

Name: Caleb McWhorter — Solutions

MATH 100

Fall 2023

HW 12: Due 12/06

“Surveys show that surveys never lie.”

–Natalie Angier

Problem 1. (10pt) For each of the following, determine whether the proposed survey method is systematic sampling, stratified sampling, cluster sampling, or convenience sampling. Be sure to justify your reasoning.

- (a) A company surveys its employees by randomly surveying a few individuals from its list of managers, drivers, administrative staff, production staff, and custodial staff.
- (b) A student asks neighbors about their political affiliation.
- (c) A government agency sends surveys to every resident in three upstate New York villages.
- (d) Every one hundred customers ordering an item are requested to fill out a customer satisfaction survey.

Solution.

- (a) This is stratified sampling. To create a representative sample, the company has broken its employees into homogeneous groups and has combined smaller samples from these groups into a larger sample. This is stratified sampling.
- (b) This is convenience sampling. The sample does not use any probabilistic methodology and rather uses what is ‘convenient’ to the student.
- (c) This is cluster sampling. To create a representative sample, the government agency grouped NYS villages into heterogeneous groups and then sampled every individual from some of those groups.
- (d) This is systematic sampling. The sales company systematically regularly surveys every one-hundredth individual to create their random sample.

Problem 2. (10pt) Determine whether each of the following measurements are nominal, ordinal, interval, or ratio. Be sure to justify your answer. You need only indicate the 'highest' level of measurement.

- (a) Placement in a race.
- (b) Type of Pain: acute, chronic, aching, burning, stabbing, etc.
- (c) The average temperature ($^{\circ}\text{F}$) of a mountain range.
- (d) The time taken to complete an assignment.

Solution.

- (a) This is an ordinal level of measurement. We can order the participants in a race by the place in which they finish. However, differences or ratios of these placement positions do not hold mathematical meaning relative to the participants individual 'qualities.'
- (b) This is a nominal level of measurement. There are named categories from which a patient can choose. However, there is no quantitative measurement whatsoever attached to these categories.
- (c) This is an interval level of measurement. The temperature can be measured and differences between these values can be taken. However, one cannot take ratios of these temperatures and there is no 'natural' zero—merely an arbitrary one.
- (d) This is a ratio level of measurement. The time can be measured and both differences and ratios of these measurements 'have meaning' relative to the quantities measured. Moreover, there is a natural choice of zero.

Problem 3. (10pt) Determine whether each of the variables underlined in the statements below are quantitative or categorical variables—be sure to justify your answer.

- (a) A company creates groupings for the various types of coffee they sell.
- (b) A professor rates the quality of a students work as ‘above average’, ‘average’, or ‘below average.’
- (c) A car manufacturer counts the number of defective products produced at a factory.
- (d) Nurses measure the concentration of medication in a patients bloodstream.

Solution.

- (a) This is a categorical variable. The coffee has a name and no numerical values.
- (b) This is a categorical variable. The quality has a ‘name’ but no numerical value associated to it.
- (c) This is a quantitative variable. The quantity has a clear numerical value associated to it that holds ‘meaning’ relative to the object(s) measured.
- (d) This is a quantitative variable. The quantity has a clear numerical value associated to it that holds ‘meaning’ relative to the object(s) measured.

Problem 4. (10pt) Being sure to fully justify your reasoning, complete the following:

- (a) Explain the difference between a random sample and a simple random sample.
- (b) If one finds a high correlation between study time and GPA, then it must be that one's study time (or lack thereof) causes one's GPA to be high or low.
- (c) If one finds that regularly taking a medication has a statistically significant chance of increasing one's cholesterol levels, then there is clear practical significance to this medication.

Solution.

- (a) A random sample is a sample in which every *individual* is equally likely to appear in the sample. A simple random sample is a sample in which every *collection* of individuals is equally likely to appear in the sample.
- (b) No, correlation does not imply causation. There may be hidden, lurking, or confounding variables that are causing the correlation between the two. For instance, it might be that more naturally 'gifted' students are more driven to study longer hours. These students would have likely done well regardless. This would likely be the variable causing the high correlation between study time and GPA. It may also be that the study directly causes one's grades to be higher—hence their overall GPA. A simple correlation analysis is not sufficient to determine one way or the other.
- (c) No, practical significance is not the same as statistical significance. For instance, it might be that the medication with a single dose causes *every* individual's cholesterol to reduce from any level to 'normal' levels. This would be practically and statistically significant. But a drug which causes *every* individual's cholesterol level to decrease by the smallest possible measurable amount is statistically significant but does not hold practical significance. Without more information, one cannot determine what is practically versus statistically significant.