

## Unit 3 – Populations and Samples

### Practice Problems

### SOLUTIONS

1. For each of the following situations, define the target population, and how you might obtain a sample. What will be your sampled population? How does this differ from the target population?

- a. A city engineer wants to estimate the average weekly water consumption for single family dwelling units in the city.

Target population: All single family dwelling units in city.

Sample: Use a census map of city dwellings to take random sample of homes and monitor weekly water consumption by reading meters on homes in sample. (I'm not sure how you would measure this in homes having a well. You might re-define the target population to include only those on city water, and take your random sample from the billing records).

Sampled Population: All homes listed as single family units on most recent census map. This list may include homes no longer in existence, or ones that have been converted to multiple units. It will miss new construction and conversions to single family units, and homes that were listed incorrectly.

- b. A physician wants to estimate the average length of time from initial diagnosis with ovarian cancer to death.

Target population: Women diagnosed with ovarian cancer, including women in future who will be diagnosed with ovarian cancer. (This is the group to which the physician is likely to want to make inferences).

Sample: Use hospital records or cancer registry to identify women diagnosed with ovarian cancer in some specified region and time period. Can include all those listed, or for larger region or time period, take a random sample from this list.

Sampled Population: All cases listed with cancer registry for the region and time period. This list may miss women who are incorrectly listed under the wrong diagnosis code, and include others for the same reason. Regional differences in treatment as well as temporal differences in treatment and disease may mean the sampled population differs from the "future" population, or from other regions of the country for the same period.

2. Which of the following estimates are likely to be biased? Why? Is the bias positive or negative? Why? (*note: Positive bias means a consistent likelihood of overestimating, negative bias is underestimating*).

- a. You estimate the average number of bank customers waiting for service whenever the bank is open by counting the number of customers whenever you go to the bank.

**Biased** – Depending on the particular time of day and week you tend to go to the bank, the bias could be positive or negative. For example, if I go to the bank on Fridays afternoons after I get a paycheck, and if other UMass employees behave similarly, that should produce a **positive bias**.

- b. You estimate the proportion of 7-12 year old children using helmets when they ride bikes by asking parents if their child wears a helmet when the child is brought to the physician's office for a "well" visit.

**Biased** – Bias is probably **positive** because 1) a sample of parents bringing kids to well visits may be parents more concerned with children's health who also enforce helmet use, and who have more money to buy helmets; 2) parents may lie about this to please authority at the doctor's office; 3) parents may think the kids are wearing helmets when they are not.

- c. A highway patrolman parks next to a highway and records speeds on his radar to estimate the percentage of people exceeding the speed limit on that highway.

**Unbiased *IF*** done throughout a long stretch of the day including rush hour and less busy times as long as patrol car is hidden. If the car is easily seen, drivers may lower speed for that stretch of highway and then increase speed later, resulting in **negative bias**.