UCLA STAT 251 / OBEE 216

Statistical Methods for the Life and Health Sciences

- **Instructor**: Ivo Dinov,

Asst. Prof. In Statistics and Neurology

University of California, Los Angeles, Winter 2003

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

UCLA STAT 251 / OBEE 216

- *Course Description*
- *Class homepage*
- *Online supplements, VOH's etc.*
- *ClassQuestionnaire.html*
- *Final Exam/Project Format*
- *Guest Lecturers*

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

UCLA STAT 251 / OBEE 216

to just hear is to forget

to see is to remember

to do it yourself is to understand …

What is Statistics? A practical example

- *Demography: Uncertain population forecasts*

by Nico Keilman, Nature 412, 490 - 491 (2001)

- Traditional population forecasts made by statistical agencies **do not quantify uncertainty**. But demographers and statisticians have developed methods to calculate probabilistic forecasts.
- The demographic future of any human population is uncertain, but some of the many possible trajectories are more probable than others. So, forecast demographics of a population, e.g., size by 2100, should include two elements: a range of possible outcomes, and a probability attached to that range.

What is Statistics?

- Together, ranges/probabilities constitute a *prediction interval* for the population. There are trade-offs between greater certainty (higher odds) and better precision (narrower intervals). Why?
- For instance, the next table shows an estimate that the odds are 4 to 1 (an 80% chance) that the world's population, now at 6.1 billion, will be in the range [5.6 : 12.1] billion in the year 2100. Odds of 19 to 1 (a 95% chance) result in a **wider** interval: [4.3 : 14.4] billion.

Table 1 Forecasted population sizes and proportions over age 60

	Median world and regional population sizes (millions)				
Year	2000	2025	2050	2075	2100
World total	6,055	7,827 (7,219–8,459)	8,797 (7,347–10,443)	8,951 (6,636–11,652)	8,414 (5,577–12,123)
North Africa	173	257 (228–285)	311 (249–378)	336 (238–443)	333 (215–484)
Sub-Saharan Africa	611	976 (856–1,100)	1,319 (1,010–1,701)	1,522 (1,021–2,194)	1,500 (878–2,450)
North America	314	379 (351–410)	422 (358–498)	441 (343–565)	454 (313–631)
Latin America	515	709 (643–775)	840 (679–1,005)	904 (647–1,202)	934 (585–1,383)
Central Asia	56	81 (73–90)	100 (80–121)	107 (76–145)	106 (66–159)
Middle East	172	285 (252–318)	368 (301–445)	413 (296–544)	413 (259–597)
South Asia	1,367	1,940 (1,735–2,154)	2,249 (1,795–2,776)	2,242 (1,528–3,085)	1,958 (1,186–3,035)
China region	1,408	1,608 (1,494–1,714)	1,580 (1,305–1,849)	1,422 (1,003–1,884)	1,250 (765–1,870)
Pacific Asia	476	625 (569–682)	702 (575–842)	702 (509–937)	654 (410–949)
Pacific OECD	150	155 (144–165)	148 (125–174)	135 (100–175)	123 (79–173)
Western Europe	456	478 (445–508)	470 (399–549)	433 (321–562)	392 (257–568)
Eastern Europe	121	117 (109–125)	104 (86–124)	87 (61–118)	74 (44–115)
European part of the former USSR	236	218 (203–234)	187 (154–225)	159 (110–216)	141 (85–218)

80 per cent prediction intervals are shown in parentheses.

Table 1 Forecasted population sizes and proportions over age 60

	Median world and regional pop		
Year	2000	2025	2050
World total	6,055	7,827 (7,219–8,459)	8,797 (7,347–10,443)
North Africa	173	257 (228–285)	311 (249–378)
Sub-Saharan Africa	611	976 (856–1,100)	1,319 (1,010–1,701)
North America	314	379 (351–410)	422 (358–498)
Latin America	515	709 (643–775)	840 (679–1,005)
Central Asia	56	81 (73–90)	100 (80–121)
Middle East	172	285 (252–318)	368 (301–445)
South Asia	1,367	1,940 (1,735–2,154)	2,249 (1,795–2,776)
China region	1,408	1,608	1,580

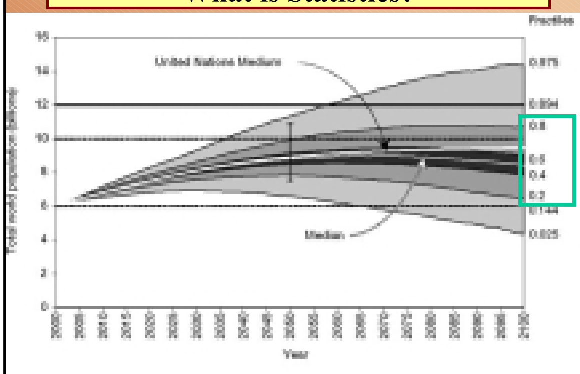
What is Statistics?

- *Demography: Uncertain population forecasts*

  by Nico Keilman, Nature 412, ,2001
- Traditional population forecasts made by statistical agencies **do not quantify uncertainty**. But lately demographers and statisticians have developed methods to calculate probabilistic forecasts.
- Proportion of population over 60yrs.

Proportion of population over age 60		
2000	2050	2100
0.10	0.22 (0.18–0.27)	0.34 (0.25–0.44)
0.06	0.19 (0.15–0.25)	0.32 (0.23–0.44)
0.05	0.07 (0.05–0.09)	0.20 (0.14–0.27)
0.16	0.30 (0.23–0.37)	0.40 (0.28–0.52)
0.08	0.22 (0.17–0.28)	0.33 (0.23–0.45)
0.08	0.20 (0.15–0.25)	0.34 (0.24–0.46)
0.06	0.18 (0.14–0.23)	0.35 (0.24–0.47)
0.07	0.18 (0.14–0.24)	0.35 (0.25–0.48)
0.10	0.30 (0.24–0.37)	0.39 (0.27–0.53)
0.08	0.23 (0.18–0.29)	0.36 (0.26–0.49)
0.22	0.39 (0.32–0.47)	0.49 (0.35–0.61)
0.20	0.35 (0.29–0.43)	0.45 (0.32–0.58)
0.18	0.38 (0.30–0.46)	0.42 (0.28–0.57)
0.19	0.35 (0.27–0.44)	0.36 (0.23–0.50)

What is Statistics?



First polio epidemic hits the U.S. 1916

- Claims hundreds of thousands of victims, mostly children!
- 1950s - several vaccines discovered, one by Jonas Salk proved safe in the lab (caused the production of antibodies against polio).
- Large-scale field trial needed to establish effectiveness outside the lab.
- 1954 Public Health Service organizes an experiment in which the subjects are children in the most vulnerable age groups, grades 1-3.
- Q: How do they assess the effectiveness?
- Q: Give the vaccine to a group of kids and compare to 1953?
- A: No. The incidence varies a lot from year to year.
- A method of comparison is needed: control group - receives placebo treatment group-receives vaccine.

First polio epidemic hits the U.S. 1916

- Compare some response (polio/no polio) among the two groups. Why a placebo? So that differences in the responses between the two groups can be attributed only to the actual treatment (vaccine) rather than the idea of treatment. Placebo effects have been shown to have substantially influence the results for some problems, such as pain relief.
- In order to eliminate other unforeseeable differences between the groups which may affect the response, called confounding factors, the subjects are randomly assigned to the two groups. For the same reason, it is best if the experiment is double-blind; neither the subjects nor the evaluators know who is in the treatment/control group.
- What happened: children could only be vaccinated with parental permission. Among the 400,000 children whose parents gave permission, half were randomized to the control group, half to the treatment group with the following results:

First polio epidemic hits the U.S. 1916

W/ Consent

Group	Size	(cases/100,000)
Treatment	200,000	28
Controls	200,000	71
No Consent	350,000	46
Grade 2 vaccine	225,000	25
Grades 1,3(control)	750,000	54
Grade 2 no consent	125,000	44

NFIP design

- Conclusion: Estimated effect of polio vaccine for children in grades 1-3 (3 strains of inactivated virus) with parents consent is rate = 14 cases/100,000 vs. 35 of controls. It can be shown that this figure is *statistically significant*.

First polio epidemic hits the U.S. 1916

- Another design that was used: NFIP - **National Foundation for Infantile Paralysis (1938 to 1960s)**.
- Treatment Group = Grade 2 with no consent Control Group = Grades 1 and 3. Conclusion: estimated effect is Change Rate = 29 cases/100,000. But this estimate is *biased* by parental consent, which is a confounding factor.
- Q: Why is consent a confounding factor?
- Q: Are children with parental consent really more susceptible to polio than children without?
- Would you believe that children from households with less income are less susceptible? That children from less hygienic surroundings are less susceptible?
- David Freedman, Robert Pisani, Roger Purves, and Ani Adhikari, *Statistics*, Second Edition (New York: W. W. Norton & Co., 1991), Table 1, p. 6. After Thomas Francis, Jr., *American Journal of Public Health* vol. 45 (1955), pp. 1-63

First polio epidemic hits the U.S. 1916

- It turns out that these children (from less hygienic surroundings) are more likely to contract polio early in childhood while still protected by antibodies from their mothers. After infection, they generate their own antibodies which protect them later.
- The NFIP study is called an observational study.
- Key components of a designed experiment:
- Randomization
- Use of placebo where possible
- Use of double-blinding where possible

Berkeley Admissions Data

- 1973, Fall quarter - Admissions by gender to the Graduate Division at the University of California, Berkeley.

Gender	Admit	Deny	Total
Men	3738(44%)	4704	8442
Women	1494(35%)	2827	4321

- Fact: more men are being admitted than women.
- Q: Is there a gender bias?
- Note that it is impossible to randomize subjects (students) to treatment (gender). This study is necessarily an observational study.

Berkeley Admissions Data – Simpson's Effect

- We are unable to conclude that gender causes a lower rate of admission though there is clearly an association between gender and rate of admission. There may be other factors that we haven't controlled for. In general, this is true of most observational studies.
- Data: Admission rates by gender for 6 largest majors

	Men		Women	
Major	Applicants	% admitted	Applicants	% admitted
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	393	6	341	7

Berkeley Admissions Data

- **Majors A & B** have much higher rates of admission than C,D,E, or F and more than half of the men considered here applied to major A or B.
- **Majors C, D, E & F** have lower rates of admission. 90% of the women applied to one of these majors.

What is Statistics?

- There is concern about the accuracy of population forecasts, in part because the rapid fall in fertility in Western countries in the 1970s came as a surprise. Forecasts made in those years predicted birth rates that were up to 80% too high.
- The rapid reduction in mortality after the Second World War was also not foreseen; life-expectancy forecasts were too low by 1–2 years; and the predicted number of elderly, particularly the oldest people, was far too low.

What is Statistics?

- So, during the 1990s, researchers developed methods for making probabilistic population forecasts, the **aim** of which is to calculate prediction intervals for every variable of interest. Examples include population forecasts for the USA, AU, DE, FIN and the Netherlands; these forecasts comprised prediction intervals for variables such as age structure, average number of children per woman, immigration flow, disease epidemics.
- We need accurate probabilistic population forecasts for the whole world, and its 13 large division regions (see Table). The conclusion is that there is an estimated 85% chance that the world's population will stop growing before 2100. Accurate?

What is Statistics?

- There are three main methods of probabilistic forecasting:
  - time-series extrapolation;
  - expert judgement; and
  - extrapolation of historical forecast errors.
- Time-series methods rely on statistical models that are fitted to historical data. These methods, however, seldom give an accurate description of the past. If many of the historical facts remain unexplained, time-series methods result in excessively wide prediction intervals when used for long-term forecasting.
- Expert judgement is subjective, and historic-extrapolation alone may be near-sighted.

Preliminaries: What is Statistics?

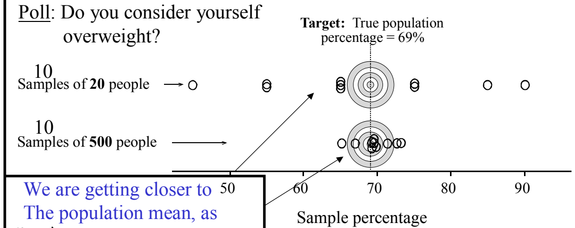
- Polls and surveys – we're all different; It's impossible or expensive to investigate every single person.
- Experimentation – sample vs. population
- Observational Studies – selection and non-response bias
- Statistics -- What is it and who uses it?

Newtonial science vs. chaotic science

- Article by Robert May, Nature, vol. 411, June 21, 2001
  - Science we encounter at schools deals with crisp certainties (e.g., prediction of planetary orbits, the periodic table as a descriptor of all elements, equations describing area, volume, velocity, position, etc.)
  - As soon as uncertainty comes in the picture it shakes the foundation of the deterministic science, because only probabilistic statements can be made in describing a phenomenon (e.g., roulette wheels, chaotic dynamic weather predictions, Geiger counter, earthquakes, etc.)
  - What is then science all about – describing absolutely certain events and laws alone, or describing more general phenomena in terms of their behavior and chance of occurring? Or may be both!

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?

Figure 1.1.1 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.

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***
- *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 above body-overweight example).
- Non-sampling errors (e.g., non-response bias) 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.

More terminology …

- What are some of the ***non-sampling errors*** that plague surveys? (non-response bias, question effects, survey format effects, interviewer effects)
- If we take a random sample from one population, can we apply the results of our survey to other populations? (It depends on how similar, in the respect studied, the two populations are. In general- No! This can be a dangerous trend.)
- Are sampling households at random and interviewing people at random on the street valid ways of sampling people from an urban population? (No, since clusters (households) may not be urban in their majority.)
- Pilot surveys – after prelim investigations and designing the trial survey Q's, we need to get a "small sample" checking clearness and ambiguity of the questions, and avoid possible sampling errors (e.g., bias).

Review

- Variations in samples
- Census
- Population parameters
- (sample) Statistics
- Sampling errors (e.g., selection bias, resulting from use of sample)
- Non-sampling errors (e.g., non-response, Q-effects)

Questions …

- How do the following lead to biases or cause differences in response:
  - ***non-response***
  - ***self-selection***
  - ***question effects***
  - ***survey-format effects***
  - ***interviewer effects***
  - ***transferring findings***?

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.

Questions …

- Why should we try to "blind" the investigator in an experiment?
- Why should we try to "blind" human experimental subjects?
- The basic rule of experimentor :

"Block *what you can and randomize what you cannot.*"

Experiments vs. observational studies for comparing the effects of treatments

- In an Experiment
  - experimenter determines which units receive which treatments. (ideally using some form of random allocation)
- Observational study – useful when can't design a controlled randomized study
  - compare units that happen to have received each of the treatments
  - Ideal for describing relationships between different characteristics in a population.
  - often useful for identifying possible causes of effects, but cannot reliably establish causation.
- Only properly designed and executed experiments can reliably demonstrate causation.

Questions …

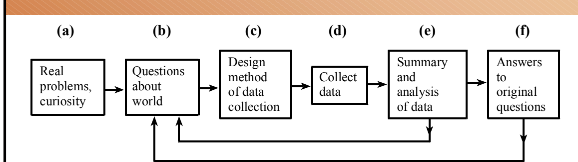
- What is the difference between a designed experiment and an observational study? (no control of the design in observational studies)
- Can you conclude causation from an observational study? Why or why not? (not in general!)
- How do we try to investigate causation questions using observational studies? In a smoking-lung-cancer study: try to divide all subjects, in the obs. study, into groups with equal, or very similar levels of all other factors (age, stress, income, etc.) – I.e. control for all outside factors. If rate of lung-cancer is still still higher in smokers we get a stronger evidence of causality.
- What is the idea of controlling for a variable, and why is it used? Effects of this variable in the treatment/control groups are similar.
- Epidemiology – science of using statistical methods to find causes or risk factors for diseases.

The Subject of Statistics

Statistics is concerned with the process of finding out about the world and how it operates -

- in the face of variation and uncertainty
- by collecting and then making sense (interpreting) of data.

The investigative process



Poll Example

- This is only a 10% response rate - the people who responded could be very unrepresentative. It could well be that the survey struck a responsive chord with stressed-out principals.

Experimental vs. Observation study

- A researcher wants to evaluate IQ levels are related to person's height. <u>100 people</u> are are randomly selected and grouped into <u>5 bins</u>: [0:50), [50;100), [100:150], [150:200), [200:250] *cm* in height. The subjects undertook a IQ exam and the results are analyzed.
- Another researcher wants to assess the bleaching effects of <u>10 laundry detergents</u> on <u>3 different colors</u> (R,G,B). The laundry detergents are randomly selected and applied to 10 pieces of cloth. The discoloration is finally evaluated.

Experimental vs. Observation study

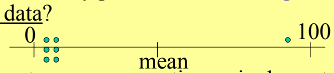
- For each study, describe what *treatment* is being compared and what *response* is being measured to compare the treatments.
- Which of the studies would be described as *experiments* and which would be described as *observational* studies?
- For the studies that are observational, could an experiment have been carried out instead? If not, briefly explain why not.
- For the studies that are experiments, briefly discuss what *forms of blinding* would be possible to be used.
- In which of the studies has *blocking* been used? Briefly describe *what* was blocked and why it was blocked.

Experimental vs. Observation study

- What is the *treatment* and what is the *response?*
  1. <u>Treatment</u> is height (as a bin). <u>Response</u> is IQ score.
  2. <u>Treatment</u> is laundry detergent. <u>Response</u> is discoloration.
- *Experiment* or *observational* study?
  1. Observational – compare obs's (IQ) which happen to have the treatment (height).
  2. Experimental – experimenter controls which treatment is applied to which unit.
- For the <u>observational</u> studies, can we conduct an experiment?
  1. This could not be done as an experiment - it would require the experimenter to decide the (natural) height (treatment) of the subjects (units).
- For the <u>experiments</u>, is there *blinding?*
  2. The only form of blinding possible would be for the technicians measuring the cloth discoloration not to know which detergent was applied.
- Is there *blocking*?
  1. & 2. No blocking. Say, if there are two laundry machines with different cycles of operation and if we want to block we'll need to randomize which laundry does which cloth/detergent combinations, because differences in laundry cycles are a known source of variation.

Mean, Median, Mode, Quartiles, 5# summary

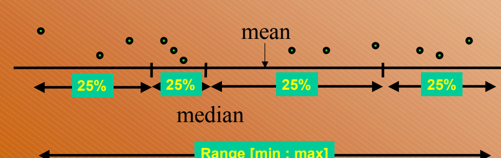
- The sample mean is the average of all numeric obs's.
- The sample median is the obs. at the index $(n+1)/2$ (note take avg of the 2 obs's in the middle for fractions like 23.5), of the observations ordered by size (small-to-large)?
- The sample median usually preferred to the sample mean for <u>skewed data</u>?



- Under what circumstances may quoting a <u>single center</u> (be it mean or median) not make sense?(multi-modal)
- What can we say about the sample mean of a qualitative variable? (meaningless)

Quartiles

The first quartile (Q_1) is the median of all the observations whose *position* is <u>strictly below the *position* of the median</u>, and the third quartile (Q_3) is the median of those above.



Five number summary

The five-number summery = (Min, Q_1, Med, Q_3, Max)