AP Statistics
Unit 03 – Sampling & Experiments
Day 01 Notes – Sampling & Surveys

Name Key
Period _____

When we sample, we are collecting data about a population in order to (hopefully) make generalizations about the population. We are not changing anything that the general population does, rather we are OBSERVING it.

Population: the entire group of individuals we want information about in a statistical study.

Census: collects data from every individual in the population.

Sample: a subset of individuals in the population from which we actually collect data.

Survey: investigation about the characteristics of a given population by collecting data from a sample of that population.

Sample Surveys:

1. Choose what population you want to describe.
2. Choose what variable(s) we want to measure.

The advantages of sampling data are:
- Reduced cost (no costly treatments to impose, less staffing, etc.)
- Reduces time needed to collect & process the data
- Enables detailed questions to be asked
- Results can be made available quickly

The disadvantages of sampling are:
- Data on sub-populations may be too unreliable
- Data for small geographical areas may be unreliable
- Tough to communicate the precision of the data
- Many biases (hidden variables included) can exist

Example 1: Sampling Hardwood and Humans

Identify the population ~~of~~ and the sample in each of the following settings:

a) A furniture maker buys hardwood in large batches. The supplier is supposed to dry the wood before shipping (wood that isn't dry won't hold its size and shape). The furniture maker chooses five pieces of wood from each batch and tests their moisture content. If any piece exceeds 12% moisture content, the entire batch is sent back.

population: all pieces of hardwood in a batch sample: 5 pieces of wood that are selected and tested.

b) Each week, the Gallup Poll questions a sample of about 1500 adult US residents to determine the national opinion on a large variety of issues.

population: all adult US residents sample: 1500 adults who respond to the survey questions.

Sometimes, bad sampling occurs and we cannot generalize our findings to the population from which we sampled. This can occur for a variety of reasons.

Convenience Sample: choosing individuals from the population who are easy to reach

* usually results in unrepresentative data

Bias: using a method that favors some outcomes over others

* underestimates/overestimates

AP explain flaw, over/under-estimate & why

Voluntary Response Sample: consists of people who choose themselves by responding to a general invitation

Example 2: Illegal Immigration

Former CNN commentator Lou Dobbs doesn't like illegal immigration. One of his shows was largely devoted to attacking a proposal to offer driver's licenses to illegal immigrants. During the show, Mr. Dobbs invited his viewers to go to loudobbs.com to vote on the question, "Would you be more or less likely to vote for a presidential candidate who supports giving driver's licenses to illegal aliens?" The result: 97% of the 7350 people who voted by the end of the show said, "Less likely."

What type of sample did Mr. Dobbs use in his poll? Explain how this sampling method could lead to bias in the poll results.

Mr. Dobbs used a voluntary response sample; people chose to go online and respond. Responders were most likely viewers of Mr. Dobbs' program and likely agree with his views. this 97% is likely an extreme overestimate of the % of people who would support a candidate with this position.

Example 3: Check Your Understanding

Identify the sampling method and explain how it could lead to bias.

a) A farmer brings a juice company several crates of oranges each week. A company inspector looks at 10 oranges from the top of each crate before deciding whether to buy all the oranges.

Convenience sampling! The ones on top are easy to access. They are likely less squished/in better condition than the oranges below them, resulting in an overestimate of orange quality.

b) The ABC program Nightline once asked whether the United Nations should continue to have its headquarters in the US. Viewers were invited to call one telephone number to response "Yes" and another for "No." There was a charge for calling either number. More than 186000 callers responded, and 67% said "No."

Voluntary response sampling! Callers self-select to participate. only those with strong opinions will pay to call and voice their opinion. The cost of the call could lead to an over or underestimate.