Unit 2

Inferential Statistics

Topics to be covered

- Inferential Statistics - Introduction

- Statistical Methods for Evalutaion:
  - Hypothesis Testing
  - Difference of Means
  - Wilcoxon Rank-Sum Test
  - Type-I & Type-II Errors

Descriptive & Inferential Statistics

Descriptive Statistics

- Organize
- Summarize
- Simplify
- Presentation of data

↓

Describing data

Inferential Statistics

- Generalize from samples to population
- Hypothesis testing
- Relationships among variables

↓

Make predictions

Introduction

- Whether the research design is experimental, quasi-experimental, or non-experimental, many researchers develop their studies to look for differences.

Inferential Statistics

- Inferential statistics are used to draw conclusions about a population by examining the sample

Or

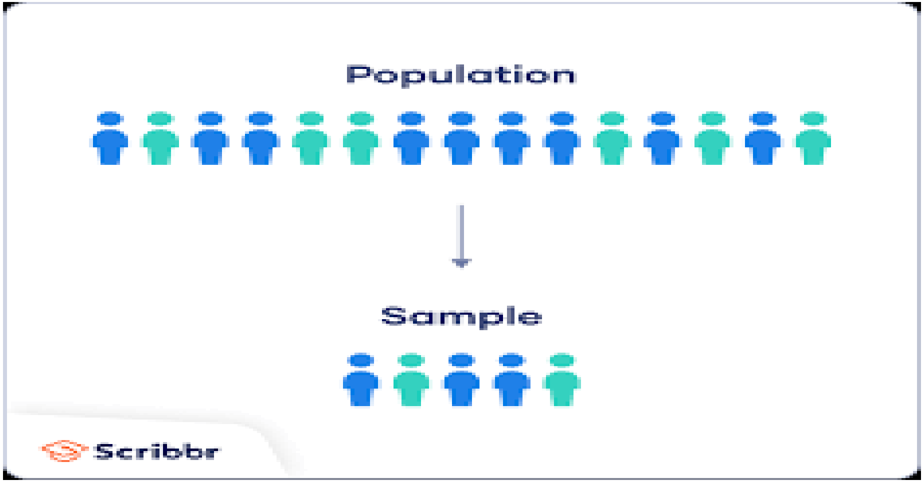
- Inferential statistics allows you to make predictions(“inferences”) from that data.

- Instead of using the entire population to gather the data, the statistician will collect a sample or samples from the million residents and make inferences about the entire population.

POPULATION

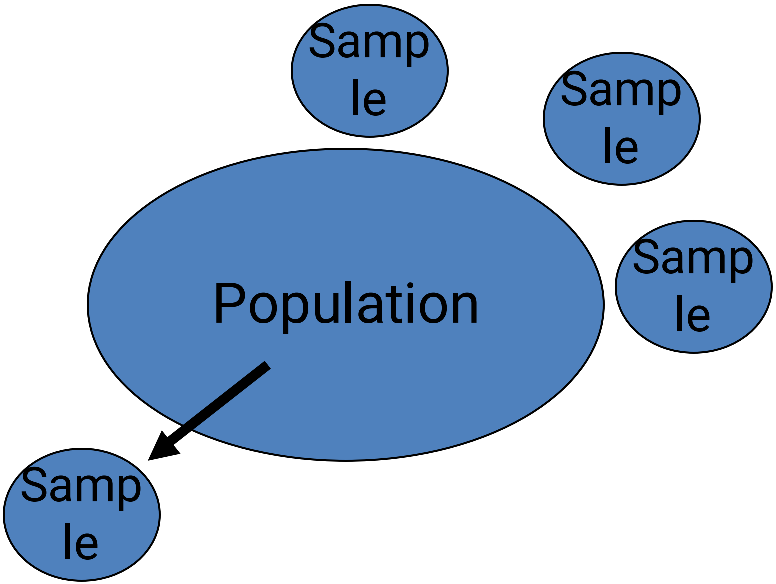
Sample (a small subset of the population)

Sampling



- It is the process of selecting units from a population of interest.
- Accuracy of inference depends on the representativeness of the sample from the population

Inferential Statistics



Draw inferences about the larger group

Simple Random Sampling

- In this sampling every member of the population has an equal chance of being selected into the sample.

- Selection of one member must be independent of the selection of every other member.

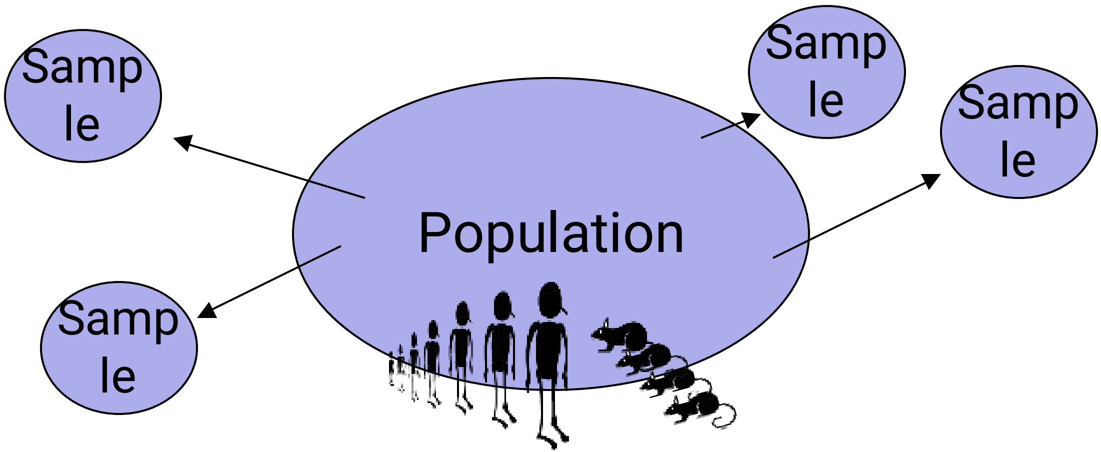
- The population consists of N objects.
- The sample consists of n objects.
- If all possible samples of n objects are equally likely to occur, the sampling method is called simple random sampling.

Stratified Sampling

- This method can be used if the population has a number of distinct “strata” or groups.
- In this sampling, first identify members of the sample who belong to each group.
- Then randomly sample from each of those subgroups in such a way that the sizes of the subgroups in the sample are proportional to their sizes in the population.

e.g. suppose in a university 70% of full-time students and 30% of part-time students are admitted. Thus, your sample of 200 students would consist of 140 full-time and 60 part-time students.

What is a Statistic????

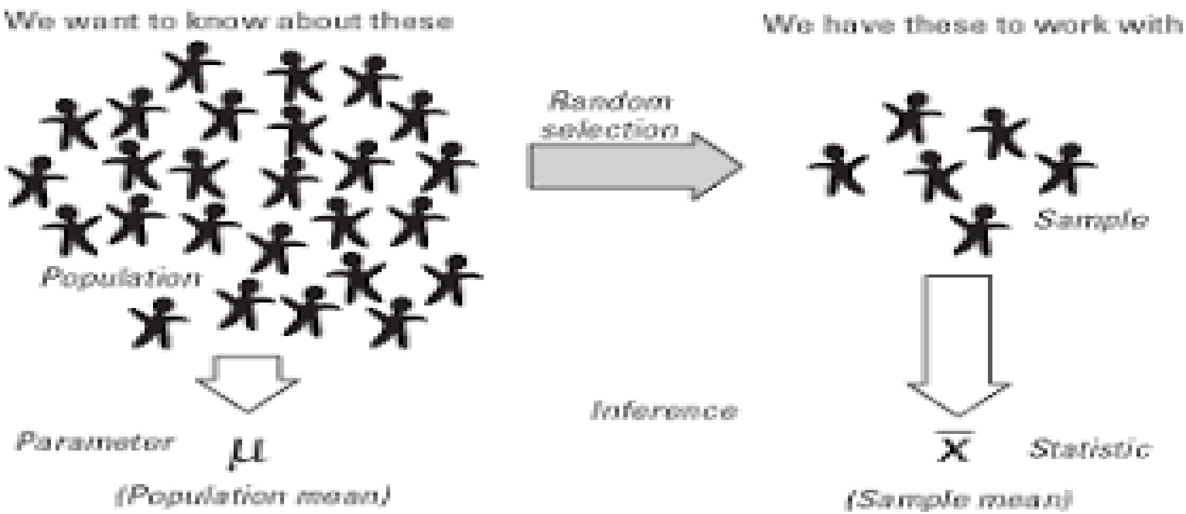


Parameter: value that describes a population

Statistic: a value that describes a sample

Difference between a Statistic & Parameter

- Statistics describe a sample.
- Statistic is a characteristic of a sample.
- Parameter describes an entire population.
- A parameter is a characteristic of a population.



Main Areas of Inferential Statistics

- Estimating Parameters – Taking statistics from your sample data(e.g. sample mean) and using it to say something about a population parameter (population mean).

e.g. Assume 37% of people in our sample said that vanilla is the favorite ice cream flavor. Can we extrapolate that 37% of all people in the world also think vanilla is the best?

can't say 100% but inferential statistical techniques can provide range of people that prefer vanilla with some level of confidence.

2. Hypothesis Testing – experimental analysis

Hypothesis testing is another way of drawing conclusions about parameters. T-test, Chi-Square, or analysis of variance (ANOVA)

Hypothesis Testing

- A hypothesis is an assumption.
- The process of deciding statistically whether the findings of an investigation reflect chance or real effects at a given level of probability.
- Two statistical data sets are compared. A hypothesis is proposed for the statistical relationship between the two data sets.

Inferential Statistics

- Inferential statistics help researchers test hypotheses and answer research questions, and derive meaning from the results
  - a result found to be statistically significant by testing the sample is assumed to also hold for the population from which the sample was drawn
  - the ability to make such an inference is based on the principle of probability

Inferential Statistics

- Researchers set the significance level for each statistical test they conduct
  - by using probability theory as a basis for their tests, researchers can assess how likely it is that the difference they find is real and not due to chance

Content

- Hypotheses testing :
- z test
- t test
- Analysis of Variance (ANOVA)

● Hypothesis Testing

A hypothesis test is a formal way to make a decision based on statistical analysis. A hypothesis test has the following general steps:

- Set up two contradictory hypothesis. One represents our "assumption".
- Perform an experiment to collect data. Analyze the data using the appropriate distribution.
- Decide if the experimental data contradicts the assumption or not.
- Translate the decision into a clear, non-technical conclusion.

● Null and Alternative Hypotheses

Hypothesis tests are tests about a population parameter (μ or p). We will do hypothesis tests about population mean and population proportion p.

The null hypothesis (H0) is a statement involving equality (=; <;>) about a population parameter. We assume the null hypothesis is true to do our analysis.

The alternative hypothesis (Ha) is a statement that contradicts the null hypothesis. The alternative hypothesis is what we conclude is true if the experimental results lead us to conclude that the null hypothesis (our assumption) is false.

The alternative hypothesis must not involve equality (≠; <;>).

The exact statement of the null and alternative hypotheses depend on the claim that you are testing.

- Examples

Example 1: Mobile App Response Time

Scenario: A developer wants to ensure that the new update of a food delivery app does not increase the average response time, which was previously 2.5 seconds.

H_0 (Null Hypothesis): The average response time after the update is equal to or less than 2.5 seconds.

H_1 (Alternate Hypothesis): The average response time after the update is greater than 2.5 seconds.

- Examples

Example 2: Website Redesign and User Retention

Scenario: A start-up redesigned its homepage and wants to check if it has improved user retention.

H_0: The new homepage design has no effect on user retention.

H_1: The new homepage design has a significant effect on user retention.

- Examples

- Example 3: Online Course Effectiveness

Scenario: A university introduces an online Python course and wants to test if it performs better than traditional classroom teaching in terms of student grades.

- Example 4: Marketing Campaign for a Tech Product

Scenario: A company wants to test if its new marketing campaign for a tech product has increased the average number of website visitors per day.

- Example 5: Performance of AI Model

Scenario: A researcher compares the accuracy of a new image classification model with an existing baseline model having 85% accuracy.

- Ans:

- Example 3: Online Course Effectiveness

H_0: The average grades of students in the online course are equal to those in the classroom course.

H_1: The average grades of students in the online course are different from those in the classroom course.

- Example 4: Marketing Campaign for a Tech Product

H_0: The average number of daily visitors is the same as before the marketing campaign.

H_1: The average number of daily visitors has increased after the marketing campaign.

- Example 5: Performance of AI Model

H_0: The new model’s accuracy is less than or equal to 85%.

H_1: The new model’s accuracy is greater than 85%.

- Outcomes and the Type I and Type II Errors

Hypothesis tests are based on incomplete information, since a sample can never give us complete information about a population. Therefore, there is always a chance that our conclusion has been made in error.

There are two possible types of error:

The first possible error is if we conclude that the null hypothesis (our assumption) is invalid (choosing to believe the alternative hypothesis), when the null hypothesis is really true. This is called a Type I error.

Type I error = { Deciding to reject the null when the null is true
incorrectly supporting the alternative

The other possible error is if we conclude that the null hypothesis (our assumption) seems reasonable (choosing not to believe the alternative hypothesis), when the null hypothesis is really false. This is called a Type II error.

Type II error = { Failing to reject the null when the null is False
incorrectly NOT supporting the alternative

TYPE I and TYPE II ERROR IN TABULAR FORM

	Decision	
	Accept H0	Reject H0
H0 True	Correct decision	Type I Error
H0 False	Type II Error	Correct Decision

When a Null hypothesis is tested, there may be four possible outcome:

- The Null Hypothesis is true but our test rejects it.
- The Null Hypothesis is false but our test accept it.
- The Null Hypothesis is true and our test accepts it.
- The Null Hypothesis is false but our test rejects it.

Type I Error : Rejecting Null Hypothesis when Null Hypothesis is true. It is called ‘α-error’.

Type II Error : Accepting Null Hypothesis when Null Hypothesis is false. It is called ‘β-error’.

- Outcomes and the Type I and Type II Errors Cont…

It is important to be aware of the probability of getting each type of error. The following notation is used:

$$\alpha = \begin{cases} \text{P(Type I error)} \\ \text{P(decide to reject null} \mid \text{null is true)} \\ \text{P(incorrectly supporting alternative)} \\ \text{the } \textbf{significance level} \text{ of the test} \end{cases}$$

$$\beta = \begin{cases} \text{P(Type II error)} \\ \text{P(decide to ``accept'' null} \mid \text{null is false)} \\ \text{P(incorrectly NOT supporting alternative)} \end{cases}$$

The significance level α is the probability that we incorrectly reject the assumption (null) and support the alternative hypothesis. In practice, a data scientist chooses the significance level based on the severity of the consequence of incorrectly supporting the alternative. In our problems, the significance level will be provided.

- Distribution Needed for Hypothesis Testing

The sample statistic (the best point estimate for the population parameter, which we use to decide whether or not to reject the null hypothesis) and distribution for hypothesis tests are basically the same as for confidence intervals.

The only difference is that for hypothesis tests, we assume that the population mean (or population proportion) is known: it is the value supplied by the null hypothesis. (This is how we \assume the null hypothesis is true" when we are testing if our sample data contradicts our assumption.)

When testing a claim about population mean μ, ONE of the following two requirements must be met, so that the Central Limit Theorem applies and we can assume the random variable, $\bar{x}$ is normally distributed:

- The sample size must be relatively large (many books recommend at least 30 samples), OR
- the sample appears to come from a normally distributed population.

It is very important to verify these requirements in real life. In the problems we
are usually told to assume the second condition holds if the sample size is

- Stating Hypotheses

The first step in conducting a test of statistical significance is to state the hypotheses.

A significance test starts with a careful statement of the claims being compared. The claim tested by a statistical test is called the null hypothesis (H0). The test is designed to assess the strength of the evidence against the null hypothesis. the null hypothesis is a statement of “no difference.”

when conducting a test of significance, a null hypothesis is used. The term null is used because this hypothesis assumes that there is no difference between the two means or that the recorded difference is not significant.

Null Hypotheses denoted by H0.

The opposite of a null hypothesis is called the alternative hypothesis. The alternative hypothesis is the claim that researchers are actually trying to prove is true.

The claim about the population that evidence is being sought for is the alternative hypothesis (Ha).

- Test Statistic
    - It is a random variable that is calculated from sample data and used in hypothesis test.
    - Test statistic compare your data with what is expected under the null hypothesis.
    - It is used to calculate P-Value.
    - A test statistic measures the degree of agreement between a sample of the data and the null hypothesis.

Different hypothesis tests use different test statistics based on the probability model assumed in the null hypothesis. Common tests and their test statistics are:

Hypothesis Test	Test Statistics
Z-test	Z-statistic
T-test	T-statistic
ANOVA	F-statistic
Chi-square tests	Chi-square statistic

- P-Value

The p-value is the probability, computed under the assumption that the null hypothesis is true, of observing a value from the test statistic at least as extreme as the one that was actually observed.

Thus, P-value is the chance that the presence of difference is concluded when actually there is none.

- When the p value is between 0.05 and 0.01 the result is usually called significant.
- When P value is less than 0.01, result is often called highly significant.
- When p value is less than 0.001 and 0.005, result is taken as very highly significant.

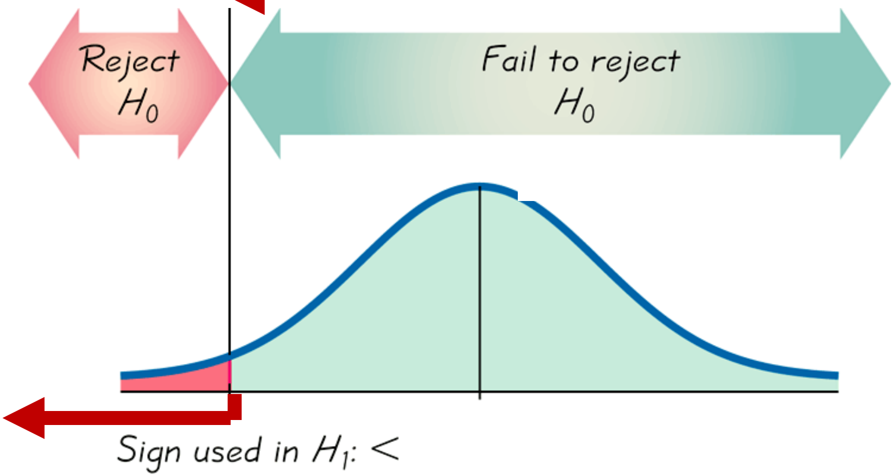
Left-tailed Test

is in the left tail

H_0: $p=0.5$

H_1: $p<0.5$

Points Left



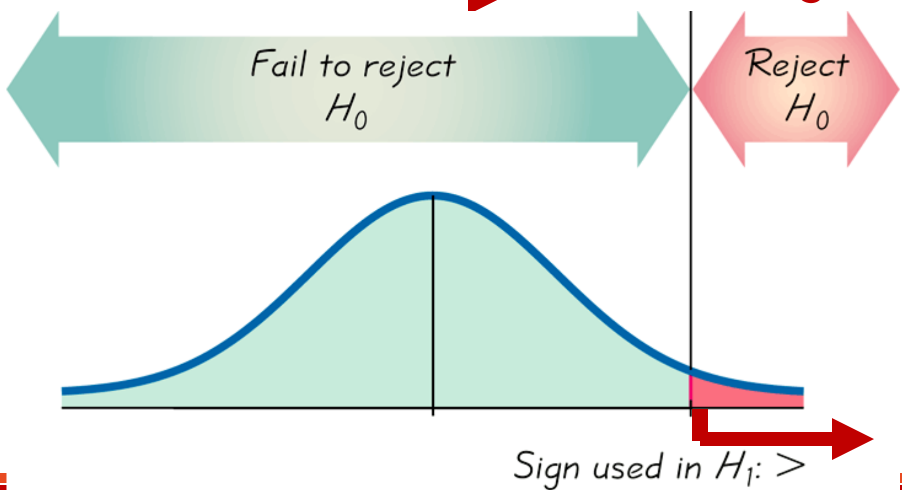
Right-tailed Test

H_0: $p=0.5$

H_1: $p>0.5$

is in the right tail

Points Right



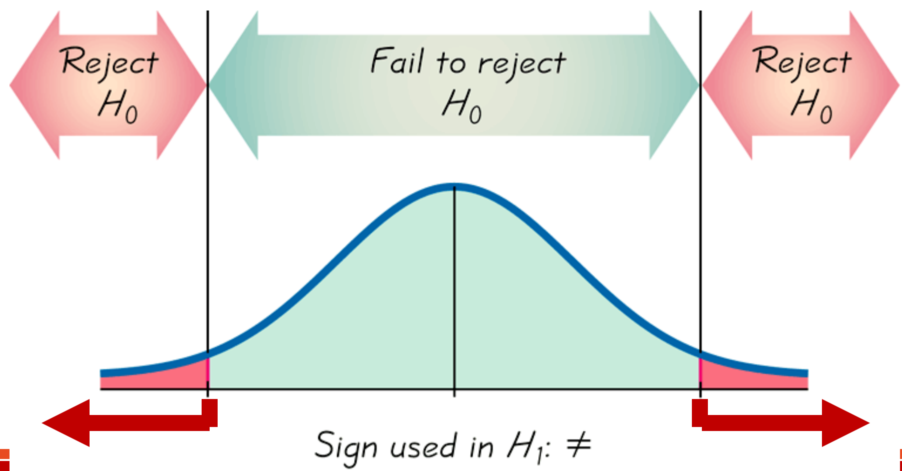
Two-tailed Test

H_0: $p=0.5$

H_1: $p\ \ 0.5$

Means less than or greater than

is divided equally between the two tails of the critical region



- Statistical test
- These are intended to decide whether a hypothesis about distribution of one or more populations should be rejected or accepted.

Statistical Test

Parametric Test applied when data concern :

- Normal distribution
- Quantitative data
- Scale of measurement is METRIC scale(i.e. interval / ratio scale
- Compare mean and standard deviation
- More powerful

Non- Parametric Test applied when data concern :

- Skewed distribution
- Qualitative data
- Nominal or ordinal scale
- Compared in terms of %, proportion, etc.
- Comparatively less powerful

- Parametric tests
  - Used for Quantitative Data
  - Used for continuous variables
  - Used when data are measured on approximate interval or ratio scales of measurements.
  - Data should follow normal distribution.
    - <u>Some parametric tests are:-</u>
      - Z test for large samples(n>30)
      - t-test
      - ANOVA (Analysis of variance)
      - Pearson’s r Correlation (r= rank)

Z-Test:

$$z=(x-\mu)/\sigma$$

This formula is used when you have a single value from a population and you want to know how far this value is from the population mean in terms of the population standarc

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

This formula is used when you have the mean of a sample and you want to know how far this sample mean is from the population mean, considering the sample size.

- Z-test

<u>ZTest</u>: The sample statistic is the sample mean of the data, $\bar{x}$. If **population standard deviation is known** (unlikely in real life), the distribution of the sample means is $\bar{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$, where μ_0 is the population mean *assumed in the null hypothesis*. The test statistic is a z-score: $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$. The p-value is the tail area under the $\bar{X}$ normal curve beyond $\bar{x}$ in the direction of the alternative hypothesis, which is the same as the tail area under $Z \sim N(0,1)$ beyond z.

- <u>Z test for large samples(n>30)</u>
  - Z-test is a statistical test where normal distribution is applied and is basically used for dealing with problems relating to large samples when the frequency is greater than or equal to 30.
  - It is used when population standard deviation is known.

Assumptions made for use:

- Population is normally distributed
- The sample is drawn at random

When the test apply?

- Population standard deviation σ is known
- Size of the sample is large (say n > 30)

Z-Test Steps:

1. Use a Z-test when:
The sample size is large ($n \geq 30$) OR the population standard deviation (σ) is known.
The data is approximately normally distributed.

2. State the Hypotheses
Example: You're testing whether a new drug affects blood pressure.
Null Hypothesis (H_0): $\mu = \mu_0$ (no effect)
Alternative Hypothesis (H_1):
Two-tailed: $\mu \neq \mu_0$
Left-tailed: $\mu < \mu_0$
Right-tailed: $\mu > \mu_0$

3. Compute the Z-Statistic by using formula

4. Use Z-Table to Find the p-value or Critical Value

If you're using the Z-table to find the p-value:

Find Z value from step 3.

Look up Z in the Z-table (typically gives area to the left of that Z).

5. Make a Decision

If $|Z| >$ Z-critical reject H_0

If $p < \alpha$ reject H_0

Problem Statement : A gym trainer claimed that all the new boys in the gym are above average weight. A random sample of thirty boys weight have a mean score of 112.5 kg and the population mean weight is 100 kg and the standard deviation is 15.

Is there a sufficient evidence to support the claim of gym trainer.

Step-1: State Null and Alternate Hypothesis

Null Hypothesis:

$H_0: \mu = 100$

Alternate Hypothesis:

$H_a: \mu > 100$

Step-2: Set the significance level (alpha-value)

Let alpha-value is 0.05, so corresponding z-score is 1.645

Step-3: Find the z-value

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{112.5 - 100}{\frac{15}{\sqrt{30}}} = 4.56$$

Step-4: Comparing with the significance level:

From step-3, we have

4.56 > 1.645

So, we have to reject the null hypothesis.

i.e. average weight of new boys are greater than 100 kg

Problem Statement 2:

You are working as a data scientist for a retail company that claims their new marketing strategy has increased the average weekly sales. Historically, the average weekly sales for the company were $5000. After implementing the new marketing strategy, a random sample of 40 weeks showed an average weekly sales of $5200 with a standard deviation of $600.

Perform a hypothesis test at a 5% significance level to determine if the new marketing strategy has significantly increased the average weekly sales.

Solution:

$$z = \frac{5200 - 5000}{\frac{600}{\sqrt{40}}} = \frac{200}{\frac{600}{6.32}} \approx \frac{200}{94.87} \approx 2.11$$

. **Make the Decision:**

- If the calculated z value is greater than the critical value, reject the null hypothesis.
- In this case, $z = 2.11$ and the critical value $z_\alpha = 1.645$.

Since $2.11 > 1.645$, we reject the null hypothesis.

. **Conclusion:**

There is sufficient evidence at the 5% significance level to conclude that the new marketing strategy has significantly increased the average weekly sales.

- Student’s t- test

Developed by Prof. W.S Gossett in 1908, who publishes statistical papers under the pen name of “student.” Thus the test is known as Student’s “t” test.

- <u>When the test is apply?</u>
  - When samples are small.
  - Population variance are not known.

- <u>Assumption made in the use of “t” test</u>
  - Samples are randomly selected.
  - Data utilized is Quantitative.
  - Variables follow normal distribution.
  - Sample variance are mostly same in both the groups under the study.
  - Samples are small, mostly lower than 30.

- Student’s t- test Cont…

  - T- test compares the difference between two means of different groups to determine whether that difference is statistically significant.

  - It is use in different – different purposes:
    - “t” test for one sample
    - “t” test for unpaired two samples.
    - “t” test for paired two samples.

- One Sample t-test
  - When compare the mean of a single group of observations

with a specified value.

  - In one sample t-test, we know the population mean. We draw a random sample from the population and then compare the sample mean with the population mean and make a statistical decision as to whether or not the sample mean is different from the population.

<u>Formula :</u>

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Where, = $\bar{x}$ sample mean
μ= population mean, = $\frac{s}{\sqrt{n}}$ standard error.

Now we compare calculate value with table value at certain level of significance (generally 5% or 1%).

- One Sample t-test Cont…

    - If absolute value of “t” obtained is grater than table value then reject the null hypothesis and if it is less than table value, the null hypothesis may be accepted.

Therefore, rule for rejecting the null hypothesis:

Reject Ho if t ≥ +ve Tabulated value
or,
Reject Ho if t ≤ -ve Tabulated value or, we can say that $p < .05$

Problem Statement:

Ten individuals are chosen randomly from a population and their heights are found to be in inches 63,63,64,65,66,69,69,70,70,71. Discuss the proposal that the mean height in the universe is 65 inches.

Problem Statement:

A shop manufacturing company distributes a particular type of brand through a large number of retail shops. Before a healthy advertising campaign, the mean weekly sales per shop was 140 dozen. After the campaign, a sample of 26 shops was taken and the mean sales were found to be 147 dozens with std. deviation 16. Can you consider the advertisement effective?

If $t > t_{critical}$, we reject the null hypothesis.

If $t \leq t_{critical}$, we fail to reject the null hypothesis.

Since 2.231>1.7082 we reject the null hypothesis.

Conclusion

There is sufficient evidence at the 5% significance level to conclude that the advertisement was effective in increasing the mean weekly sales per shop.

- T- test for unpaired two samples

  - Used when the two independent random samples come from the normal populations having unknown or same variance.

  - We test the null hypothesis that the two population means are same i.e., $\mu1 = \mu2$

  - <u>Assumption made for use</u>
    - Populations are distributed normally
    - Samples are drawn independently and at random

- <u>When the test is apply?</u>

  - Standard deviations in the populations are same and not known
  - Size of the sample is small

One-Sample T-Test

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$ = obersved mean of the sample
μ = assumed mean
s = standard deviation
n = sample size

Two-Sample T-Test

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$ = observed mean of 1st sample
$\bar{x}_2$ = observed mean of 2nd sample
s_1 = standard deviation of 1st sample
s_2 = standard deviation of 2nd sample
n_1 = sample size of 1st sample
n_2 = sample size of 2nd sample

Image by Microbe Notes

- T- test for unpaired two samples Cont…

If two independent samples xi (i = 1,2,….,n1) and yj (j = 1,2, …..,n2) of sizes n1and n2 have been drawn from two normal populations with means μ1 and μ2 respectively.

Null hypothesis

H0 : μ1 = μ2

Under H0, the test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$ = observed mean of 1st sample

$\bar{x}_2$ = observed mean of 2nd sample

s_1 = standard deviation of 1st sample

s_2 = standard deviation of 2nd sample

n_1 = sample size of 1st sample

n_2 = sample size of 2nd sample

e.g. A researcher wants to test whether two teaching methods lead to different average student scores.
Method A scores: 85, 90, 88, 75, 95
Method B scores: 78, 82, 80, 76, 79
At 5% significance level, test whether the average scores differ.

- T- test for paired two samples

Used when measurements are taken from the same subject before and after some manipulation or treatment.

Ex: To determine the significance of a difference in blood pressure before and
after administration of an experimental pressure substance.

- T- test for paired two samples Cont…
  - <u>Assumptions made for the test</u>
    - Populations are distributed normally
    - Samples are drawn independently and at random
  - <u>When the test apply</u>
    - Samples are related with each other.
    - Sizes of the samples are small and equal.
    - Standard deviations in the populations are equal and not known.

Null Hypothesis:

H0: μd = 0

Under H0, the test statistic

= |¯ | s/ n

Where, d = difference between ~~x1 and~~ x2

d = Average of d

s = Standard deviation

n = Sample size

e.g. The water diet requires you to drink 2 cups of water every half hour from when you get up until you go to bed but eat anything you want. Four adult volunteers agreed to test this diet. They are weighed prior to beginning the diet and 6 weeks after. Their weights in pounds are

Conduct a one-sample t-test using the difference with the following hypotheses:
H0: Diff = 0, Ha: Diff ≠ 0 Report the test statistic with the P-value, then summarize your conclusion.

Person	1	2	3	4	mean	S.d.
Weight before	180	125	240	150	173.75	49.56
Weight after	170	130	215	152	166.75	36.09

- <u>ANOVA (Analysis of Variance)</u>
  - Developed by R.A.Fischer.
  - Analysis of Variance (ANOVA) is a collection of statistical models

used to analyze the differences between group means or variances.

  - Compares multiple groups at one time.

- One way ANOVA

Compares two or more unmatched groups when data are categorized in one factor.

Example :

- Comparing a control group with three different doses of aspirin
- Comparing the productivity of three or more employees based on working hours in a company

- Two way ANOVA

- Used to determine the effect of two nominal predictor variables on a continuous outcome variable.
- It analyses the effect of the independent variables on the expected outcome along with their relationship to the outcome itself.

Example :

Comparing the employee productivity based on the working hours and working conditions.

Assumptions of ANOVA :

- The samples are independent and selected randomly.
- Parent population from which samples are taken is of normal distribution.
- Various treatment and environmental effects are additive in nature.
- The experimental errors are distributed normally with mean zero and variance σ2

ANOVA compares variance by means of F-ratio

- F = /
- It again depends on experimental designs

Null hypothesis:

Ho = All population means are same

- If the computed Fc is greater than F critical value, we are likely to reject the null hypothesis.
- If the computed Fc is lesser than the F critical value , then the null hypothesis is accepted.

Basics of ANOVA

- ANOVA stands for Analysis of Variance.
- ANOVA enables us to test for significance of difference among more than two sample means.

- Assumption for ANOVA
  - Samples follow normal distribution.
  - Samples have been selected randomly and independently.
  - Each group should have common variance.
  - Data are independent.

Basics of ANOVA

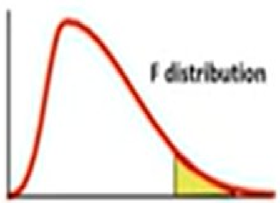
Null Hypothesis – The means for all groups are the same (equal).

$$H_0 : \mu_1 = \mu_2 = \mu_3 \ldots\ldots\ldots\ldots \mu_n$$

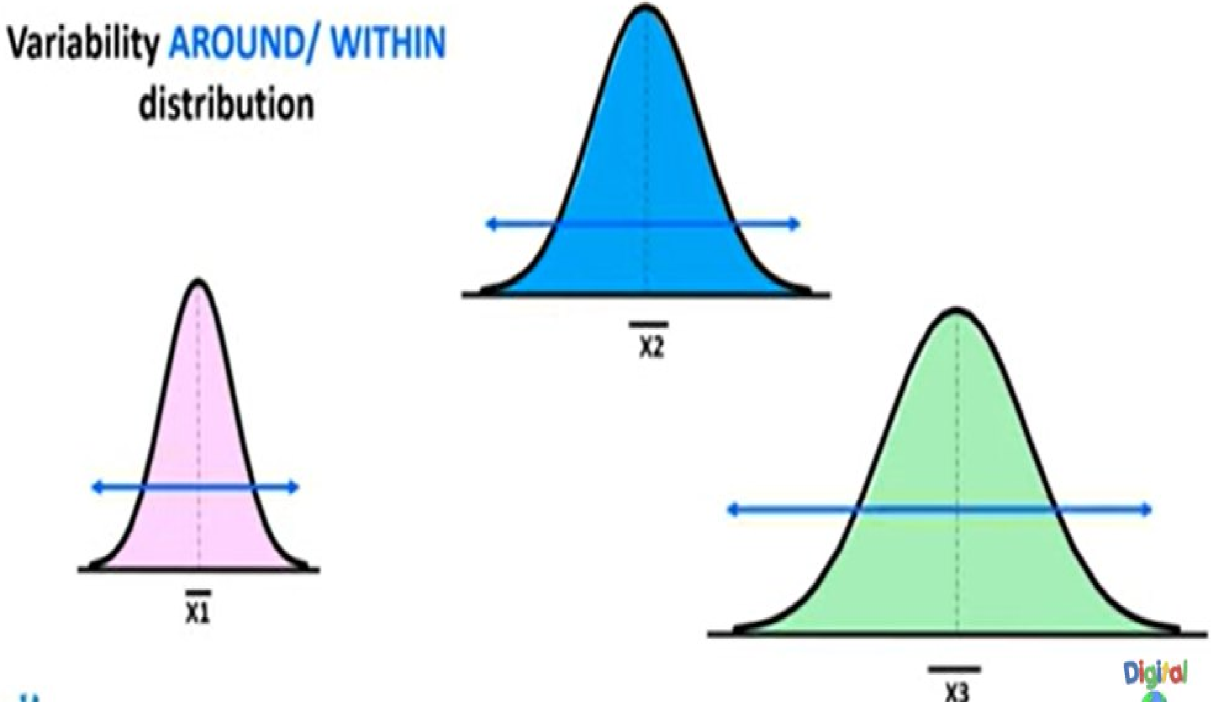
Alternate Hypothesis – The means are different for at least one pair of groups.

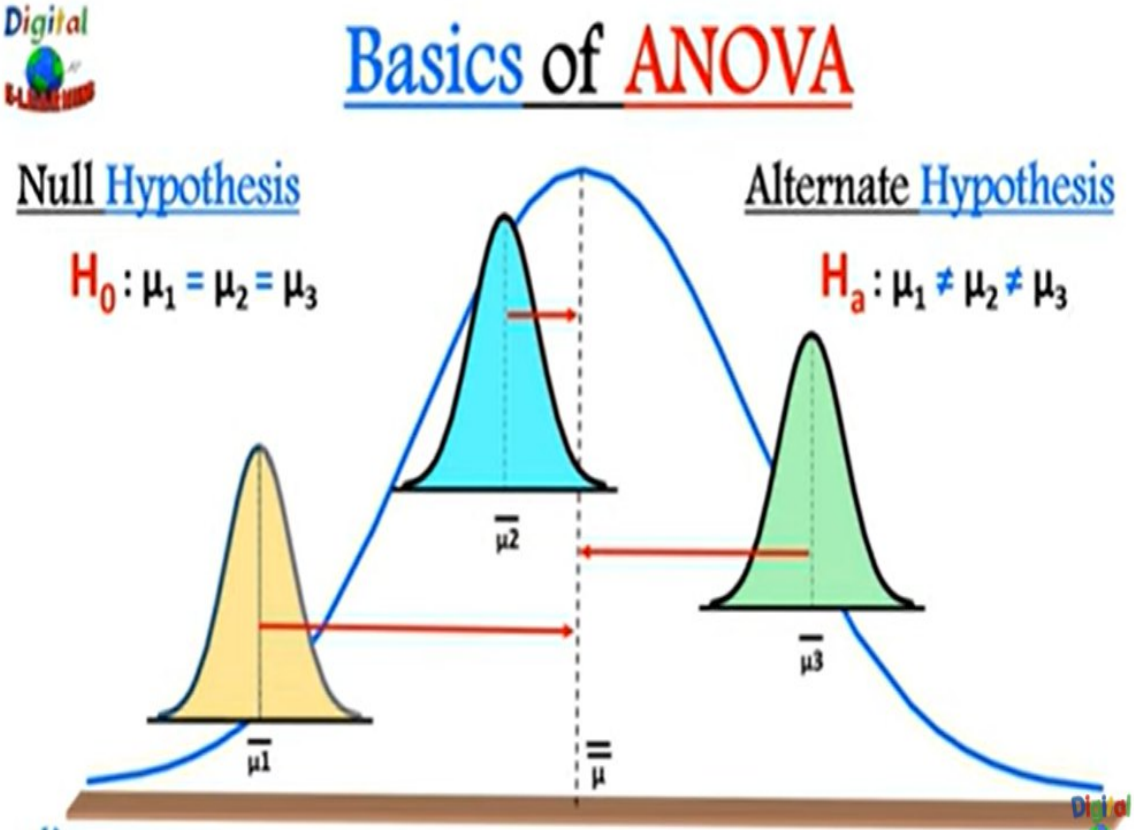
$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots\ldots\ldots \mu_n$$

$$\text{ANOVA} = \frac{\text{Variance Between}}{\text{Variance Within}}$$



Basics of ANOVA




Digital
Basics of ANOVA
Null Hypothesis
$H_0 : \mu_1 = \mu_2 = \mu_3$
Alternate Hypothesis
$H_a : \mu_1 \neq \mu_2 \neq \mu_3$
$\overline{\mu 2}$
$\overline{\mu 3}$
$\overline{\mu 1}$
$\overline{\overline{\mu}}$

Basics of ANOVA

$$\text{ANOVA} = \frac{\text{Variance Between}}{\text{Variance Within}}$$

Total Variance = Variance Between + Variance Within

$$\frac{\text{Variance Between}}{\text{Variance Within}} > 1$$ Reject H_0

$$\frac{\text{Variance Between}}{\text{Variance Within}} < 1$$ Fail to Reject H_0

$$\frac{\text{Variance Between}}{\text{Variance Within}} = 1$$ Fail to Reject H_0

Basics of ANOVA

We want to see if three different studying methods can lead to different mean exam scores or not. To test this, we select 30 students and randomly assign 10 each to use a different studying method.

We will solve this using 2 different method. Let look at Method 1

Basics of ANOVA

Sno	Method A	Method B	Method C
1.	10	8	9
2.	9	9	8
3.	8	10	7
4.	7.5	8	10
5.	8.5	8.5	9
6.	9	7	8
7.	10	9.5	7
8.	8	9	10
9.	8	7	9
10.	9	10	8
Group Mean	8.7	8.6	8.5

Overall Mean 8.6

Between Group Variation =10*(8.7-8.6)^2+10*(8.6- 8.6)^2+10*(8.5-8.6)^2

Between Group Variation = 0.2

Within Group Variation: $\Sigma(X_{ij} - X_j)^2$

Where:

Σ: a symbol that means "sum"

X_{ij}: the i^{th} observation in group j

X_j: the mean of group j

Method A: $(10-8.7)^2 + (9-8.7)^2 + (8-8.7)^2 + (7.5-8.7)^2 + (8.5-8.7)^2 + (9-8.7)^2 + (10-8.7)^2 + (8-8.7)^2 + (8-8.7)^2 + (9-8.7)^2 = 6.6$

Method B: $(8-8.6)^2 + (9-8.6)^2 + (10-8.6)^2 + (8-8.6)^2 + (8.5-8.6)^2 + (7-8.6)^2 + (9.5-8.6)^2 + (9-8.6)^2 + (7-8.6)^2 + (10-8.6)^2 = 10.9$

Method C: $(9-8.5)^2 + (8-8.5)^2 + (7-8.5)^2 + (10-8.5)^2 + (9.5-8.5)^2 + (8-8.5)^2 + (7-8.5)^2 + (10-8.5)^2 + (9-8.5)^2 + (8-8.5)^2 = 10.5$

Within Group Variation: 6.6+10.9+10.5 = 28

Basics of ANOVA

$$\frac{\text{Variance Between}}{\text{Variance Within}} = \frac{0.2}{28} = 0.0071 < 1$$

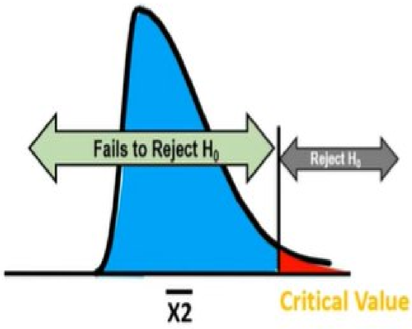
Fail to Reject H_0

"Means are very close to overall mean and distribution overlap is hard to distinguish".

Basics of ANOVA

$F_{Critical} > F_{Stat}$ Fail to Reject H_0

$F_{Critical} < F_{Stat}$ Reject H_0



Assuming $\alpha = 0.05$

$$F_{Stat} = \frac{\text{Variance Between}}{\text{Variance Within}}$$

$$\frac{0.2}{28} = 0.0071$$

$$F_{Critical} = \frac{\text{Numerator Degree of Freedom}}{\text{Denominator Degree of Freedom}}$$

Numerator Degree of Freedom = No. of Samples - 1 = 3-1 = 2

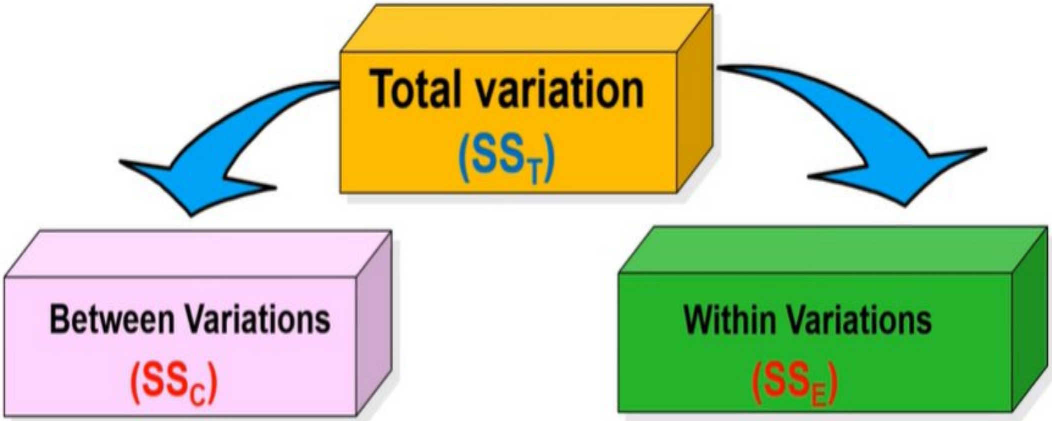
Denominator Degree of Freedom = $\sum(nj-1) = n_T - k = 30 - 3 = 27$

$F_{Critical} = F_{(2,27)} = 3.35$

$F_{Critical} > F_{Stat}$

Fail to Reject H_0

Basics of ANOVA



Sum of Squares (SS): Reflects variation. Depend on sample size.

Degrees of freedom (df): Number of population averages being compared.

Mean Square (MS): SS adjusted by df. MS can be compared with each other. (SS/df)

One way ANOVA

Source	df Degree of Freedom	SS (Sum of Squares) variation	MS (Mean Square) (variance)	F (or F Ratio)
Factor (Between)	$a-1$	SS_B	$MS_B = \dfrac{SS_B}{a-1}$	$F = \dfrac{MS_B}{MS_E}$
Error (Within)	$a(n-1)$	SS_E	$MS_E = \dfrac{SS_E}{a(n-1)}$	
Total	$a-1+a(n-1)$	$SS_T = SS_B + SS_E$		

Example-1

The exam scores for each group are shown below:

Group 1	Group 2	Group 3
85	91	79
86	92	78
88	93	88
75	85	94
78	87	92
94	84	85
98	82	83
79	88	85
71	95	82
80	96	81

Perform one – way ANOHA on this data.

Result of One-way ANOHA

Source	SS	df	MS	F
Treatment	192.2	2	96.1	2.358
Error	1100.6	27	40.8	
Total	1292.8	29		

Example-2

Exam score of 5 students given below, apply one way ANOHA on this data.

X1	X2	X3	X4
8	12	18	13
10	11	12	9
12	9	16	12
8	14	6	16
7	4	8	15

Answer:

Source of Variance	Sum of Square	Degree of Freedom (df)	Mean sum of square	F-Ratio
Variation Between Sample	SSc=50	C-1=3	MSC=50/3 =16.67	F=MSC/MSE = 16.67/13 =1.28
Variation Within Sample	SSe= 208	N-c=16	MSE= 208/16 =13	
Total	SST= 258	N-1=19		

- Non- Parametric Test
  - Non-parametric tests can be applied when:

– Data don’t follow any specific distribution and no assumptions about the population are made.
– Data measured on any scale applied when data concern.

- Commonly used Non Parametric Tests are:
  - Chi Square test
  - Mann-Whitney U test
  - Kruskal-wallis one-way ANOVA
  - Friedman ANOVA
  - The Spearman rank-order correlation test.

Parametric Vs Non-parametric

Properties	Parametric	Non-parametric
Assumptions	Yes	No
central tendency Value	Mean value	Median value
Correlation	Pearson	Spearman
Probabilistic distribution	Normal	Arbitrary
Population knowledge	Requires	Does not require
Used for	Interval data	Nominal data
Examples	z-test, t-test, etc	Wilcoxon Rank Sum or Mann-Whitney, Kruskal-Wallis

Parametric Test	Corresponding Nonparametric test	Purpose of test
t test for independent samples	Mann-Whitney U test; Wilcoxon rank-sum test	Compares two independent samples
Paired t test	Wilcoxon matched pairs signed-rank test	Examines a set of differences
Pearson correlation coefficient	Spearman rank correlation coefficient	Assesses the linear association between two variables.
One way analysis of variance (F test)	Kruskal-Wallis analysis of variance by ranks	Compares three or more groups
Two way analysis of variance	Friedman Two way analysis of variance	Compares groups classified by two different factors

Types of Data and Analysis

Nominal

Ordinal

Discrete

Continuous

Non-parametric

Parametric

Types of Data

Nominal: No numerical value
Ordinal: Order or rank
Discrete: Counts
Continuous: Interval, ratio

Nominal Data

- Non numerical value
- Blood grouping: A, B, AB, O
- Grades in PHL 541: A^+
- Urates in urine: ++, +++, +

Ordinal Data

Items on an ordinal scale are set into some kind of order by their position on the scale. This may indicate such as temporal position, superiority, etc.

The order of items is often defined by assigning numbers to them to show their relative position. Letters or other sequential symbols may also be used as appropriate.

You cannot do arithmetic with ordinal numbers -- they show sequence only.

Example
The first, third and fifth person in a race.

Pay bands in an organization, as denoted by A, B, C and D.

Discrete Data

A type of data is discrete if there are only a finite number of values possible or if there is a space on
the number line between each 2 possible values.

Exmple. A 5 question quiz is given in PHL 541 class. The number of correct answers on a student's quiz is an example of discrete data. The number of correct answers would have to be one of the following : 0, 1, 2, 3, 4, or 5. There are not an infinite number of values, therefore this data is discrete. Also, if we were to draw a number line and place each possible value on it, we would see a space between each pair of values.

Exmple. In order to obtain a taxi license in Riyadh, a person must pass a written exam regarding different locations in the city. How many times it would take a person to pass this test is also an example of discrete data. A person could take it once, or twice, or 3 times, or 4 times, or… . So, the possible values are 1, 2, 3, … . There are infinitely many possible values, but if we were to put them on a number line, we would see a space between each pair of values.

Discrete data usually occurs in a case where there are only a certain number of values, or when we are counting something (using whole numbers).

Continuous Data

Continuous data makes up the rest of numerical data. This is a type of data that is usually associated with some sort of physical measurement.

Example. The height of trees at a nursery is an example of continuous data. Is it possible for a tree to be 76.2" tall? Sure.
How about 76.29"? Yes. How about 76.2914563782"? Yes.

One general way to tell if data is continuous is to ask yourself if it is possible for the data to take on values that are fractions or decimals. If your answer is yes, this is usually continuous data.

Example. The length of time it takes for a light bulb to burn out is an example of continuous data. Could it take 800 hours? How about 800.7? 800.7354? The answer to all 3 is yes

Classify each set of data as discrete or continuous.

1) The number of suitcases lost by an airline.

2) The height of corn plants.

3) The number of ears of corn produced.

5) The time it takes for a car battery to die.

6) The production of tomatoes by weight.

Answers

1) Discrete: The number of suitcases lost must be a whole number.

2) Continuous: The height of corn plants can take on infinitely many values (any decimal is possible).

3) Discrete: The number of ears of corn must be a whole number.

4) Continuous: The amount of time can take on infinitely many values (any decimal is possible).

5) Continuous: The weight of the tomatoes can take on infinitely many values (any decimal is possible).

What is a parameter and why should I care?

Most statistical tests, like the t test, assume some kind of underlying distribution, like the normal distribution

If you know the mean and the standard deviation of a normal distribution then you know how to calculate probabilities

Means and standard deviations are called Parameters; all theoretical distributions have parameters.

Statistical tests that assume a distribution and use parameters are called parametric tests

Statistical tests that don't assume a distribution or

Parametric Test Procedures

1- Involve Population Parameters

Example: Population Mean

2- Require Interval Scale or Ratio Scale

Whole Numbers or Fractions

Example: Height in Inches (72, 60.5, 54.7)

3- Have Stringent Assumptions

Example: Normal Distribution

Nonparametric Test Procedures

A nonparametric test is a hypothesis test that does not require any specific conditions about the shape of the populations or the value of any population parameters.

Tests are often called "distribution free" tests.

Why non-parametric statistics?

Need to analyse ‘Crude’ data (nominal, -ordinal)

Data derived from small samples

Data that do not follow a normal distribution

Data of unknown distribution

Wilcoxon rank sum test (or the Mann-Whitney U test)

In statistics, the Mann-Whitney U test (also called the Mann-Whitney-Wilcoxon (MWW), Wilcoxon rank-sum test, or Wilcoxon-Mann-Whitney test) is a non-parametric test for assessing whether two samples of observations come from the same distribution.

It requires the two samples to be, independent and the observations to be ordinal or continuous measurements, i.e. one can at least say, of any two observations, which is the greater.

It is one of the best-known non-parametric significance tests.

It was proposed initially by Wilcoxon (1945), for equal sample sizes, and extended to arbitrary sample sizes and in other ways by Mann and Whitne (1947).

MWW is virtually identical in performing an ordinary parametric two-sample t test on the data after ranking over the combined samples.

15-4 WILCOXON RANK-SUM TEST

Suppose that we have two independent continuous populations X_1 and X_2 with means μ_1 and μ_2. Assume that the distributions of X_1 and X_2 have the same shape and spread and differ only (possibly) in their locations. The Wilcoxon rank-sum test can be used to test the hypothesis H_0: $\mu_1 = \mu_2$. This procedure is sometimes called the Mann-Whitney test, although the Mann-Whitney test statistic is usually expressed in a different form.

15-4.1 Description of the Test

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be two independent random samples of sizes $n_1 \leq n_2$ from the continuous populations X_1 and X_2 described earlier. We wish to test the hypotheses

$$H_0\text{: } \mu_1 = \mu_2$$
$$H_1\text{: } \mu_1 \neq \mu_2$$

The test procedure is as follows. Arrange all $n_1 + n_2$ observations in ascending order of magnitude and assign ranks to them. If two or more observations are tied (identical), use the mean of the ranks that would have been assigned if the observations differed.

Let W_1 be the sum of the ranks in the smaller sample (1), and define W_2 to be the sum of the ranks in the other sample. Then,

$$W_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - W_1 \tag{15-7}$$

Now if the sample means do not differ, we will expect the sum of the ranks to be nearly equal for both samples after adjusting for the difference in sample size. Consequently, if the sums of the ranks differ greatly, we will conclude that the means are not equal.

Appendix Table X contains the critical value of the rank sums for $\alpha = 0.05$ and $\alpha = 0.01$ assuming the two-sided alternative above. Refer to Appendix Table X with the appropriate sample sizes n_1 and n_2, and the critical value w_α can be obtained. The null H_0: $\mu_1 = \mu_2$ is rejected in favor of H_1: $\mu_1 \neq \mu_2$ if either of the observed values w_1 or w_2 is less than or equal to the tabulated critical value w_α.

The procedure can also be used for one-sided alternatives. If the alternative is H_1: $\mu_1 < \mu_2$, reject H_0 if $w_1 \leq w_\alpha$; for H_1: $\mu_1 > \mu_2$, reject H_0 if $w_2 \leq w_\alpha$. For these one-sided tests, the tabulated critical values w_α correspond to levels of significance of $\alpha = 0.025$ and $\alpha = 0.005$.

The mean axial stress in tensile members used in an aircraft structure is being studied. Two alloys are being investigated. Alloy 1 is a traditional material, and alloy 2 is a new aluminum-lithium alloy that is much lighter than the standard material. Ten specimens of each alloy type are tested, and the axial stress is measured. The sample data are assembled in Table 15-3. Using $\alpha = 0.05$, we wish to test the hypothesis that the means of the two stress distributions are identical.

We will apply the eight-step hypothesis-testing procedure to this problem:

1. The parameters of interest are the means of the two distributions of axial stress.
2. H_0: $\mu_1 = \mu_2$
3. H_1: $\mu_1 \neq \mu_2$
4. $\alpha = 0.05$
5. We will use the Wilcoxon rank-sum test statistic in Equation 15-7,

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1$$

6. Since $\alpha = 0.05$ and $n_1 = n_2 = 10$, Appendix Table X gives the critical value as $w_{0.05} = 78$. If either w_1 or w_2 is less than or equal to $w_{0.05} = 78$, we will reject H_0: $\mu_1 = \mu_2$.

Table 15-3 Axial Stress for Two Aluminum-Lithium Alloys

Alloy 1		Alloy 2	
3238 psi	3254 psi	3261 psi	3248 psi
3195	3229	3187	3215
3246	3225	3209	3226
3190	3217	3212	3240
3204	3241	3258	3234

Alloy Number	Axial Stress	Rank
2	3187 psi	1
1	3190	2
1	3195	3
1	3204	4
2	3209	5
2	3212	6
2	3215	7
1	3217	8
1	3225	9
2	3226	10
1	3229	11
2	3234	12
1	3238	13
2	3240	14
1	3241	15
1	3246	16
2	3248	17
1	3254	18
2	3258	19
2	3261	20

The sum of the ranks for alloy 1 is

$$w_1 = 2 + 3 + 4 + 8 + 9 + 11 + 13 + 15 + 16 + 18 = 99$$

and for alloy 2

$$w_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} - w_1 = \frac{(10 + 10)(10 + 10 + 1)}{2} - 99 = 111$$

8. Conclusions: Since neither w_1 nor w_2 is less than or equal to $w_{0.05} = 78$, we cannot reject the null hypothesis that both alloys exhibit the same mean axial stress.


THANK
YOU