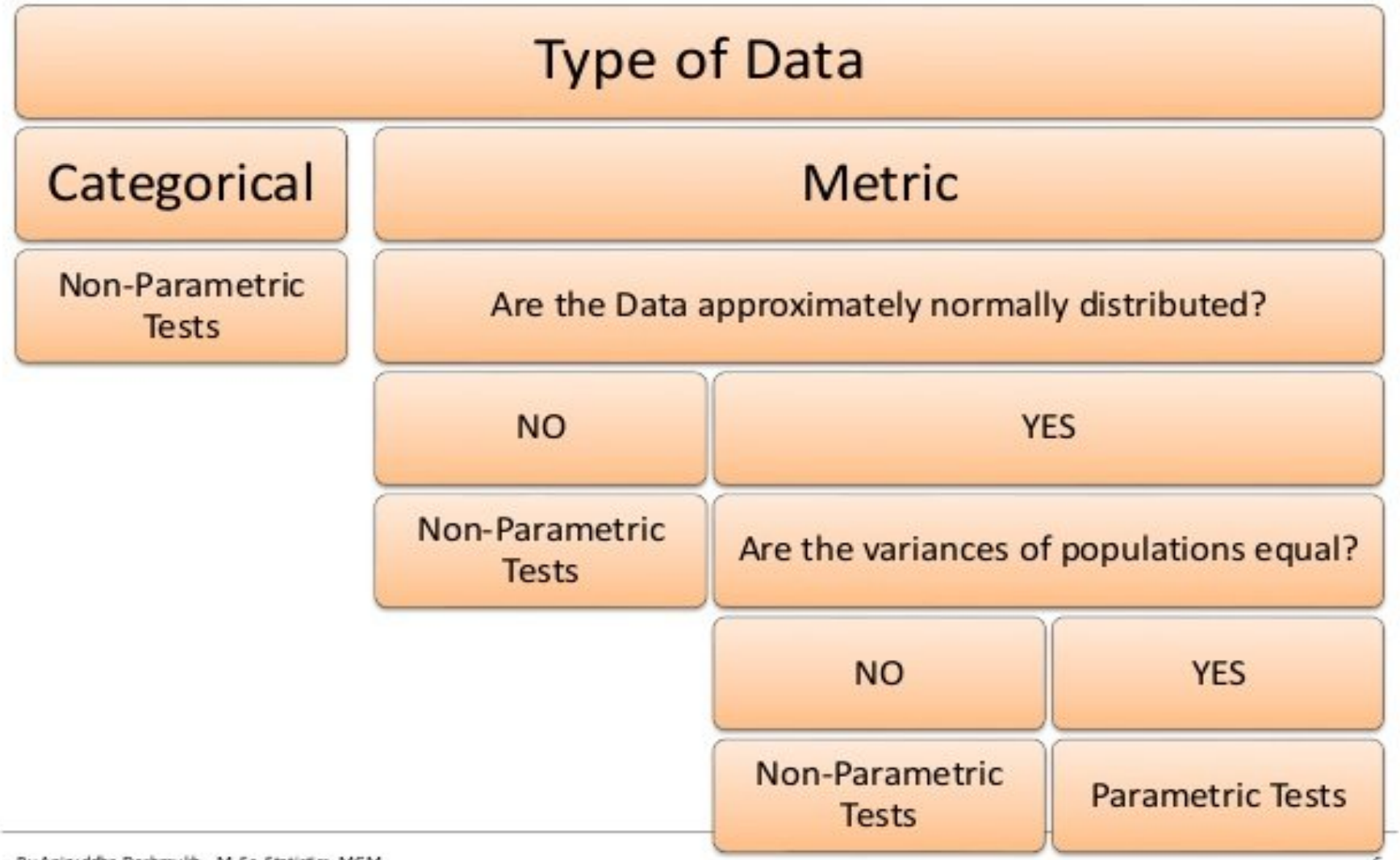
Biostatistics

Sneha prkash v
Asst. professor
DAMCOP

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 (how likely is it that our hypothesized parameter is correct. To test the validity of our assumption we determine the difference between the hypothesized parameter value and the sample value.)

Parametric or Non-Parametric Determination



Conclusive Thoughts

	Parametric	Non-parametric
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Any
Typical data	Ratio or Interval	Ordinal or Nominal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers
Tests		
Choosing	Choosing parametric test	Choosing a non-parametric test
Correlation test	Pearson	Spearman
Independent measures, 2 groups	Independent-measures t-test	Mann-Whitney test
Independent measures, >2 groups	One-way, independent-measures ANOVA	Kruskal-Wallis test
Repeated measures, 2 conditions	Matched-pair t-test	Wilcoxon test
Repeated measures, >2 conditions	One-way, repeated measures ANOVA	Friedman's test

Background

- Introduced in 1908 by William Sealy Gosset.
- Gosset published his mathematical work under the pseudonym “Student”.
- Definition of t test: “ It’s a method of testing hypothesis about the mean of small sample drawn from a normally distributed population.”

Assumptions of t-Test

- Dependent variables are interval or ratio.
- The population from which samples are drawn is normally distributed.
- Samples are randomly selected.
- The groups have equal variance (Homogeneity of variance).
- The t-statistic is robust (it is reasonably reliable even if assumptions are not fully met.)

Applications of t test

- The calculation of a confidence interval for a sample mean.
- To test whether a sample mean is different from a hypothesized value.
- To compare mean of two samples.
- To compare two sample means by group.

Types of “t” test

- Single sample t test – we have only 1 group; want to test against a hypothetical mean.
- Independent samples t test – we have 2 means, 2 groups; no relation between groups, Eg: When we want to compare the mean of T/T group with Placebo group.
- Paired t test – It consists of samples of matched pairs of similar units or one group of units tested twice. Eg: Difference of mean pre & post drug intervention.

One Sample t-test

- It is used in measuring whether a sample value significantly differs from a hypothesized value.
- For example, a research scholar might hypothesize that on an average it takes 3 minutes for people to drink a standard cup of coffee.
- He conducts an experiment and measures how long it takes his subjects to drink a standard cup of coffee.
- The one sample t-test measures whether the mean amount of time it took the experimental group to complete the task varies significantly from the hypothesized 3 minutes value.

Equation for a one-sample t-test

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

where

t = the t statistic

$\bar{x}$ = the mean of the sample

μ = the comparison mean

s = the sample standard deviation

n = the sample size

One sample t test

- Q. 10 individuals had taken an exam and we want to test whether their scores, all together, are significantly different from the score of 100

- Set up null hypothesis and alternative hypothesis

Example

- 10 individuals had taken an exam and we want to test whether their scores, all together, are significantly different from the score of 100.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{107.8 - 100}{5.35/\sqrt{10}} = 4.61$$

•We need to calculate the degrees of freedom.
•Here, the degrees of freedom is simply the sample size minus one.
•Therefore, Degrees of freedom = n – 1 = 10 – 1 = 9
•Now, we will refer to a t table to determine the critical t value for 9 degrees of freedom at the .05 level of significance.
•Looking at a t table, this value is 2.26 .
•Since our calculated t value of 4.61 is greater than the critical t value of 2.26, we can say that the scores of our sample of 10 individuals differ significantly from the score of 100.

T table

df	.05	.01	.001
1	12.706	63.657	636.619
2	4.303	9.925	31.598
3	3.182	5.841	12.924
4	2.776	4.604	8.610
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587

Independent t test

- • The independent sample t-test consists of tests that compare mean value(s) of continuous-level (interval or ratio data), in a normally distributed data.
- • The independent sample t-test compares two means.
- • The independent samples t-test is also called unpaired t-test/ two sample t test.
- • It is the t-test to be used when two separate independent and identically distributed variables are measured.
- • Eg: 1. Comparision of quality of life improved for patients who took drug Valporate as opposed to patients who took drug Levetiracetam in myoclonic seizures.
- • 2.Comparasion of mean cholesterol levels in treatment group with placebo group after administration of test drug.

Assumptions

- A random sample of each population is used.
- The random samples are each made up of independent observation.
- Each sample is independent of one another.
- The population distribution of each population must be nearly normal, or the size of the sample is large

- Q.The following are the Comparasion of mean cholesterol levels in treatment group after administration of test drug with placebo group

Cholestrol levels in mg/dl	
Test group(x)	Placebo group(y)
100	180
130	200
140	210
155	140
120	130

1.Set up the null hypothesis and alternative hypothesis

H_0=there is no significant difference in the mean cholestrol level after taking test drug compared with placebo $\bar{d}$rug $\bar{x}=\bar{y}$

H_A=there is significant difference in the mean cholestrol level after taking test drug compared with placebo drug $\bar{x}\neq\bar{y}$

This is one tailed test

We are testing t the 95%or 5% (.05)level

1. Set up the null hypothesis and alternative hypothesis
2. Calculate the mean and standard deviation for the data sets

Cholestrol levels in mg/dl	
Test group(x)	Placebo group(y)
100	180
130	200
140	210
155	140
120	130
645	860
129	172
18.54	31.87

total
mean
s.d

1. Set up the null hypothesis and alternative hypothesis
2. Calculate the mean and standard deviation for the data sets
3. Calculate the difference between the means

   172-129=43
4. Calculate the standard error in the difference

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

square root of $\frac{18.54^2}{5} + \frac{32.87^2}{5}$ =16.87

1.Set up the null hypothesis and alternative hypothesis

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

3.Calculate the difference between the means

172-129=43

4. Calculate the standard error in the difference

5.Calculate t valu

$$t = \frac{|\overline{x_1} - \overline{x_2}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t=43/16.87=2.54

1.Set up the null hypothesis and alternative hypothesis

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

3.Calculate the difference between the means

172-129=43

4. Calculate the standard error in the difference

5.Calculate t value

6.Calculate the number of degrees of freedom $df=n_1 + n_2 -2$,

10-2 =8

And find the critical value

1.Set up the null hypothesis and alternative hypothesis
2.Calculate the mean and standard deviation for the data sets
3.Calculate the difference between the means
4. Calculate the standard error in the difference
5.Calculate t value
6.Calculate the number of degrees of freedom df=n_1 + n_2 -2,

10-2

=8

And find the critical value t value =2.306

df	PROPORTION IN TWO TAILS COMBINED 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	
8	0.706	1.397	1.860	2.306
9	0.703	1.383	1.833	

1. Set up the null hypothesis and alternative hypothesis
2. Calculate the mean and standard deviation for the data sets
3. Calculate the difference between the means
4. Calculate the standard error in the difference
5. Calculate t value
6. Calculate the number of degrees of freedom df=$n_1 + n_2 - 2$, 10-2 =8

And find the critical value t value =2.306

Critical value (table value)=2.54 calculated t value=2.306

$T_{critical} < T_{calculated\ value}$

So we reject the null hypothesis and accept alternate hypothesis

there is significant difference in the mean cholestrol level after taking test drug compared with placebo drug

Paired t test

- A paired t-test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample.
- A comparison of two different methods of measurement or two different treatments where the measurements/treatments are applied to the same subjects.

Eg: 1.pre-test/post-test samples in which a factor is measured before and after an intervention,

2.Cross-over trials in which individuals are randomized to two treatments and then the same individuals are crossed-over to the alternative treatment,

3.Matched samples, in which individuals are matched on personal characteristics such as age and sex

Paired t test

Q…sample of 10 subjects were given an antihypertensive drug we want to check blood pressure before and after treatment . We want to find out the effectiveness of the treatment by comparing mean blood pressure pre & post treatment

Systolic Blood pressure in mmof Hg										
Pre treatment	140	150	130	120	130	180	190	190	170	180
Post treatment	120	130	100	110	110	130	140	130	120	130

1.Set up the null hypothesis and alternative hypothesis

H_0=there is no significant difference in the mean BP before and after treatment $\bar{x}=\bar{y}$

H_A=there is significant difference in the mean BP before and after treatment $\bar{x} \neq \bar{y}$

This is one tailed test

We are testing t the 95%or 5% (.05)level

1.Set up the null hypothesis and alternative hypothesis

2.Calculate the difference (di = xi − yi) between the two observations on each pair.

Pre(x)	Post(y)	Difference(x-y)
140	120	20
150	130	20
130	100	30
120	110	10
130	110	20
180	130	50
190	140	50
190	130	60
170	120	50
180	130	50

1.Set up the null hypothesis and alternative hypothesis

2.Calculate the difference (di = xi − yi)between the two observations on each pair.

3.Calculate the mean difference, $\bar{D}$.

pre	post	difference
140	120	20
150	130	20
130	100	30
120	110	10
130	110	20
180	130	50
190	140	50
190	130	60
170	120	50
180	130	50
∑D		360

$$\frac{\sum D}{N} = \frac{360}{10} = 36$$

1.Set up the null hypothesis and alternative hypothesis
2.Calculate the difference (di = xi − yi)between the two observations on each pair.
3.Calculate the mean difference, D.
4.Calculate the standard deviation

pre	post	difference	$\overline{D}$-D	$(\overline{D}-D)^2$
140	120	20	16	256
150	130	20	16	256
130	100	30	6	36
120	110	10	26	676
130	110	20	16	256
180	130	50	-14	196
190	140	50	-14	196
190	130	60	-24	576
170	120	50	-14	196
180	130	50	-14	196
∑D		360	$\sum(D-D)^2$	3096

$$s.d=\sqrt{\frac{\sum(D-D)^2}{n}}$$

s.d=17.59

1. Set up the null hypothesis and alternative hypothesis
2. Calculate the difference (di = xi − yi)between the two observations on each pair.
3. Calculate the mean difference, D.
4. Calculate the standard deviation
5. Calculate the standard error of mean

$$SE = \frac{S.D}{\sqrt{n}} = 17.59/3.16 = 5.56$$

1.Set up the null hypothesis and alternative hypothesis

2.Calculate the difference (di = xi − yi)between the two observations on each pair.

3.Calculate the mean difference, D.

4.Calculate the standard deviation

5.Calculate the standard error of mean

6.calculate the t value

T value=_D/S.E =36/5.56 = 6.47

7.Calculate the number of degrees of freedom df=n-1,10-1 =9

And fin

	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

1.Set up the null hypothesis and alternative hypothesis

2.Calculate the difference (di = xi − yi)between the two observations on each pair.

3.Calculate the mean difference, D.

4.Calculate the standard deviation

5.Calculate the standard error of mean

6.calculate the t value

T value= D/S.E =36/5.56 = 6.47

7.Calculate the number of degrees of freedom df=n-1,10-1 =9

And find the critical value

Critical value (table value)=1.855 calculated t value=5.56

$$T_{criticalvalue} < T_{calculated\ value}$$

1.Set up the null hypothesis and alternative hypothesis

2.Calculate the difference (di = xi − yi)between the two observations on each pair.

3.Calculate the mean difference, D.

4.Calculate the standard deviation

5.Calculate the standard error of mean

6.calculate the t value

T value= D/S.E =36/5.56 = 6.47

7.Calculate the number of degrees of freedom df=n-1,10-1 =9

And find the critical value

Critical value (table value)=1.855 calculated t value=5.56

$$T_{criticalvalue} < T_{calculated\ value}$$

So we reject the null hypothesis and accept alternate hypothesis

There is significant difference in the mean BP values pre and post treatment

Home work

- These are blood sugar values of patients before and after taking the drug .find out whether the drug is effective

Blod glucose level in mg/dl	
Before taking	After taking
170	150
200	190
120	120
180	140
130	110

Home work

- The following are the serum cretinine levels test drug receiving patients and placebo receiving patients ,find out whether the test drug is effective compared to placebo

Serum creatinine values	
Test group	Placebo group
1.6	2.1
1.8	1.5
2.2	1.7
1.4	1.8
1.5	2.9