MODULE 6

Statistical Methods for Evaluation-

- Hypothesis testing,
- difference of means,
- wilcoxon rank–sum test,
- type 1 type 2 errors,
- power and sample size,
- ANNOVA

What is Hypothesis?

- A hypothesis is an **educated guess** about something in the world around you. It should be testable, either by experiment or observation. For example:
  - A new medicine you think might work.
  - A way of teaching you think might be better.

What is a Hypothesis Statement?

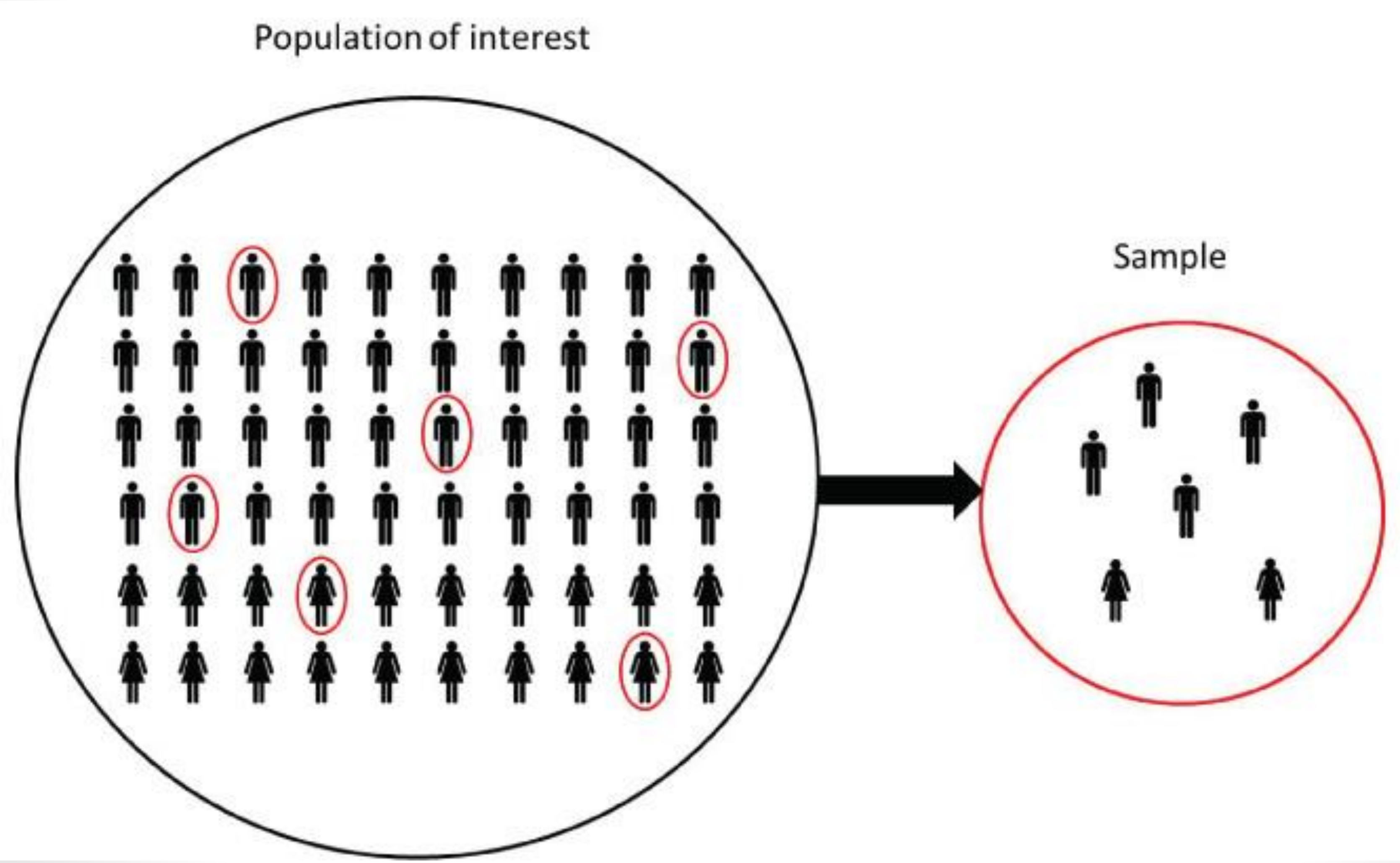
- Hypothesis statement will look like this:
- **"If I…(do this to an independent variable)….then (this will happen to the dependent variable)."**

- **For example:**
- If I (decrease the amount of water given to herbs) then (the herbs will increase in size).
- If I (give patients counseling in addition to medication) then (their overall depression scale will decrease).

What is Hypothesis Testing?

- Hypothesis testing refers to
  1. Making an assumption, called hypothesis, about a population parameter.
  2. Collecting sample data.
  3. Calculating a sample statistic.
  4. Using the sample statistic to evaluate the hypothesis

Hypothesis Testing :Population & sample



Hypothesis Testing

HYPOTHESIS TESTING

Null hypothesis, H_0

- State the hypothesized value of the parameter before sampling.
- The assumption we wish to test (or trying to reject)
- E.g
- $\mu = 20$
- There is no difference between coke and diet coke

Alternative hypothesis, H_A

All possible alternatives other than the null hypothesis.

E.g

$\mu \neq 20$

$\mu > 20$

$\mu < 20$

There is a difference between coke and diet coke

Hypothesis Testing

Basic concept is to form an assertion and test it with data

Common assumption is that *there is no difference between samples* (default assumption)

Statisticians refer to this as the *null hypothesis (H_0)*

The *alternative hypothesis (H_A)* is that *there is a difference between samples*

What is the Null & alternate Hypothesis?

- The null hypothesis is always the accepted fact or accepted as being true are:
  - ***DNA is shaped like a double helix.***
  - ***There are 8 planets in the solar system (excluding Pluto).***

- Given a population, the initial (assumed) hypothesis to be tested ,Ho , is called the null hypothesis.

- Rejection of null hypothesis causes another hypothesis,H1,is called the alternative hypothesis, to be made.

Statistical Methods for Evaluation-

- Hypothesis testing,
- difference of means,
- wilcoxon rank–sum test,
- type 1 type 2 errors,
- power and sample size,
- ANNOVA

mean, variance , standard deviation

Mean
(or Average)
denoted by

- **μ** if working with population
- **$\bar{X}$** if working with samples

Variance
denoted by

- **σ^2** (for population)
- **s^2** (for sample)

Standard deviation
denoted by

- **σ_X** or σ (for population)
- **s_X** or s (for sample))

Mean – is a simple average of given data values

- Example

- 4,5,9,2,14,6

- Mean $\bar{X}$= (4+ 5+9+3+15+6) /6
- = 42/6
- = 7

Variance: a measure of how data points differ from the mean

- Data Set 1: 3, 5, 7, 10, 10
  Data Set 2: 7, 7, 7, 7, 7
- What is the mean of the above data set?
  - Data Set 1: mean = 7
  - Data Set 2: mean = 7
- But we know that the two data sets are not identical! The **variance** shows how they are different.
- We want to find a way to represent these two data set numerically.

How to Calculate variance?

- If we conceptualize the spread of a distribution as the extent to which the values in the distribution differ from the mean and from each other, then a reasonable measure of spread might be the average deviation, or **difference, of the values from the mean.**

$$\frac{\sum(x-\bar{X})}{N}$$

How to Calculate variance?

- The average of the squared deviations about the mean is called the variance.

For population variance

$$\sigma^2 = \frac{\sum\left(x - \bar{X}\right)^2}{N}$$

For sample variance

$$s^2 = \frac{\sum\left(x - \bar{X}\right)^2}{n - 1}$$

Example 1- **Variance**

	Score X	$X - \overline{X}$	$(X - \overline{X})^2$
1	3		
2	5		
3	7		
4	10		
5	10		
Total	**35**		

The mean is 35/5 = 7.

Example 1- **Variance**

	Score X	$X - \overline{X}$	$(X - \overline{X})^2$
1	3	3-7=-4	
2	5	5-7=-2	
3	7	7-7=0	
4	10	10-7=3	
5	10	10-7=3	
Totals	35		

Example 1- **Variance**

	Score X	$X-\overline{X}$	$(X-\overline{X})^2$
1	3	3-7=-4	16
2	5	5-7=-2	4
3	7	7-7=0	0
4	10	10-7=3	9
5	10	10-7=3	9
Totals	35		38

Example 1- **Variance**

	Score X	$x-\overline{x}$	$(x-\overline{x})^2$
1	3	3-7=-4	16
2	5	5-7=-2	4
3	7	7-7=0	0
4	10	10-7=3	9
5	10	10-7=3	9
Totals	35		38

$$s^2 = \frac{\sum\left(x-\bar{X}\right)^2}{n} = \frac{38}{5} = 7.6$$

Example 1- **Variance**

	Score X	$x - \bar{x}$	$(x - \bar{x})^2$
1	7	7-7=0	0
2	7	7-7=0	0
3	7	7-7=0	0
4	7	7-7=0	0
5	7	7-7=0	0
Totals	35		0

$$s^2 = \frac{\sum\left(x - \bar{X}\right)^2}{n} = \boxed{\mathbf{0/5 = 0}}$$

Example 2- **Variance**

Drive	**Mark**	**Myrna**
1	28	27
2	22	27
3	21	28
4	26	6
5	18	27

Which diver was more consistent?

Example 2- **Variance**

Dive	Mark's Score X	$X - \overline{X}$	$(X - \overline{X})^2$
1	28	5	25
2	22	-1	1
3	21	-2	4
4	26	3	9
5	18	-5	25
Totals	**115**	**0**	**64**

Mark's Variance = 64 / 5 = 12.8
Myrna's Variance = 362 / 5 = 72.4

Conclusion: Mark has a lower variance therefore he is more consistent.

standard deviation - a measure of variation of scores about the mean

- Can think of standard deviation as the average distance to the mean
- Higher standard deviation indicates higher spread, less consistency, and less clustering.

- sample standard deviation: $s = \sqrt{\frac{\sum\left(x - \bar{X}\right)^2}{n-1}}$

- population standard deviation: $\sigma = \sqrt{\frac{\sum\left(x - \mu\right)^2}{N}}$

Example – **Standard Deviation**

Dive	Mark's Score X	$X - \overline{X}$	$(X - \overline{X})^2$
1	28	5	25
2	22	-1	1
3	21	-2	4
4	26	3	9
5	18	-5	25
Totals	115	0	64

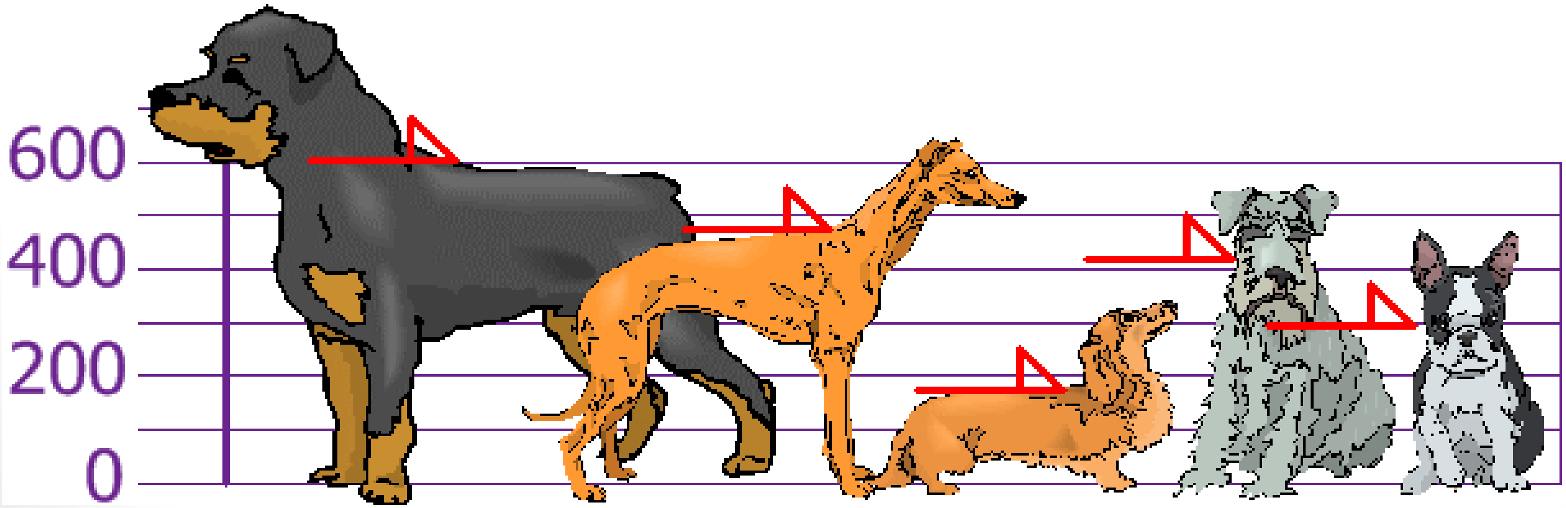
Mark's Variance = 64 / 5 = 12.8

Mark's Standard Deviation for population = $\sqrt{\frac{12.8}{5}} = 1.6$

Mark's Standard Deviation for sample $\sqrt{\frac{12.8}{4}} = 1.78$

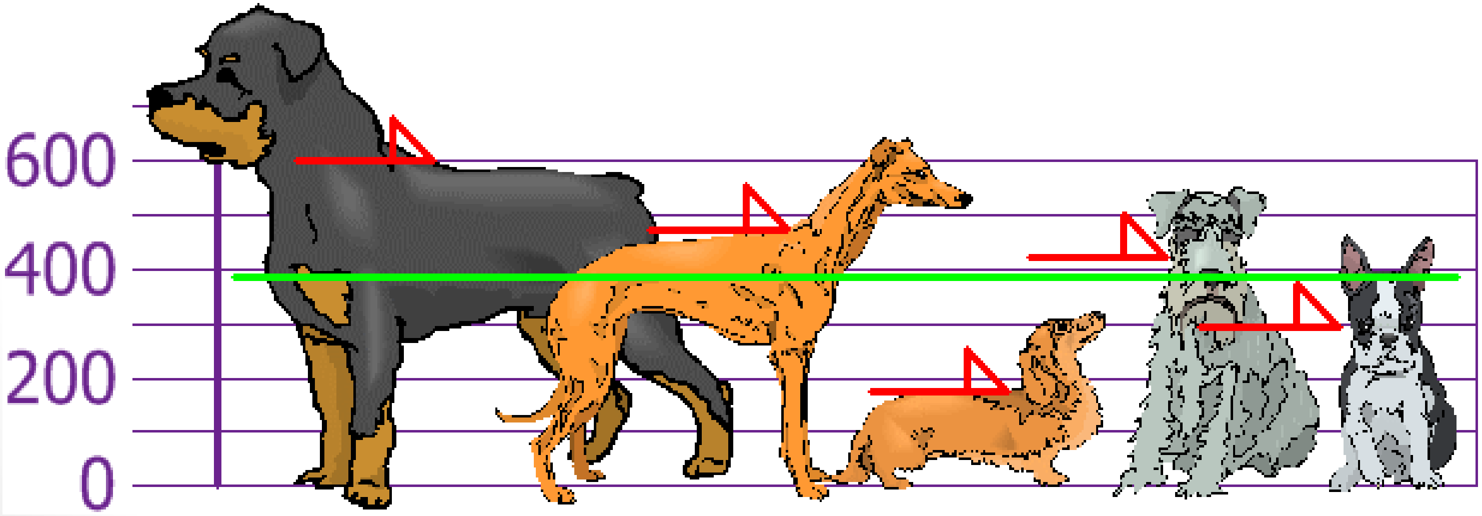
Example- Variance & Standard Deviation

- You have just measured the heights of your dogs (in mm)
- **The heights (at the shoulders) are: 600mm, 470mm, 170mm, 430mm and 300mm.**
- Find out the Mean, the Variance, and the Standard Deviation.



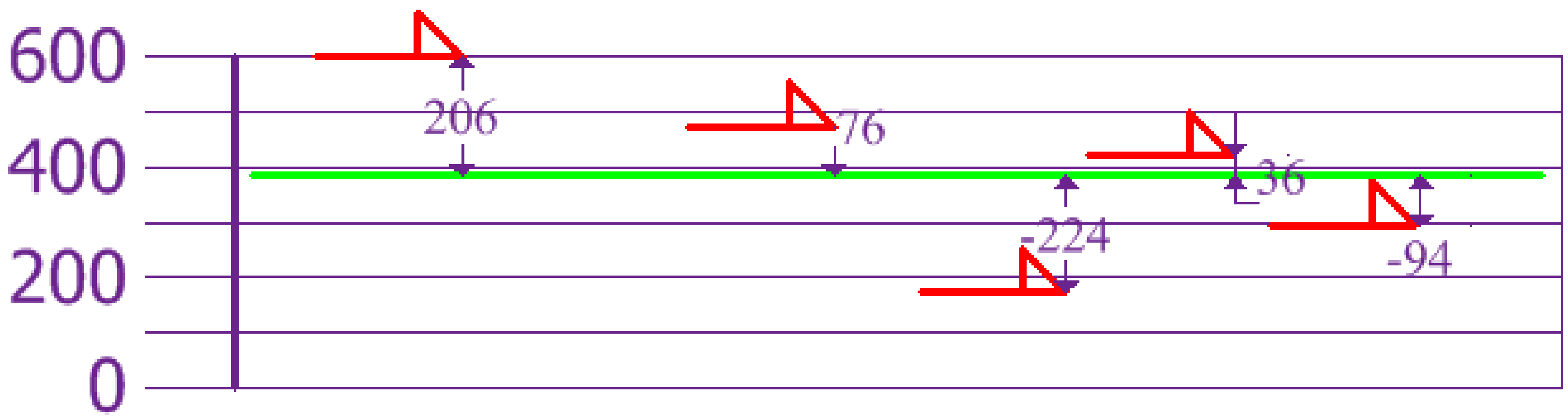
Example- **Variance & Standard Deviation**

- Your first step is to find the Mean:
- Mean = (600 + 470 + 170 + 430 + 300)5
- Mean = 1970/5
- **Mean = 394**



Example- **Variance & Standard Deviation**

- Now we calculate each dog's difference from the Mean



Example- Variance & Standard Deviation

- To calculate the Variance, take each difference, square it, and then average the result:
- Variance

$$\sigma^2 = 206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2 / \mathbf{5}$$

$$= 42436 + 5776 + 50176 + 1296 + 8836 / \mathbf{5}$$

$$= 108520 / \mathbf{5}$$

$$= 21704$$

- **So the Variance σ^2 is 21,704**

Example- Variance & Standard Deviation

- And the Standard Deviation is just the square root of Variance, so:
- **Standard Deviation**

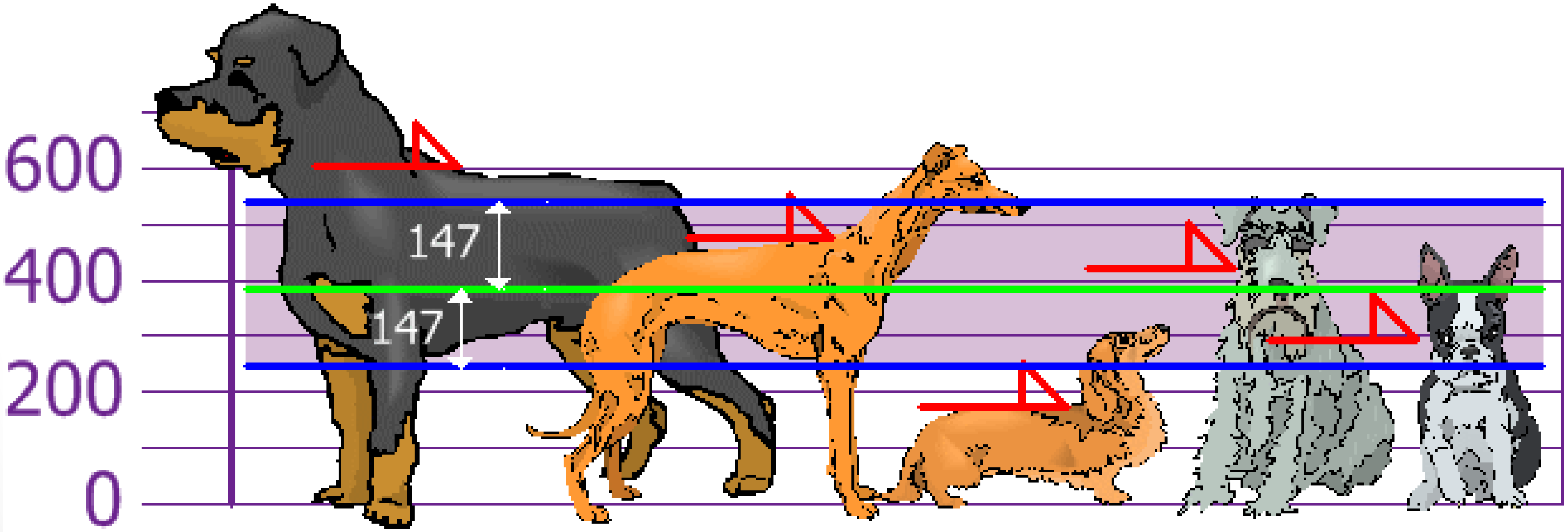
$$\sigma = \sqrt{21704}$$

$$= 147.32...$$

$$= \mathbf{147} \text{ (to the nearest mm)}$$

Example- Variance & Standard Deviation

- And the good thing about the Standard Deviation is that it is useful. Now we can show which heights are within one Standard Deviation (147mm) of the Mean:



- **So, using the Standard Deviation we have a "standard" way of knowing what is *normal*, and what is *extra large* or *extra small*.**

difference of means

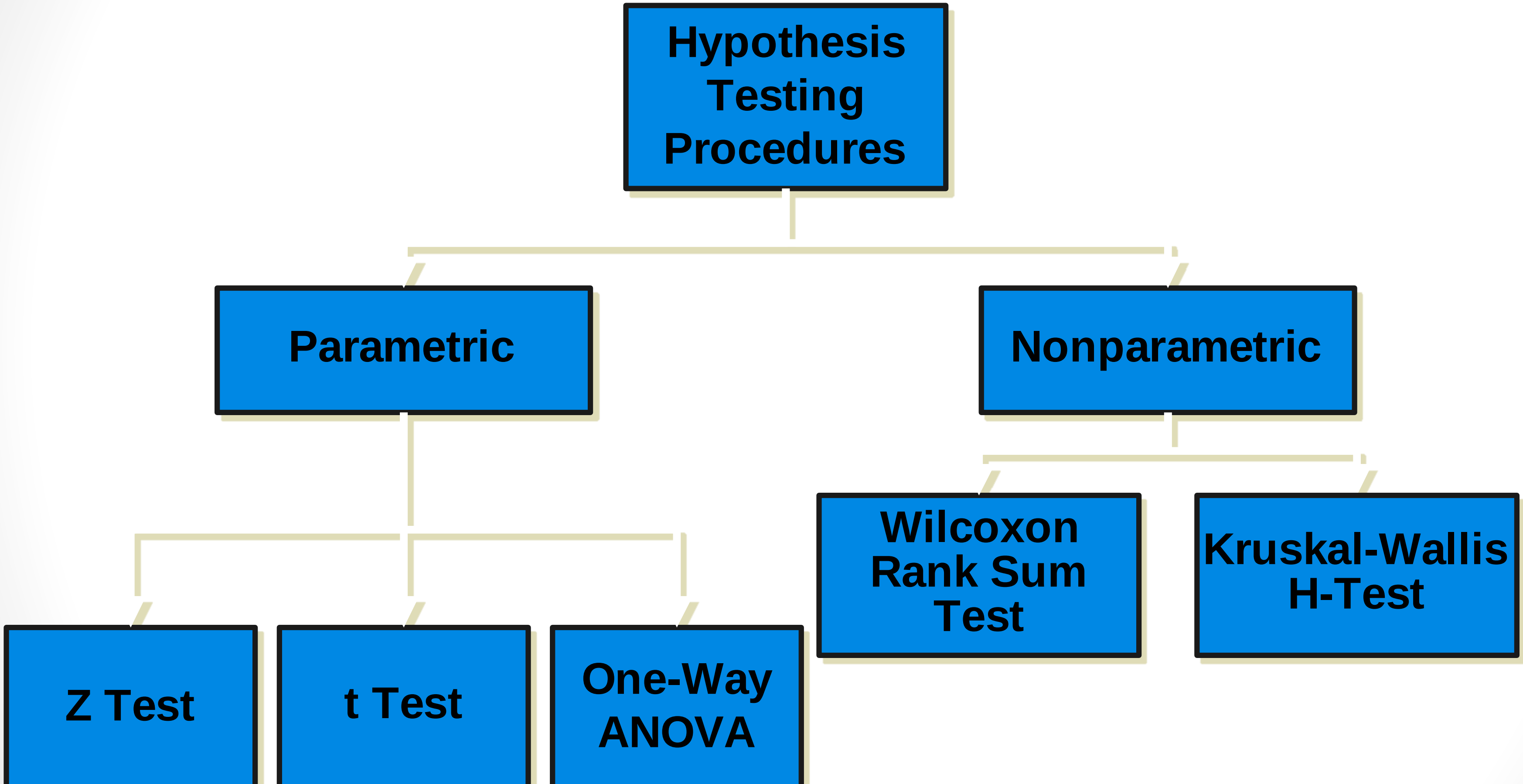
State the hypotheses

Formulate an analysis plan

Analyze sample data using **hypothesis test**

Interpret results.

Hypothesis Testing Procedures



Many More Tests Exist!

Parametric Test Procedures

1. Involve Population Parameters (Mean)
2. Have Stringent(strict) Assumptions (Normality)
3. Examples: Z Test, t Test, c^2 Test, F test

Nonparametric Test Procedures

1. Do Not Involve Population Parameters

- Example: Probability Distributions, Independence

2. Data Measured on Any Scale (Ratio or Interval, Ordinal or Nominal)

3. Example: Wilcoxon Rank Sum Test

Parametric Test Procedures

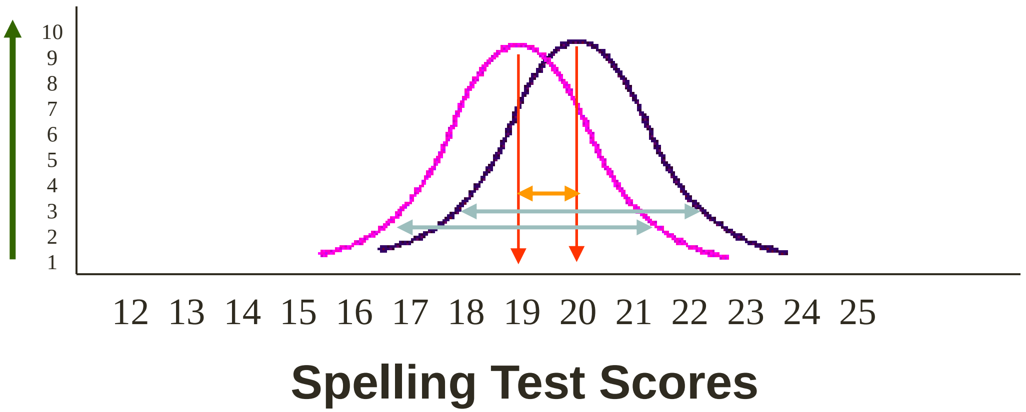
A *t* test allows us to compare the means of two groups

The calculations for a *t* test requires three pieces of information:

- the difference between the means (mean difference)

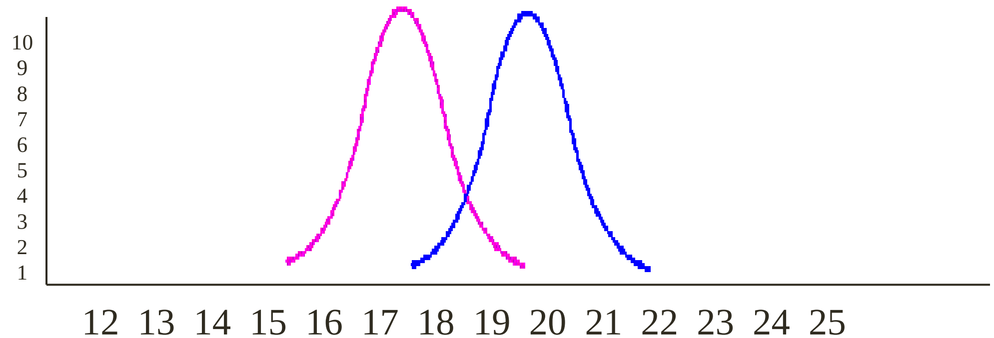
- the standard deviation for each group

- and the number of subjects(samples) in each group.

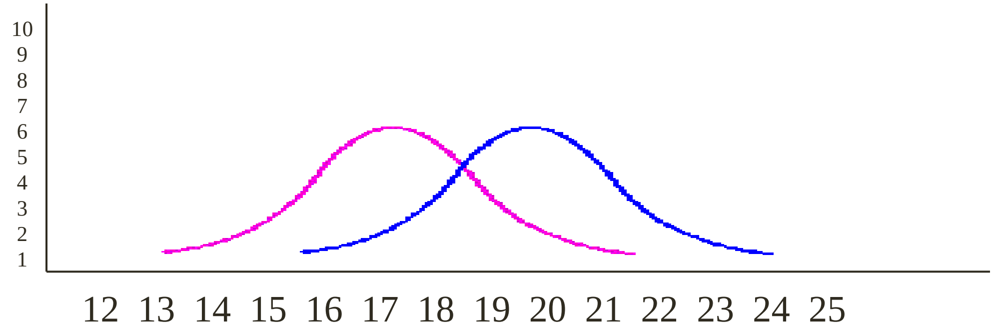


The size of the standard deviation also influences the outcome of a *t* test.

Given the same difference in means, groups with smaller standard deviations are more likely to report a significant difference than groups with larger standard deviations.

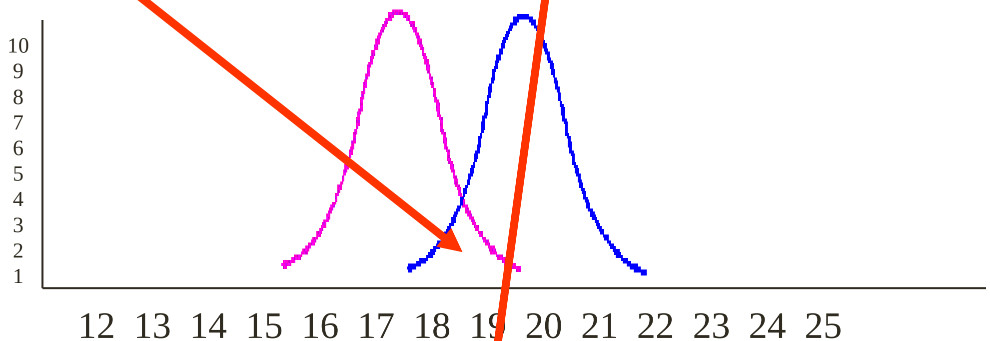


Spelling Test Scores

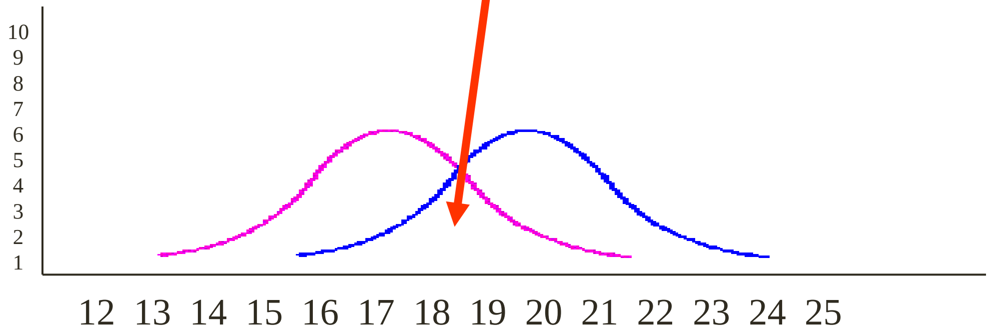


Spelling Test Scores

From a practical standpoint, we can see that smaller standard deviations produce less overlap between the groups than larger standard deviations. ***Less overlap would indicate that the groups are more different from each other.***



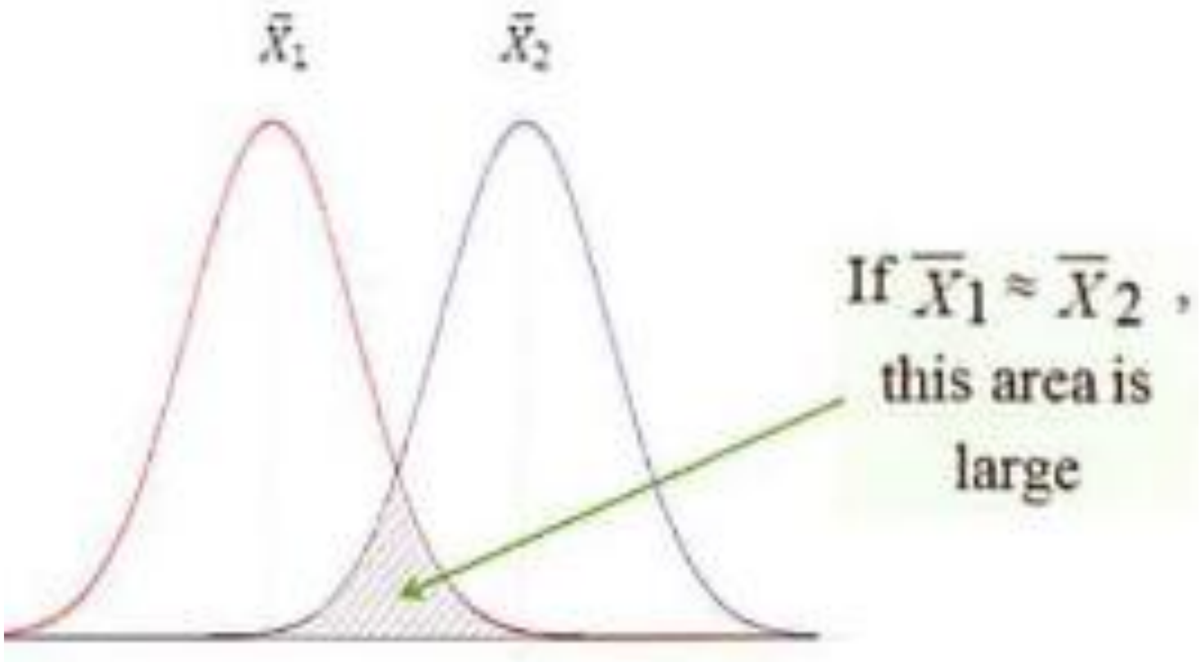
Spelling Test Scores



Spelling Test Scores

Difference of Means

Two populations – same or different?



How do we determine which *t* test to use…

Are the scores for the two means from the same subject (or related subjects)?

Yes → **Paired *t* test** (Dependent t-test; Correlated t-test)

No → **Are there the same number of people in the two groups?**

Yes → **Equal Variance Independent *t* test** (Pooled Variance Independent t-test)

No → **Are the variances of the two groups same?**

yes (Significance Level for Levene (or F-Max) is $p<.05$) → **Equal Variance Independent *t* test** (Pooled Variance Independent *t* test)

No (Significance Level for Levene (or F-Max) is $p>=.05$) → **Unequal Variance Independent t-test** (Separate Variance Independent *t* test)

Difference of Means
Two Parametric Methods

Student's t-test

- Assumes two normally distributed populations, and that they have equal variance

Welch's t-test

- Assumes two normally distributed populations, and they don't necessarily have equal variance

Student’s t-test

Student’s t-test assumes that distributions of the two populations have **equal but unknown variances**.

Suppose **n1 and n2 samples** are randomly and independently selected from **two populations, pop1 and pop1**, respectively.

If each population is normally distributed with the **same mean ($\mu 1=\mu 2$) and with the same variance**, then

T (the t-statistic), follows a t-distribution with **degrees of freedom (df)**

Student's t-test

$$T = \frac{\bar{x}_1 - \bar{x}_2}{Sp\sqrt{\frac{1}{n1} + \frac{1}{n2}}}$$

Where

$$Sp = \frac{(n1-1)S1^2 + (n2-1)S2^2}{n1+n2-2}$$

- significance level $\boldsymbol{\alpha = 0.05}$
- degree of freedom **df =n1+n2-2**
- **T*- is critical value found using df (from table)**

Student's t-test

$$T = \frac{\bar{X}1 - \bar{X}2}{Sp\sqrt{\frac{1}{n1} + \frac{1}{n2}}}$$

Where

$$Sp = \frac{(n1-1)s1^2 + (n2-1)s2^2}{n1+n2-2}$$

If T> =T*
the null hypothesis
is rejected

- Sp is pooled variance
- significance level $\boldsymbol{\alpha = 0.05}$
- degree of freedom **df =n1+n2-2**
- **T*- is critical value found using df (from table)**

Welch’s t-test

When the equal population variance assumption is not justified in performing Student’s t-test for the difference of means, Welch’s t-test can be used based on

Also known as unequal variances t-test

Welch's t-test

$$T_{welch} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s1^2}{n1} + \frac{s2^2}{n2}}}$$

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Where $\bar{x}$, s^2, n correspond to the sample mean, sample variance, and sample size.

Notice that Welch's t-test uses the sample variance (s^2) for each population instead of the pooled sample variance.

Example

t-test independent samples

Example

Some brown hairs were found on the clothing of a victim at a crime scene.

The five of the hairs were measured: 46, 57, 54, 51, 38 μm.

A suspect is the owner of a shop with similar brown hairs. A sample of the hairs has been taken and their widths measured: 31, 35, 50, 35, 36 μm.

Is it possible that the hairs found on the victim were left by the suspect‟s? Test at the %5 level.

[From D. Lucy *Introduction to Statistics for Forensic Scientists* Chichester: Wiley, 2005 p. 44.]

t-test independent samples

1. Calculate the mean and standard deviation for the data sets

t-test independent samples

1. Calculate the mean and standard deviation for the data sets

	A	B
	46	31
	57	35
	54	50
	51	35
	38	36
Total		
Mean		
Standard deviation		

t-test independent samples

1. Calculate the mean and standard deviation for the data sets

	Dog A	Dog B
	46	31
	57	35
	54	50
	51	35
	38	36
Total	**246**	**187**
Mean	**49.2**	**37.4**
Standard deviation	**7.463**	**7.301**

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means

$$49.2 - 37.4 = 11.8$$

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means.
3. Calculate the standard error in the difference

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\sqrt{\frac{7.463^2}{5} + \frac{7.301^2}{5}} = \sqrt{18.56\ldots + 10.66\ldots}$$

$= 4.669 \approx 4.67$ (3 sf)

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *T*

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of T

T = difference between the means ÷ standard error in the difference

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *T*:

T = difference between the means ÷ standard error in the difference

$11.8 \div 4.669 = 2.527$

≈ 2.53 (3 sig fig)

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *T*:
5. Calculate the degrees of freedom

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *T*:
5. Calculate the degrees of freedom = $n_1 + n_2 - 2$

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *T*:
5. Calculate the degrees of freedom = $n_1 + n_2 - 2$

$$5 + 5 - 2 = 8$$

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of T
5. Calculate the degrees of freedom
6. Find the critical value for the particular significance you are working to from the table

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of T:
5. Calculate the degrees of freedom
6. Find the critical value T* for the particular significance you are working to

from the table

	PROPORTION IN ONE TAIL			
	0.25	0.10	0.05	0.025
	PROPORTION IN TWO TAILS COMBINED			
df	0.50	0.20	0.10	0.05
1	1.000	3.078	6.314	12.706
2	0.816	1.886	2.920	4.303
3	0.765	1.638	2.353	3.182
4	0.741	1.533	2.132	2.776
5	0.727	1.476	2.015	2.571
6	0.718	1.440	1.943	2.447
7	0.711	1.415	1.895	2.365
8	0.706	1.397	1.860	2.306
9	0.703	1.383	1.833	2.262

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *t*:
5. Calculate the degrees of freedom
6. Find the critical value for the particular significance you are working to and find the critical value from the table

	PROPORTION IN ONE TAIL			
	0.25	0.10	0.05	0.025
	PROPORTION IN TWO TAILS COMBINED			
df	0.50	0.20	0.10	0.05
1	1.000	3.078	6.314	12.706
2	0.816	1.886	2.920	4.303
3	0.765	1.638	2.353	3.182
4	0.741	1.533	2.132	2.776
5	0.727	1.476	2.015	2.571
6	0.718	1.440	1.943	2.447
7	0.711	1.415	1.895	2.365
8	0.706	1.397	1.860	2.306
9	0.703	1.383	1.833	2.262

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *t*:
5. Calculate the degrees of freedom
6. Find the critical value for the particular significance you are working to and find the critical value from the table

	PROPORTION IN ONE TAIL			
	0.25	0.10	0.05	0.025
	PROPORTION IN TWO TAILS COMBINED			
df	0.50	0.20	0.10	0.05
1	1.000	3.078	6.314	12.706
2	0.816	1.886	2.920	4.303
3	0.765	1.638	2.353	3.182
4	0.741	1.533	2.132	2.776
5	0.727	1.476	2.015	2.571
6	0.718	1.440	1.943	2.447
7	0.711	1.415	1.895	2.365
8	0.706	1.397	1.860	2.306
9	0.703	1.383	1.833	2.262

t-test independent samples

1. Calculate the mean and standard deviation for the data sets

2. Calculate the magnitude of the difference between the two means

3. Calculate the standard error in the difference

4. Calculate the value of *t*:

5. Calculate the degrees of freedom

6. Find the critical value for the particular significance you are working to from the table

	PROPORTION IN ONE TAIL			
	0.25	0.10	0.05	0.025
	PROPORTION IN TWO TAILS COMBINED			
df	0.50	0.20	0.10	0.05
1	1.000	3.078	6.314	12.706
2	0.816	1.886	2.920	4.303
3	0.765	1.638	2.353	3.182
4	0.741	1.533	2.132	2.776
5	0.727	1.476	2.015	2.571
6	0.718	1.440	1.943	2.447
7	0.711	1.415	1.895	2.365
8	0.706	1.397	1.860	2.306
9	0.703	1.383	1.833	[illegible]

At the 0.05 level $t_{crit} = 2.306$

t-test independent samples

1. Calculate the mean and standard deviation for the data sets
2. Calculate the magnitude of the difference between the two means
3. Calculate the standard error in the difference
4. Calculate the value of *t*:
5. Calculate the degrees of freedom
6. Find the critical value for the particular significance you are working to and find the critical value from the table

If T < T* (critical value) then there is no significant difference between the two sets of data ,i.e. null hypothesis is Accepted

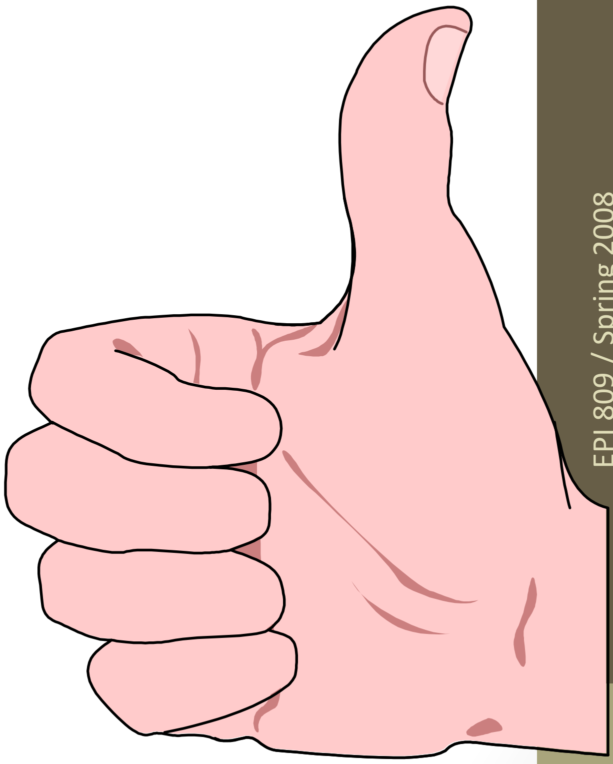
If *T* >=T* (critical value) then there is a significant difference between the two sets of data i.e. null hypothesis is Rejected

Statistical Methods for Evaluation-

- Hypothesis testing,
- difference of means,
- wilcoxon rank–sum test,
- type 1 type 2 errors,
- power and sample size,
- ANNOVA

Advantages of Nonparametric Tests

1. Used With All Scales
2. Easier to Compute
3. Make Fewer Assumptions
4. Need Not Involve Population Parameters
5. Results May Be as Exact as Parametric Procedures



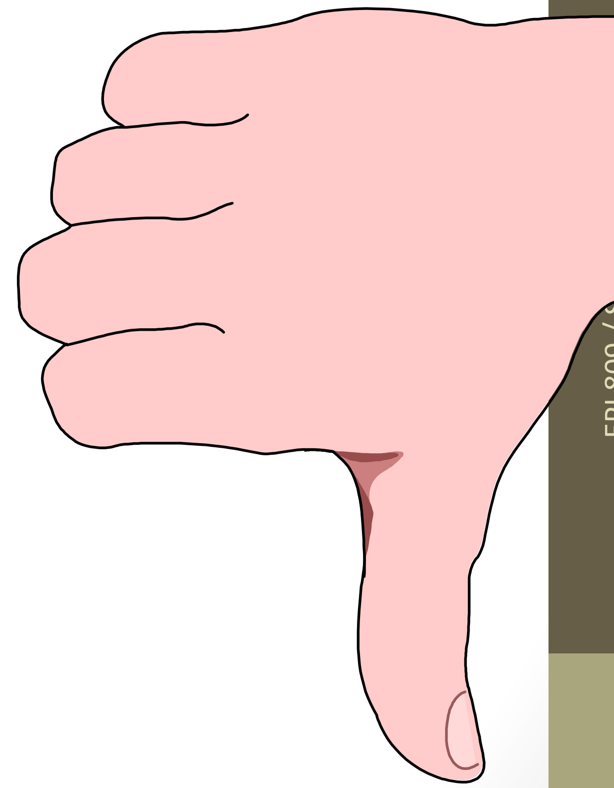
Disadvantages of Nonparametric Tests

1. May Waste Information

   Parametric model more efficient if data Permit

2. Difficult to Compute by hand for Large Samples

3. Tables Not Widely Available



Popular Nonparametric Tests

1. Sign Test
2. Wilcoxon Rank Sum Test
3. Wilcoxon Signed Rank Test

Wilcoxon Rank Sum Test

Wilcoxon Rank-Sum Test
A Nonparametric Method

- Makes no assumptions about the underlying probability distributions

Wilcoxon Rank Sum Test

1. Tests Two Independent Population Probability Distributions

2. Corresponds to t-Test for 2 Independent Means

3. Assumptions
    - Independent, Random Samples
    - Populations Are Continuous

4. Can Use Normal Approximation If $n_i \geq 10$

Wilcoxon Rank Sum Test
Procedure

1. Assign Ranks, R_i, to the $n_1 + n_2$ Sample Observations

   If Unequal Sample Sizes, Let n_1 Refer to Smaller-Sized Sample

   Smallest Value = 1

2. Sum the Ranks, T_i, for Each Sample

Test Statistic Is T_A (Smallest Sample)

Null hypothesis: both samples come from the same underlying distribution

Distribution of T is not quite as simple as binomial, but it can be computed

Wilcoxon Rank Sum Test **Example**

- You're a production planner.
- You want to see if the operating rates for 2 factories is the same.
- For factory 1, the rates are
  - **71**, **82**, **77**, **92**, **88**.
- For factory 2, the rates are
  - **85**, **82**, **94** & **97**.
- Do the factory rates have the same **probability distributions** at the **.05** level?

Wilcoxon Rank Sum Test Solution

- H_0:
- H_a:
- α =
- n_1 = n_2 =
- Critical Value(s):

Test Statistic:

Decision:

Conclusion:

Σ Ranks

Wilcoxon Rank Sum Test Solution

- H_0: Identical Distrib.
- H_a: Shifted Left or Right
- α =
- n_1 = n_2 =
- Critical Value(s):

Test Statistic:

Decision:

Conclusion:

Σ Ranks

Wilcoxon Rank Sum Test Solution

- H_0: Identical Distrib.
- H_a: Shifted Left or Right
- α = .05
- $n_1 = 4 \quad n_2 = 5$
- Critical Value(s):

Test Statistic:

Decision:

Conclusion:

Σ Ranks

Wilcoxon Rank Sum
Table 12 (Rosner) (Portion)

α = .05 two-tailed

		n_1						
		4		5		6		..
		T_L	T_U	T_L	T_U	T_L	T_U	..
	4	10	26	16	34	23	43	..
n_2	5	11	29	17	38	24	48	..
	6	12	32	18	42	26	52	..
	:	:	:	:	:	:	:	:

Wilcoxon Rank Sum Test Solution

- H_0: Identical Distrib.
- H_a: Shifted Left or Right
- $\alpha = .10$
- $n_1 = 4 \quad n_2 = 5$
- Critical Value(s):

Reject	Do Not Reject	Reject

12   28   Σ Ranks

Test Statistic:

Decision:

Conclusion:

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71		85	
82		82	
77		94	
92		97	
88		…	…
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	
82		82	
77		94	
92		97	
88		…	…
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	
82		82	
77	2	94	
92		97	
88		…	…
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	
82	3	82	4
77	2	94	
92		97	
88		…	…
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	
82	~~3~~ 3.5	82	~~4~~ 3.5
77	2	94	
92		97	
88		…	…
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	5
82	~~3~~ 3.5	82	~~4~~ 3.5
77	2	94	
92		97	
88		...	...
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	5
82	~~3~~ 3.5	82	~~4~~ 3.5
77	2	94	
92		97	
88	6	…	…
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	5
82	~~3~~ 3.5	82	~~4~~ 3.5
77	2	94	
92	7	97	
88	6	...	...
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	5
82	~~3~~ 3.5	82	~~4~~ 3.5
77	2	94	8
92	7	97	
88	6	...	...
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	5
82	~~3~~ 3.5	82	~~4~~ 3.5
77	2	94	8
92	7	97	9
88	6	...	...
Rank Sum			

Wilcoxon Rank Sum Test Computation Table

Factory 1		Factory 2	
Rate	Rank	Rate	Rank
71	1	85	5
82	~~3~~ 3.5	82	~~4~~ 3.5
77	2	94	8
92	7	97	9
88	6	...	...
Rank Sum	19.5		25.5

Wilcoxon Rank Sum Test Solution

- H_0: Identical Distrib.
- H_a: Shifted Left or Right
- $\alpha = .05$
- $n_1 = 4 \quad n_2 = 5$
- Critical Value(s):

Reject	Do Not Reject	Reject

12 28 Σ Ranks

Test Statistic:

$T_2 = 5 + 3.5 + 8 + 9 = 25.5$

(Smallest Sample)

Decision:

Conclusion:

Wilcoxon Rank Sum Test Solution

- H_0: Identical Distrib.
- H_a: Shifted Left or Right
- α = .05
- $n_1 = 4$ $n_2 = 5$
- Critical Value(s):

Reject	Do Not Reject	Reject

12 28 Σ Ranks

Test Statistic:

$T_2 = 5 + 3.5 + 8 + 9 = 25.5$

(Smallest Sample)

Decision:

Do Not Reject at α = .05

Conclusion:

Wilcoxon Rank Sum Test Solution

- H_0: Identical Distrib.
- H_a: Shifted Left or Right
- α = .05
- $n_1 = 4 \quad n_2 = 5$
- Critical Value(s):

Reject	Do Not Reject	Reject

12 28 Σ Ranks

Test Statistic:

$T_2 = 5 + 3.5 + 8 + 9 = 25.5$

(Smallest Sample)

Decision:

Do Not Reject at α = .05

Conclusion:

There is No evidence for unequal distrib

Statistical Methods for Evaluation-

- Hypothesis testing,
- difference of means,
- wilcoxon rank–sum test,
- type 1 type 2 errors,
- power and sample size,
- ANNOVA

Type I and Type II errors

Type I error refers to the situation when ***we reject the null hypothesis when it is true*** (H_0 is wrongly rejected). **Denoted by α**

Type II error refers to the situation when ***we accept the null hypothesis when it is false.*** (H_0 is wrongly Accepted). **Denoted by β**

		Conclusion about null hypothesis from statistical test	
		Accept Null	Reject Null
Truth about null hypothesis in population	True	Correct	Type I error Observe difference when none exists
	False	Type II error Fail to observe difference when one exists	Correct

Type I and Type II errors

Which one is more dangerous Type I or Type II error ? Justify your answer.

Statistical Methods for Evaluation-

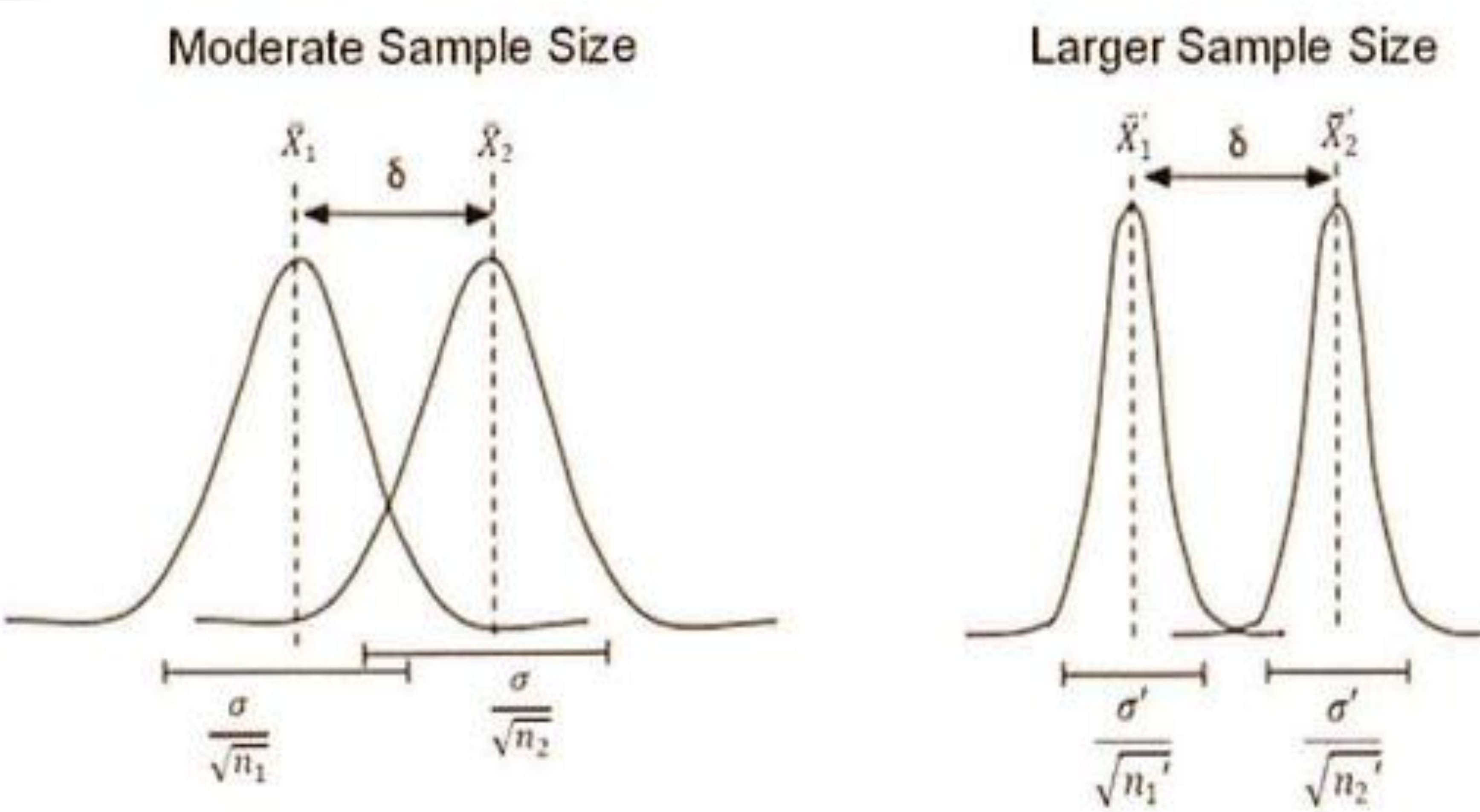
- Hypothesis testing,
- difference of means,
- wilcoxon rank–sum test,
- type 1 type 2 errors,
- power and sample size,
- ANNOVA

Power and Sample Size

- The *power of a test* is the *probability of correctly rejecting the null hypothesis*
- It is denoted by β , where *(1- β)* is the probability of a *type II error*.
- The power of a test improves as the sample size increases
- power is used to determine the necessary sample size.
- power of a hypothesis test depends on the true difference of the population means.
- A larger sample size is required to detect a smaller difference in the means.
- In general, *Effect size* δ = difference between the means
- It is important to consider an appropriate effect size for the problem at hand

Power and Sample Size

A larger sample size better identifies a fixed effect size



Statistical Methods for Evaluation-

- Hypothesis testing,
- difference of means,
- wilcoxon rank–sum test,
- type 1 type 2 errors,
- power and sample size,
- ANNOVA

ANOVA (Analysis of Variance)

A generalization of the hypothesis testing of the difference of two population means

Good for analyzing more than two populations

ANOVA tests if any of the population means differ from the other population means

ANOVA (Analysis of Variance)

Find the **mean for each of the groups**.

Find the **overall mean** (the mean of the groups combined).

Find the **Within Group Variation**; the total deviation of each member's score from the Group Mean.

Find the **Between Group Variation**: the deviation of each Group Mean from the Overall Mean.

Find the **F critical** and **F statistic**: the ratio of Between Group Variation to Within Group Variation.

F statistic < F critical accept Ho else reject H0 and accept Ha

References

- https://www.slideshare.net/darlingjunior/hypothesis-testing?from_action=save
- https://www.mathsisfun.com/data/standard-deviation.html
- *http:/www2.aueb.gr/users/koundouri/resees/uploads/Chapter10.ppt*
- *https://researchbasics.education.uconn.edu/wp-content/uploads/sites/1215/.../ttest.pps*
- *https://msu.edu/~fuw/teaching/Fu_ch9_Nonpara.ppt*
- http://www.statisticshowto.com/probability-and-statistics/t-test/