Today's plan:

- Section 4.4.4: Clinical studies

Example

Clinical studies determine the effectiveness of a drug or medical procedure.

A typical clinical study involves a drug and a **placebo** (looks like the drug, but does nothing).

A typical clinical study involves a drug and a **placebo** (looks like the drug, but does nothing).

- Some patients get the real drug

A typical clinical study involves a drug and a **placebo** (looks like the drug, but does nothing).

- Some patients get the real drug
- The others get the placebo

A typical clinical study involves a drug and a **placebo** (looks like the drug, but does nothing).

- Some patients get the real drug
- The others get the placebo
- At the end the two groups are compared.

Remarks:

- If a patient knows she has the placebo, that may affect the outcome

Remarks:

- If a patient knows she has the placebo, that may affect the outcome
- Knowing she has the real drug may also affect things.

Remarks:

- If a patient knows she has the placebo, that may affect the outcome
- Knowing she has the real drug may also affect things.

Thus, clinical studies are done so the patients don't know which group they're in.

Remarks:

- If a patient knows she has the placebo, that may affect the outcome
- Knowing she has the real drug may also affect things.

Thus, clinical studies are done so the patients don't know which group they're in.

Definition

These are called **blind studies**.

Moreover:

- If the doctors know which group a patient is in

Moreover:

- If the doctors know which group a patient is in
- this may affect their interactions with the patients.

So this also needs to be controlled.

Clinical studies are usually done so

Clinical studies are usually done so

- neither patients

Clinical studies are usually done so

- neither patients
- nor medical personnel

Clinical studies are usually done so

- neither patients
- nor medical personnel

know who belongs to which group.

Clinical studies are usually done so

- neither patients
- nor medical personnel

know who belongs to which group.

Definition

These are called **double blind studies**.

Example

A clinical study is done to determine the effectiveness of a new drug.

Example

A clinical study is done to determine the effectiveness of a new drug.

- 2,500 volunteers are picked

Example

A clinical study is done to determine the effectiveness of a new drug.

- 2,500 volunteers are picked
- they're divided into two groups of 1,250

Example

During the course of the study, some patients leave the study for various reasons.

At the end of the study there are

- 1,180 patients in the placebo group

Example

During the course of the study, some patients leave the study for various reasons.

At the end of the study there are

- 1,180 patients in the placebo group
- 1,047 in the drug group.

Example (Continued)

At the end, for each patient it's determined whether the symptoms improved, or not.

Example (Continued)

At the end, for each patient it's determined whether the symptoms improved, or not.

Symptoms\ Group	**Placebo**	**Drug**
Improve	805	785
Not	375	262
Total	1,180	1,047

Example

Determine, with a confidence level of 99.7%, if the drug is more effective than the placebo.

- Estimate what percentage p of patients improved under the placebo versus the drug.

- Estimate what percentage p of patients improved under the placebo versus the drug.
- This will be based on the observed percentage $\hat{p}$ for each group.

- Estimate what percentage p of patients improved under the placebo versus the drug.
- This will be based on the observed percentage $\hat{p}$ for each group.
- To have a confidence level of 99.7%, we need to move $3\frac{\sigma}{n}$ away from $\hat{p}$.

Placebo	**Drug**
Improve $= 805$	Improve $= 785$
Not $= 375$	Not $= 262$
$n = 1,180$	$n = 1,047$
$\hat{p} = \frac{805}{1180} \approx 0.6822$	$\hat{p} = \frac{785}{1047} \approx 0.7498$

Standard error

$\frac{\sigma}{n} \approx 0.01355$	$\frac{\sigma}{n} \approx 0.01339$

$$\hat{p} - 3\frac{\sigma}{n} \leq \; p \; \leq \hat{p} + 3\frac{\sigma}{n}$$

$0.6415 \leq p \leq 0.7228$	$0.7096 \leq p \leq 0.7899$

- the intervals overlap

- the intervals overlap
- so we can't conclude that p for the drug group is higher than p for the placebo group

- the intervals overlap
- so we can't conclude that p for the drug group is higher than p for the placebo group

In other words, there is **not statistically significant** evidence that the drug is better than the placebo.

Remark:

If we worked under a lower confidence level, say 95% confidence, then the conclusion would be different.

Simpson's paradox

Here is a real-life example from a medical study, comparing the success rates of two treatments for kidney stones (1984).

Example

- Two treatments A and B for kidney stones are compared

Example

- Two treatments A and B for kidney stones are compared
- The drugs are tested on Small Stones and Large stones

Example

- Two treatments A and B for kidney stones are compared
- The drugs are tested on Small Stones and Large stones

The table shows the corresponding success rates:

Example

	Treatment A	**Treatment B**
Small Stones	Group 1 **93% (81/87)**	Group 2 87% (234/270)
Large Stones	Group 3 **73% (192/263)**	Group 4 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

The paradoxical conclusion is that

- treatment A is more effective when used on small stones

The paradoxical conclusion is that

- treatment A is more effective when used on small stones
- treatment A is more effective when used on large stones

The paradoxical conclusion is that

- treatment A is more effective when used on small stones
- treatment A is more effective when used on large stones
- yet treatment B is more effective when considering both sizes at the same time

Question

What treatment should we use?

Question

What treatment should we use?

- **There is no easy answer that always works**

Question

What treatment should we use?

- **There is no easy answer that always works**
- You have to analyze the hidden internal relations between groups

Factors that led to Simpson's paradox:

Factors that led to Simpson's paradox:

- The sizes of the groups are very different

Factors that led to Simpson's paradox:

- The sizes of the groups are very different
- Both treatments work better for small stones

Factors that led to Simpson's paradox:

- The sizes of the groups are very different
- Both treatments work better for small stones
- Treatment B was used mostly by people with small stones (higher success rates)

Factors that led to Simpson's paradox:

- The sizes of the groups are very different
- Both treatments work better for small stones
- Treatment B was used mostly by people with small stones (higher success rates)
- Treatment A was used mostly by people with large stones (lower success rates)

Factors that led to Simpson's paradox:

- The sizes of the groups are very different
- Both treatments work better for small stones
- Treatment B was used mostly by people with small stones (higher success rates)
- Treatment A was used mostly by people with large stones (lower success rates)

So here treatment A seems to be better.

Simpler example: Alice and Bob go fishing.

Simpler example: Alice and Bob go fishing.

One week Alice goes 3 times and catches fish 0 times.

Simpler example: Alice and Bob go fishing.

One week Alice goes 3 times and catches fish 0 times.
Bob goes 7 times and catches fish 1 time.

Simpler example: Alice and Bob go fishing.

One week Alice goes 3 times and catches fish 0 times.
Bob goes 7 times and catches fish 1 time.

So in week one, Alice has a 0% average and Bob has 14.2%. (Bob wins)

The next week, Alice goes 7 times and catches fish 5 times.

The next week, Alice goes 7 times and catches fish 5 times.
Bob goes 3 times and catches fish all 3 times.

The next week, Alice goes 7 times and catches fish 5 times.
Bob goes 3 times and catches fish all 3 times.

In week two, Alice has a 71.4% average and Bob has 100%. (Bob wins again)

For both weeks considered together though, Alice has 5/10 and Bob has 4/10.

For both weeks considered together though, Alice has $5/10$ and Bob has $4/10$.

So Alice is actually better at catching fish, when we consider all the data.