

Assignment 1- Solution

1. Random Experiment vs Observational Study.

- a. The scientific question of whether the flu vaccine prevented the flu can be abstracted as *cause-and-effect* analysis. A **randomized controlled experiment** is the best choice to study *cause-and-effect*, as it ensures that the differences in groups being compared are either due to treatment or due to chance (randomness).
- b. The question of whether the support for same-sex marriage has increased over time can be answered using an **observational/descriptive study**, since the study involves merely observing the inclination of subjects in a random sample towards same-sex marriage, rather than investigating the cause for support.
- c. To answer whether banning laptops in class lead to improvements in exam scores, we need to conduct a cause-and-effect analysis. The study involves checking whether banning laptops in the most likely cause which lead to the effect of increase in exam scores. As with any study that involves cause-and-effect analysis, the current scientific question can also be best answered using a **randomized controlled experiment**, as it ensures that the differences in groups being compared are either due to the cause or due to chance (randomness).
- d. **Observational study**- Since the survey involves studying rather than controlling the subject in the sample, an observation study can better answer the question of whether voters with college degrees are more likely to support Hillary Clinton than voters without college degrees.
- e. **Randomized controlled experiment**- Since the study involves examining the cause, whether bacon, for colorectal cancer, a randomized controlled experiment could best answer the study.

2.

- a. The study done at Kaiser Permanente in Walnut Creek, California, to determine the effect of oral contraceptives on cervical cancer can be categorized under **observational study**.
- b. No, the study most likely doesn't prove that the pill causes cancer, as it's biased- the study made some adjustments related to age, education and marital status in the sample, without any empirical evidence. It is possible that the adjusted factors, i.e., age, education and marital status might have biased the conclusion towards the hypothesis that pill causes cancer.
- c. Factors such as Smoking, Immune system deficiency, Socioeconomic factors might be different between women who use the pill and women who don't. (Source- <https://www.cancer.org/cancer/cervical-cancer/causes-risks-prevention/risk-factors.html>).

3.

a. No, the survey fails to provide an unbiased answer to the question, because the sample on which the study was conducted doesn't represent a perfectly balanced group, and the sampling method involves "purposeful sampling of respondents who meet relevant criteria"- similar to the 1948 presidential election polls where the results were incorrect due to human bias in quota sampling.

i. Unbalanced sample- The survey indicates that 44.1 percent of its respondents were between the ages of 18 and 29, 38.3 percent of its respondents were between the ages of 30 and 44, and 17.8 percent of its respondents were above the age of 44. So, there is an obvious bias between different age groups. It could be possible that the 44.1% of the respondents who are young tend to follow sports more than the other age groups and, hence, are sympathetic of sportsmen or sports teams.

ii. More than 50% of the respondents have democratic inclinations. So, the results could be biased, in that the respondents were inline with some democratic leader who made a statement in favor of sports teams expressing opinions on social causes.

b.

i. Since this is an observational study, my first effort would be to have a sample that is representative of the entire population- the sample must not be biased towards an age group, gender, socioeconomic status, etc. So, I would not select the respondents based on a fixed criterion. Random sampling of respondents to the survey could generate a better representation of the population.

ii. Before random sampling, effort must be made to identify the segments within population that are most correlated to or influence the outcome of our study.

iii. Then, stratified random sampling must be employed on the total respondents to generate an appropriate sample which is representative of the entire population.

iv. The randomness ensures that there's no human bias, and the outcome is either actual or due to randomness. So, the generated results would be more reliable.

4. $S = \{1,2,3,4,5,6,7,8,9,10\}$, $A = \{1,3,5,7,9\}$, $B = \{1,2,3,5,7\}$

a. $A^c = S - A = \{2,4,6,8,10\}$

b. $B^c = S - B = \{4,6,8,9,10\}$

c. $(A \cup B)^c = S - (A \cup B) = \{1,2,3,4,5,6,7,8,9,10\} - \{1,2,3,5,7,9\} = \{4,6,8,10\}$

d. $(A \cap B)^c = S - (A \cap B) = \{1,2,3,4,5,6,7,8,9,10\} - \{1,3,5,7\} = \{2,4,6,8,9,10\}$

5. A fair Penny is tossed 10 times, A fair quarter is tossed 5 times.

a. The number of sequences of 15 tosses that result in a total of exactly 6 heads=
 ${}^{15}C_6 = \frac{15!}{6! \cdot 9!} = 5005$.

b. The number of sequences of 15 tosses that result in exactly 2 heads in 10 tosses of the penny and exactly 4 heads in 5 tosses of the quarter is ${}^{10}C_2 * {}^5C_4 = 45 * 5 = 225$.

6.

a. $\Phi(6) = 2^6 = 64$

b. $\Phi(-3) = 2^{-3} = 1/8$

c. $\Phi(R) = 2^R = R$ (Real-valued number)

d. $\Phi^{-1}(16) = \log_2(16) = 4$

e. $\Phi^{-1}(1/4) = \log_2(2^{-2}) = -2.$