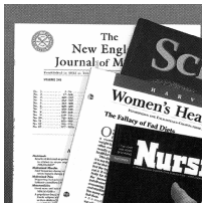
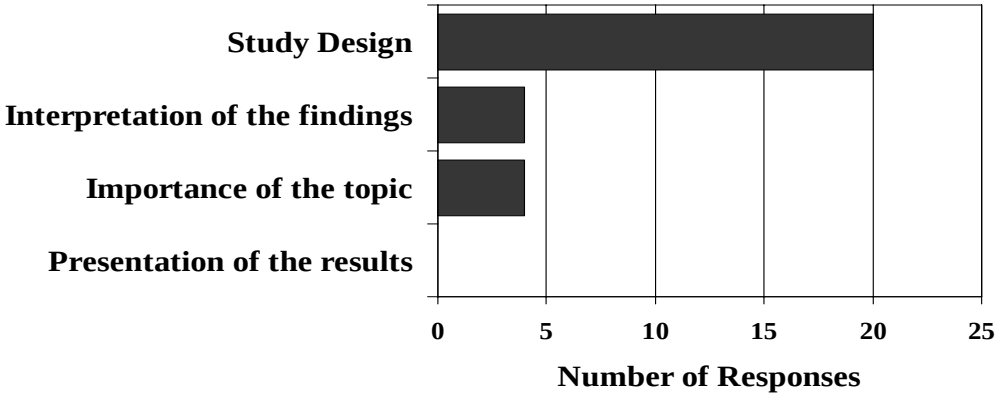
Common Errors in Statistics How to Avoid Them



Kim, Soo-Nyung
Konkuk University Medical School

The Most Common Type of Flaw



의학논문에서 흔한 통계적오류

- 외국 학술지 45-50%,국내 학술지 48-100%
- 통계적 검정법의 활용빈도
  t-test, chi-square test, ANOVA test
- 국제의학학술지 편집인협의회 (1988)
  의학학술지에 투고하는 원고의 통일양식
- 한국통계학회 홈페이지 http://www.kss.or.kr

Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication

Updated February 2006
International Committee of Medical Journal Editors

IV.A.6.c. Statistics

- Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.
- When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals).

IV.A.6.c. Statistics

- Avoid relying solely on statistical hypothesis testing, such as the use of P values, which fails to convey important information about effect size.
- References for the design of the study and statistical methods should be to standard works when possible (with pages stated).
- Define statistical terms, abbreviations, and most symbols.
- Specify the computer software used.

IV.A.7. Results

- When data are summarized in the Results section, give numeric results not only as derivatives (for example, percentages) but also as the absolute numbers.
- Specify the statistical methods used to analyze them.
- Use graphs as an alternative to tables with many entries; do not duplicate data in graphs and tables.

IV.A.7. Results

- Avoid nontechnical uses of technical terms in statistics, such as “random” (which implies a randomizing device), “normal,” “significant,” “correlations,” and “sample.”
- Where scientifically appropriate, analyses of the data by variables such as age and sex should be included.

통계적 오류의 종류

- ❖부적절한 통계적 기법 적용
- ❖통계학적 방법론 서술의 과오
- ❖통계분석 결과 해석의 과오
- ❖통계용어 및 기호 사용 과오

의학논문

방 법: 40명의 아토피성 천식 소아에게 흡입용 스테로이드를 12주 동안 투여하고 투여 전과 투여 후에 PC_{20}의 변화를 비교 분석하였다.

통계 분석: 결과는 mean±SE로 제시하였다. 각 군간의 비교는 Student t-test로 시행하였으며 통계적 유의수준은 P value가 0.05 이하로 하였다.

의학논문

결 과: PC_{20}은 치료전 1.89 mg/mL에서 치료후 4.56 mg/mL로 의미있게 증가하였다($P<0.01$).

Table 1. Comparison of PC_{20}

	Before Tx	After Tx
PC_{20}(mg/mL)*	1.89±0.45 †	4.56±0.98 †

*Mean±SE, † $P<0.01$

Basic Statistics Without Tears

Kim, Soo-Nyung M.D., Ph.D.
Konkuk University College of Medicine

Statistics Without Tears
for Non-Mathematicians

- Do NOT Remember
- Just Understand
- Look Up

Population and Sample

- Parameter
- Statistic

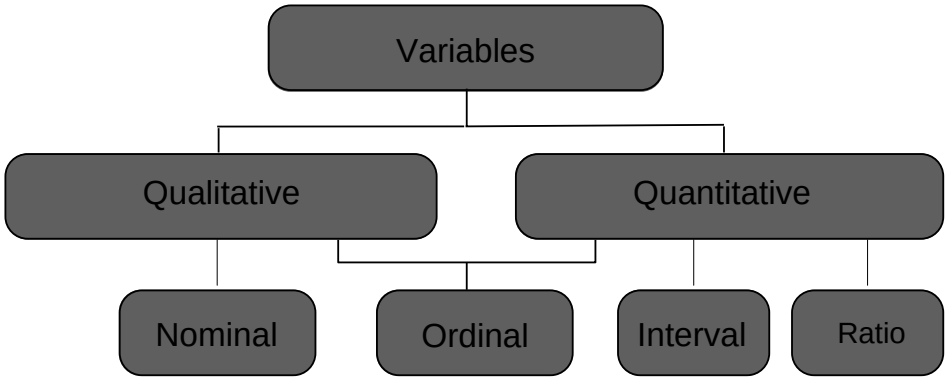
Data Type

- ✓ **Nominal**
- ✓ **Ordinal**
- ✓ **Interval**
- ✓ **Ratio**

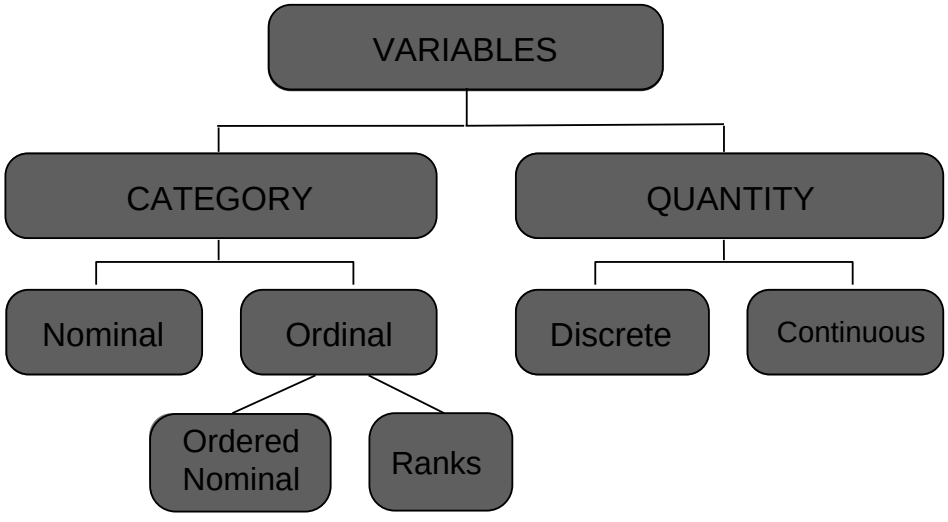
Level of Measurement

- Nominal - categorical
  - gender, race, blood type, hypertensive
- Ordinal - categories that can be ranked
  - rating scale, anxiety score, grade, tumor stage
- Interval - continuous
  - temperature, calendar year, IQ, psychological test
- Ratio – interval with true zero (none of the quantity)
  - blood pressure, age, days in the hospital

Statistical Variables



Statistical Variables



Statistical Analysis

- Descriptive Statistics
- Inferential Statistics

Descriptive vs. Inferential

- Descriptive statistics summarize your group.
  - average age 78.5, 89.3% white.
- Inferential statistics use the theory of probability to make inferences about larger populations from your sample.
  - White patients were significantly older than black and Hispanic patients, $P<0.001$.

Descriptive Statistics

- How are data described and summarized?
- What are measures of central tendency and dispersion?
- Which measures of central tendency do I use and when?

Central Tendency

- Mean
- Median
- Mode

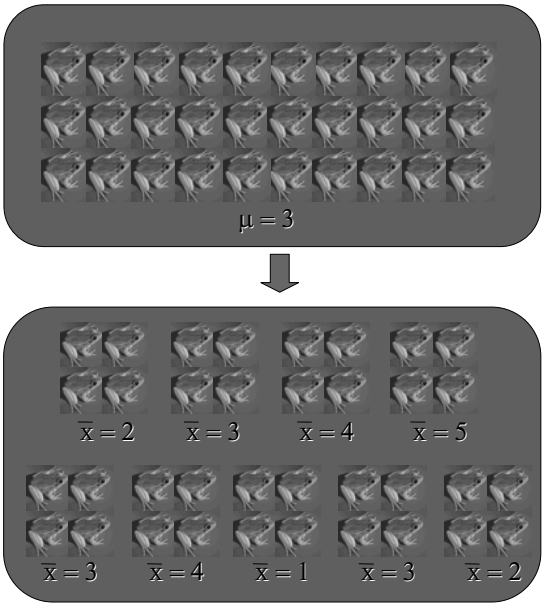
Dispersion

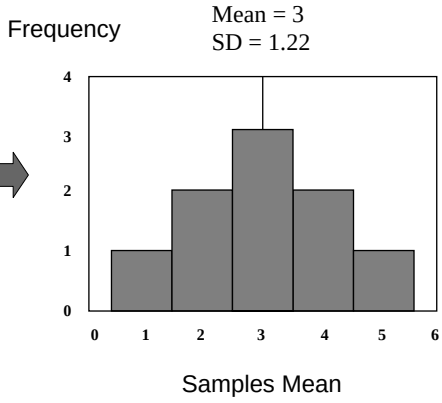
- SD (Variance)
- Range
- IQR (Inter-quartile range)

What is the Standard Error?

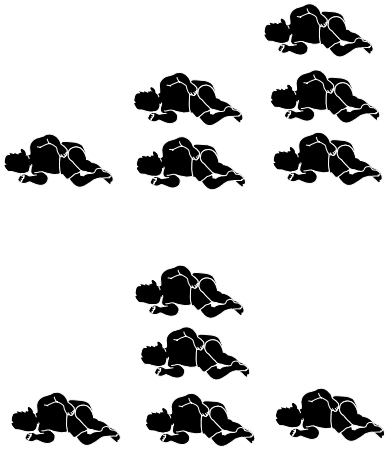
- SD of sample means (SEM)
- Prediction of how close a sample mean is to the true population mean
- Parametric statistic
- Calculating confidence limits

Standard Error



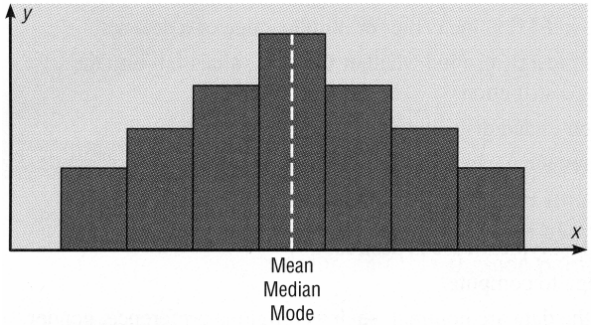


Shape

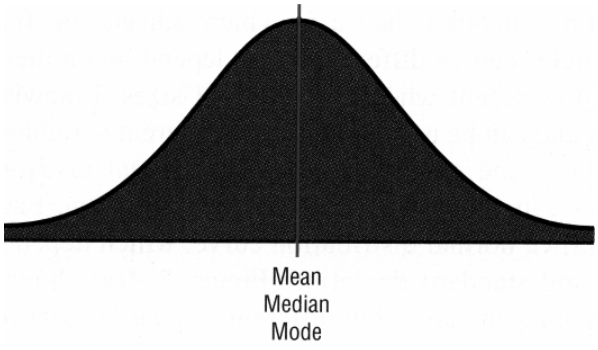


- Skewness
- Kurtosis

Symmetrical Distribution

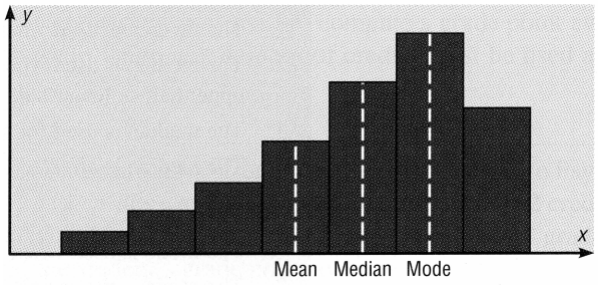


Normal Distribution

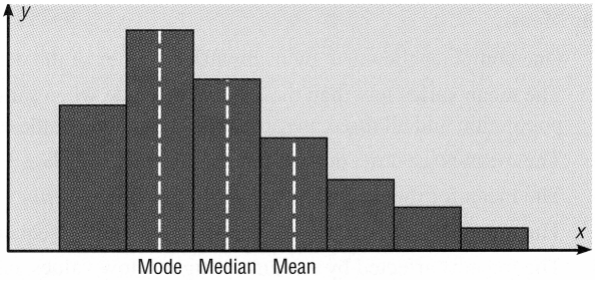


Skewed Distribution

- Negatively skewed



- Positively skewed



Descriptive Statistics

Data Analysis

Central Tendency	Dispersion	Shape
Mean	Variance	Skewness
Median	SD	Kurtosis
Mode	Range	



Descriptive Statistics

Data Type

Nominal	Ordinal	Interval	Ratio
Frequency	Median	Mean	GM
Mode	Range	SD	CV

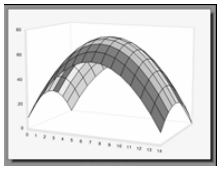


Example
Descriptive Statistics

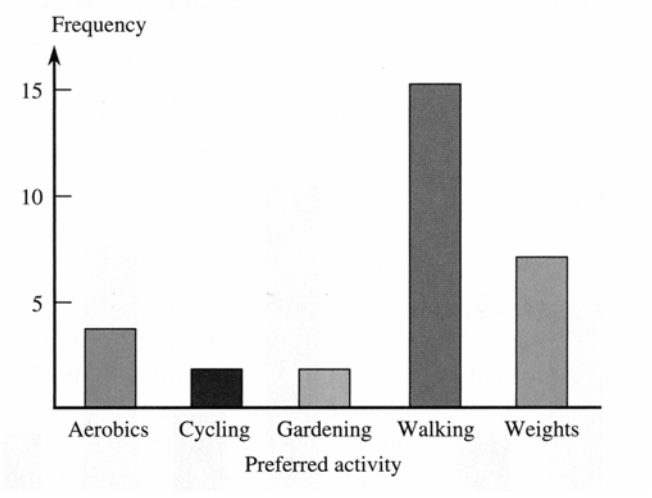
3265	3323	2581	2759
3260	3649	2841	3248
3245	3200	3609	3314
3484	3031	2838	3101
4146	2069	3541	2834

weight (gm)

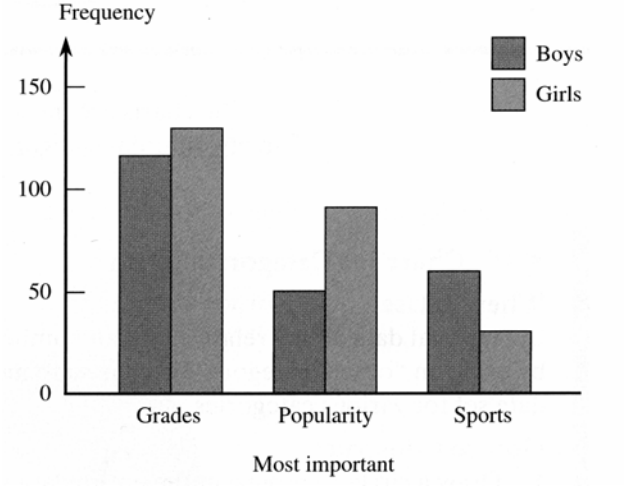
Graphical Methods for Describing Data



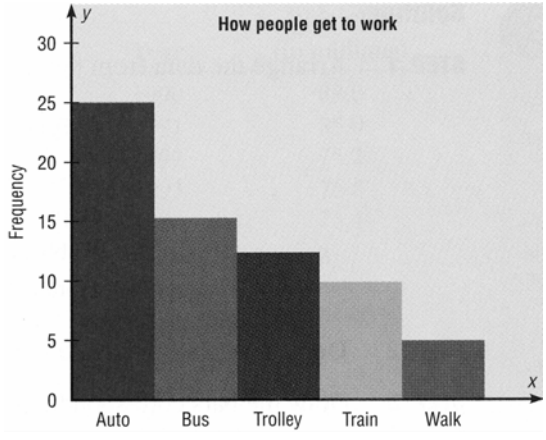
Nominal Data
Bar Chart



Nominal Data
Bar Chart

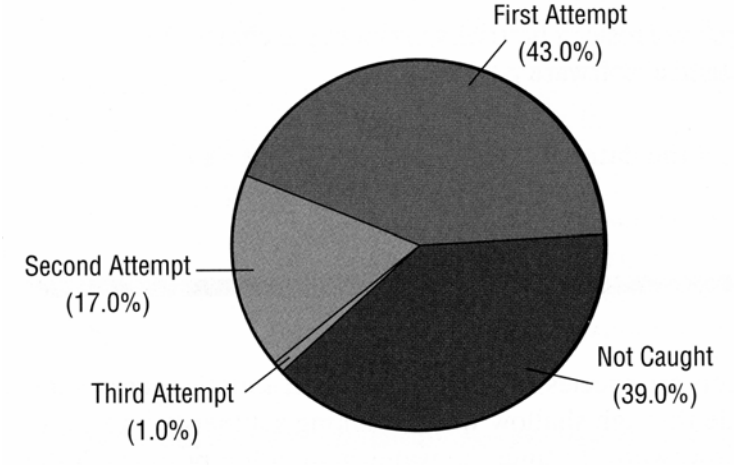


Nominal Data
Pareto Chart

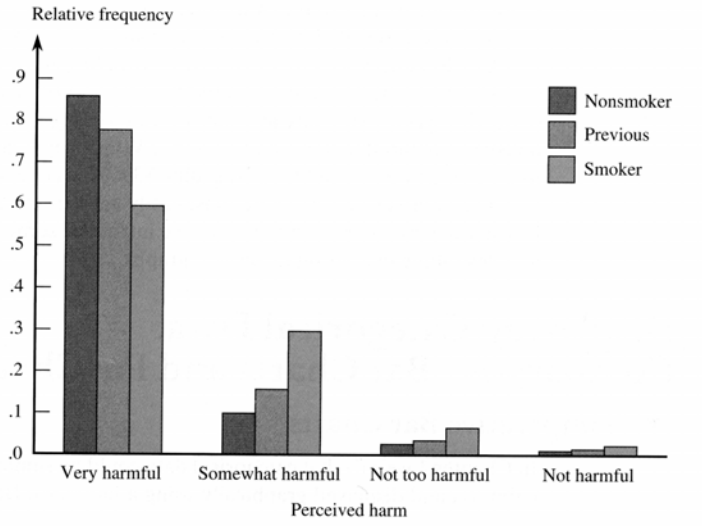


Pareto Chart

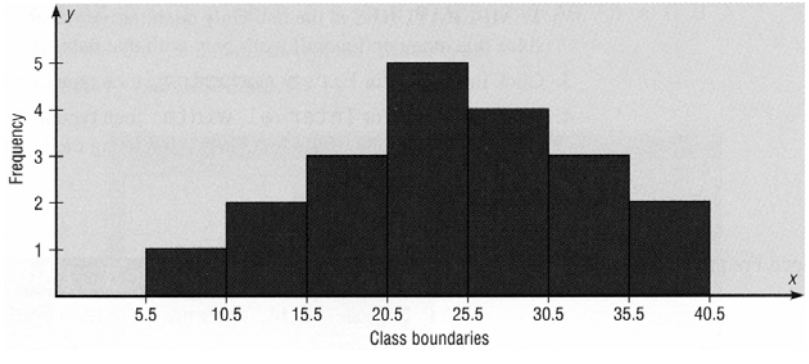
Nominal Data
Pie Chart



Ordinal Data
Bar Chart

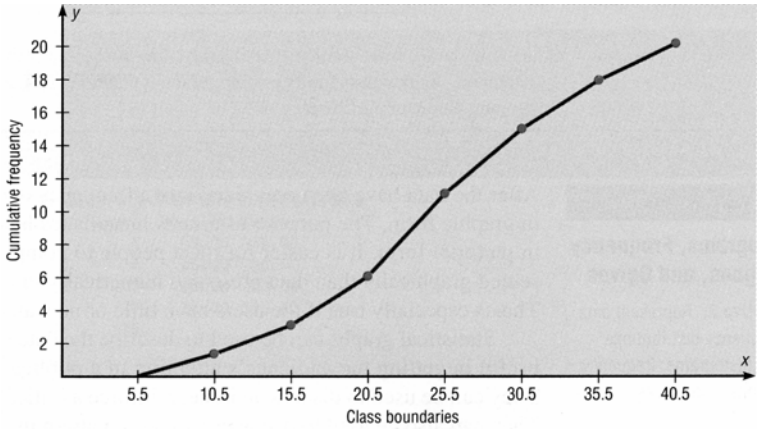


Interval/Ratio Data
Histogram



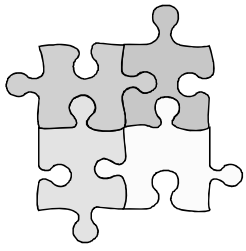
Interval/Ratio Data
Frequency Polygon

Interval/Ratio Data
Cumulative Frequency Graph



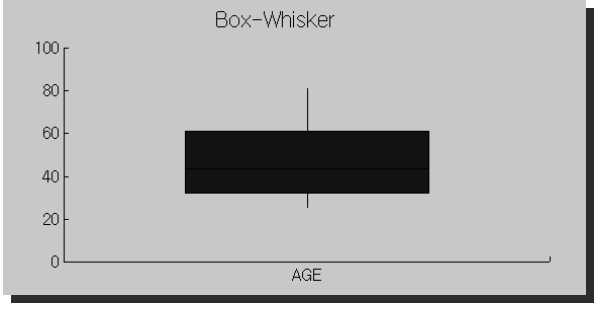
Ogive

Data Screening for Statistical Analysis



Five-Number Summary

- Min, Q1, Q2, Q3, Max
- IQR = Q3 – Q1
- Boxplot
- Box-and-Whisker Diagram

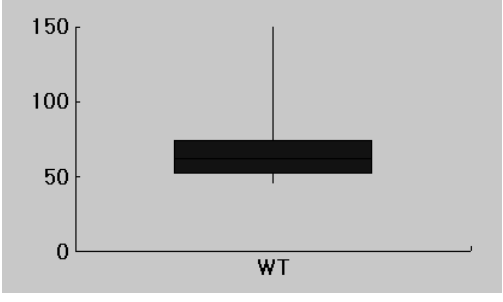


Outliers

- Recording error
- Different population
- Unusual extreme observation

Dramatic effect on

- Mean, SD
- Histogram

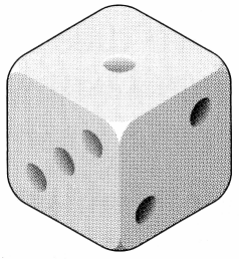


Upper Limit = Q3 + 1.5 IQR

Lower Limit = Q1 – 1.5 IQR

You Will Need to Learn Inferential Statistics

Distribution and Probabilities



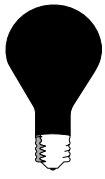
- What are probabilities?
- How are probabilities and distributions linked?
- What types of distributions are there?

P Value

- A P value is an estimate of the probability of results such as yours could have occurred by chance alone if there truly was no difference or association.
- $P < 0.05$ = 5% chance, 1 in 20.
- $P < 0.01$ = 1% chance, 1 in 100.
- Alpha is the threshold. If P is < this threshold, you consider it statistically significant.

Probability Distribution

- Continuous
- Discontinuous



Probability Distribution

Continuous

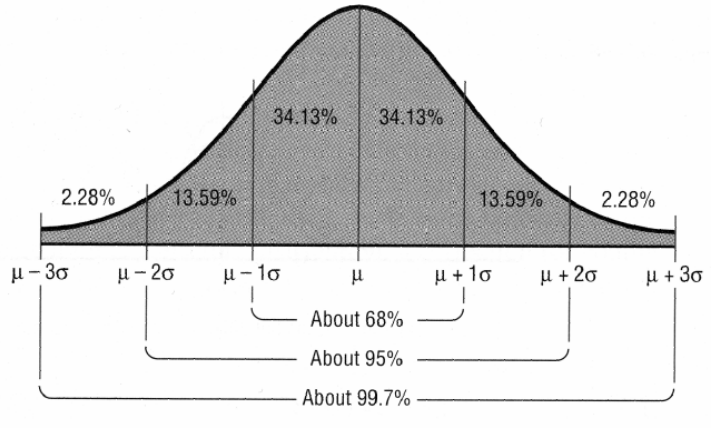
- **Normal**
- **t**
- **F**
- **X^2**

Probability Distribution

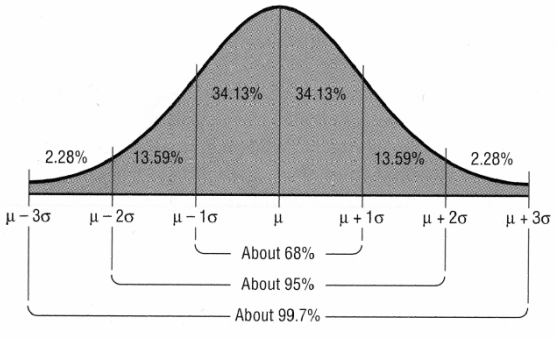
Discontinuous

- **Binomial**
- **Poisson**
- **Bernoulli**
- **Hypergeometric**
- **Multinomial**

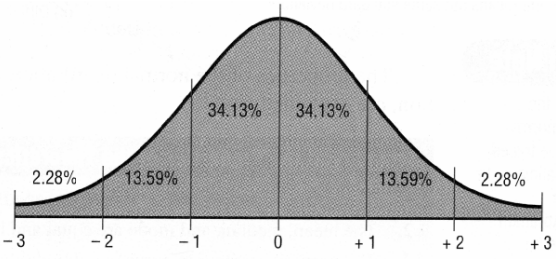
Normal Distribution



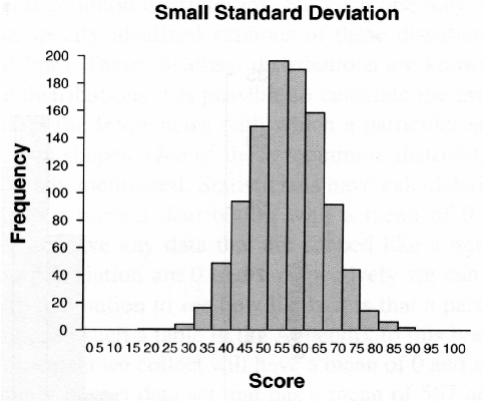
Normal Distribution

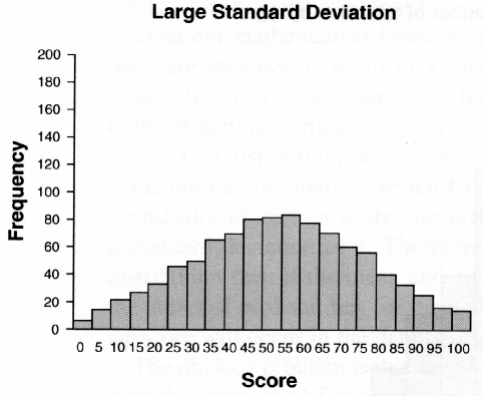


Standard Normal Distribution



Standard Deviation





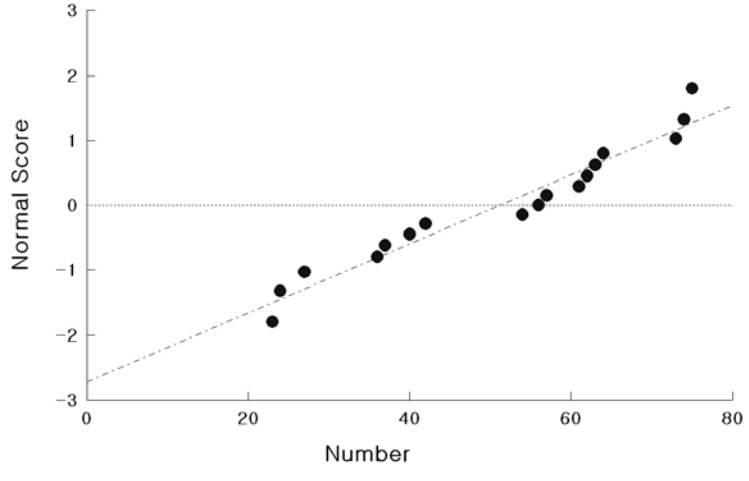
Example

Normal Distribution

23	24	36	40
64	56	61	62
73	74	23	27
37	42	54	57
61	63	73	75

number

Normal Probability Plot



Test for Normality

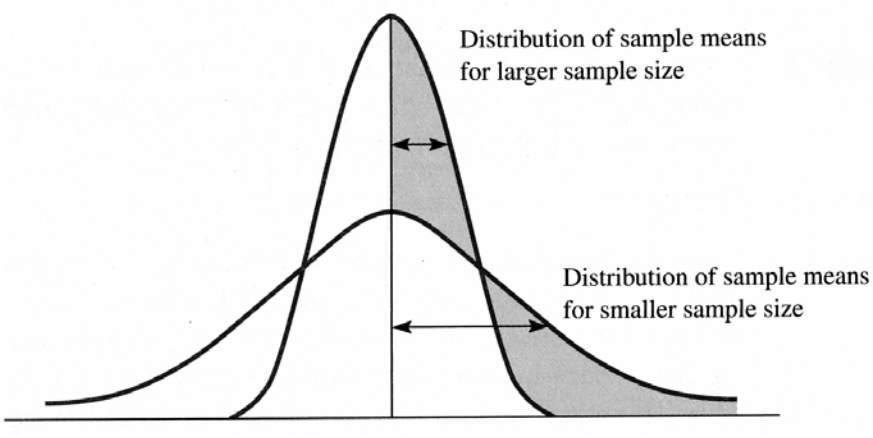
- √ Kolmogorov-Smirnov
- √ Shapiro-Wilk test
- √ Lilliefors test
- √ Cochran test
- √ Chi-square
- √ Anderson-Darling test

Central Limit Theorem

The distribution of sample means will be approximately normal if sample size is sufficiently large.

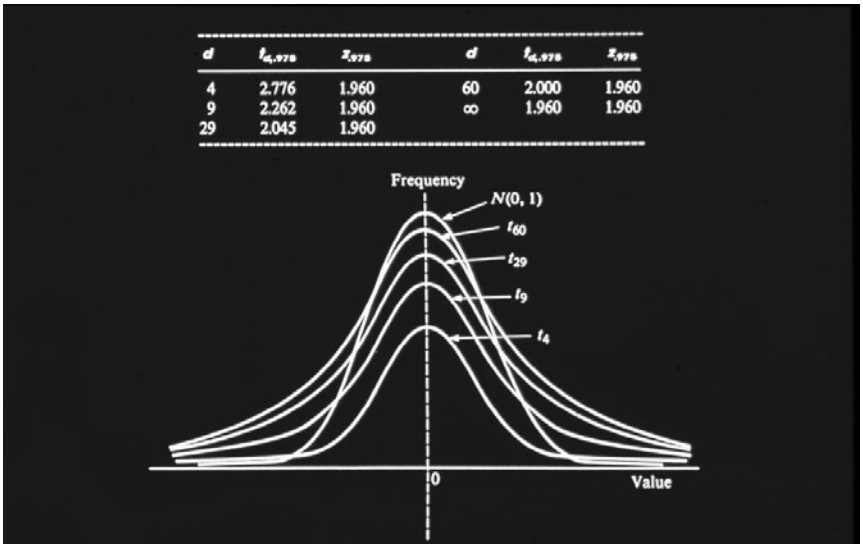
$N > 30$

Distribution of Sample Means



t Distribution

d	$t_{d,.975}$	$z_{.975}$	d	$t_{d,.975}$	$z_{.975}$
4	2.776	1.960	60	2.000	1.960
9	2.262	1.960	∞	1.960	1.960
29	2.045	1.960			



Statistical Hypothesis

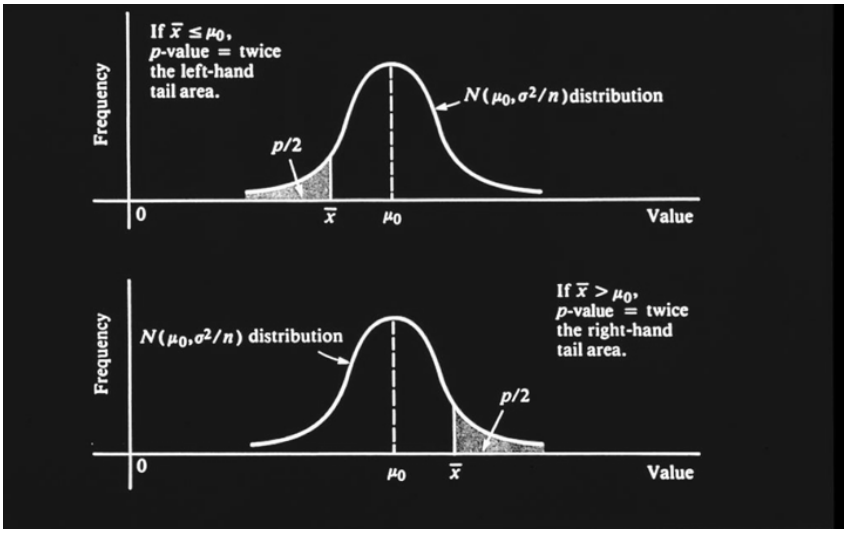
- Null Hypothesis
- Alternate Hypothesis

Statistical Analysis

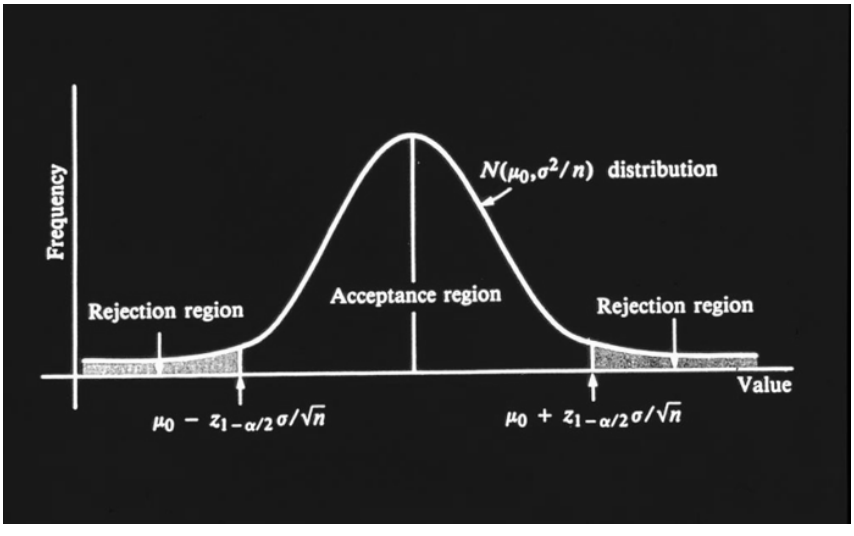
Rejection Region

- One-tailed Test
  - Right-tailed
  - Left-tailed
- Two-tailed Test

One-tailed Test



Two-tailed Test



Statistical Analysis

Level of Significance

- Type I error

  Reject a true null hypothesis

- Type II error

  Accept a false null hypothesis

Type I and Type II Errors

	Null Hypothesis (Ho)	
	True	False
Reject	Type I	Correct
Accept	Correct	Type II

Confidence Interval

- Range of values used to estimate some population parameter with a specific confidence level.
- t Confidence Interval (CI)
  - Random sample
  - Normal distribution
- 95% CI for Sample Means
  - $\bar{X} - 1.96\ SE < m < \bar{X} + 1.96\ SE$

Level of Significance

- α Level
- Probability of a type I error
- Rejecting a true null hypothesis
- $\alpha = 0.05$ (5%) in biostatistics

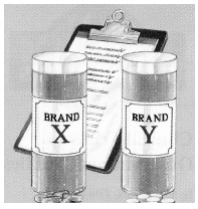
Statistical Significance

1. Significance Level (α)
2. Compute P Value
3. If $p < \alpha$, reject null hypothesis

Statistical Analysis

1. Hypothesis
2. Statistical Test
3. Level of Significance
4. Test Statistic
5. Rejection Region
6. Decision

Statistical Methods for Comparative Studies



Use Inference Statistics to Test for Differences and Associations

- There are hundreds of statistical tests.
- A clinical researcher does not need to know them all.
- Learn how to perform the most common tests.
- Learn how to use the statistical flowchart to determine which test to use.

Commonly Used Statistical Methods

- Student's t-test
- Paired *t*-test
- One-way analysis of variance (ANOVA)
- Chi-square test
- Fisher's exact test
- Mann-Whitney U (Wilcoxon rank-sum) test
- Wilcoxon signed-rank test
- Kruskal-Wallis test

Commonly Used Statistical Methods

- Pearson correlation
- Spearman rank-order correlation
- Linear regression analysis
- Repeated-measures analysis of variance
- Analysis of covariance (ANCOVA)
- Discriminant analysis
- Logistic regression
- Kaplan-Meier method
- Log-rank test

Statistical Analysis

- Parametric
- Nonparametric

Parametric Test
Assumption

- Normal Distribution
- Equal Variance
- Interval Data
- Independence

Parametric Test

- Z test
- t test
- F test
- ANOVA

Nonparametric Test

- Kolmogorov-Smirnov
- Chi-square
- Fisher
- Mann-Whitney
- Wilcoxon signed rank
- Kruskal-Wallis
- Friedman

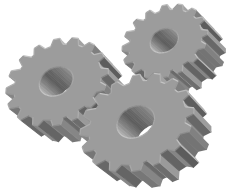
When to Use Parametric Test

1. Interval or Ratio Scale
2. Normal Distribution
3. Equal Variance

When to Use Nonparametric Test

1. Ordinal Scale
2. Lack Normality
3. Significant Different Variance

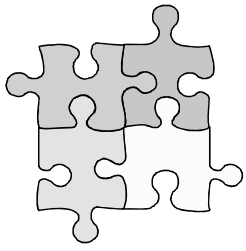
Statistical Terminology Required to Select the Proper Inferential Test



Unmatched vs. Matched

- Some statistical tests are designed to assess groups that are unmatched or independent.
  - Is the admission systolic blood pressure different between men and women?
- Some statistical tests are designed to assess groups that are matched or data that are paired.
  - Is the systolic blood pressure different between admission and discharge?

You Will Need to Know Which Statistical Test to Use



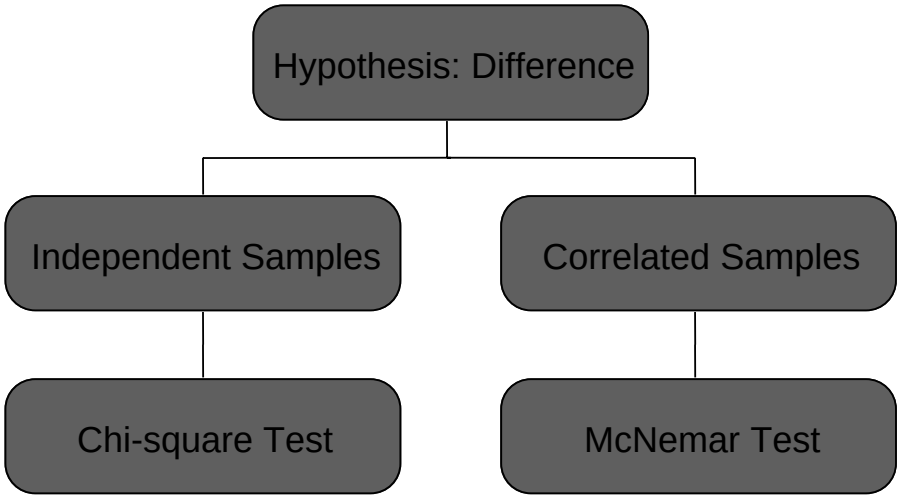
Analyzing the Methodology

The Critical Decisions

1. Scale of measurements?
2. Hypothesis?
3. Independent or correlated?
4. How many sets?

Statistical Test

Nominal Data



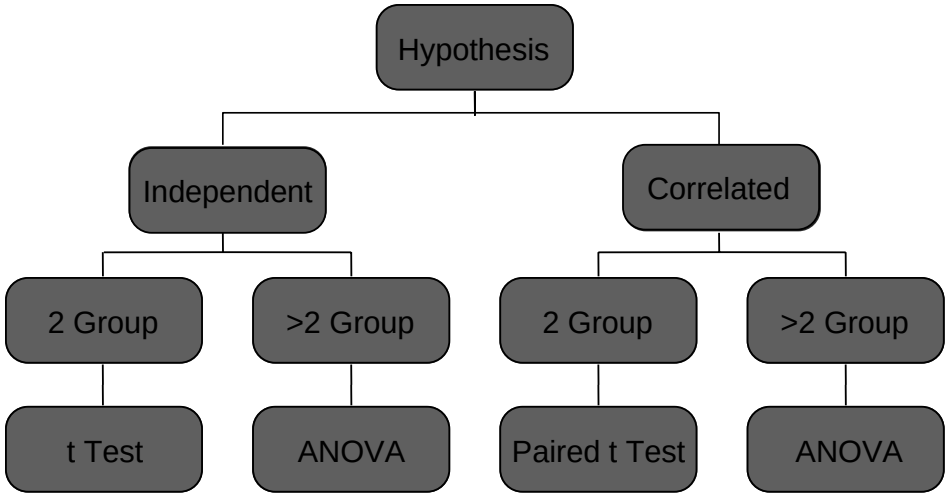
Example

Cross tabulation

	Smoker	Non-Smoker
Cancer (+)	18	6
Cancer (-)	162	114

Statistical Test

Interval Data



Example

Student's t Test

Fish A		Fish B	
38	28	34	27
40	42	34	28
40	44	37	
42	35	37	
39	32	29	

cm

When to Use Student's t Test

Difference - Means of 2 Samples

- Interval or Ratio Scale
- Normal Distribution
- Equal Variance - F test
- Two Independent Samples

Example

Paired t Test

Number	Score Before Education	Score After Education
1	45	49
2	52	56
3	34	31
4	38	46
5	47	54
6	42	39
7	61	68
8	53	55
9	52	50
10	49	55

Score

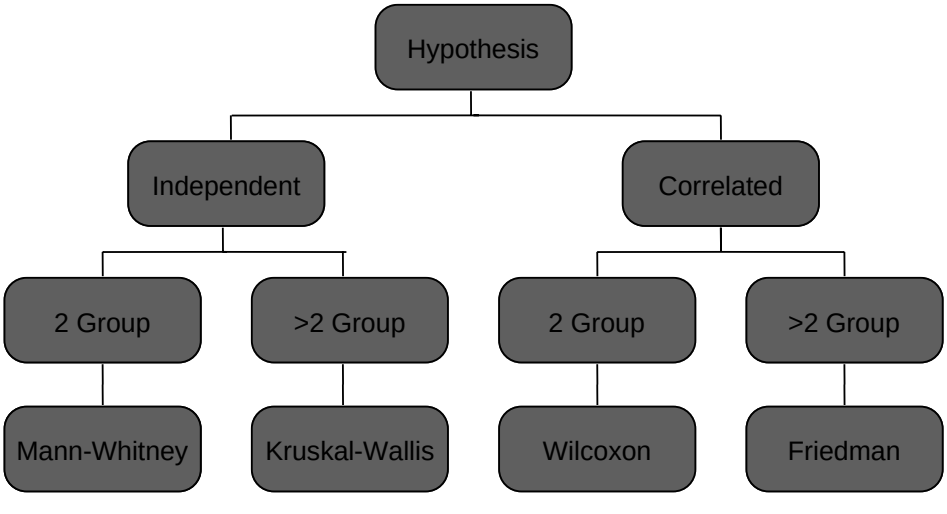
When to Use Paired t Test

Difference – Means of 2 Samples

- Interval or Ratio Scale
- Two Paired Samples
- Normal Distribution of Differences

Statistical Test

Ordinal Data



Example

Mann-Whitney U Test

Man		Woman	
70	67	85	72
60	70	67	71
82	69	79	78
68		72	
80		90	

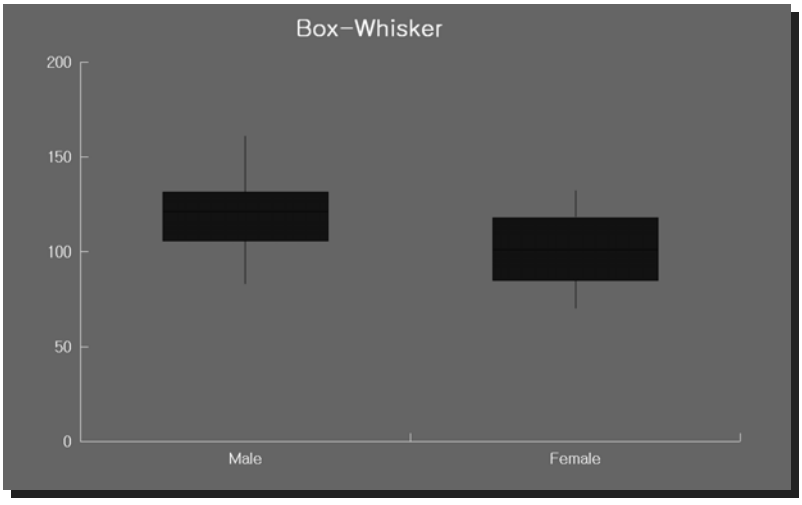
Score

When to Use Mann-Whitney Test

Difference - Distribution of 2 Samples

- Ordinal Scale
- Two Independent Samples
- Same-shape populations
- Wilcoxon Rank Sum Test - Median

Interval/Ratio Data
Non-Normally Distributed Variables



Example
Wilcoxon Signed Rank Test

Number	Symptom Before Drug	Symptom After Drug
1	19	22
2	11	18
3	14	17
4	17	19
5	23	22
6	11	12
7	15	14
8	19	11
9	11	19
10	8	7

Score

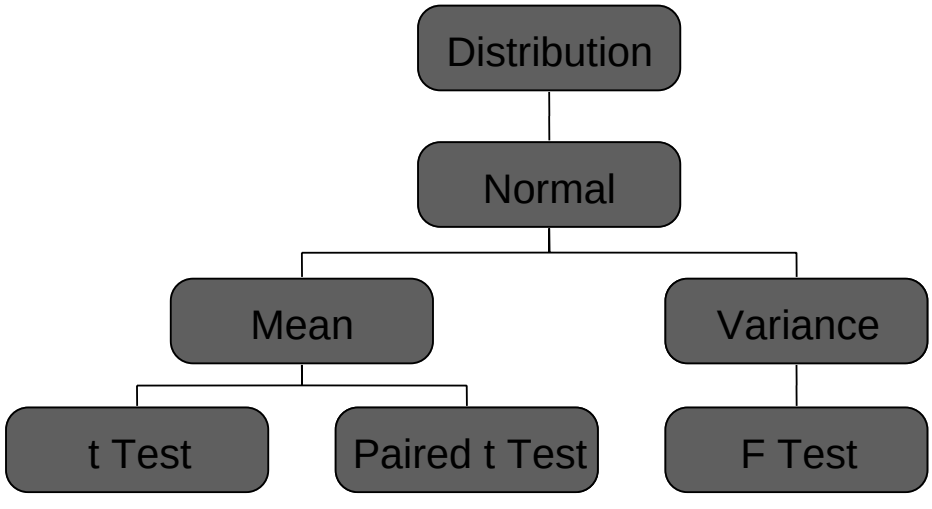
When to Use Wilcoxon Signed Rank Test

Difference - Distribution of 2 Samples

- Ordinal Scale
- Two Paired Samples
- Symmetric Distribution of Differences
- Differences are Independent

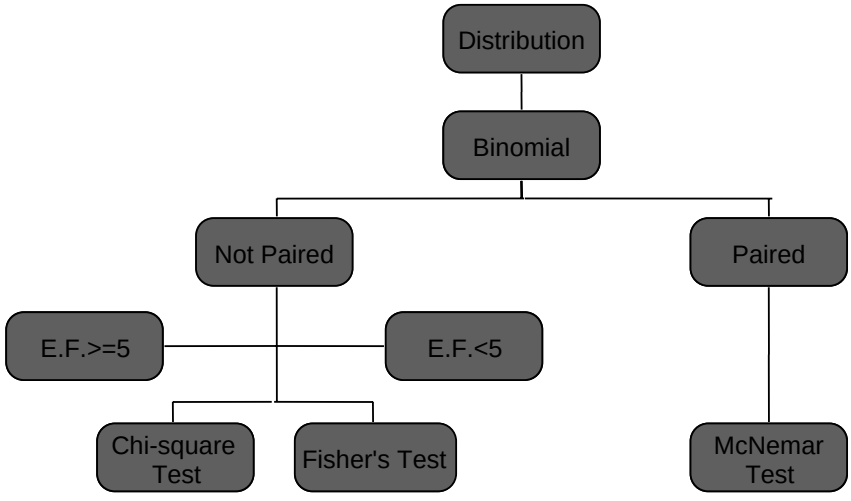
Statistical Test

Two Samples



Statistical Test

Two Samples



Example

Contingency Coefficient

City Character	City Size	City Character	City Size	City Character	City Size
1	1	1	2	2	2
1	1	1	2	2	2
1	1	1	3	2	2
1	1	1	3	2	3
1	1	1	3	2	3
1	2	2	1	2	3
1	2	2	1	2	3
1	2	2	1	2	3
1	2	2	1	2	3
1	2	2	1	2	3

Character: 1=industrial, 2=commercial
Size: 1=large, 2=middle, 3=small

Example
Chi-square Test

Sex	Smoking	Sex	Smoking	Sex	Smoking
1	1	1	2	2	2
1	1	1	2	2	2
1	1	2	1	2	2
1	1	2	1	2	2
1	1	2	2	2	2
1	1	2	2	2	2
1	1	2	2	1	1
1	1	2	2		
1	1	2	2		

When to Use Chi-square Test

Association – Proportion of 2 Samples

- Norminal Scale
- Expected Frequency ≥ 5
- A Random Sample

Example

Fisher's Exact Test

Sex	Smoking	Sex	Smoking
1	1	2	2
1	1	2	2
1	2	2	2
2	1	2	2
2	1	1	2
2	1	1	1

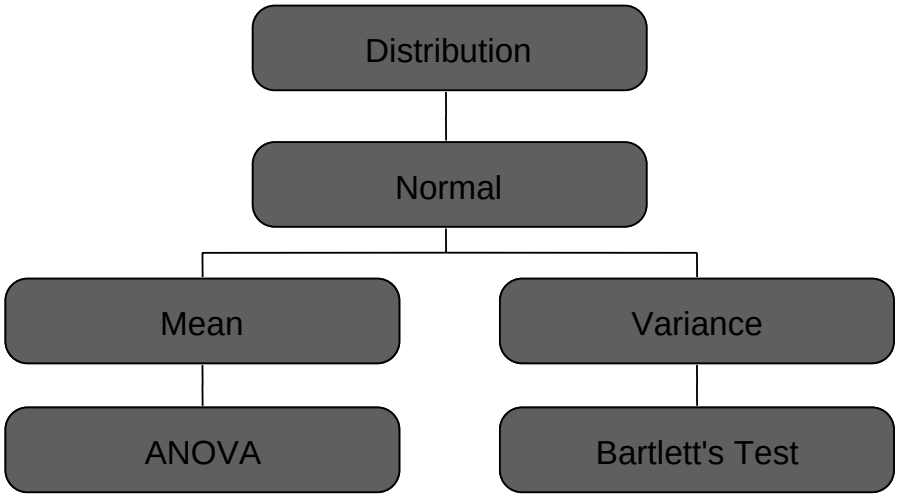
When to Use Fisher Exact Test

Association – 2 x 2 Contingency Table

- Norminal Scale
- Sample Size $N \leq 20$
- $20 < N < 40$, Expected Frequency < 5
- A Random Sample

Statistical Test

More than Two Samples



Test for Variance

- √ Bartlett test
- √ Hartley test
- √ Levene test

Example
One-Way ANOVA

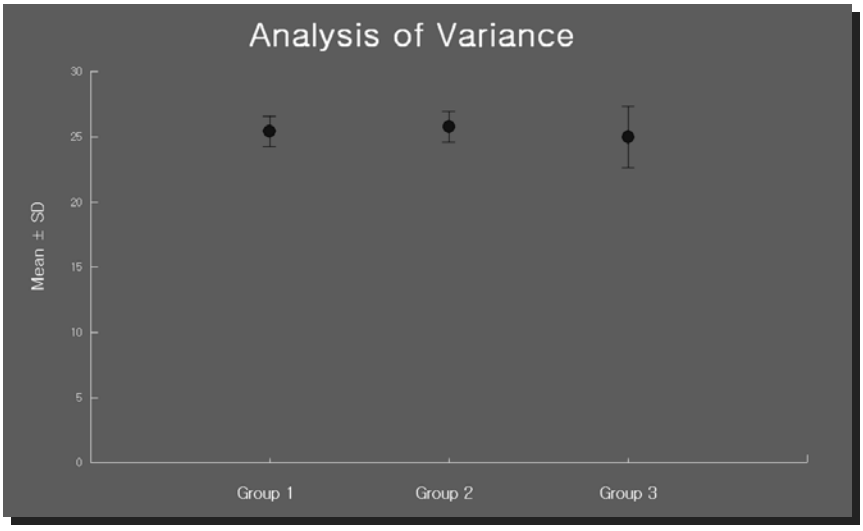
Drug A	Drug B	Drug C	Drug D
200	240	180	260
260	245	220	260
220	260	245	280
235	235	274	230
240	258	248	245

When to Use One-Way ANOVA

Difference - Means of $\geq$3 Samples

- Interval or Ratio Scale
- Normal Distribution
- Equal Variance - Bartlett's test
- k Independent Samples

Normally Distributed Variables



Multiple Comparison After ANOVA

Do not use multiple t test

- Scheffe's test
- Duncan's multiple range test
- Tukey test
- Student-Newman-Keuls (SNK) test
- Least Significance Difference (LSD) test
- Bonferroni test

Example

Kruskal-Wallis Test

Small City	Middle City	Large City
96	82	115
128	124	149
83	132	166
61	135	147
101	109	175

When to Use Kruskal-Wallis Test

Difference - Medians of $\geq$ 3 Samples

- Ordinal Scale
- k Independent Samples
- One-Way ANOVA by Ranks
- All Sample Sizes $\geq$ 5

Multiple Comparison After KW

Do not use multiple Mann-Whitney test

- Dunn test
- Miller test
- Mann-Whitney test with Bonferroni correction

Example

Repeated Measures ANOVA

Mouse	2hr	12hr	24hr
1	23	31	21
2	17	25	21
3	16	22	19
4	20	28	24
5	19	24	20

When to Use Repeated ANOVA

Difference - Means of $\geq$3 Samples

- Interval or Ratio Scale
- Normal Distribution
- Equal Variance of Differences
- k Related Samples

Multiple Comparison After Repeated ANOVA

Do not use multiple t test

- Bonferroni test
- Scheffe's test
- Duncan's multiple range test
- Tukey test
- Student-Newman-Keuls (SNK) test

Example
Friedman Test

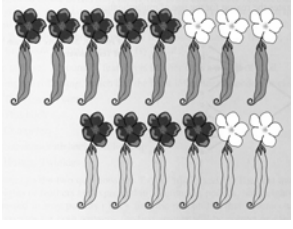
Therapist	MODEL1	MODEL2	MODEL3
1	2	3	1
2	2	3	1
3	2	3	1
4	1	3	2
5	3	2	1
6	1	2	3
7	2	3	1
8	1	3	2
9	1	3	2

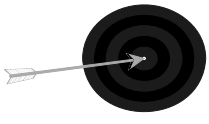
When to Use Friedman Test

Difference - Medians of $\geq$3 Samples

- Ordinal Scale
- Two-Way ANOVA by Ranks
- No Interaction – Blocks and Treatments
- k Related Samples

Summary of Inferential Tests

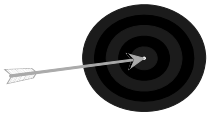




Statistical Test

Parametric vs Nonparametric

	Parametric	Nonparametric
Scale	Interval / Ratio	Ordinal / Norminal
Distribution	Normal	Normal or Not
Sample Size	Large (>10)	Small
Selection	Random Sample	Random or Not
Power	More Powerful	Less Power



Statistical Test

Parametric vs Nonparametric

Sample	Parametric	Nonparametric
One	t test	Kolmogorov-Smirnov
Two		
Independent	Student t test	Mann-Whitney
Paired	Paired t test	Wilcoxon signed rank
Three		
Independent	ANOVA	Kruskal-Wallis
Repeated	ANOVA	Friedman

A Simple Example

Blood Cholesterol Level Changes

Cholesterol Level Two Days After	Cholesterol Level Four Days After
270	218
236	234
210	214
142	116
280	200
272	276
160	146
220	182
226	238
242	288
186	190
266	236