# UCLA STAT 110A
## Applied Statistics

- **Instructor**: **Ivo Dinov,**
  **Asst. Prof. In Statistics and Neurology**
- **Teaching Assistants:** **Helen Hu, UCLA Statistics**

**University of California, Los Angeles, Spring 2002**

***http://www.stat.ucla.edu/~dinov/***



## Chapter 3: Design of Experiments and Data collection

- Designs
- Populations
- Sampling, sampling & non-sampling errors
- Bias and precision
- Blocking
- Confounding
- Controlled experiments



## Types of variates (variables)
(variate =data, variable = model)

**We distinguish between two broad types of variables: qualitative and quantitative (or numeric). Each is broken down into two sub-types: qualitative data can be ordinal or nominal, and numeric data can be discrete (often, integer) or continuous.**

**Qualitative data always have a limited number of alternative values, such variables are also described as discrete. All qualitative data are discrete, while some numeric data are discrete and some are continuous.**

**For statistical analysis, quantitative data can be converted into discrete numeric data by simply counting the different values that appear.**



## Controlled Experiments

When a new drug is introduced its effectiveness needs to be evaluated. The basic method is comparison. Drug is administered to subjects in a treatment group and a second groups of subjects are used as controls (two groups should be randomly chosen).

Most of these experiments are carried as double-blind designs – neither the subjects taking the medicine nor the physicians who measure the response know which subject is in which group – to avoid biasing of the observed data.

Note: treatment and control groups need to be as similar (demographically) as possible, except for treatment.



## Controlled Experiments

If the two groups differ in some factor, other than the treatment, we get this other factor possibly effecting the outcome of the study - this is called confounding effect. This should be avoided.

Example: Some disease fall more heavily on the poor. Hence, if a study tests the efficacy of a disease of hygiene (say) we'd need to match the subjects in the two groups for their socio-economic status (income), age, perhaps education level etc. to avoid the effects of these factors crippling into the results of the test.

Most often we use: randomized, controlled, double-blind studies

- reduces the bias to a minimum.



## Controlled Experiments

Randomized, controlled, double-blind studies

are very hard to do, however. As a result sometimes we need to use designs that are not so perfect, but are more economical. Examples – using historical control groups.

Placebo groups: groups of subjects (patients) who receive fake treatment, sugar-pill, (not no-treatment, as in the treatment-control design). This design factors out the implicit psychological effects of been treated.

## Randomization, Replication and Blocking

The use of chance to allocate experimental units into groups is called randomization. Randomization is the major principle of the statistical *design of experiments*.

Randomization produces groups of experimental units that are more likely to be similar in all respects before the treatments are applied than using non-random methods. At the end of the study if the differences in the outcome variable between the two groups is too large to attribute to chance, then the difference is called statistically significant. The decision about how large a difference is required to be ***significant*** depends on statistical inference using the laws of probability. This will be discussed in later sections.



## Randomization, Replication and Blocking

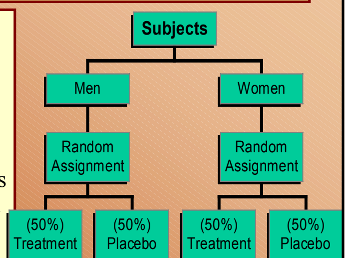
Another principle is that experiments with more subjects are more likely to detect differences than those with fewer subjects. Repeating an experiment on many subjects (or over many times) is called replication and increases the power of a statistical test.

If it is known, before the experiment is carried out, that other variables of no interest influence the outcome (e.g. age or sex of a patient), then randomization can be carried out within subsets of experimental units defined by these variables. This is called a block design.



## Controlled Experiments - blocking

For example, the response to treatment for a given type of cancer is *expected* to depend on the sex of the patient. Ideally, equal numbers of males or females are required in the control or treatment groups and this can be achieved by randomizing 50% of males to the treatment group as shown below.




Compare separately the treatment and placebo groups within individual blocks



## Errors in Samples …

- Selection bias: Sampled population is not a representative subgroup of the population really investigated.
- Non-response bias: If a particular subgroup of the population studied does not respond, the resulting responses may be skewed.
- Question effects: Survey questions may be slanted or loaded to influence the result of the sampling.
- Is quota sampling reliable? Each interviewer is assigned a fixed quota of subjects (subjects district, sex, age, income exactly specified, so investigator can select those people as they liked).
- *Target population* –entire group of individuals, objects, units we study.
- *Study population* –a subset of the target population containing all "units" which could possibly be used in the study.
- *Sampling protocol – procedure used to select* ***the sample***
  - Read the **Methods** section on sampling protocol online.
- *Sample* – the subset of "units" about which we actually collect info.



## More terminology …

- ***Census – attempt to sample the entire population***
- ***Parameter – numerical characteristic of the population, e.g., income, age, etc. Often we want to estimate population parameters.***
- ***Statistic – a numerical characteristic of the sample. (Sample) statistic is used to estimate a corresponding population parameter.***
- Why do we sample at random? We draw "units" from the study population at random to avoid bias. Every subject in the study sample is equally likely to be selected. Also random-sampling allows us to calculate the likely size of the error in our sample estimates.

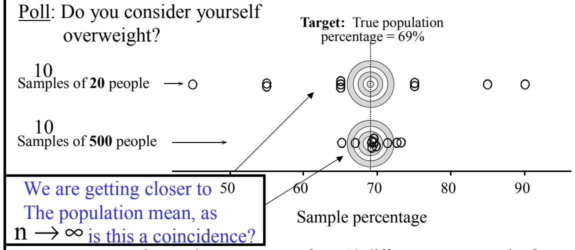


## More definitions …

- How could you implement the lottery method to randomly sample 10 students from a class of 250? – list all names; assign numbers 1,2,3,…,250 to all students; Use a random-number generator to choose (10-times) a number in range [0;250]; Process students drawn.
- Random or chance error is the difference between the sample-value and the true population-value (e.g., 49% vs. 69%, in the following body-overweight example).
- Non-sampling errors (e.g., non-response bias, technical difficulties, misinterpretations) in the census may be considerably larger than in a comparable survey, since surveys are much smaller operations and easier to control.
- Sampling errors–arising from a decision to use a sample rather than entire population
- ***Unbiased procedure/protocol:*** (e.g., using the proportion of left-handers from a random sample to estimate the corresponding proportion in the population).
- **Cluster sampling**- a cluster of individuals/units are used as a sampling unit, rather than individuals.

## Variation in sample percentages

Poll: Do you consider yourself overweight?




We are getting closer to The population mean, as $n \to \infty$ is this a coincidence?

Comparing percentages from 10 different surveys each of 20 people with those from 10 surveys each of 500 people (all surveys from same population).

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.



## Questions …

- Give an example where non-representative information from a survey may be useful. Non-representative info from surveys may be used to estimate parameters of the actual sub-population which is represented by the sample. E.g., Only about 2% of dissatisfied customers complain (most just avoid using the services), these are the most-vocal reps. So, we can not make valid conclusions about the stereotype of the dissatisfied customer, but we can use this info to tract down changes in levels of complains over years.
- Why is it important to take a pilot survey?
- Give an example of an unsatisfactory question in a questionnaire. (In a telephone study: What time is it?

  Do we mean Eastern/Central/Mountain/Pacific?)



## Questions …

- Random allocation – randomly assigning treatments to units, leads to representative sample only if we have large # experimental units.
- Completely randomized design- the simplest experimental design, allows comparisons that are unbiased (not necessarily fair). Randomly allocate treatments to all experimental units, so that every treatment is applied to the same number of units. E.g., If we have 12 units and 3 treatments, and we study treatment efficacy, we randomly assign each of the 3 treatments to 4 units exactly.
- Blocking- grouping units into blocks of similar units for making treatment-effect comparisons only within individual groups. E.g., Study of human life expectancy perhaps income is clearly a factor, we can have high- and low-income blocks and compare, say, gender differences within these blocks separately.



## Using chance in surveys

- When a ***sample selection procedure*** is biased taking a larger sample size does not help (just repeats the error on a larger scale). However, if the sample selection is unbiased, then large sample sizes improves the survey/results.
- Non-responders could form a completely different subject category from the responders to a survey. Beware of non-response bias.
- Bad/good samples: Selection bias? Non-response bias?
- Quota sampling: the sample is hand picked to resemble the population on some key characteristic. Seems reasonable, but does not work well in practice because of unintentional bias on part of the interviewers.

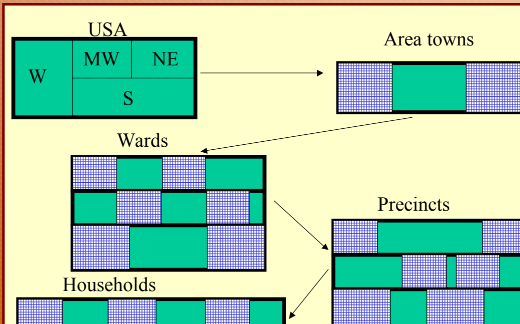


## Use *objective* impartial chance mechanism to select sample units at each stage of the selection protocol.

- An Interviewer should have no discretion as to whom they interview;
- There should be a pre-defined protocol for selecting the sample, prior to surveying, and it should involve the planned use of chance/random selection.



## Use *objective* impartial chance mechanism to select sample units at each stage of the selection protocol.



## The Role of Randomization

Well designed statistical studies employ randomization to avoid subjective and other biases.

- Surveys and observational studies should use random sampling to obtain representative samples.
- Experiments should use random assignment of experimental subjects to treatment groups
  - to ensure comparisons are fair i.e., treatment groups are as similar as possible in every way except for the treatment being used.



## Sources of error in surveys

- Random sampling leads to ***sampling errors***, sampling-size (as we saw for the overweight survey), arising for the choice to use a sample, as opposed to census.
- ***Non-sampling errors*** can be much larger than the sampling errors. Selection bias, non-response bias, survey/question/interview format are all non-sampling errors.



## Dealing with errors

- ***Statistical methods*** are available for estimating the likely size of ***sampling errors***.
- All we can do with ***non-sampling errors*** is to try to minimize them at the study-design stage.
- ***Pilot survey***:
  One tests a survey on a relatively small group of people to try to identify any problems with the survey design before conducting the survey proper.