. Hence with this, it's better the experiment is conducted using observational study.

2. According to a study done at Kaiser Permanente in Walnut Creek, California, women who use oral contraceptives (“the pill”) have a higher rate of cervical cancer than women who do not use the pill, even after adjusting for age, education, and marital status.

(a) Was this study a blind randomized experiment or an observational study?

Answer: Since the expected end result is finding a probable cervical cancer patient number, the experiment would not be conducted using blind/randomized approach. The reason for this is because the random sample who are chosen to swallow oral contraceptives would just not be ready as cancer is a life taking disease. Hence this is an observational study which involves the past few years of data and this helps in making probable conclusion.

(b) Does the study prove that the pill causes cervical cancer? Explain why or why not.

Answer: The study gives a very high probable number that the pill causes cervical cancer. Since the study is made via observations considering many factors like age, education and marital status and for over the course of time, it proves that the pill causes cervical cancer.

(c) Besides age, education, and marital status, what other factor(s) related to cervical cancer are different (on average) between women who use the pill and women who don’t?

Answer: The other factors like dosage and time lapse can be an important factor in measuring this. Pill taken up to a restricted dosage might yield different results. Similarly, how often (time gap) the pill is consumed can also be a good factor for this measure.

3.The Center for American Progress to study public attitudes toward sports teams that ex-pressed opinions on issues that could be controversial, such as LGBT issues. One research questions they were interested in was: “To what degree do people believe that professional sports teams should take public stances on social causes?”
In general, respondents either “somewhat agreed” (33.2 percent) or “strongly agreed” (19.3 percent) that “professional sports teams should utilize their platforms to advocate for causes they believe in.” One in three respondents (30 percent) stated they were neutral on this issue or had “no opinion.” Among respondents who identified as men, 46.5 percent either “somewhat agreed” or “strongly agreed” that sports teams should use their platforms to advocate for causes, while 33 percent had “no opinion.” Among respondents who identified as women, 60.4 percent either “somewhat agreed” or “strongly agreed” with that statement, with another 25.6 percent stating that they were neutral or had “no opinion.”

The study also gave the following information about their survey:

The results presented above are from a convenience sample of 367 respondents recruited using Amazon Mechanical Turk, an online platform that allows for the purposeful sampling of respondents who meet relevant criteria. This survey-hosting website has been shown to be an efficient platform for gathering reliable data from diverse populations. . . 44.1 percent of respondents were between the ages of 18 and29, 38.3 percent were between the ages of 30 and 44, and 17.8 percent were above

the age of 44. 10.4 percent of respondents identified as "strong Democrat," 26.0percent as "Democrat," 16.7 percent as "independent-lean Democrat," 18.9 percent as "independent," 8.7 percent as "independent-lean Republican," 10.4 percent as "Republican," and 4.6 as "strong Republican."

(a) Was this survey a statistically unbiased answer to the question "To what degree do people believe that professional sports teams should take public stances on social causes?" If so, explain why; if not, describe and explain the likely direction of the bias.

Answer: The survey is a very good example of statistically unbiased experiment because of the quality of samples which were chosen to conduct the experiment. There is an equal propionate of all kinds of people (different ages) who were selected and the factors like these gives a good probable values for the experiment. Also as the survey mentions, the mechanical turk tool also helps in providing a randomized sample choices which helps the experiment go in a right order.

(b) Suppose an interested party gives you a reasonable budget to carry out a more rigorous study of the Center for American Progress' research question. Describe briefly the study you would perform.

Answer: I would include other factors like choosing people from different countries, different origin, different languages and also people with different professions. This would help the survey in a very big way because of the changes in culture and professions would impact their decisions and hence yielding a better probable result.

---

4. A deck of cards contains 52 different cards. For each of the following situations, give the number of equally likely outcomes.
(a) I draw the top card.

Answer: Since there are 52 cards in the deck, there are 52 possible ways of picking a card. Hence the number of equally likely outcomes are **52**.

(b) I draw the top two cards. The order matters

Answer: Since I draw the top 2 cards and the order matters, the number of ways selecting 2 cards from a deck of 52 cards would be

$_{52}P_2 = 52!/50! = 52*51 = 2652$

Hence the number of equally likely outcomes are **2652**.

(c) I draw the top two cards. The order doesn't matter.

Answer: Since I draw the top 2 cards and the order doesn't matter, the number of ways selecting 2 cards from a deck of 52 cards would be

$_{52}C_2$ = 52!/(50!*2!)=(52*51)/2=2652/2=1326

Hence the number of equally likely outcomes are **1326**.

(d) I draw the top card. Without replacing the card, I reshuffle the remaining cards in the

deck, and draw the top card. (The order matters.)

Answer: As we solved in the first question, the number of ways of selecting the top card is 52 ways. Since we do not replace and have to select one more (here shuffling doesn't matter), the number of ways of selecting 51 cards is 51.

Hence the number of equally likely outcomes are 52*51 = **2652**

(e) I draw the top card. Then I replace the card, reshuffle the deck, and draw the top card.

Answer: As we solved in the previous question, the number of ways of selecting the top card is 52 ways. Since we replace the card and have to select one more (after shuffling), the number of ways of selecting 52 cards again is 52.

Hence the number of equally likely outcomes are 52*52 = **2704**

---

5. I toss four fair coins: a penny, a nickel, a dime, and a quarter.

(a) What is the probability that the penny and the nickel show heads, while the dime and the quarter show tails?

Answer: Since it's a fair coin, all the coins has 2 sides (heads and tails).

The probability that penny shows heads is ½.

Similarly, the probability that nickel shows heads is ½.

Similarly, the probability that dime shows tails is ½.

Similarly, the probability that quarter shows tails is ½.

Hence the total probability is ½*½*½*½ = 1/16 = 0.0625 (approximately 6.25 %)

(b) What is the probability exactly two of the coins show heads?

Answer: Since we have the condition that exactly 2 coins should be head, we have 4 slots to fit 2 heads. And the total number of events if 2*2*2*2 = 16

Hence the probability is

$_4C_2$ / 16 = (4! / ((4-2)! * 2!)) / 16 = ((4*3*2)/ (2*2)) / 16 = **0.375** (Approximately 37.5%)

(c) What is the probability exactly two of the coins show heads and one of them is the penny?

Answer: Since we have the condition that exactly 2 coins should be head and one among them is penny, we have 3 more slots to fit a single head. And the total number of events if 2*2*2*2 = 16

Hence the probability is

$_3C_1$ / 16 = (3! / ((3-1)! * 1!)) / 16 = ((3*2)/ (2*1)) / 16 = 3/16 = **0.1875** (Approximately 19%)

---

6. I roll three fair six-sided dice.

(a) What is the probability none of the dice show six?

Answer: By axioms, we know that

P (none of dice show six) = 1 – P (atleast one show six)

P (none of dice show six) = (1 – 1/6) * 3 dice

P (none of dice show six) = 5/6 * 5/6 * 5/6

P (none of dice show six) = 125/216 = 0.5787 (approximately 58%)

(b) What is the probability at least one of the dice shows six?

Answer: By axioms, we know that

P (none of dice show six) = 1 – P (atleast one show six)

And we calculated P (none of dice show six) as 125/216

Hence P (atleast one show six) = 1 – P (none of dice show six)

Hence P (atleast one show six) = 1 – (125/216) = 91/216 = 0.4212 (approximately 42%)

(c) What is the probability at least two of the dice show an even number?

Answer: Solution1 - We can write out all the combinations for 3 dice. It can be either even or odd.

Hence the combinations are:

EEE, EEO,EOE,OEE,OOO,OOE,OEO,EOO

There are 8 total outcomes and 4 of them at least have 4 even numbers in them.

Hence probability is 4/8 = ½ = 50%

Solution2 – Since there should be at least 2 even numbers, only one dice will be left. Hence the probability of getting an even number or an odd number out of one die is ½ .

Hence probability is ½ = 50%