Lesson Plan

- Go over Diagnostic Quizzes.
- Designed Experiments
- Observational Studies.
- Preliminary
Preliminaries

Inference

Probability, chance, variability, Modeling

Relationships

Description or Summary

Design

Statistics today - Complex processes

Statistics - "Inference in the face of uncertainty"
Chapters 1 and 2 contrast designed experiments (DES) with observational studies (OSS).
1.1 Designed Experiments

The gold standard for a statistical study is the double-blind, randomized, controlled experiment. A study is double-blind if neither the subjects nor the scientists know who is assigned to which group until after the data are collected. This prevents scientists from introducing any unconscious bias into the data collection process.

A study is controlled if one group receives the treatment and another group does not. (In medicine, that group usually gets either a placebo, or standard medical care, or both.) A study is controlled experimentation.

Prevents subjects in different groups from behaving in different ways.

The gold standard for a statistical study is the double-blind, randomized controlled experimentation.
A study is randomized if the control group and the treatment group are chosen at random.

current patients get better nursing care (a confounding factor).

A study is randomized if the control group and the treatment group are chosen at random.

Withou...
Consider the Salk vaccine trials.

- Why did randomization make a difference? What other assignment strategies were considered, and how might they have affected the results?
- Why did blinding make a difference? Thatis, how might children have behaved differently if they had known whether or not they had been vaccinated? How might physicians have behaved differently if they had known whether or not they had been vaccinated?
- Why did controlling make a difference? What would be an alternative, and why is that less definitive?
- Why did randomization make a difference? That is, how might children have behaved differently if they had known whether or not they had been vaccinated?
Jonas Salk was depressed for years after the trial. The plan was to give the vaccine at random to half the children in the study. But later he realized that they could have given the vaccine to 90% of the children in the study, rather than 50%.

How would you go about designing an experiment to test the validity of astrology? To make this concrete, assume you have a cooperative astrologer who believes he can use information on birthdays to describe people's personalities very accurately.

How would you go about designing an experiment to test the validity of astrology? To make this concrete, assume you have a cooperative astrologer who believes he can use information on birthdays to describe people's personalities very accurately.

Of course, one might argue that the 50-50 split reflected a concern that the vaccine would harm children rather than protect them. But there had already been substantial Phase I clinical trials that showed that the vaccine was safe, so the only real question was whether or not it was effective. The study had still had sufficient sample size to prove its effectiveness. But later he realized that they could have given the vaccine to 90% of the children in the study at random to half the children in the study. But later he realized that they could have given the vaccine to 90% of the children in the study.
One design is:

1. Find a volunteer and get their birthday.
2. Randomly generate a false birthday for that person.
3. Have the astrologer generate horoscopes for both the true and false birthdays, without knowing which is which.
4. Have the volunteer (or his spouse, who may be a better judge) determine which horoscope gives the best description, without knowing which corresponds to the true and false birthdays.
5. Repeat this for about 30 people.

If the astrologer is a fake, then about half of the time the correct birthdate will be chosen. If the astrologer is real, then the correct birthdate will be chosen significantly more than half the time.

Recollect this for about 30 people.
Later in the course we shall learn how to decide what level of success is significant. For instance, 18 successes out of 30 is not surprising, but 28 successes out of 30 is remarkable and provides strong evidence that the astrolger is successful.

Class discussion questions: Explain whether this experiment is:

- Double-blind
- Randomized
- Controlled

If you cannot find 30 volunteers, how could you still perform this experiment with a smaller sample size? Justify your answer.
The portacaval shunt study of a treatment for cirrhosis of the liver is telling. Physicians reported 50 experiments on the procedure in the medical literature (most of these experiments were small, involving only about ten or so patients). Physicians reported 50 experiments on the procedure in the medical literature (most of these experiments were small, involving only about ten or so patients). The following table breaks out the 50 studies according to their design and their conclusions:

<table>
<thead>
<tr>
<th>Degree of Enthusiasm</th>
<th>Randomized, Controlled</th>
<th>Control, Not Randomized</th>
<th>No Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>7</td>
<td>24</td>
</tr>
</tbody>
</table>

The portacaval shunt study of a treatment for cirrhosis of the liver is telling. Physicians reported 50 experiments on the procedure in the medical literature (most of these experiments were small, involving only about ten or so patients).
In an observational study, the researcher does not get to determine who receives the treatment. The tobacco lobby used to say no, arguing that there might be a gene that predisposes people to both enjoy smoking and get cancer. For example, people who smoke get lung cancer at a higher rate than those who do not smoke. Does smoking cause lung cancer? The real cause of lung cancer: (e.g., less exercise, high alcohol consumption), and that may be the people who like to smoke may tend to follow unhealthy lifestyles. People who like to smoke may tend to follow unhealthy lifestyles. Does smoking cause lung cancer? The real cause of lung cancer: (e.g., less exercise, high alcohol consumption), and that may be the people who like to smoke may tend to follow unhealthy lifestyles. People who like to smoke may tend to follow unhealthy lifestyles.
In this case, two possible confounding factors are genes and lifestyle.

Confounding factor.

and non-smokers may be due to smoking, or they may be due to a

weight. The differences between lung cancer rates in the smokers

mammals and birds.

But animal studies strongly indicate that smoking causes lung cancer in

double-blind—people know if they smoke.

And it would be hard to make this

experiment (one would have to assign 14 year-olds at random to smoke

Obviously, it would be unethical to do a randomized controlled
A confounding factor is associated with both outcome group membership.

For example, one might argue that lung cancer is caused by matches, not tobacco. Or one might argue that cholesterol does not cause heart disease, but rather is a result of poor circulation or breakdown of heart muscle tissue.

One way to try to handle confounding is to make subgroup comparisons that control for possible confounding effects. For example, one could compare the lung cancer rates for smokers who use matches against smokers who use lighters.

A confounding factor is associated with both outcome group membership.
Do seatbelts save lives?

Seatbelt studies are usually observational. One compares the fatality rates among seatbelt wearers and non-wearers in similar cars, or cars that are thought designed with safety in mind (e.g., those Corvette characters).

So people try to control for this by comparing the fatality rates among seatbelt wearers and non-wearers in similar cars, or cars thought to have been traveling at the same speed. But this is awkward to do and invites criticism.

But one has to worry about confounding factors. For example, people who don't wear seatbelts may tend to drive faster and cars that are not seatbelts wearers may tend to drive faster.

Seatbelt studies are usually observational (why?). One compares the fatality rates in accidents in which seatbelts were worn to the fatality rate in accidents without seatbelts.

People who don't wear seatbelts may tend to drive cars that are not designed with safety in mind (e.g., those Corvette characters). People who don't wear seatbelts may tend to drive faster.

Doseatbelts savelives?
In order to control for a confounding factor, one has to guess what it is. But that can be hard and you are never sure that you have thought of everything. In order to control for a confounding factor, one has to guess what it is.

But that can be hard and you are never sure that you have thought of everything. In order to control for a confounding factor, one has to guess what it is.

In contrast, with a randomized design, the random assignment of people to the treatment and control groups ensures that there is almost no chance of a systematic difference between the groups. You are unlikely to get most of the safe drivers in one group and the reckless in the other, or most of the people with good genes for lung cancer in one group and all those with bad genes in the other.

Health experts say that exercise increases one’s lifespan. What kinds of data might they have, and what would be the statistical issues regarding the validity of their claim?
a larger proportion of women than men. Overall, Berkeley was rejecting a larger proportion of women than men. This looked like sex bias. But when the deans asked each department to report their admission rates separately, it turned out that each department accepted a larger proportion of women than men. This looked like sex bias. Another way to do this is to use a weighted average.

Recall the Berkeley Graduate admissions data. Overall, Berkeley was rejecting a larger proportion of women than men. Another way to do this is to use a weighted average.
This apparent reversal of a pattern is sometimes called Simpson's Paradox. It happens when there is a third variable (major) which affects the relationship between the other two (admission and gender).
Suppose 100 men and 100 women apply. Consider three scenarios.

- All the men apply to A, all the women apply to B.
- All the men apply to A, all the women apply to B.
- Half the men and half the women apply to A, the rest to B.

To see how the weighted average works, imagine that we have just two majors. Major A accepts 80% of all applicants, but Major B accepts just 10%.
In the first scenario, the raw number of men who are accepted is 850+. So $\frac{8}{100}$ are accepted. The same is true for women.

In the second scenario, the raw number of men who are accepted is 150. So $\frac{1}{100}$ is the rate. This looks like gender bias, but actually it is not—the admission policy is completely gender blind. So the rate is $\frac{10}{100}$. This looks like gender bias, but actually it is $\frac{1}{10}$. So the rate is $\frac{10}{100}$. The raw number of women who are accepted is $1 \times 100 = 80$. So $\frac{80}{100}$ are accepted, which is the same rate for women.

In the third scenario, the proportion of men who are accepted is $\frac{10}{100}$, but the proportion of women is $\frac{80}{100}$. It looks like women are favored, but actually, the admission policy is completely gender blind.
Now suppose we use a weighted average for men, This weight the acceptance rates for men in each major by the fraction of people applying to each major.

The weights give the same answer as one has from the raw counts when equal numbers of men and women apply to each major.

\[
\text{Weighted proportion of men accepted is} = \frac{0.45 \times 200}{(0.8 \times 100) + (1 \times 100)} = 0.45
\]

\[
\text{Weighted proportion of women accepted is} = \frac{0.45 \times 200}{(0.8 \times 100) + (1 \times 100)} = 0.45
\]

In all three scenarios, the weighted average proportion of men accepted is the same.

Now suppose we use a weighted average for men, This weight the