STA305/1004 - Class 5

January 25, 2016

Today's Class

SD vs. SE

Each measurement has SD σ but if x_1, x_2, x_3, x_4

$$\sqrt{Var\left(\frac{x_1+x_2+x_3+x_4}{4}\right)} = \text{standard error}$$

$\sqrt{Var(\hat{\beta})}$ = SE of $\hat{\beta}$

- REMINDER: Assignment #1 due Jan. 26 on portal by 22:00
- Answers to last week's in-class problem
- Introduction to Phase III Clinical Trials
- Introduction to power

t-test assumption:

1. normal distr'n
2. independent observation

STA305/1004 – Week 2 Additional Problems

A scientist would like to test a new drug that will inhibit a mouse's ability to run through a maze. Two mice are randomly chosen to receive the drug and another two mice don't receive the drug (control group). The time each mouse takes to go through the maze is measured in seconds. The results of the experiment are as follows:

Drug		Control	
30	25	18	21

The average time for the drug group is 27.5s and the average time for the control group is 19.5s. The mean difference in times is 27.5-19.5=8.0s.

Answer the following questions:

1. If the drug does not really influence times then the split of the observations into two groups was essentially random. Give an example of how the outcomes of the experiment could have been distributed into the two groups? Is the mean difference in your example the same or different than the observed mean difference?
2. How many possible ways could the treatments be randomized into the two groups?
3. Explicitly write out all possible distributions of the observations in the table below.

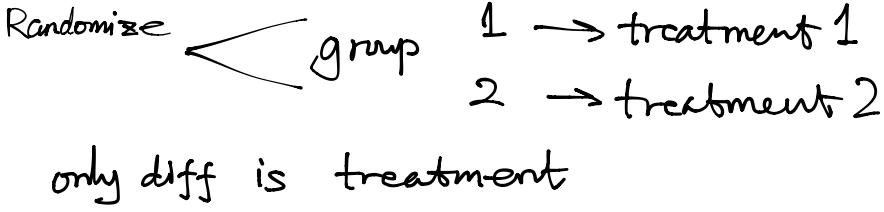
$\binom{4}{2} = 6$

Drug		Control		$\bar{X}_D$	$\bar{X}_C$	$d_i = \bar{X}_D - \bar{X}_C$
30	25	18	21	27.5	19.5	8
30	18	25	21	24	23	1
30	21	18	25	25.5	21.5	4
25	18	30	21	21.5	25 5	-4
25	21	30	18	23	24	-1
18	21	30	25	19.5	27.5	-8

4. Let μ_D and μ_C be the mean time in the drug and control groups respectively. Calculate the P-value of the test $H_0 : \mu_D = \mu_C$ versus $H_0 : \mu_D > \mu_C$ Pvalue = $\frac{1}{6}$
5. What has been assumed in calculating the P-value?
6. What can you conclude about the effectiveness of the drug on time to complete the maze? $= 0.17$

What are clinical trials?

Clinical trials are prospective intervention studies with human subjects to investigate experimental drugs, new treatments, medical devices, or clinical procedures (Yin, 2012).



Phases of clinical trials

Developing a new drug for cancer.

Preclinical studies: In vitro (e.g. slides, test tubes) and in vivo (living organism such as rodents) studies on wide range of doses of experimental agents. This stage of study provides preliminary toxicity and efficacy data including pharmacokinetics (PK) and pharmacodynamics (PD) information.

Phase I: Usually first study in humans to investigate the toxicity and side effects of the new agent. Identify MTD. maximum tolerated dose

Phase II: Assess if drug has sufficient efficacy. The drug is usually administered around the MTD. If drug does not show efficacy or is too toxic then further testing is discontinued.

Phases of clinical trials

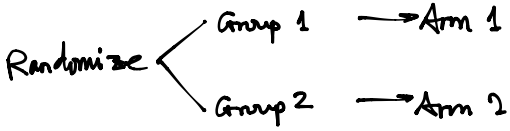
Phase III: If drug passes phase II testing then it is compared to the current standard of care or placebo. These are long-term, large scale randomized studies that may involve hundreds or thousands of patients.

If the drug is proven to be effective (e.g. two positive phase III trials required for FDA approval) the company will file an application with regulatory agencies to sell the drug. If approved then the drug will be available to the general population in the country where it was approved.

Phase IV: After approval a study might follow a large number of patients over a longer period of time to monitor side effects and drug interactions. For example, findings from these studies might add a warning label to the drug.

Phases of clinical trials

- The four phases are usually conducted sequentially and separately.
- Each trial requires an independent study design and a study protocol.
- Every aspect of trial design, monitoring, and data analysis call upon statistical methods.
- In randomized clinical trials a treatment group is often referred to as an **arm**.



Phases of clinical trials

- Experimental design plays a very important role in the design of clinical trials.
- Two arm clinical trials use all of theory of randomization that we learned about last week. Randomization is used to design phase III clinical trials since causation can usually be assessed using a randomized design.

How can causation be assessed using a randomized design?

- Suppose that patients are randomized in a two arm clinical trial where one of the arms is the standard treatment and the other arm is an experimental treatment
- A statistically significant difference in the outcome between the two arms is observed showing the experimental treatment is more efficacious. outcome = mortality – 5 month
- The interpretation is that the experimental treatment *caused* patients to have a better outcome since the only difference between the two arms is the treatment. Randomization is supposed to ensure that the groups will be similar with respect to all the factors measured in the study and all the factors that are not measured.

Baseline Measurements

	Trtmt 1	Trtmt 2
Age	$\bar{x}_1$ 30.5	$\bar{x}_2$ 31.7
sex	63%	57%

Measurements before treatments

unmeasured covariates should also be balanced due to randomization

How many patients should be enrolled in a Phase III clinical trial?

- In a phase III trial sample size is the most critical component of the study design. The sample size has implications for how many subjects will be exposed to a drug that has no proven efficacy.
- The investigator needs to specify type I, II error rates, and the effect sizes. → the difference expected between 2 groups (smallest)
- Standard practice is to compute the smallest sample size required to detect a clinically important/significant treatment difference with sufficient.

How many patients should be enrolled in a Phase III clinical trial?

- If the sample size is too small then the trial might fail to discover a truly effective drug because the statistical test cannot reach the significance level (5%) due to a lack of power.
- If the sample size is overestimated then resources wasted and drug development delayed since patient enrollment is often the main factor in time to complete a trial.

Statistical hypotheses

Suppose that subjects are randomized to treatments A or B with equal probability. Let μ_A be the mean response in the group receiving drug A and μ_B be the mean response in the group receiving drug B. The null hypothesis is that there is no difference between A and B, the alternative claims there is a clinically meaningful difference between them.

$$H_0 : \mu_A = \mu_B \text{ versus } H_1 : \mu_A \neq \mu_B$$

is standard treatment better than experimental

or

is experimental better than standard

} 2 sided (pick 1 then 1-sided)

Statistical hypotheses

The type I error rate is defined as:

.

$$\begin{aligned}\alpha &= P\left(\text{type I error}\right) \\ &= P\left(\text{Reject } H_0 | H_0 \text{ is true}\right).\end{aligned}$$

Statistical hypotheses

The type II error rate is defined as:

$$\begin{aligned}\beta &= P\left(\text{type II error}\right) \\ &= P\left(\text{Accept } H_0 | H_1 \text{ is true}\right).\end{aligned}$$

Statistical hypotheses

Power is define as:

$$\begin{aligned} \text{power} &= 1-\beta \\ &= 1 - P(\text{Accept } H_0 | H_1 \text{ is true}) \\ &= P(\text{Reject } H_0 | H_1 \text{ is true}). \end{aligned}$$

fail to reject

↑ Fixed

Power

The probability that a fixed level α test will reject H_0 when a particular alternative value of the parameter is true is called power of the test to detect that alternative.

$H_0: \mu = 0$

$H_a: \mu = 1$

Calc power of this test ?

Power

Can a 6-month exercise program increase the total body bone mineral content (TBBMC) of young women? Based on results of a previous study $\sigma = 2$ for the percent change in TBBMC over the 6-month period. A change in TBBMC of 1% would be considered important. Is 25 subjects a large enough sample size for this project?

$H_0: \mu = 0$ for $\mu = 1$

$H_1: \mu > 0$ is 25 subjects enough to detect a difference?

Reject H_0:

$$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \geq 1.645 \qquad \alpha = 0.05 = P(\text{type I})$$

$$\frac{\bar{X}-0}{2/5} \geq 1.645 \iff \bar{X} \geq 0.658$$

rejection region

rejection H_0

$$P(\bar{X} \geq 0.658 \mid H_1 \text{ is true}) = \text{Power}$$

$$= P(\bar{X} \geq 0.65 \mid \mu = 1)$$

under H_1, $\bar{X} \sim N(1, \frac{4}{\sqrt{25}})$

$$P\left(\frac{\bar{X}-1}{2/\sqrt{25}} \geq \frac{0.658-1}{2/\sqrt{25}} \mid \mu = 1\right)$$

$$= P(Z \geq -0.835) = 0.80 = \text{power}$$

In repeated sampling the test will reject 80% of the time at the 5% level.

Power of the one sample z-test

Let $X_1, ..., X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. A test of the hypothesis

$$H_0 : \mu = \mu_0 \text{ versus } H_0 : \mu \neq \mu_0$$

will reject at level α if and only if

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2},$$

or

$$\left| \bar{X} - \mu_0 \right| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)^{th}$ percentile of the $N(0, 1)$.

Power of the one sample z-test

The power of the test at $\mu = \mu_1$ is

$$\begin{aligned} 1-\beta &= 1 - P(\text{type II error}) \\ &= P(\text{Reject } H_0 | H_1 \text{ is true}) \\ &= P(\text{Reject } H_0 | \mu = \mu_1) \\ &= P\left(\left|\bar{X} - \mu_0\right| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2} | \mu = \mu_1\right) \end{aligned}$$

Subtract the mean μ_1 and divide by $\sigma/\sqrt{n}$ to obtain (why?):

$$1-\beta = 1 - \Phi\left(z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)\right) + \Phi\left(-z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)\right),$$

where $\Phi(\cdot)$ is the $N(0,1)$ CDF.

Power of the one sample z-test

The power function of the one-sample z-test is:

$$1 - \beta = 1 - \Phi\left(z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)\right) + \Phi\left(-z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)\right).$$

What is the limit of the power function as:

- $n \to \infty$
- $\mu_1 \to \mu_0$