AP Statistics Additional Summary

Updated June 19, 2011

1 Sampling and Experiment

1.1 Terms

- Population: Entire set of individuals that we are interested in
- Sample: The part of the population that is actually being examined

1.2 Methods of Data Collection

- Census: A complete enumeration of an entire population.
  eg.: US Bureau of the Census conducts a census every 10 years.
- Sample survey: Obtain information about a whole population by studying a part of it (sample). Randomness is needed to minimize bias.
- Experiment: A planned study to generate data. We attempt to influence some (or all) observations. The primary concern is the design of the experiment.
- Observational Study: There is no deliberate attempt to influence any observations. A primary concern is obtaining a sample from the population.

1.3 Methods of Sampling

- Simple random sampling: Each member has an equal chance of being selected.
  - Sampling with replacement
  - Sampling without replacement
- Systematic sampling: Selecting every k-th item
- Stratified sampling: Divide population into homogeneous groups called strata, and randomly sample from each stratum.
  Example of strata: sex, age group, race, region, etc.
- Convenience sampling: The subjects are chosen based on ease of access.
- Proportional sampling: Divide the population into strata, and simple random sample of size proportional to the stratum size is obtained from each stratum.
- Cluster sampling: Divide population into nonhomogeneous groups called clusters, and a simple random sample is obtained from each cluster.
- Multistage sampling: Involves two or more steps

1.4 Sources of Bias in Surveys

- Response bias: Caused by the behavior of the interviewer or respondent
- Nonresponse bias: The person selected for an interview cannot be contacted or refuses to answer.
- Undercoverage bias: Part of the population is left out of the selection process
- Wording bias: May occur if confusing or misleading questions are asked
- Household bias: When a sample includes only one member of any given household, members of large households are underrepresented.
- Quota sampling bias: This results when interviewers are given free choice in picking people.
- Selection bias: In 1936, Literary Digest opinion poll selected only people with cars and telephones which were only the wealthy minority.
- Size bias: Throwing darts at a map to decide in which states to sample would bias in favor of geographically large states.
- Voluntary response bias: Samples based on individuals who offer to participate typically give too much emphasis to people with strong opinions.

1.5 Placebo Effect

- Comparison of a control group and treatment group
- There may be a psychological effect if the people know which group they belong to.
- Placebo group: A control group that receives a placebo (fake drug) in experiments involving medicines

1.6 Blinding

- Measurements may be biased if the person taking the measurements knows whether a patient received a placebo or not.
- Single blind: Either the patient or the person measuring the patient's reaction does not know which treatment was given.
- Double blind: Both the patient and the person measuring the patient's reaction do not know which treatment the patient was given.

1.7 Randomization

- Blocking: used to control the effect of known factors, for example, gender.
- Completely randomized design: Treatments are assigned randomly to all experimental units
- Randomize block design: randomized within each block
- Matched-pair design: Both treatments are applied within each block, with each experimental unit receiving only one treatment.

1.8 Terms and Concepts in Experiments

- Response (dependant) variable: Variable to be measured in the experiment
- Explanatory (independent) variable: A variable that may explain the differences in responses
- Experimental unit: The smallest unit of the population to which a treatment is applied.
- Confounding variable: A variable whose effect on the response cannot be separated from the effect of the explanatory variable
- Factor: A variable whose effect on the response is of interest in the experiment
- Levels: The values of a factor used in the experiment

2 Simpson's Paradox

Example

Bob claims that there has been discriminations between males and females in admissions at his professional school for the last 5 years. The data are as below.

Admissions at Bob's Professional School (2005 - 2009)

		Law School	Medical School
Male	Number of admissions	64(80%)	128(40%)
	Number of applicants	80	320
Female	Number of admissions	143(65%)	20(25%)
	Number of applicants	220	80

Do you think Bob is right?

3 Combining Independent Random Variables

Let X and Y be independent random variables such that $\mathrm{E}(X) = \mu_X$, $\mathrm{E}(Y) = \mu_Y$, $\mathrm{Var}(X) = \sigma_X^2$, and $\mathrm{Var}(Y) = \sigma_Y^2$. If a and b are constants, then

Random variable	Mean	Variance	SD
aX	$a\mu_X$	$a^2\sigma_X^2$	$a\sigma_X$
$X + b$	$\mu_X + b$	σ_X^2	σ_X
$aX + b$	$a\mu_X + b$	$a^2\sigma_X^2$	$a\sigma_X$
$X + Y$	$\mu_X + \mu_Y$	$\sigma_X^2 + \sigma_Y^2$	$\sqrt{\sigma_X^2 + \sigma_Y^2}$
$X - Y$	$\mu_X - \mu_Y$	$\sigma_X^2 + \sigma_Y^2$	$\sqrt{\sigma_X^2 + \sigma_Y^2}$
$aX + bY$	$a\mu_X + b\mu_Y$	$a^2\sigma_X^2 + b^2\sigma_Y^2$	$\sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}$

Let $X_1, X_2, \cdots, X_n$ are independent random variables such that $\mathrm{E}(X_i) = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$ for $i = 1, 2, \cdots, n$. If a_i is a constant for $i = 1, 2, \cdots, n$, then

$$\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} a_i \mu_i, \\
\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} a_i^2 \sigma_i^2, \\
\mathrm{SD}\left(\sum_{i=1}^{n} a_i X_i\right) &= \sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2}.
\end{aligned}$$