Quick Review of Last Week's Lab

A Clinical Trial on Preventing Depression

- During the 1980's the National Institutes of Health (NIH) sponsored a clinical trial to **evaluate two drugs to prevent the recurrence of depression** in patients who have had at least one previous episode of the illness (Prien et al., *Archives of General Psychiatry*, 1984).

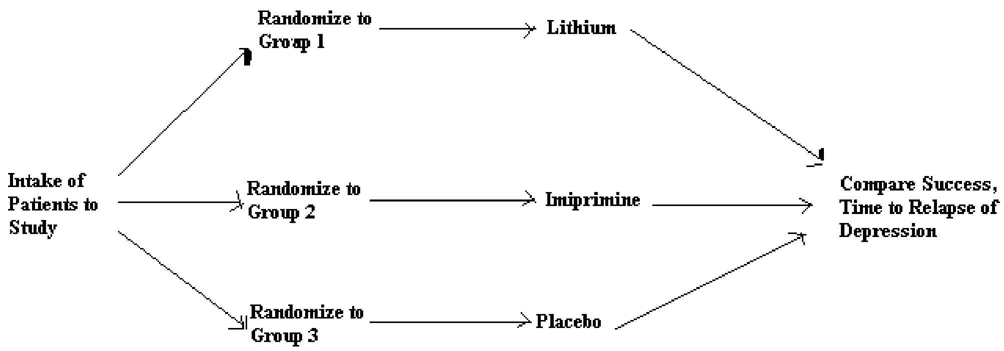
Design of the Study

- The study was **multi-centered**.
- Patients were **randomized** to one of the 3 treatment groups, *Imipramine* (Imip), *Lithium* (Li), or a *Placebo* (Pl).
- Patients were followed from 2-4 years to see whether or not they had a recurrence of depression.
- The study was **double-blinded**.

Variables Measured in the Study

- HOSPT: Which hospital: 1, 2, 3, 5 or 6.
- TREAT: 0=*Lithium*; 1=*Imipramine*; 2=*Placebo*.
- OUTCOME: 0=*Success* (no recurrence) 1=*Failure*.
- TIME: number of weeks until a recurrence.
- GAS: a measure of social functioning from 0-100.
- ACUTET: How many days depressed before the study.
- AGE: Age in years.
- GENDER: 1=*Female* 2=*Male*.

In lab: You looked at this study as a simple randomized clinical trial (RCT):

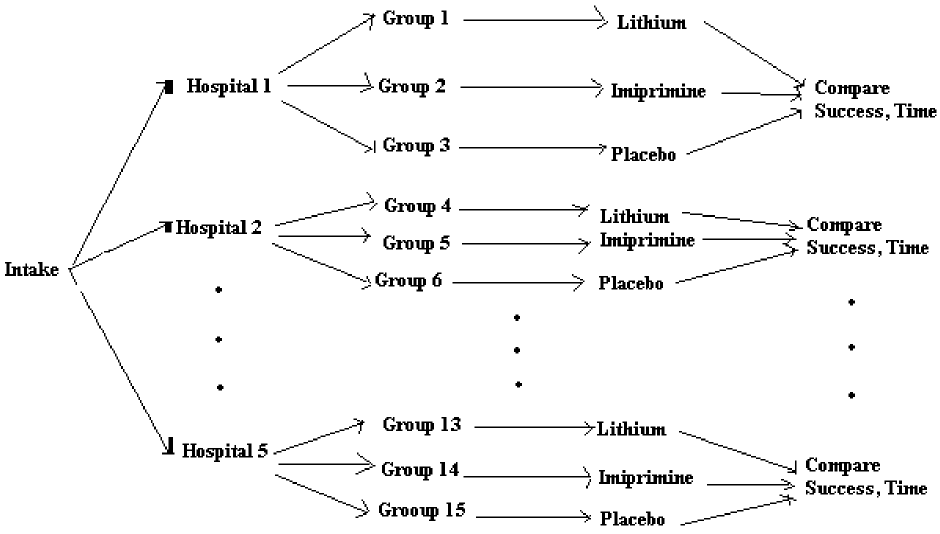


Randomization: "Balance" representation of all the variables (measured or not) that preceed treatment (everything except SUCCESS and TIME across the three treatment arms.

Blocked Design

- Mashing the HOSPITALs together might produce misleading results (analogy: Simpson's paradox).
- The patient populations at the different hospitals might be different. Maybe:
  - Lots more patients were at HOSPITAL 1, and they responded well to IMIPRIMINE.
  - Fewer patients were at all the other hospitals combined, but they responded well to LITHIUM.

To detect this sort of problem, *Block* the study on hospital. A blocked design looks like this:

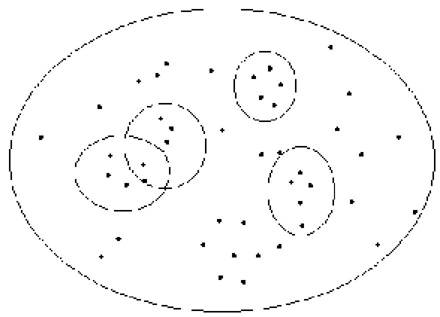


Might also block on GENDER, AGE, ACUTET, GAS, etc., depending on the purpose of the study.

Sample vs. Population

Sample Statistic, Population Parameter

- A *statistic* is a number describing a *sample*.
- A *parameter* is a number describing a *population*.



- I take a sample of 100 students and average their GPA's to estimate the average GPA at Carnegie Mellon. The average in my sample is a *statistic*, the average at Carnegie Mellon is a *parameter*.

- I survey 1000 US adults to estimate the fraction who believe that the US Government budget surplus should be spent on new social programs, rather than rebated in a tax cut. The percent in my sample is a *statistic*, the percent of all US adults who feel this way is a *parameter*.

- To find the difference in time to recurrence of depression among patients taking Imiprimine, vs. Lithium, I find 100 patients at risk of depression, give half Imiprimine and half Lithium. The average difference in time to recurrence among all patients at risk of depression is a ________________; the average difference in my two groups of 50 patients is a ________________.

- To find out the fraction of registered voters intending to vote in the next election, I conduct a random-digit-dialing telephone survey, and in each household where there is a registered voter I ask whether that person intends to vote. The fraction of voters intending to vote in my survey is a ________________; the fraction of all registered voters intending to vote is a ________________.

Bias vs. Precision

- *Bias* is difficult to measure directly. It is usually reduced by careful design and implementation of the study.

- *Precision* is easier to measure, *if the subjects were selected by an SRS*. It is usually increased by increasing the size of the SRS.

Measuring Precision: Margin of Error

- For an SRS involving a binary variable (1/0, right/wrong, yes/no, ...), we know that

$$\text{Sample Mean} \;=\; (x_1 + \cdots + x_n)/n \;=\; \frac{(\text{\# Yes's})}{n} \;=\; \hat{p}$$

  $\hat{p}$ is a *statistic* that is estimating the *parameter* p, the true proportion of *Yes's* in the population.

- The variability in $\hat{p}$ (how much it bounces around p, from sample to sample) can be measured with

$$\text{Standard error} \;=\; \sqrt{\frac{p(1-p)}{n}} \;=\; SE_p$$

  (A standard error (SE) is an SD for a mean or fraction.)

- We don't know p (that's why the poll!) so we make a guess. There are two approaches:

  – $p \approx \hat{p}$: $\quad SE_p \approx \sqrt{\hat{p}(1-\hat{p})/n}$ $\qquad$ (*smaller*).

  – $p \approx 0.50$: $\quad SE_p \approx \sqrt{.5(1-.5)/n}$ $\quad$ (*larger*).

  Usually you want the second approach, to protect yourself from making errors.

- 68–95–99.7 rule: *margin of error*, is $\pm 2 \cdot SE_p$.
- The true percent is likely to be somewhere between $\hat{p} - 2 \cdot SE_p$ and $\hat{p} + 2 \cdot SE_p$.

Example. *A poll of 500 voters across the state of Pennsylvania found that 44% (220 of the 500 voters surveyed) are in favor of selling off state liquor stores to set up a trust fund to pay for pro sports stadiums and other public works. Use $p \approx 50\%$ to estimate the margin of error of the poll.*

Answer: The problem states that $\hat{p} = 0.44$, and asks us to assume that the true p is 0.50. Therefore the standard error is $SE_p = \sqrt{p(1-p)/n} = \sqrt{(0.50)(0.50)/500} = 0.0224$. Using the middle case of the 68–95–99.7 rule, the <u>*margin of error*</u> is $\pm 2 \cdot SE_p = \pm 0.045$, or about $\pm 4.5\%$.

- Thus the percent of all Pennsylvania voters (parameter) is likely to be between $44\% - 4.5\% = 39.5\%$ and $44\% + 4.5\% = 48.5\%$. It is probable that a majority of voters does not favor this plan.
- If we had used $p \approx \hat{p} = 0.44$, we would have gotten $\sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{(0.44)(0.56)/500} = 0.0222$, which is slightly smaller than the 0.0224 we got above.

- Statewide and national opinion polls are not exactly SRS's.
  - Difficult to take a true SRS of voters in a geographical area as large as the U.S.
  - Multistage Sampling
  - Stratified Sampling
  - Probability Sampling
- However, we can get a feel for the uncertainty in a poll by using the SRS-based calculation.
- Usually the "real" calculation will lead to a larger margin of error than the SRS calculation.
- If the sample was not conducted using an SRS or another probability-based mechanism, *you cannot calculate a margin of error.*

Ethics and Uses

- Confidentiality
- Institutional review boards
- Informed consent
- Ethics of placebos for extreme illness (first do no harm)
- Behavioral and social science experiments (informed consent)

Believing a Survey or Experiment/Observational Study

- Who conducted the study?
- What was the population?
- How were the subjects selected?
- How many subjects were there?
- When was the study conducted?
- For surveys:
  - What was the response rate?
  - How were subjects contacted?
  - What were the exact questions asked?
- For experiments and observational studies:
  - What was the design of the experiment?
  - What variables were measured before treatment? After treatment?
  - Exactly how were the variables measured?