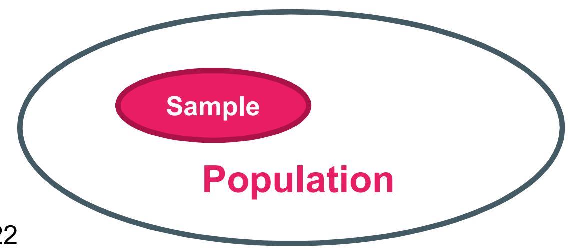
Inferential Statistics & Hypothesis Testing

V. Sridhar
IIIT Bangalore

Inferential Statistics

Sampling and Sampling Distributions

- **In many instances, one cannot study an entire population. Why?**
- **In such a case, one selects a subset of the population, called a sample, and inspects the attributes of each item in the sample**
- **Based on the findings from the sample, one makes conclusion about the entire population.**
    - **For example, if one finds 3 percent of the items in the sample as defective, the conclusion is made that 3 percent of the items in the population is defective**



Sample and Point Estimation

- Now that we know how to select a sample, let's use the sample to estimate population characteristics (mean, and proportion)
  - Using sample data to estimate a population mean or proportion is known as Point Estimate
- Estimation: Any sample statistic that is used to estimate a population parameter is called an *estimator*. An *estimate* is a specific observed value of a statistic.
- Central Limit Theorem:
  - The relationship between the shape of the population distribution and the shape of the sampling distribution of the sample statistic

Poll

- Suppose heights of all women have a standard deviation of 2.7 inches, and a random sample of 100 women's heights yields a standard deviation of 4 inches.
- Which one is the population parameter?
    - 2.7" or 4"

Difference between SD and SE

- **standard deviation measures the variability in the data, while standard error measures the variability in point estimates from different samples of the same size and from the same population, i.e. measures the sampling variability.**

Sampling and Sampling Distributi

- **MRF Tyres wants to know the mean (or average) life of its new brand of ZLX tyres**
  - **One way is testing and wearing out each tire manufactured. Obviously, this does not make sense.**
- **MRF in their tyres stock, takes a <u>sample of tyres</u>, tests and wears out each of these tyres and then calculates the mean (or average) life of the sampled tyres**
- **Suppose, the mean life is calculated as 42,000 kms**
  - **Based on this sample, it is concluded that the mean life <u>all new brand of tyres</u> (that is population) is 42,000 kms.**

Central Limit Theorem

CENTRAL LIMIT THEOREM FOR THE SAMPLE MEAN

When we collect a sufficiently large sample of n independent observations from a population with mean μ and standard deviation σ, the sampling distribution of $\bar{x}$ will be nearly normal with

$$\text{Mean} = \mu \qquad\qquad \text{Standard Error } (SE) = \frac{\sigma}{\sqrt{n}}$$

- **Independence: The sample observations must be independent**
  - **The most common way to satisfy this condition is when the sample is a simple random sample from the population**
- **Normality: When a sample is small, we also require that the sample observations come from a normally distributed population**
  - **We can relax this condition more and more for larger and larger sample sizes.**

Determining the Sample Size in Estimation

- **If the sample size is too small, we may fail to achieve the objective of our analysis, but if it is too large, that will be waste of resources.**
  - **Some sampling error will arise because we have not studied the whole population. Sampling error is controlled by selecting a sample that is adequate in size.**
- **If we want a high level of precision, we have to sample enough of the population to provide the required information.**

Confidence Intervals

Interval Estimates

- An interval estimate describes a range of values within which a population parameter is likely to appear.
  - Interval estimate is constructed using point estimate, standard error and corresponding probability
- The probability that we associate with an interval estimate is called the ***confidence level***
  - The ***confidence interval*** is the range of the estimate we are making. ***Confidence limits*** are the upper and lower limits of the confidence interval
- 95% confidence interval means: “That if we select many random samples of the same size and calculate a confidence interval for each of these samples, then in about 95 percent of these cases, the population parameter will lie within that interval.”

Poll

- If the confidence interval of sample mean is large, then the probability of population means falling within the interval is:
  a) High
  b) Low
- If the confidence interval of sample mean is large, then the accuracy of sample mean relative to population mean is:
  a) High
  b) Low

Interval Estimates

- General formula for interval estimate (Large sample):

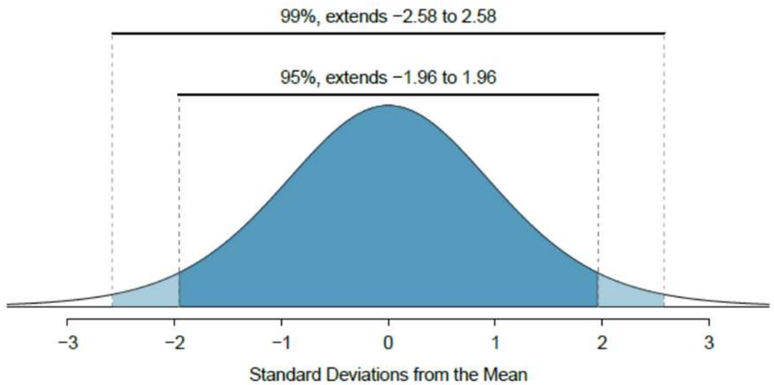
  Statistic + Z. Standard error (Statistic): Upper confidence limit

  Statistic – Z. Standard error (Statistic): Lower confidence limit

- Interval estimate for mean of large sample (Population s.d. is known):

$$\bar{x} + \left(z \times \frac{\sigma}{\sqrt{n}}\right), \bar{x} - \left(z \times \frac{\sigma}{\sqrt{n}}\right)$$

Confidence Interval from Point Estimate



- 99% CI -> $\overline{x} \pm 1.96 \times SE = \overline{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$
- Margin of Error: $z \times SE$
- Confidence Intervals are always about population estimates
  - They provide plausible range of population parameters

Quiz

- Medical experts all over the world as well as WHO are considering whether a vaccination booster shot is needed to improve the immunity levels of individuals. However, on the contrary there are views, including from WHO that a booster shot should not be administered until all individuals are fully vaccinated with 2 shots of the vaccine.
- A TV station conducted a poll of 1,042 individuals in City X to find our the orientation of the people in favour or against the booster shot.
    - The poll revealed that 82% of the samples who were interviewed were in favor of booster shot
- What is the sample statistic in this case?
- If the sample standard deviation is 0.384, what is the SE?
- What is the 95% CI for the point estimate?
    - And what can you say about the general public opinion about the booster shot?
- Can we say with 95% confidence that 80% of the City are likely to opine in favour of the booster shot?
    - What happens if we change the confidence level from 95% to 99%
        - To: Confidence Interval and Accuracy of our decision

Confidence Interval

Point estimate $\bar{p} = 0.82$

SE of estimate $= 0.012$

95% CI for $\bar{p}$

$\bar{p} \pm Z \times SE$

$0.82 \pm 1.96 \times 0.012$

$\{0.7965,\ 0.8435\}$

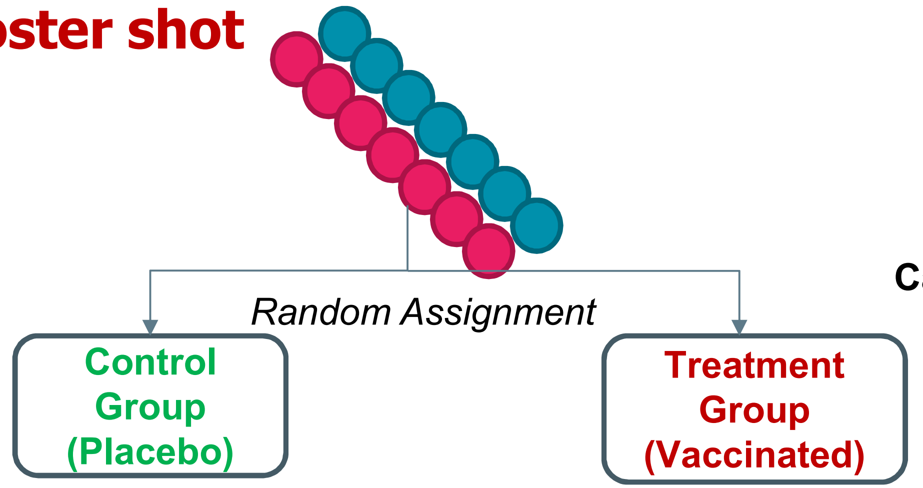
We can say with 95% confidence that μ lies between $\{0.7965, 0.8435\}$

Hypothesis Testing

Testing Hypothesis Procedure

- **Hypothesis testing begins with an assumption, called a hypothesis, that we make about a population parameter.**
  - **Hypothesis testing is about making inferences about a population from only a small sample**
- **In hypothesis testing, we must state the assumed or hypothesized value of the population parameter before we begin sampling**
  - **The assumption we wish to test is called the null hypothesis.**
  - **Whenever we reject the hypothesis, the conclusion we do accept is called alternative hypothesis.**

Efficacy of the Booster shot



Cause → Effect

Measure difference in values of Response variable
(COVID positive) before and after Treatment

- Does vaccination reduce the chances of one getting COVID?
- Null Hypothesis: Proportion of COVID affected (T) = Proportion of COVID affected (C)
- Alternative Hypothesis: Proportion of COVID affected (T) < Proportion of COVID affected (C)
- Is it by chance or indeed it varied
    - Significance level: p – very small -> not by chance; reject Null

Hypothesis Testing

Decision on the Test	In Reality (The population)	
	H_0	H_1
Accept: H_0		β :Type II Error
Reject: H_0	α : Type I Error	$(1-\beta)$ Power

- **Ideally α & β should both be small**
  - **α & β are inversely related**
- **Power of the test indicate how well your test works**
  - **Large sample size increases power of the test**

Which one should we reduce α or β?

Decision on Conviction	Reality of the Defendant	
	Innocent	Perpetrator
Accept H_0 Let Defendant Free		β :Type II Error
Reject: H_0 Convict the Defendant	α : Type I Error	$(1-\beta)$ Power

- **H0: Accused is Innocent**
- **H1: Accused is a Guilty**

Should $\alpha > \beta$; or $\alpha < \beta$

The **presumption of innocence** is a legal principle that **every person accused of any crime is considered innocent until proven guilty**

Which one should we reduce α or β?

Decision on Vaccination Phase-1 Trial	Reality of the Safety of the Vaccine	
	Safe	Unsafe
Accept H_0 : Approve the vaccine		β :Type II Error
Reject: H_0 : Do not give approval to the vaccine	α : Type I Error	$(1-\beta)$ Power

- **H0: Vaccine is safe**
- **H1: Vaccine is unsafe**

Which should be minimized: α or β

Which one should we reduce α or β?

Decision on Vaccination Phase-2 Trial	Reality of the Efficacy of the Vaccine	
	Efficacy is High	Efficacy is Low
Accept H_0 : Approve for mass vaccination		β :Type II Error
Reject: H_0 : Do not approve for mass vaccination	α : Type I Error	$(1-\beta)$ Power

- **H0: Vaccine is efficacious**
- **H1: Vaccine is not efficacious**

Which should be minimized: α or β

What are the Null and Alternate Hypothesis

- A tutoring company would like to understand if most students tend to improve their grades (or not) after they use their services.
    - They sample 200 of the students who used their service in the past year and ask them if their grades have improved or declined from the previous year
    - Let x_{pre} and x_{post} be the test scores of the students before and after they are trained by the tutoring service

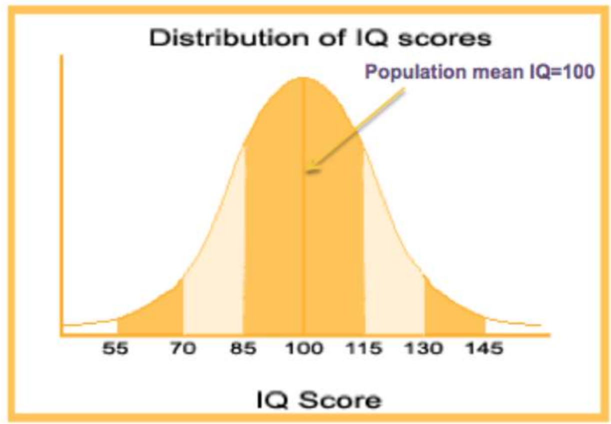
Tests of Significance

Tests of Significance

https://www.calculator.net/z-score-calculator.html

Population Mean: IQ=100

Is the sample mean significantly different than the population mean?



Population mean: IQ=100

Population st dev=16

Sample mean: IQ=108

Sample size: N=16

- **In order to determine if two numbers are *significantly different*, a statistical test must be conducted to provide evidence**
  - **Researchers must collect statistical evidence to make a claim, and this is done by conducting a test of statistical significance.**

Tests of Significance

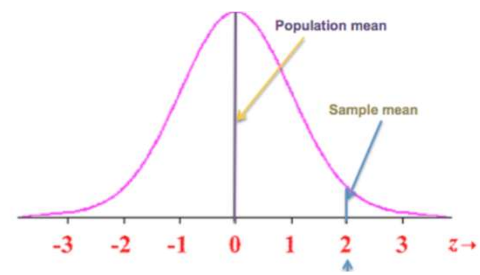
https://www.calculator.net/z-score-calculator.html

1. State the Hypothesis
   - H_0: $\mu = 100$
   - H_1: $\mu > 100$
2. Which Test to perform?
   - Should we do one-sided test? Or two-sided?
   - If H_1: $\mu \neq 100$, which test?
3. What test statistic to use
   1. *Z* or *t*?
4. Compute the test statistic

Population mean: IQ=100
Population st dev=16
Sample mean: IQ=108
Sample size: N=16

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

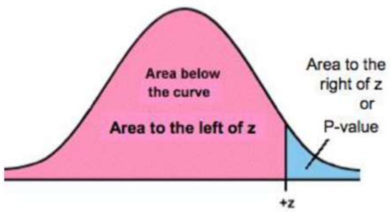
Test Statistic



$$\frac{(108 - 100)}{\left(16 / \sqrt{16}\right)} = 2$$

Tests of Significance

$Ho{:}\ \boldsymbol{\mu} = 100)$
$H1{:}\ \boldsymbol{\mu} > 100)$



- **Assuming H_0 is TRUE**
  - **Find out the probability of obtaining this score when the null hypothesis is true**
    - **= p value**
- **Area to the left of z = the probability of obtaining scores lower than z**
  - **Area to the right of z (p-value) = the probability of obtaining scores higher than z**
  - **The smaller the p-value, the stronger the evidence against H_0 provided by the data.**

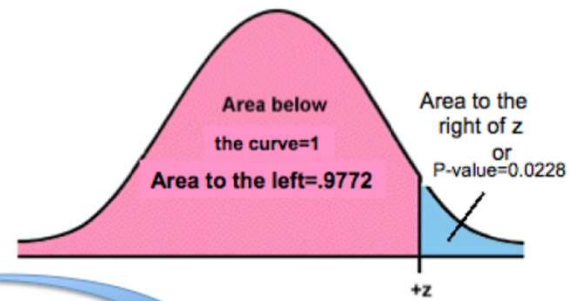
The p-value

- The p-value represents the probability of obtaining scores that are at the z level or higher when the null hypothesis is true
  - In other words, what percent chance exists of getting this specific sample mean score if it is actually no different from the population mean.
  - If z is far away from the mean, the p-value is small. The larger the test statistic (the farther from the mean), the smaller the p-value.
- When the p-value is very small, researchers can say they have strong evidence that the null hypothesis is FALSE
  - This is because if the p-value is very small, it means that the probability of obtaining a score that is so extreme or even higher is very small, if indeed the Null Hypothesis is TRUE
    - Hence decision is to Reject the Null Hypothesis in favour of Alternative Hypothesis
- How should we decide on the threshold value of *p* when we do the hypothesis testing

Tests of Significance

Example:

Table A provides the area to the left of z:



Z	[illegible]	.01	.02
1.9	[illegible]	.9719	9726
2.0	.9772	.9778	.9783
2.1	.9821	.9826	.9830

z=2.00 => Area to the left=0.9772

P_value=1-0.9772 => P_value=0.0228

=> "If the population mean IQ is equal to 100, there is a 2% probability of recording a sample mean of 108.

- **This means that there is only a 2% chance that the null hypothesis is true.**
- **In other words, if the population mean is 100, then there is only a 2% chance of having a sample mean equal to 108.**

Relationship between p-value and Significance level

- Choosing a significance level α for a test is important in many contexts, and the traditional level is $\alpha = 0.05$
- If making a Type 1 Error is dangerous or especially costly, we should choose a small α level (e.g. 0.01)
  - Under this scenario we want to be very cautious about rejecting the null hypothesis,
    - so we demand very strong evidence favoring HA before we would reject H0.
- If a Type 2 Error is relatively more dangerous or much more costly than a Type 1 Error, then we might choose a higher α (e.g. 0.10)
  - Here we want to be cautious about failing to reject H0 when the alternative hypothesis is actually true

Summary

1. State the null and alternative hypotheses.
2. Calculate the test statistic.
3. Find the *P*-value (using a table or statistical software).
4. Compare P-value with the threshold Significance Level: α and decide whether the null hypothesis should be rejected or accepted
    1. If the p-value is less than α, then there is a strong evidence for rejecting the Null Hypothesis
    2. Otherwise, there is not a stronger evidence and hence cannot reject the Null Hypothesis

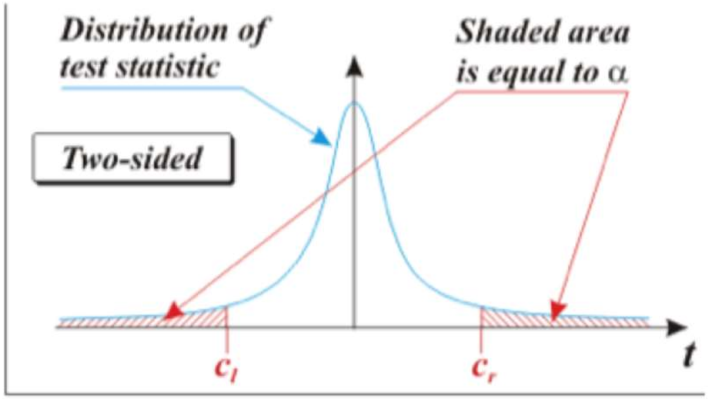
Two-Sided Test

When the alternative hypothesis is two sided:

P-value=Area to the right of the curve (from Table A)*2

Example:
Null Hypothesis: Sample IQ=100
Alternative Hypothesis: Sample IQ≠100

Z=2
Area to the right=.0228
P_Value=.0228*2= .0456



Example:

α =.05
p-value=0.04

p-value<α => Decision: Reject the null hypothesis & accept the alternative

Conclusion: "The test of significance provided evidence that the sample IQ is significantly different than 100."

Relationship between p value and α

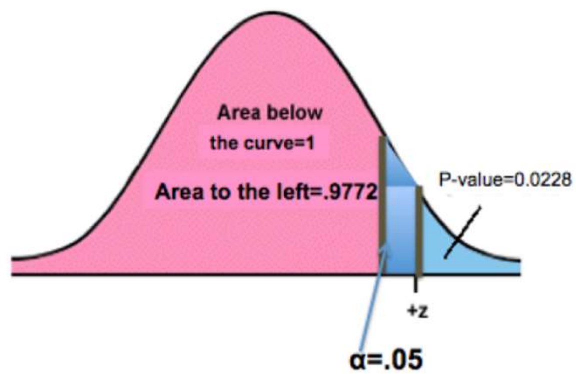
How small should the P_value be to reject the null hypothesis?

α (alpha) – (significance level: the probability of rejecting Ho when Ho is true)

Is the P-value smaller than α ?

P-value > α => accept the null hypothesis

P-value <= α => reject the null hypothesis & accept the alternative hypothesis

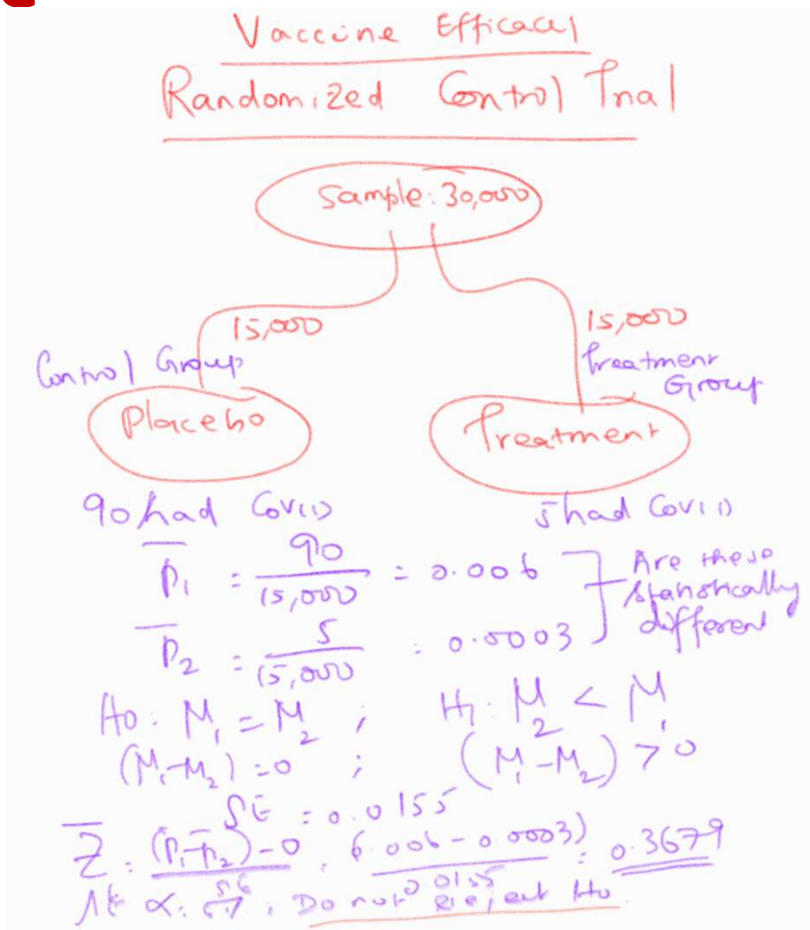


- **Poll: If your p-value is greater than α then what should you do:**
  - **Accept the null hypothesis**

Polls

- **Decreasing the significance level α will increase the probability of making a Type 1 Error Yes/ No**

Vaccine Example

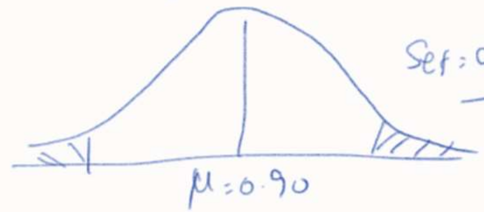


TV Survey Example

Relationship between
CI and Hyp Testing

$H_0: \mu = 0.90$; $H_1: \mu \neq 0.90$

Set: $\alpha = 5\%$



point estimate: $\bar{p} = 0.82$

Test statistic $Z = \frac{\bar{p} - \mu}{SE} = \frac{0.82 - 0.90}{0.012} = -6.67$

For $Z = -6.67$, $p \simeq 0$

Reject H_0, Accept: H_1

CI calculated for the sample

$\bar{p} = 0.82$, $SE = 0.012$ for 95% Conf. level

$\{0.7965, 0.8435\}$ $p = 0.90$ outside

References

- Diez, D. M., Barr, C. D., & Çetinkaya-Rundel, M. (2015). OpenIntro Statistics. OpenIntro. Retrieved April, 18, 2019.
- Levin R., and Rubin, D. (1998). Statistics for management (7th Edition), Pearson.
- Gujarati, D. (2012). Econometrics by example. Macmillan
- R-Tutorial on Elementary Statistics. Available at: http://www.r-tutor.com/elementary-statistics
- Coursera: Inferential Statistics course offered by Mine Çetinkaya-Rundel, Duke University
- Distribution App: https://gallery.shinyapps.io/dist_calc/