**Applied Statistical Inference and Experimental Design**

**Homework 1**

* 1. **For each of the following, describe the population and, if relevant, the sample. For each number presented, determine if it is a parameter or a statistic (or something else).**

**a) A survey of 1500 high school students finds that 47% watch the cable show “Game of Thrones.”**

**b) The 2010 US Census reports that 9.6% of the US population was between the ages of 18 and 24 years.**

**c) Based on the rosters of all National Basketball Association teams for the 2006–2007 season, the average height of the players was 78.93 in.**

**d) A March 2016 Harris poll consisting of 2106 national adults, age 18 years or older, found that 19% strongly or somewhat disagree with the statement that the US has come a long way toward reaching gender equality.**

a) Population: High school students.

Sample: 1500 high school students.

Statistic is 47%.

b) Population: US Population

Parameter:9.6%

c) Population: All National Basketball Association teams

Parameter:78.93

d) Sample: 2106 national adults

Statistics: 19%

* 1. **Review the description of the Iowa recidivism case study in Section 1.4.**

**a) Does this data represent a population or a sample?**

**b) In this data set, 19.4% of the offenders were originally convicted of a misdemeanor. Does this number represent a parameter or a statistic?**

a) Population of all offenders convicted of either a misdemeanor or felony who were released from an Iowa prison during the 2010 fiscal year

b) Parameter as it represents Population

**1.4 The journal Molecular Psychiatry reported on a study claiming that playing the video game Tetris reduces the formation of bad memories after a traumatic event (Iyadurai et al. (2017)). Seventy-one patients who were involved in a motor vehicle accident and admitted to a British emergency room were recruited. After completing some baseline assessments, they were randomly assigned to either play Tetris for at least 10 uninterrupted minutes or fill out a simple log detailing their activities while waiting in the emergency room. The patients who played Tetris reported having fewer intrusive memories about their accident than the patients who had completed a log.**

**a) In this experiment, what were the treatments?**

**b) Was this a double-blind study?**

**c) Can the researchers conclude that playing Tetris causes the reduction of painful memories?**

**d) Can we generalize these results to a population?**

a) Patients played Tetris or fill out a log.

b) No.

c) Yes, treatments were assigned randomly.

d) No, patients were recruited.

**1.8 In the Google mobile ads case study (Section 1.12),**

**a) Why is this study described as an experiment and not an observational study?**

**b) Can Google claim that their recommendations “caused” the outcome of the study?**

**c) Can Google generalize their results to all advertisers who advertise on Google?**

1. For observational study we would only have the pre results. For an experiment we can find both the pre and post results for mobile and desktop.
2. Partially No, as Google were recommending increasing the mobile multiplier.
3. No because each row corresponds to a single combination of campaign and ad group: These could be for different products, a different set of ads, target a different population,

**1.11 In the mobile ads case study (Section 1.12), the variables m.cpr and d.cpr, which measure cost/value (how much it costs a company to advertise per how much they make), are recorded as 0 if value is 0. The error is defined by error = m.cpr-d.cpr. If m.cpr=10 and d.cpr=1, then error is 9. However, if on the desktop mobile, d.value is 0, that is, the company did not make any money, then d.cpr is defined to be 0. So the error is −1 which is smaller in absolute value than the first case.**

**a) Do you think that this accurately reflects the magnitude of the difference in these two scenarios?**

**b) If you were a consultant for Google, can you recommend other ways of defining cpr when the denominator value is 0?**

a) Based on textbook calculation, CPR = Total amount paid by advertisers/Value of conversion, which is inaccurate.

For example, in Row 1, the pre\_cpr for mobile is 0.489, this value moves closer to 0 as the value of pre\_conversion increases. Also, for the same row, the post\_cpr for mobile is 0 since the value of post\_conversion is 0.

In one situation a higher conversion value will move cpr towards 0 and in another instance a 0 value in conversion will result in 0 cpr as well. Due to this flaw in methodology the true magnitude of differences cannot be calculated.

b) I would recommend, the below

CPR = (value of conversion-amount paid by advertiser)/amount paid by advertisers\*100(%)

By doing so, the true magnitude can be calculated.

* For situations where value of conversion is 0, the result is -ve100%, meaning the advertisers has not made any profits.
* For situations where amount paid by advertiser is 0, the cpr is 0.
* For situations where value of conversion < amount paid by advertiser, the resulting value is -ve. Showing the loss in terms of negative %.
* For situations where value of conversion > amount paid by advertiser, the resulting value is +ve. Showing the gain in terms of positive%.