Probability, Statistics, Error Analysis and Risk Assessment
An example online poll

- Rate my professor [link]

- What can we learn from this poll?
Types of surveys

- **Voluntary surveys**
  - Sample likely biased (motivation to participate)
  - Don’t measure true population

- **Random sampling**
  - Truly select a random (representative) group

- **“Predictive” polling**
  - Election polls must sample, and then attempt to predict behavior.
  - Public opinion does not necessarily match an election outcome.

- **Push polling**
  - A “fake” poll to get your attention.
Americans' Satisfaction With the Way the Nation Is Being Governed

On the whole, would you say you are satisfied or dissatisfied with the way the nation is being governed?

- % Satisfied
- % Dissatisfied

GALLUP
Pew says: “In addition to sampling error, one should bear in mind that question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls.”

**Questions:**
- Are these poll results consistent with one another?
- What is the “uncertainty” in those numbers?
- What does “margin of error” mean?
- Is the uncertainty statistical? Systematic? Both?
Statistical/sampling Uncertainties

- Normal or Gaussian distribution
  - Mean (average) is $\mu$
  - Statistical uncertainty is quantified by $\sigma$
  - Bigger sample size $\rightarrow$ smaller statistical uncertainty

IQ | 55 | 70 | 85 | 100 | 115 | 130 | 145
---|----|----|----|-----|-----|-----|-----

- 68% of people have IQ between 85 and 115
- 95% of people have IQ between 70 and 130
Margin of Error

- Say a poll says Romney 47% with margin of error 3%
  - Margin of error is the $2\sigma$ number.
  - So 95% of all polls should be within the margin of error.

Romney at 47% with a 3% margin of error means that 95 polls out of 100 will have Romney between 44% and 50%
A fudged poll

- **Gallup tracking**
  - Expect week to week variations due to sampling.

- **Research 2000 tracking.**
  - Does this look like normal variation?
Comments

- **Margin of error is a “statistical” or “sampling” uncertainty.**
  - The bigger the size of the poll (or experiment) the smaller the statistical uncertainty.
  - Statistical uncertainty does not account for bias in the questions or bias in the sample. These are called “systematic” uncertainties.
  - Response rate for polls is 10-15%. To survey 1000 voters, need to call 10,000.

- **Types of systematic uncertainties**
  - Sample composition.
  - Likely versus registered voters
  - Wording of questions
  - Cell phone versus land line.
    - Gallup: Within each contacted household reached via landline, an interview is sought with an adult 18 years of age or older living in the household who has had the most recent birthday. (This is a method pollsters commonly use to make a random selection within households without having to ask the respondent to provide a complete roster of adults living in the household.) Gallup does not use the same respondent selection procedure when making calls to cell phones because they are typically associated with one individual rather than shared among several members of a household.
  - Pew says: “In addition to sampling error, one should bear in mind that question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls.”

- **Combining polls**
  - Effectively makes for a larger sample size which reduces the statistical uncertainty.
  - But what if two polls have different systematic uncertainties? Is it ok to combine them?
Experiment

- Take a piece of paper.

- “Fake” an experiment where you flip a coin 10 times, mark H for heads and T for tails.
  - Don’t flip the coin, just write down an example of a fake experiment.

- Now, do the real experiment: flip a coin 10 times, mark H for heads and T for tails.
  - Does your fake experiment look like your real experiment?
Flipping Coins

- I’m willing to bet that at least one person had either 5 heads or 5 tails in a row to start their list.
  - Am I making a good bet? Will you bet me?
  - On average, 1 out of every 16 people will flip either 5 heads or 5 tails in a row to start their list.
  - I’m making a good bet, since there are ~30 people in this room, it is very likely that somebody (maybe 2 people) will flip 5 in a row.
    - It’s also possible that nobody will. Although it’s likely that somebody (or more than one person will flip 5 in a row) there is no guarantee that anybody will.
    - I’m making a good bet, but it’s not a sure bet.

- Each flip has a 50/50 chance of coming up heads.
  - Each new flip is totally independent of the previous flip.

- Q: If you flip 5 heads in a row, are you “due” to flip a tail on the next try?
Risk

- With your neighbor, try to think of some examples of places where you evaluate risk on a daily basis.

- What is involved in risk assessment?
Risk Assessment Quiz

1. Which should you worry more about: avian flu or the common flu?

2. Which should you worry more about: mad cow pathogens in your hamburger or the cholesterol in your hamburger?

3. Which should you fear more: AIDS or a heart attack?

4. Which should you fear more: cigarettes or terrorism?

5. Should you fly or drive?
Risk Assessment Quiz

1. Which should you worry more about: avian flu or the common flu?
2. Which should you worry more about: mad cow pathogens in your hamburger or the cholesterol in your hamburger?
3. Which should you fear more: AIDS or a heart attack?
4. Which should you fear more: cigarettes or terrorism?
5. Should you fly or drive?

How do these risks compare to our "knee jerk" concerns??
Risk Assessment Quiz

1. Which should you worry more about: avian flu or the common flu?
   • 1,000-17,000 Americans died from common flu each year, 0 from avian flu

2. Which should you worry more about: mad cow pathogens in your hamburger or the cholesterol in your hamburger?
   • 700,000 Americans died of heart disease last year, 0 from mad cow.

3. Which should you fear more: AIDS or a heart attack?
   • 25,000 people in North America died of AIDS in 2008. 910,000 people died from heart disease in 2006.

4. Which should you fear more: cigarettes or terrorism?
   • 3000 people died in the World Trade Centers on 9/11/2001. 3000 people die from smoking related illnesses every two days.

5. Should you fly or drive?
   • 44,000 automobile deaths last year, a few hundred died in airplane crashes.
Lottery

- Illinois Mega-Millions
  - Must match 5 numbers drawn from 1-52
  - Odds of winning are 1 in 135,145,920 which is about 1 in 100 million or about 1 in $10^8$.

- There are about 300 million people in the US
  - About 100 people per year get hit by lightning
  - This means that your chances of getting hit by lightning this year is about 1 in 3 million.
  - In other words, you are 30 times more likely to get hit by lightning than to win the lottery on a single set of numbers.

- This analogy is not really accurate. What's wrong with it?
What are the odds?

- Two seemingly uncorrelated things occur more than once.
  - I thought of my friend and a specific song came on the radio.
- What are the odds that this occurred by chance?

- Are you an unbiased observer?
  - Have you carefully noted every time you thought of your friend?
  - Is it possible that you only remember the hits and forget the misses?
Vegas is Crooked

- “You can’t win in Vegas because all of the games are rigged.”
  - **Wrong.** Why would you rig a game and risk going to jail when you can make money hand over foot legally?
- All casino games have odds that are in favor of the “house”
  - One person can come along and win money. Just like it was possible that nobody flipped four heads (or tails) in a row.
  - But when many, many people gamble...on average, they are going to lose...and that’s how casinos make money.
- **Craps**
  - Probability of rolling a 7 is more likely than any other number.
- **Blackjack**
  - If you bust, you lose your money before the dealer has a chance to bust.
- **Roulette**
  - Numbers 0 and 00 tilt the odds in favor of the house.
Smoking Causes Cancer

- "Thus the argument that since there are heavy smokers who do not have lung cancer (and, of course, the majority do not) and because there are some rare cases of non-smokers who do have lung cancer then smoking does not cause lung cancer, is totally fallacious. ....Thus for male smokers in the U.S., the U.S.A. and several other countries from the epidemiological evidence alone it can be concluded that smoking cigarettes causes lung cancer and some other respiratory diseases."
  - Dr. S.J. Green, Brown and Williamson Tobacco Company
  - Date unknown, but probably in the 1960’s.
Smoking and Cancer

- Is a smoker guaranteed to get cancer? No.

- If a smoker gets cancer, does that mean it was caused by smoking? Not necessarily.

- If a person smokes, are they more likely to get cancer? Yes.
  - How do we know? Many studies identifying both the correlation and causation of tobacco and cancer.
Risk

• To weigh the risks of anything, you need to assess:
  - The probability of a specific outcome
  - The consequences of that outcome.

• ... Wall Street failed to properly assess the probability of failure of the subprime mortgage market and also failed to properly assess the consequences...

• In general, humans do pretty well at assessing risk when the probability of an outcome is pretty high. We do poorly when the probability of an outcome is very low.
Pattern Recognition
Face of Satan?

- Photographer Mark D. Phillips (unmodified)
Et = 37.17 GeV
Astrology
Wikipedia I

- Astrology is a group of systems, traditions, and beliefs which hold that the relative positions of celestial bodies and related details can provide information about personality, human affairs, and other terrestrial matters. A practitioner of astrology is called an astrologer. Astrologers believe that the movements and positions of celestial bodies either directly influence life on Earth or correspond to events experienced on a human scale. Modern astrologers define astrology as a symbolic language, an art form, or a form of divination. Despite differences in definitions, a common assumption of astrologers is that celestial placements can aid in the interpretation of past and present events, and in the prediction of the future.
Astrology is generally considered a **pseudoscience** or **superstition** by the **scientific community**, which cites a lack of **statistically significant** astrological predictions, while **psychology** explains much of the continued faith in astrology as a matter of **cognitive biases**. In 2006 the **U.S. National Science Board** published a statement which said it considers belief in ten survey items, astrology among them, to be "**pseudoscientific**".
Can planet position matter?

- The force on you from mars is at most: 0.000000005 pounds (usually it’s less)

- This is the same force exerted by the gravitational field of a person standing near you.

- How can this affect your personality?
"The most incomprehensible thing about the world is that it is comprehensible."
Summary

• Statistics and Probabilities are all around us.
  – You evaluate probabilities every day.

• When you are given “measured” information (studies, polls) be careful to evaluate the uncertainties on those measurements.
  – If the experimenter doesn’t provide uncertainties, then you cannot put any stock in the measurement.
  – If the experimenter provides statistical uncertainty but not systematic (e.g. polling) be careful in evaluating the data.

• In studies of human health, it is very difficult to identify cause and effect in a single person.
  – But it is possible to identify cause and effect in samples of people.
  – Your behavior has consequences, although the consequences are not known in advance on a case by case basis.