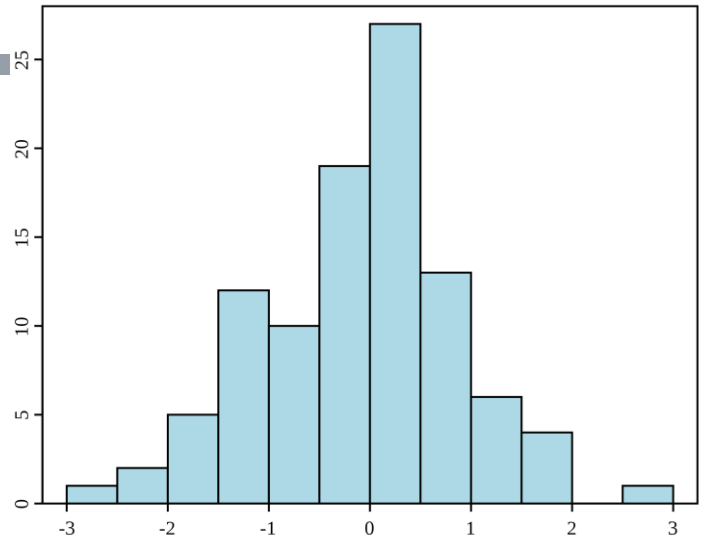
STATISTICS FOR WRITING PAPER

2021. 12. 03.

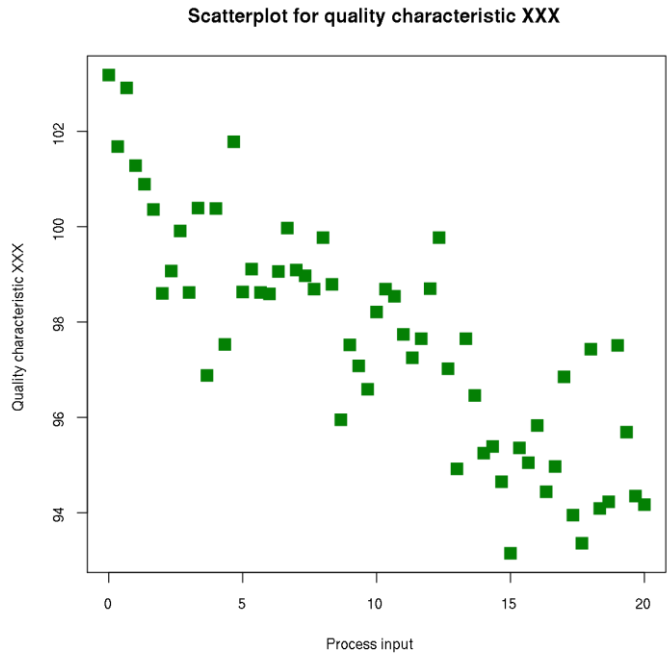
Hyunsoo, Yu

INDEX

1. DESCRIPTIVE STATISTICS

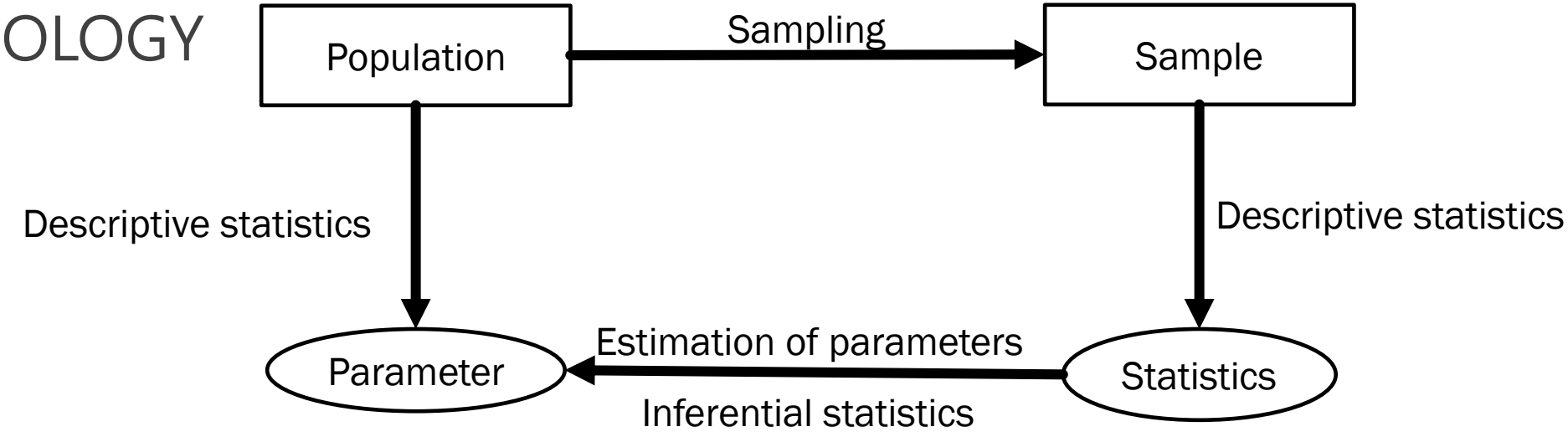


- Statistics which explain the attributes of materials
  - Measure of central tendency
    - Mean, median, mode
  - Variance
    - Standard deviation, quantile
  - Graph
    - Bar graph, histogram
  - Scatter



2. INFERENTIAL STATISTICS

2.1 TERMINOLOGY



Parameter name	Population parameter symbol	Sample statistic
Number of cases	N	n
Mean	μ (mu)	$\bar{x}$ (Sample mean)
Proportion	π (Pi)	P (Sample proportion)
Variance	σ^2 (Sigma-square)	s^2 (Sample variance)
Standard deviation	σ (Sigma)	s (sample standard deviation)
Correlation	ρ (rho)	r (Sample correlation)
Regression Coefficient	β (beta)	b (sample regression coefficient)

2. INFERENTIAL STATISTICS
2.2 HYPOTHESIS TEST

1. Sort
    1. One-sided test (Bigger or smaller)
        1. $H_0: \mu_1 \leq \mu_2,\ H_1: \mu_1 > \mu_2$
        2. $H_0: \mu_1 \geq \mu_2,\ H_1: \mu_1 < \mu_2$
    2. Two-sided test (Different)
        1. $H_0: \mu_1 = \mu_2,\ H_1: \mu_1 \neq \mu_2$
2. Order
    1. Set a 'null hypothesis(H_0)' and a 'alternative hypothesis(H_1)'
    2. Set a significance level (alpha)
    3. Acquire sampling (physical experiment)
    4. Compute 'test statistic' (p-value)
    5. Based on p-value, accept one of 2 hypothesis

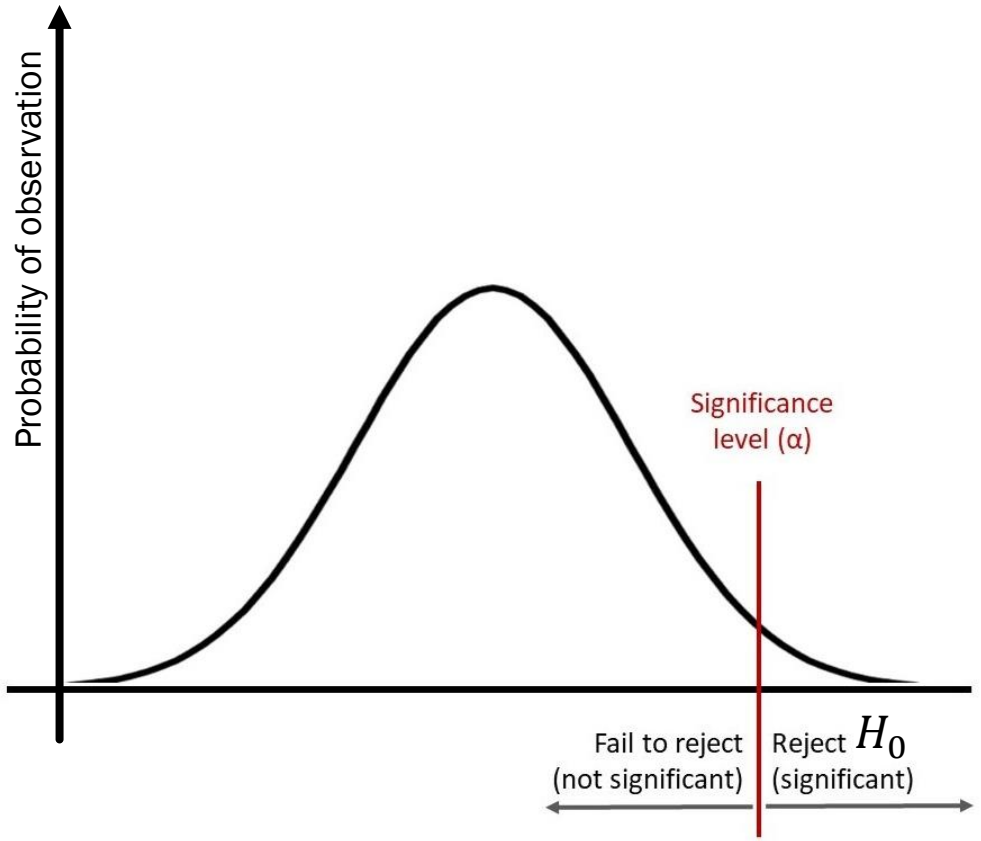
2. INFERENTIAL STATISTICS
2.2 HYPOTHESIS TEST
2.2.1 SET A 'NULL HYPOTHESIS' AND A 'ALTERNATIVE HYPOTHESIS'

- Alternative hypothesis (H_1)
  - The hypothesis that the researcher wants to prove based on the data obtained
  - Ex>H_1: $\mu_1 \neq \mu_2$

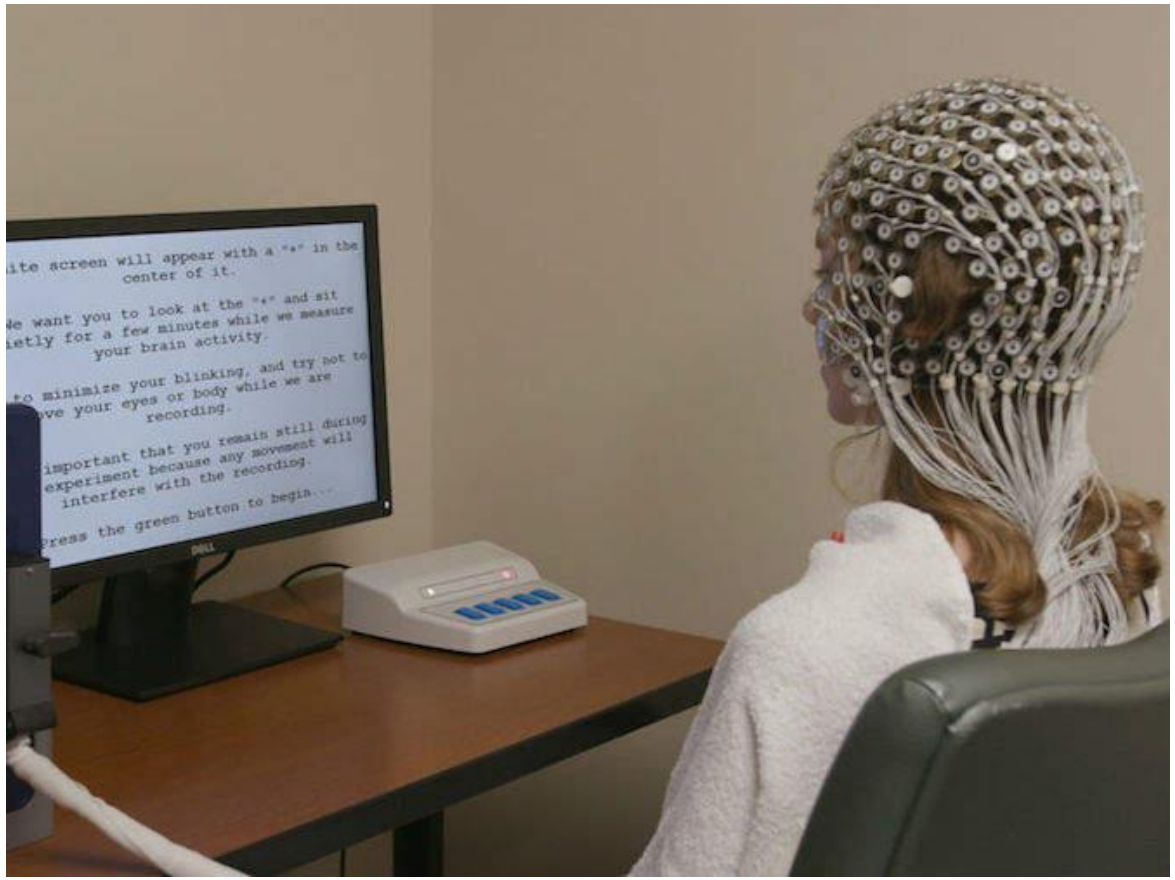
- Null hypothesis (H_0)
  - The opposite of the alternative hypothesis
  - The subject of the test
  - Ex> H_0: $\mu_1 = \mu_2$,

2. INFERENTIAL STATISTICS
2.2 HYPOTHESIS TEST
2.2.2 SET A SIGNIFICANCE LEVEL (ALPHA)

- Significance level (α)
  - Maximum allowable range of error for rejecting the null hypothesis
  - Commonly 0.05

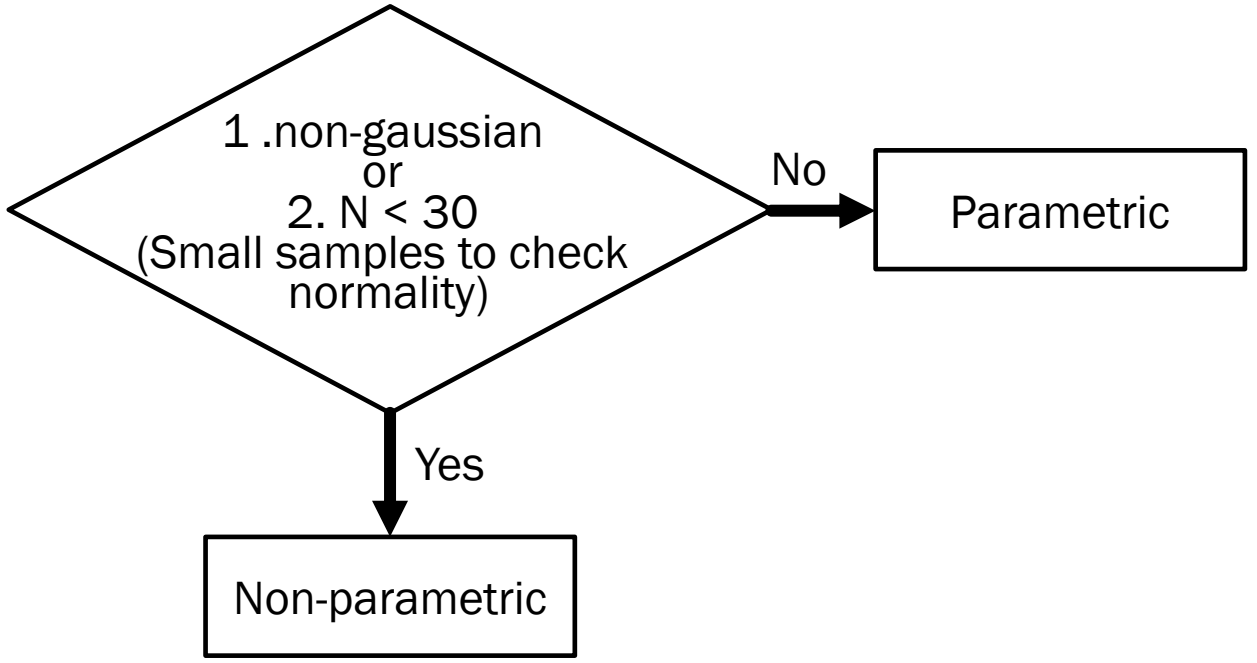


2. INFERENTIAL STATISTICS
2.2 HYPOTHESIS TEST
2.2.3 ACQUIRE SAMPLING (PHYSICAL EXPERIMENT)



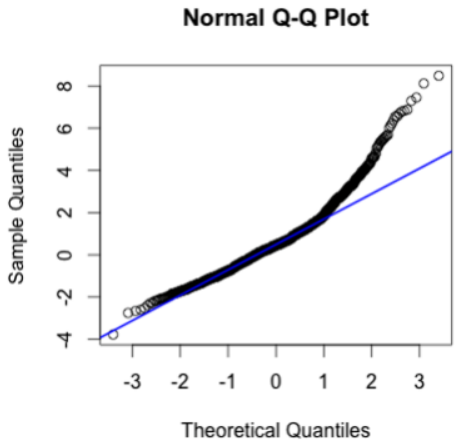
2. INFERENTIAL STATISTICS
2.2 HYPOTHESIS TEST
2.2.4 COMPUTE 'TEST STATISTIC' (P-VALUE)

1. **Test statistic**
    1. **Difference test**
        1. **Parametric test**
        2. **Non-parametric test**
    2. **Relation test**
        1. **Parametric test**
        2. **Non-parametric test**

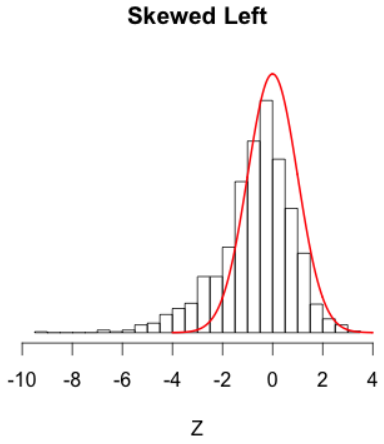


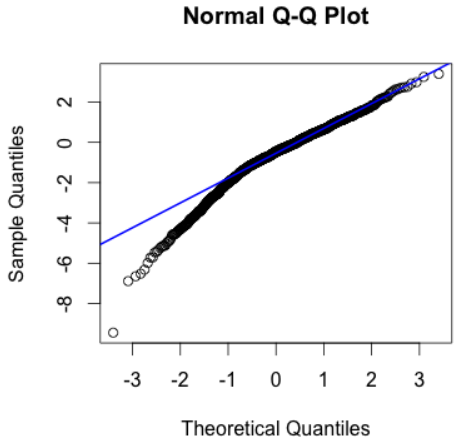
2.2.5 BASED ON P-VALUE, ACCEPT THE HYPOTHESIS

2. INFERENTIAL STATISTICS
2.3 TEST STATISTICS
* NORMALITY TEST

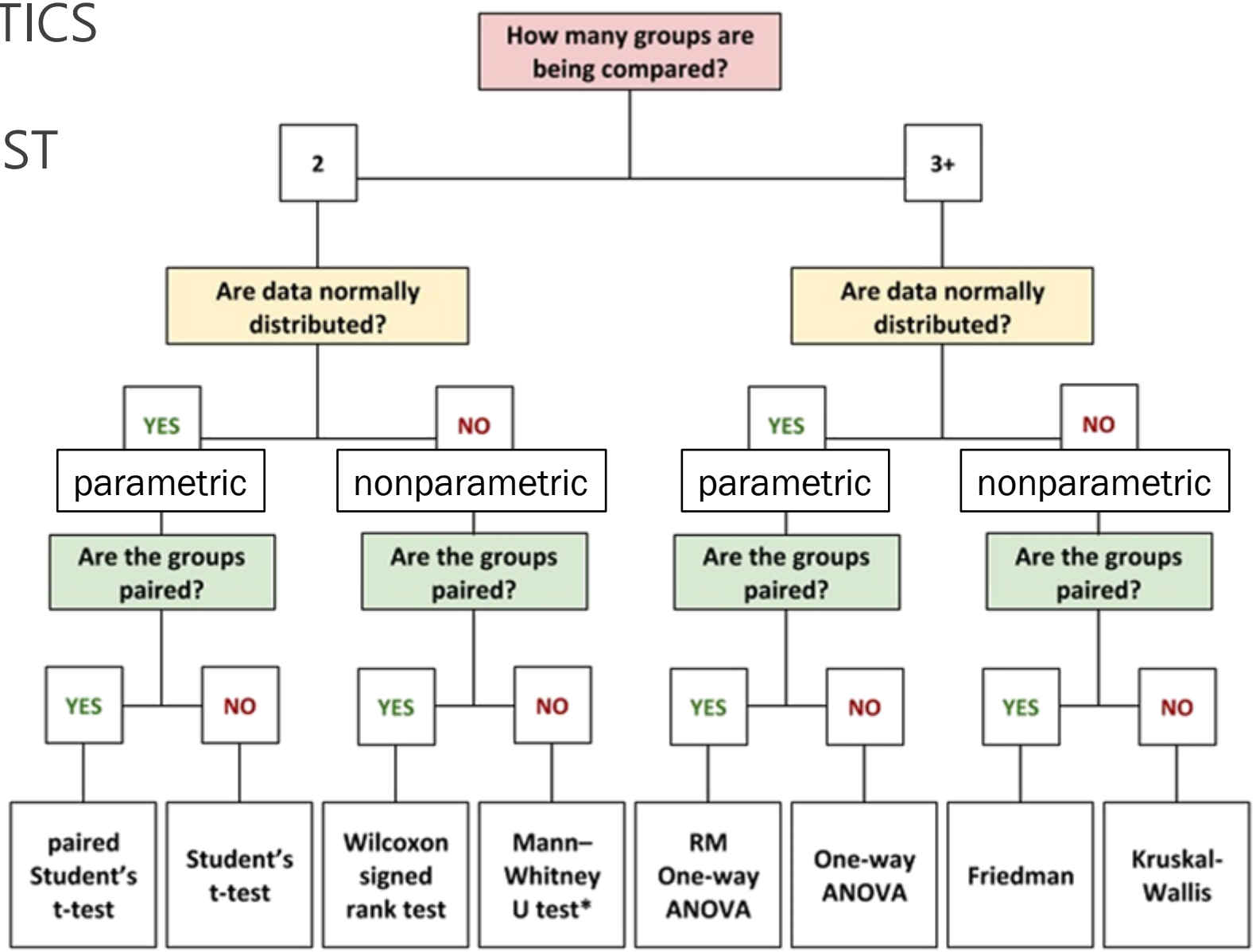


- **Shapiro-Wilk test**
  - N < 2000
- Kolmogorove-Smirnov test
  - N > 2000
- Quantile-Quantile plot (Graphic test)
  - Visual analysis method
  - If N is too small, it is intuitive to check normality with visual methods.



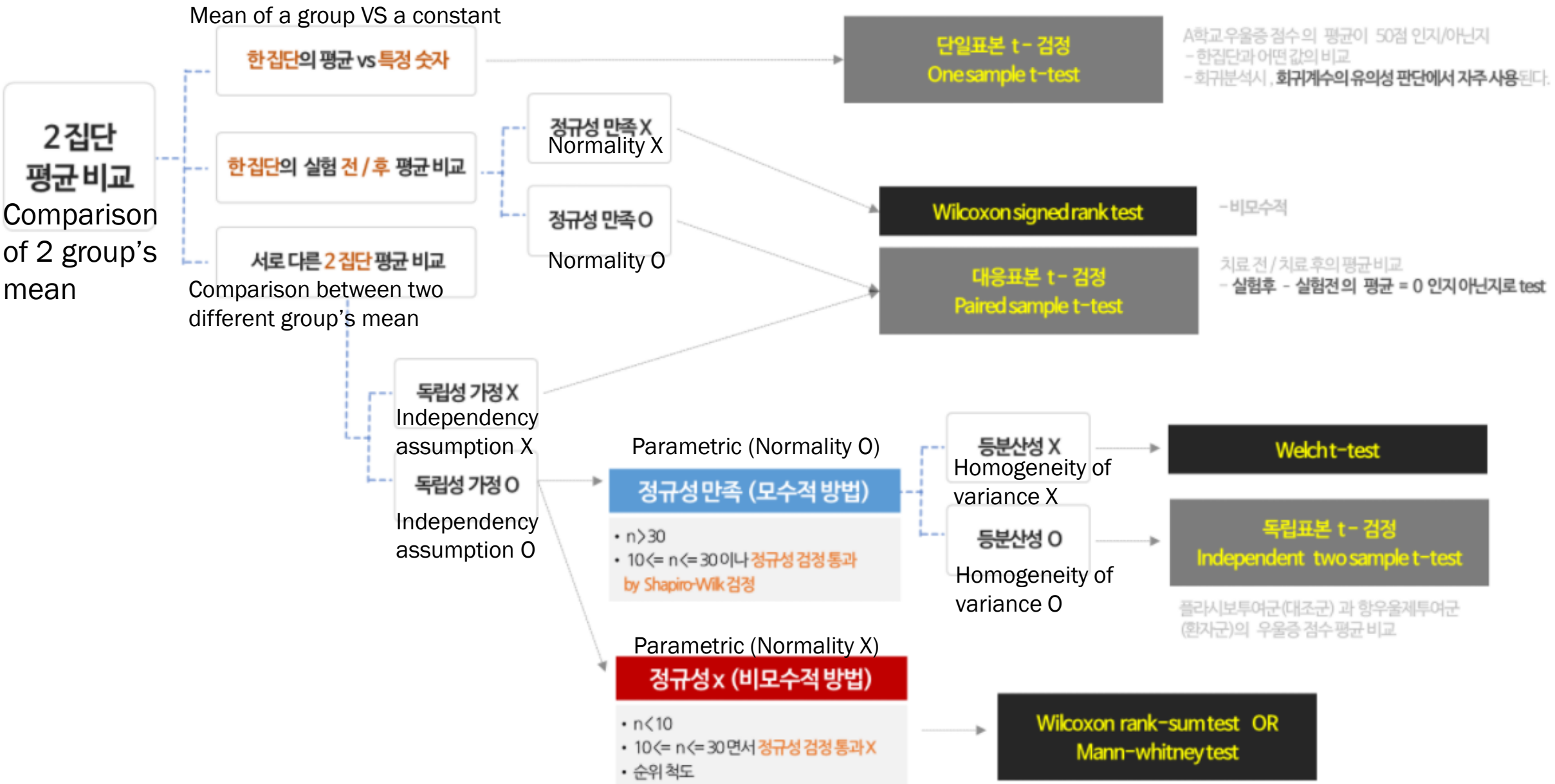


2. INFERENTIAL STATISTICS
2.3 TEST STATISTICS
2.3.1 DIFFERENCE TEST

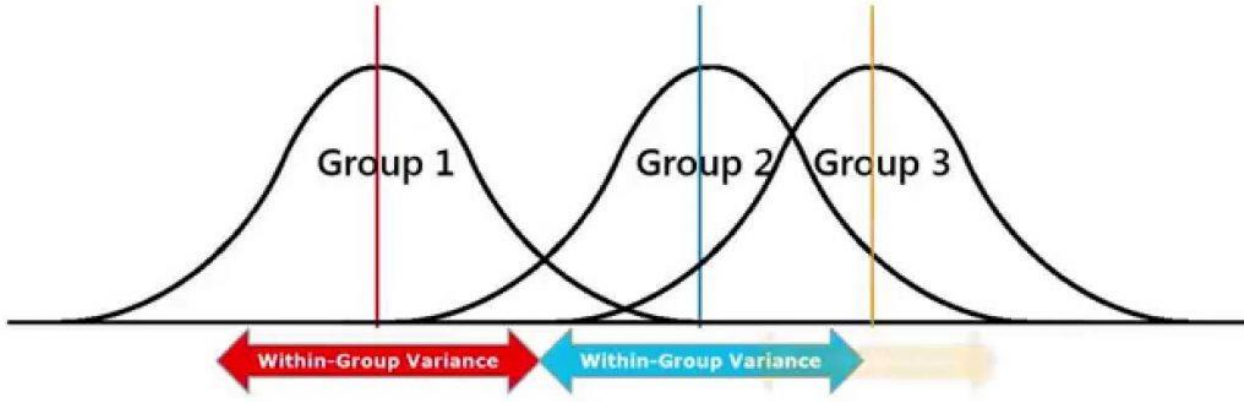


2.3 TEST STATISTICS
2.3.1 DIFFERENCE TEST
2.3.1.1 PARAMETRIC TEST

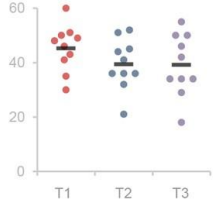
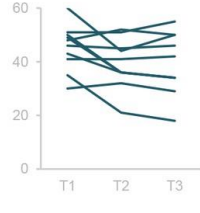
- 2 groups
  - T-test : Check whether two groups are different
    - One sample t-test : mean of a group vs a constant
    - Paired sample t-test : mean of a group vs mean of a group
    - Independent two sample t-test : mean of an independent group vs mean of an independent group
      - When the sample passes **the homogeneity of variance test**
    - Welch t-test : mean of an independent group vs mean of an independent group
      - When the sample doesn't pass **the homogeneity of variance test**
- 3+ groups
  - ANOVA(ANalysis Of VAriance)


2집단 평균 비교
Comparison of 2 group's mean
Mean of a group VS a constant
한 집단의 평균 vs 특정 숫자
단일표본 t-검정
One sample t-test
A학교 우울증 점수의 평균이 50점 인지/아닌지
-한집단과 어떤 값의 비교
-회귀분석시, 회귀계수의 유의성 판단에서 자주 사용된다.
한 집단의 실험 전/후 평균 비교
정규성 만족 X
Normality X
정규성 만족 O
Normality O
Wilcoxon signed rank test
-비모수적
대응표본 t-검정
Paired sample t-test
치료 전/치료 후의 평균비교
- 실험후 - 실험전의 평균 = 0 인지 아닌지로 test
서로 다른 2집단 평균 비교
Comparison between two different group's mean
독립성 가정 X
Independency assumption X
독립성 가정 O
Independency assumption O
Parametric (Normality O)
정규성 만족 (모수적 방법)
• n>30
• 10<= n<=30 이나 정규성 검정 통과 by Shapiro-Wilk 검정
등분산성 X
Homogeneity of variance X
Welch t-test
등분산성 O
Homogeneity of variance O
독립표본 t-검정
Independent two sample t-test
플라시보투여군(대조군) 과 항우울제투여군 (환자군)의 우울증 점수 평균 비교
Parametric (Normality X)
정규성 x (비모수적 방법)
• n<10
• 10<= n<=30 면서 정규성 검정 통과 X
• 순위 척도
Wilcoxon rank-sum test OR
Mann-whitney test

2.3 TEST STATISTICS
2.3.1 DIFFERENCE TEST
2.3.1.1 PARAMETRIC TEST



- 2 groups
  - T-test
- 3+ groups
  - ANOVA(ANalysis Of VAriance) : Check whether three or more groups are different
    - One-way ANOVA : # of factor is one
    - Two-way ANOVA : # of factor is two
    - MANOVA (Multivariate ANOVA) : # of factor is three or more
    - rmANOVA(Repeated Measures ANOVA) : Deal with repeatedly measured data (same group)

ANOVAs with vs. without repeated measures…	ANOVA without repeated measures	Repeated Measures ANOVA
…see the data differently	Compares **3 independent groups** (n = 30, 10/group) 	Compares **10 pairs of related observations** (n = 10) 

Factor : variable which cause the observation

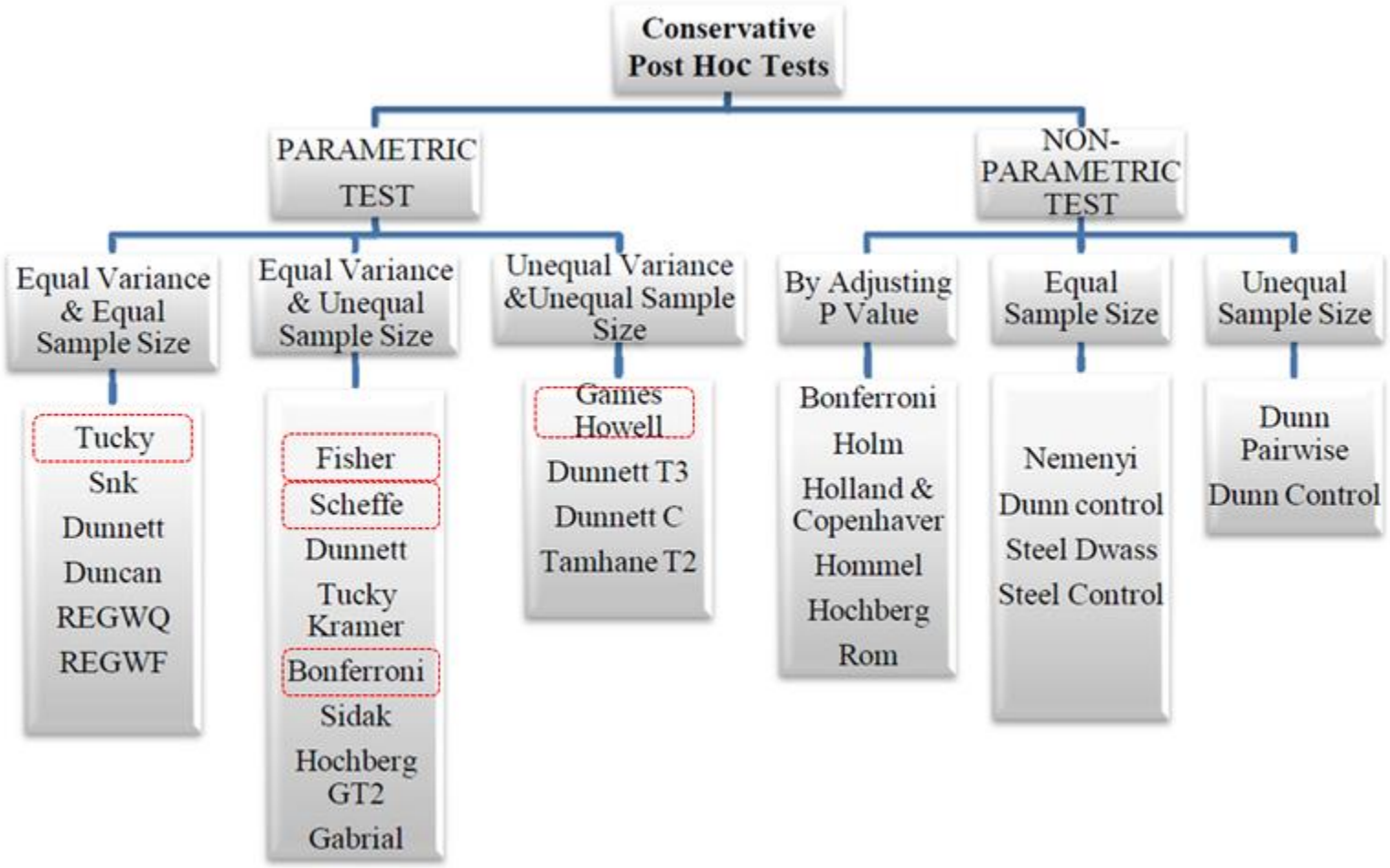
2.3 TEST STATISTICS
2.3.1 DIFFERENCE TEST
2.3.1.2 NON-PARAMETRIC TEST

- 2 groups
  - Wilcoxon signed rank test (alternatives to paired t-test)
  - Mann-Whiteney U test ((alternatives to independent t-test)
- 3+ groups
  - Kruskal-Wallis test (alternatives to ANOVA)
  - Freidman test (alternatives to rmANOVA)

		Criterion / Measure / Dependent Variable (Continuous)	
		Non-Parametric Test	Parametric Equivalent
Categorical	1 Variable 2 Categories Between-subjects	**Mann-Whitney U Test** (Nonparametric Tests → Legacy Dialogs → 2 Independent Samples)	**Independent *t* Test**
	1 Variable 2 Categories Within-subjects	**Wilcoxon Signed Rank Test** (Nonparametric Tests → Legacy Dialogs → 2 Related Samples)	**Paired *t* Test**
	1 Variable >2 Categories Between-subjects	**Kruskal-Wallis H Test** (Nonparametric Tests → Legacy Dialogs → K Independent Samples)	**One-Way ANOVA**
	1 Variable >2 Categories Within-subjects	**Friedman Test** (Nonparametric Tests → Legacy Dialogs → K Related Samples)	**Repeated Measures ANOVA**

2.3 TEST STATISTICS
2.3.1 DIFFERENCE TEST
2.3.1.3 POST HOC

- After variance test with three or more groups, Post hoc analysis is required.
  - We only know the difference among three groups, not the difference each other.
- Commonly **pair-wise t-test** is used.
- There are various way to solve this problem.



2. INFERENTIAL STATISTICS
2.3 TEST STATISTICS
2.3.2 RELATION TEST

- Parametric
  - Pearson's correlation coefficient
    - Assess the linear relation of two variables
- Non-parametric
  - Spearman's rho
    - Assess the linear relation of two variables based on rank