Common Errors in Statistics How to Avoid Them

Kim, Soo-Nyung

Konkuk University Medical School

Introduction

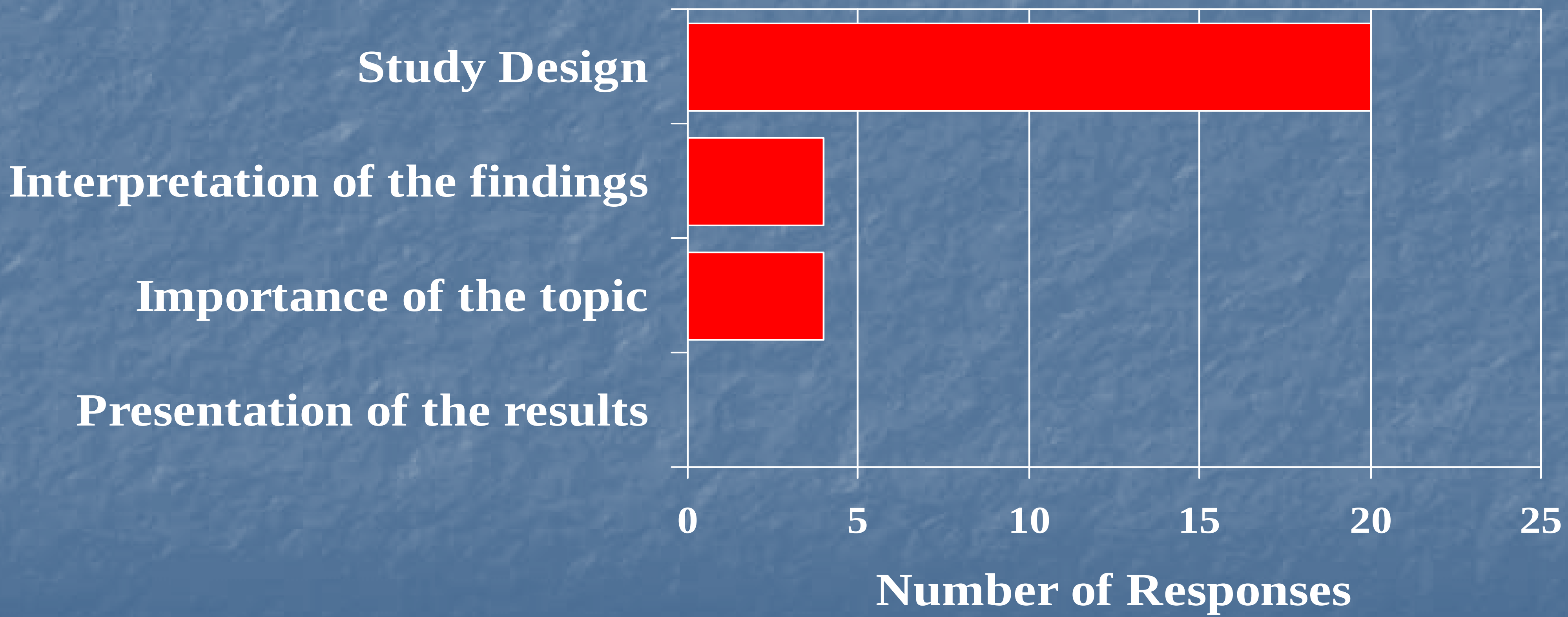
- Any errors in statistical analyses will mean that the conclusions of the study may be incorrect.
- In this era of evidence-based health care, both clinicians and researchers need to critically appraise the statistical aspects of published articles.

Introduction

"Bad statistics leads to bad research and bad research is unethical." - Altman (1982)

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write." - Wells

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
- 대한의학학술지편집인협의회 www.kamje.or.kr
- 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.

Golden Rules for Reporting Numbers

Rule	Correct Expression
Use a zero before the decimal point	The P value was 0.013
Report percentages to only one decimal place if the sample size is larger than 100	In the sample of 212 children, 10.4% had diabetes
Report percentages with no decimal places if the sample size is less than 100	In the sample of 44 children, 11% had diabetes
Do not use percentages if the sample size is less than 20	In the sample of 18 children, 2 had diabetes
Do not imply greater precision than the measurement instrument	Only use one decimal place more than the basic unit of measurement when reporting statistics (mean, median, SD, 95% CI, etc.) e.g. mean height was 143.2 cm
For ranges use ‘to’ or a comma but not ‘-’ to avoid confusion with a minus sign	The mean height was 162 cm (95% CI 156 to 168)
P values between 0.001 and 0.05 should be reported to three decimal places	There was a significant difference in blood pressure between the two groups ($t = 3.0$, $df = 45$, $P = 0.004$)
P values shown on output as 0.000 should be reported as <0.0001	There was a significant difference in blood pressure between the two groups ($t = 5.47$, $df = 78$, $P < 0.0001$)

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.

Neuron Numbers and Volume of the Amygdala in Subjects Diagnosed with Bipolar Disorder or Schizophrenia

Sabina Berretta, Harry Pantazopoulos, and Nicholas Lange

Statistical Analysis

Differences between groups relative to the main outcome measures, that is, total number, numeric density, and somata size of neurons, as well as volume, of the LN, BN, ABN, and CO were assessed for statistical significance using a stepwise linear regression process. A logarithmic transformation was uniformly applied to all original values because the data were not normally distributed. Statistical analyses were performed using JMP v5.0.1a (SAS Institute, Cary, North Carolina). The BD and SZ subjects were first compared separately with normal control subjects. Subsequently, the three groups were considered together using an analysis of variance (ANOVA) version of the same linear regression process to test for differences between diagnostic groups while correcting for multiple comparisons. Age, gender, postmortem time interval (PMI), hemisphere, cause of death, brain weight, alcohol exposure, and lifetime as well as last 6 months' exposure to antipsychotic drugs and lithium treatment were tested systematically for their effects on total number and numeric density of neurons, somata size, and volume; these were included in the model if they significantly improved the model's goodness-of-fit. Values relative to the *t* ratio and *p* value for main outcome measure differences found to be statistically significant are reported in the Results section. Any and all covariates found to significantly affect an outcome measure are also reported, together with their *p* values.

Cause of death was categorized as acute (e.g., myocardial infarction) or chronic (e.g., cancer). Data on alcohol exposure was only available for subjects with SZ or BD and were categorized on the basis of the subjects' record as high, moderate, and absent as well as present or absent during the last 10 years of life. We analyzed medical records for exposure to various classes of psychotropic and neurotropic drugs. Estimated daily milligram doses of antipsychotic drugs were converted to the approximate equivalent of chlorpromazine (CPZ-eq) as a standard comparator (52) and corrected on the basis of a qualitative assessment of treatment-adherence based on taking prescribed psychotropic medicines more or less than approximately half of the time, as indicated by the extensive antemortem clinical records. These values are reported as lifetime as well as last 6 months of life (CPZ-eq/6m), grams per patient (Supplemental Tables 2 and 3, respectively). Exposure to lithium salt was estimated and reported in the same manner. The use of other classes of psychotropic drugs was reported as present or absent and could be not tested reliably because the number of subjects exposed to each was too low. However, exposure to these drugs was taken into account as a possible explanation when apparent clustering of subjects was observed (Supplemental Table 2). In addition to testing the potential effects of exposure to antipsychotics and lithium salt within our stepwise linear regression process, the effects of these variables, together with other psychotropic and neurotropic drugs, adherence to pharmacologic treatment (good or poor; see Supplemental Table 2), age of onset of the disease, and duration of the illness, were tested directly in separate ANOVA.

통계적 오류의 종류

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

의학논문 1

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 1

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

Population and Sample

- Parameter
- Statistic

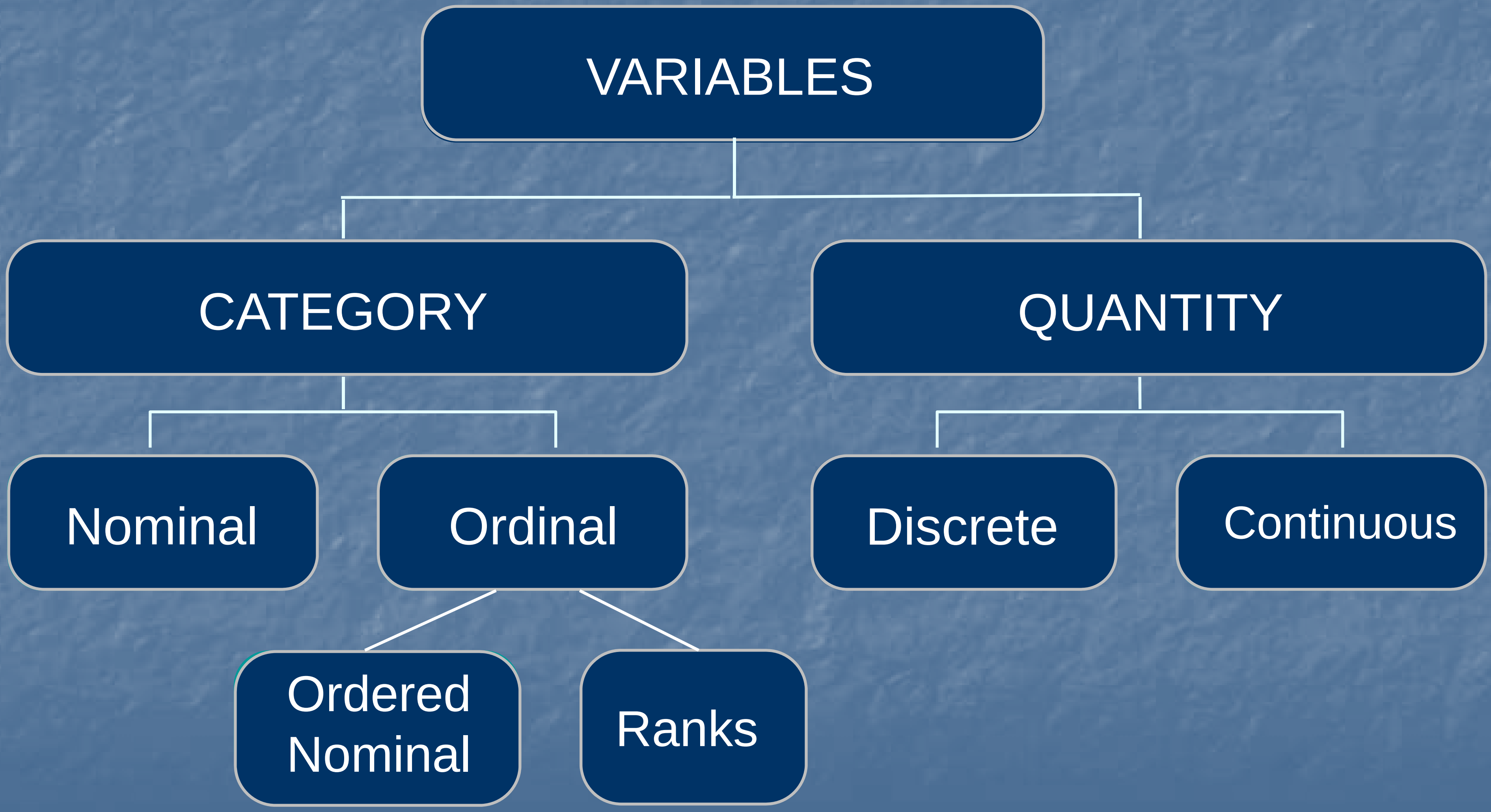
Data Type

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

Level of Measurement

Data	Character	Example
Nominal	categorical	gender, race, blood type, hypertensive
Ordinal	categorical ranked	rating scale, anxiety score, grade, tumor stage
Interval	continuous	temperature, calendar year, IQ, psychological test
Ratio	interval with true zero	blood pressure, age, days in the hospital

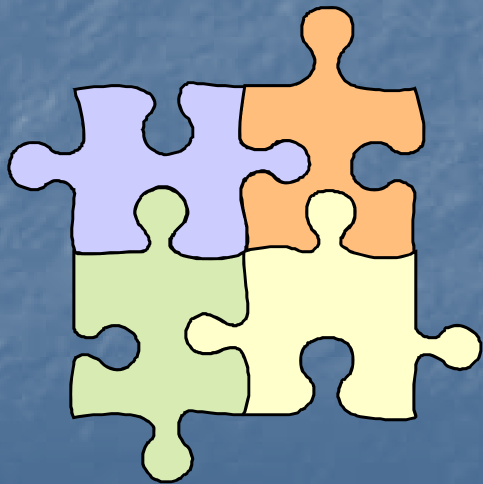
Statistical Variables



Statistical Analysis

1. Descriptive Statistics
2. Inferential Statistics

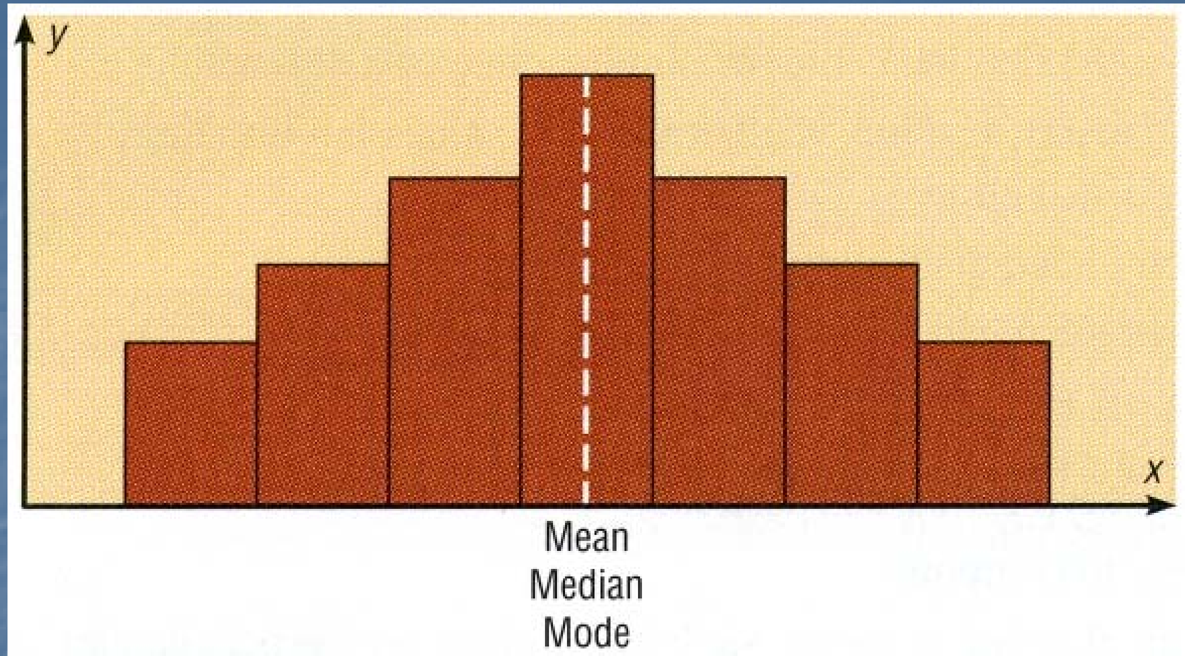
How to Summarize Data



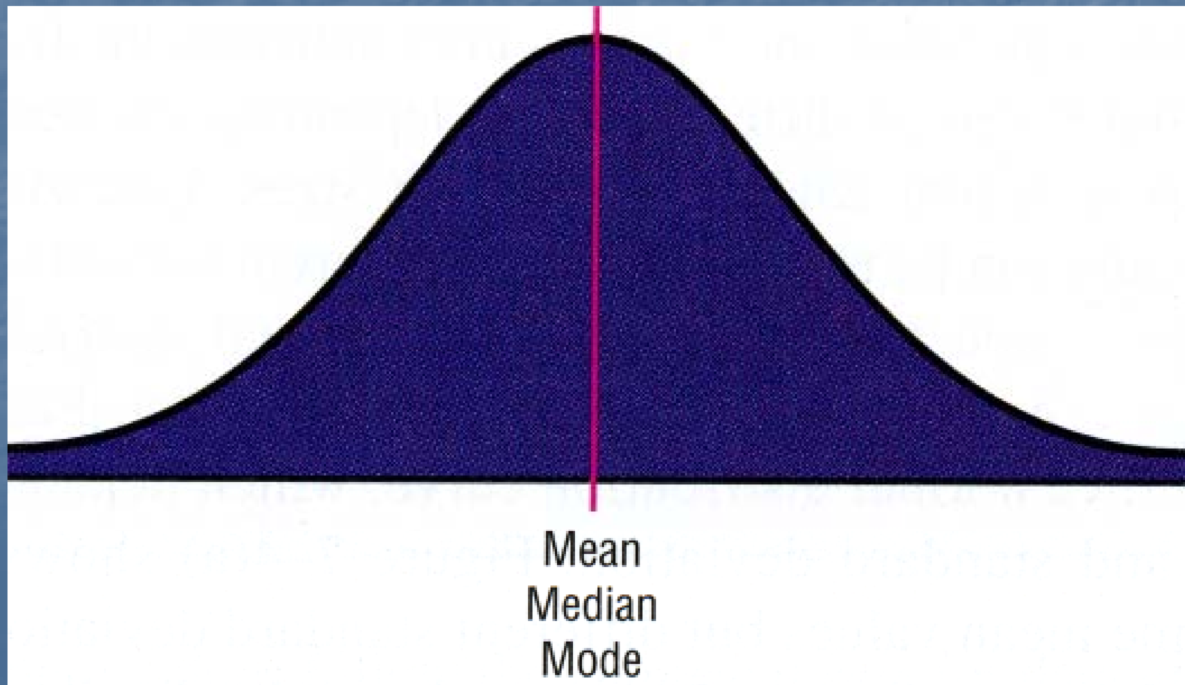
Central Tendency

- Mean
- Median
- Mode

Symmetrical Distribution



Normal Distribution

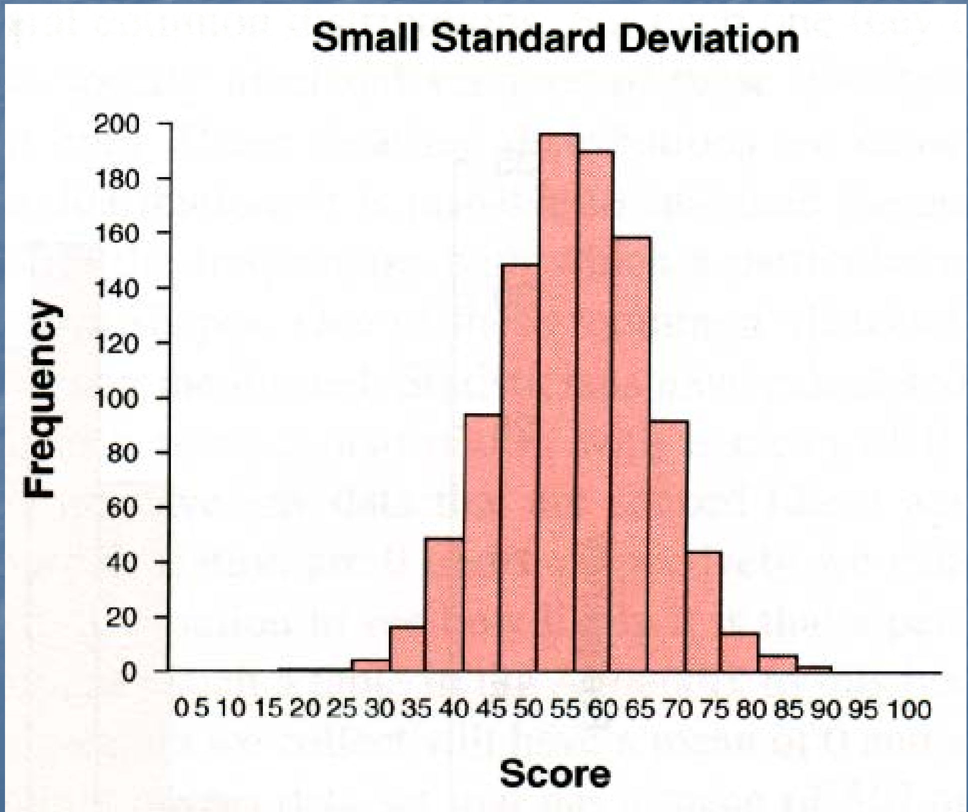


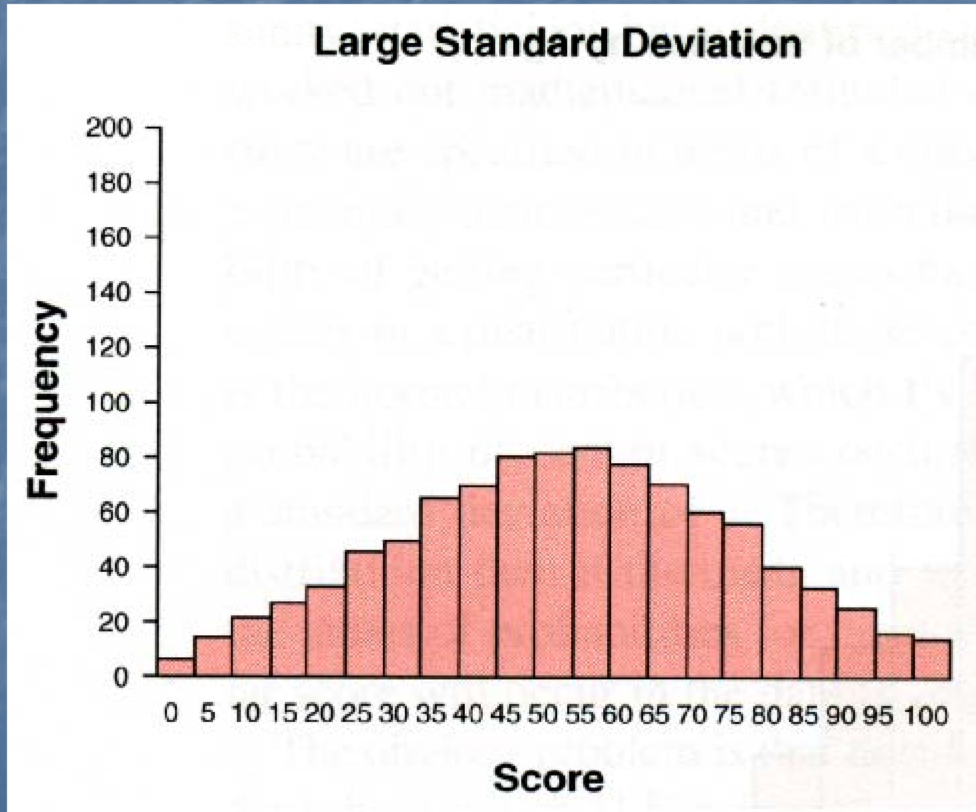
Dispersion

Variability of a single sample values

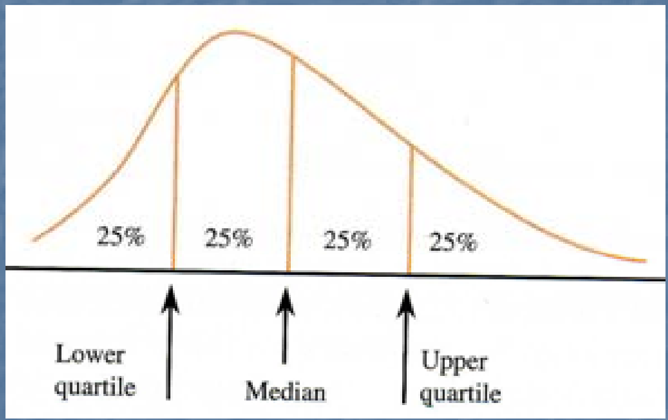
- SD (Variance)
- Range
- IQR (Inter-quartile range)

Standard Deviation





Interquartile Range



Q_1 Q_2 Q_3

- Quantiles (Percentiles)
- Quartiles (25^{th}, 50^{th}, 75^{th})
- IQ Range = $Q_3 - Q_1$

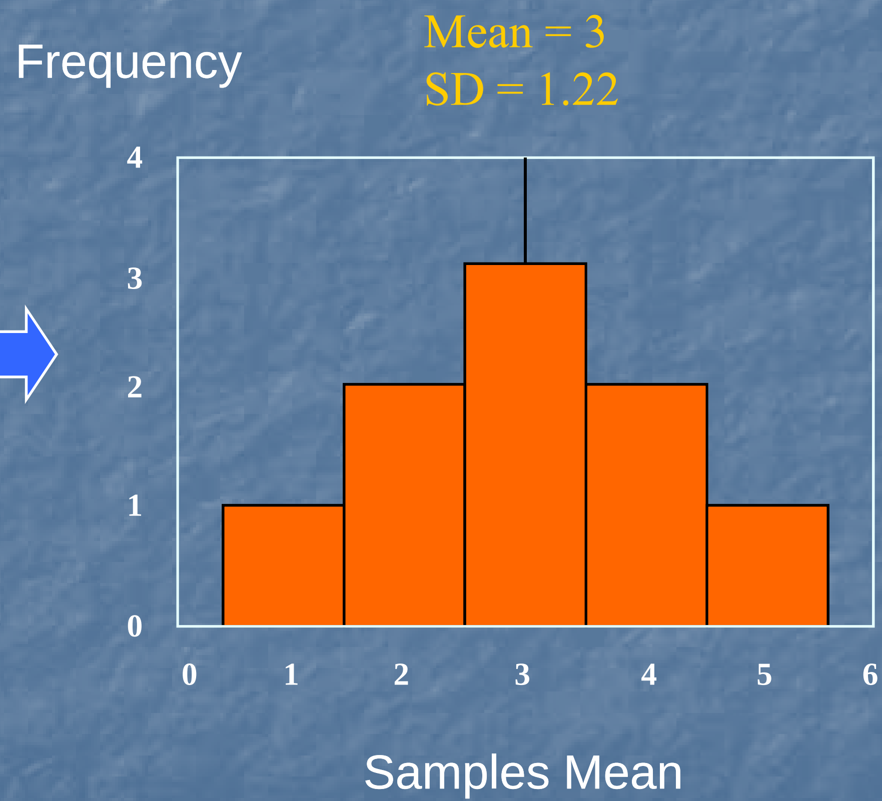
The percentiles divide a data set into 100 equal parts.

The quartiles divide a data set into quarters (four equal parts).

What is the Standard Error?

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

Standard Error



Measures of Variability

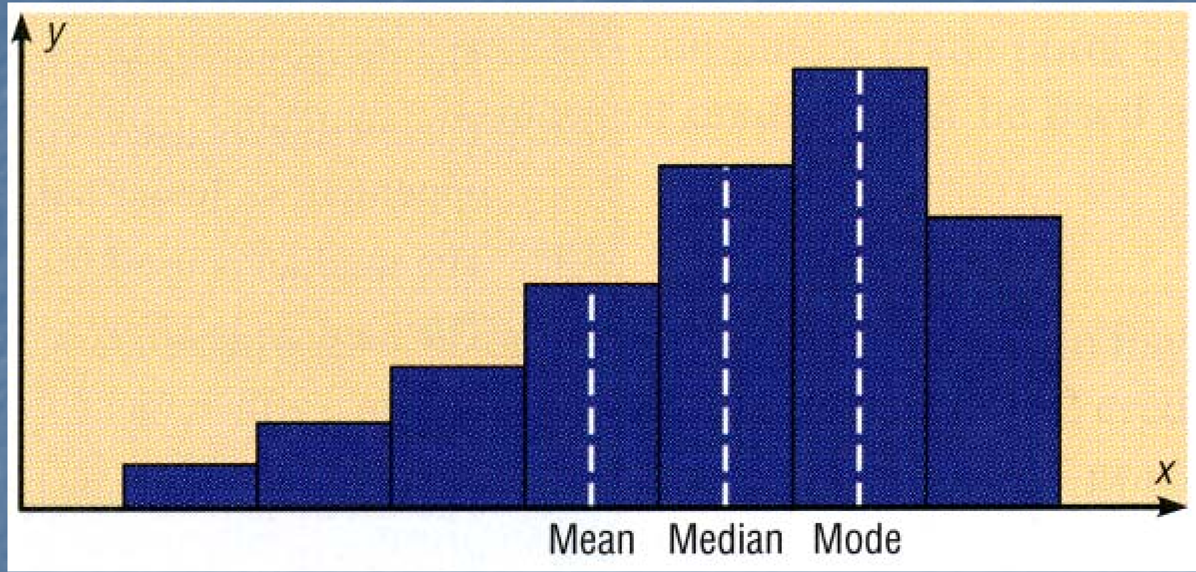
Statistic	Variability	Mean ± 2SD (2SE)
Standard Deviation	Single Sample's Values	95% of Population Values
Standard Error	Sample Means	95% of Population Mean

Shape

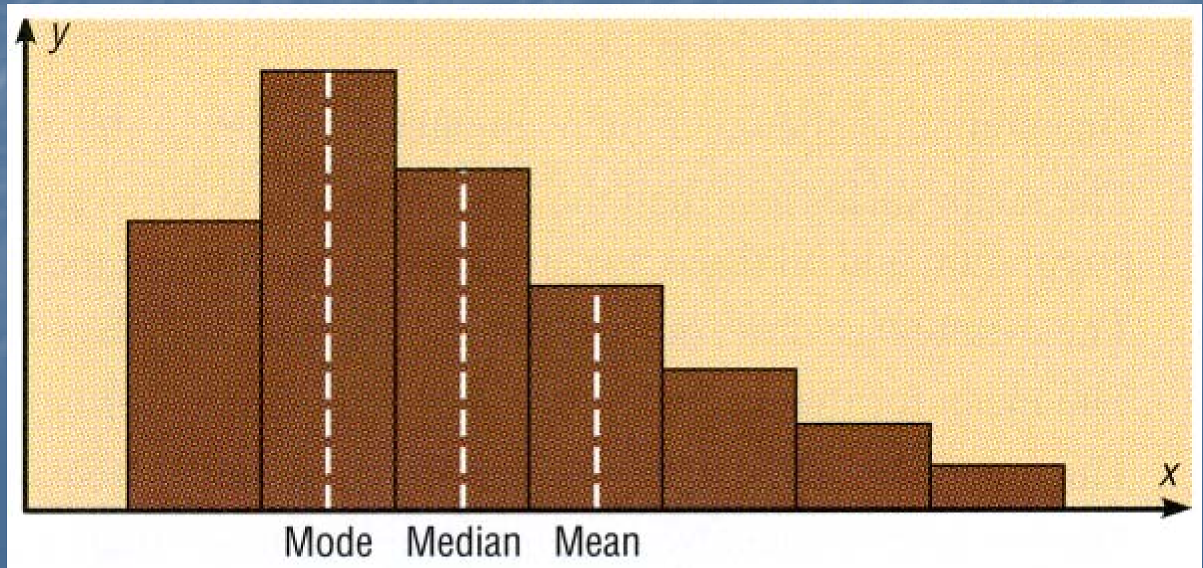
- Skewness
- Kurtosis

Skewed Distribution

- Negatively skewed



- Positively skewed



Descriptive Statistics

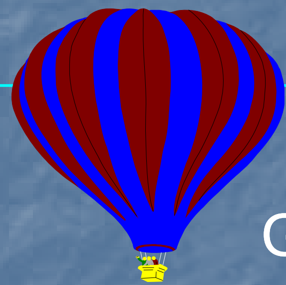
Data Analysis

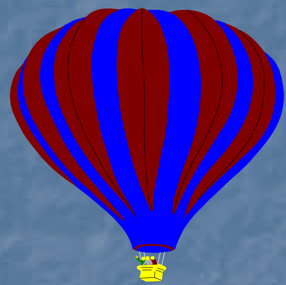
Central Tendency	Dispersion	Shape
Mean	Variance	Skewness
Median	SD	Kurtosis
Mode	Range	

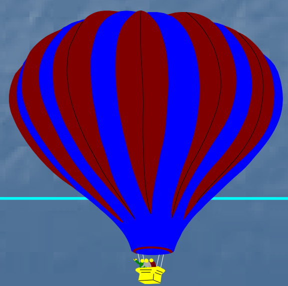
Descriptive Statistics

Data Type

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







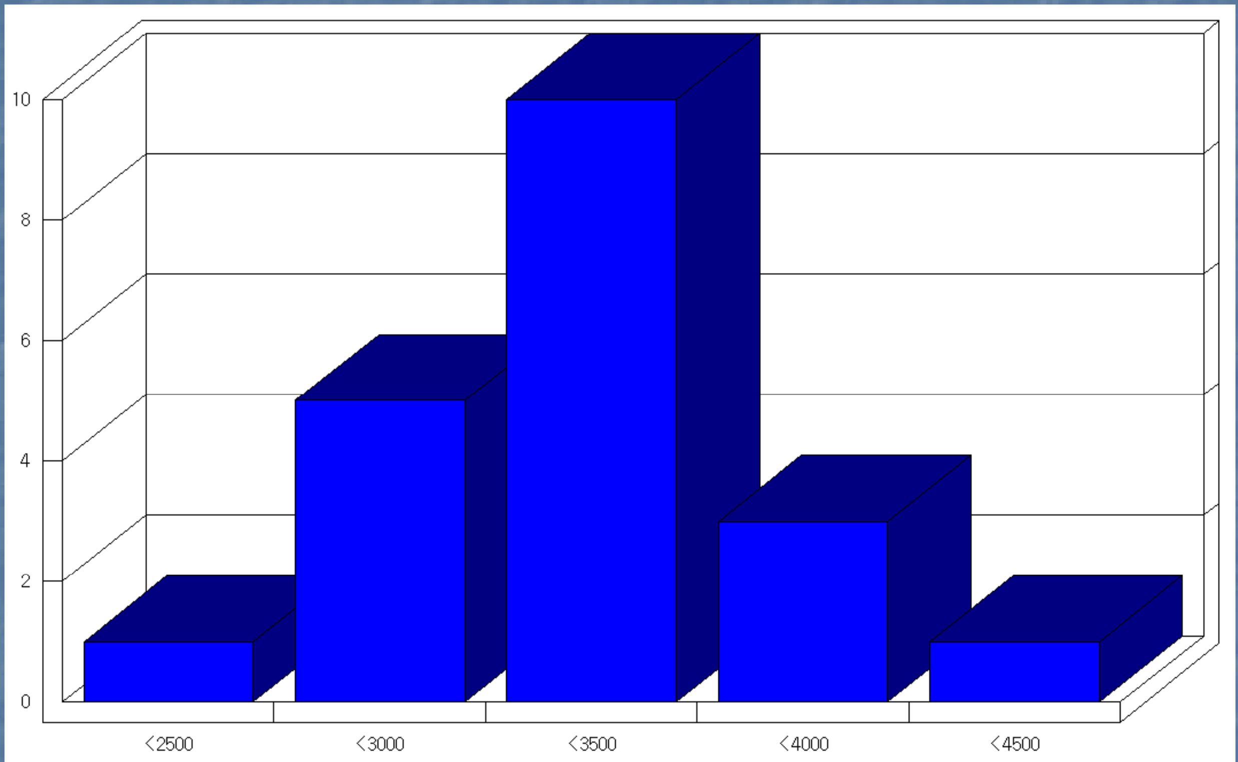
Example: Descriptive Statistics

The following values were obtained from 20 newborn babies.

ID	Weight	ID	Weight
001	3265	011	2581
002	3260	012	2481
003	3245	013	3609
004	3484	014	2838
005	4146	015	3541
006	3323	016	2759
007	3649	017	3248
008	3200	018	3314
009	3031	019	3101
010	2069	020	2834

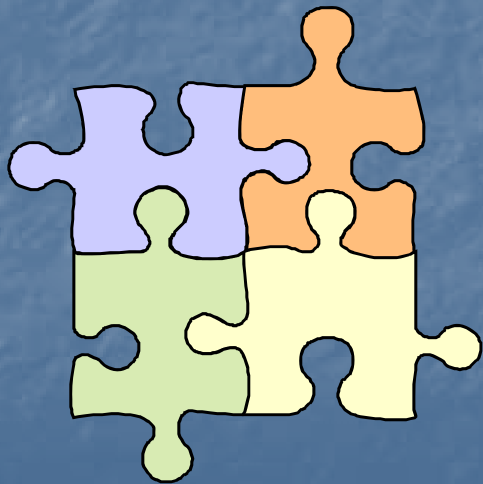
weight (gm)

Distribution in Sample



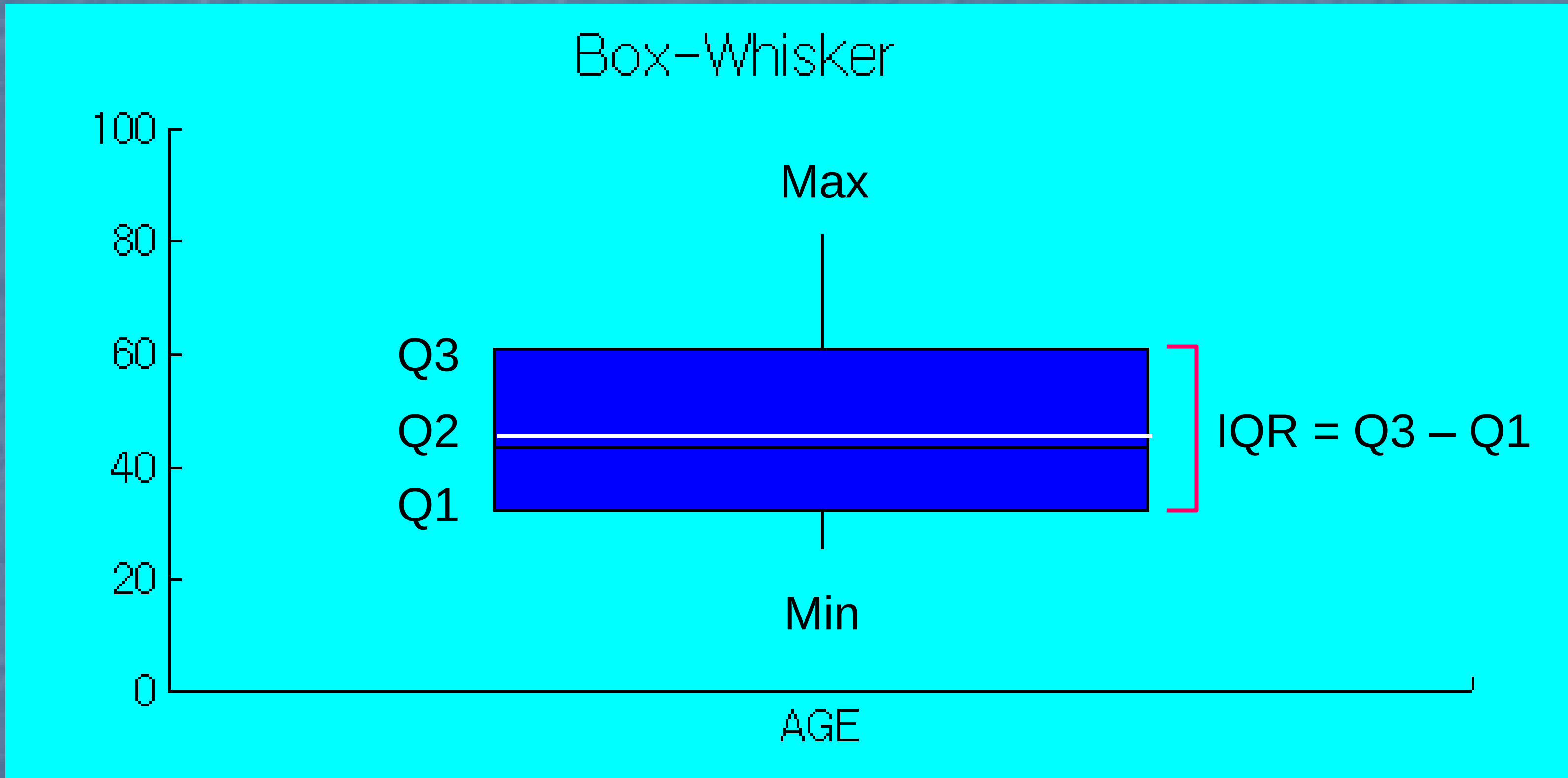
Variable	N	Mean (SD)	Median (IQ range)
Birth weight	20	3166.9 (445.3)	3246.5 (2839.5 to 3403.5)

Data Screening for Statistical Analysis



Five-Number Summary

Box Plot (Box-and-Whisker)

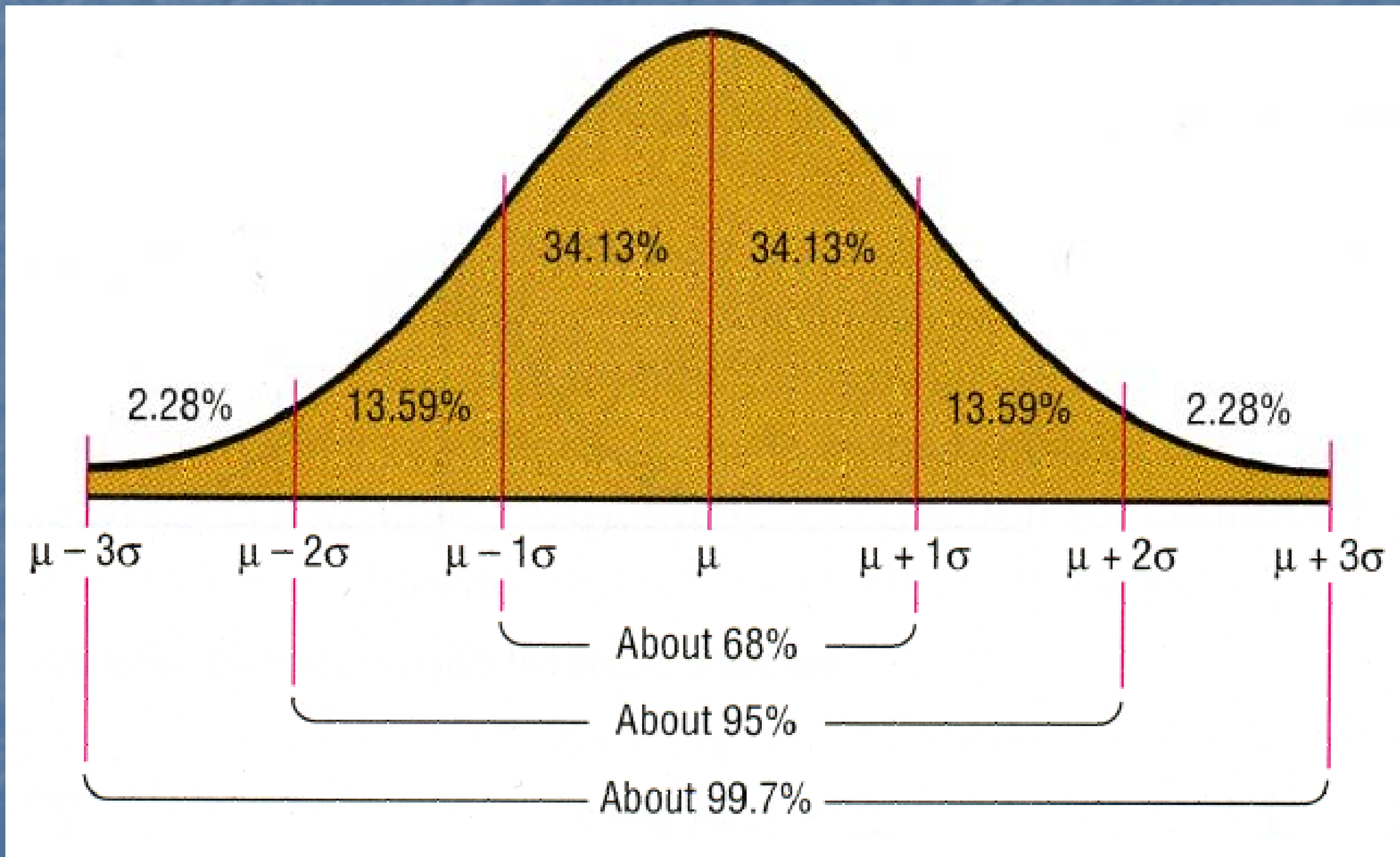


You Will Need to Learn Inferential Statistics

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.

Normal Distribution



Tests of Normality

1. 95% Range (mean $\pm$ 2SD)
2. Skewness and Kurtosis
3. Outliers
4. Statistical Tests
5. Normality Plots

95% Range of Variables

- 95% of the data value = mean ± 1.96 SD
- Mean < 2 SD

  non-normally distributed

  assuming that negative values are impossible.

Variable	Mean ± 2 SD	95% range	Minimum and Maximum
Age	38 ± (2 x 6.1)	26 to 50	20 to 57
Weight	54.2 ± (2 x 4.0)	46.2 to 62.2	41.0 to 63.0
Length of Stay	7.5 ± (2 x 5.0)	-2.5 to 17.5	0 to 42

Practical Example

Variable	Normotensive group (n=31)	Hypertensive group (n=11)	P
Plasma renin activity (ng/mL/hr)	0.5±0.8	0.5±0.7	0.945
Aldosterone (pg/mL)	490.2±268.6	452.8±261.9	0.692
Potassium (mmol/L)	2.9±0.8	3.2±0.4	0.073
Creatinine (mg/dL)	0.9±0.2	1.4±1.2	0.187
Size of the tumor (mm)	19.1±6.9	17.9±4.7	0.610
Side of the tumor (Right/Left)	10/21	2/9	0.464

Test for Normality

- √ Kolmogorov-Smirnov
- √ Shapiro-Wilk test
- √ Lilliefors test

- √ Cochran test
- √ Chi-square
- √ Anderson-Darling test

Normality Plots

- √ Histograms
- √ Normal Q-Q plot
- √ Box plot
- √ Detrended normal Q-Q plot

Example: Normal Distribution

The serum cortisol levels for 30 patients with diabetes is shown below.

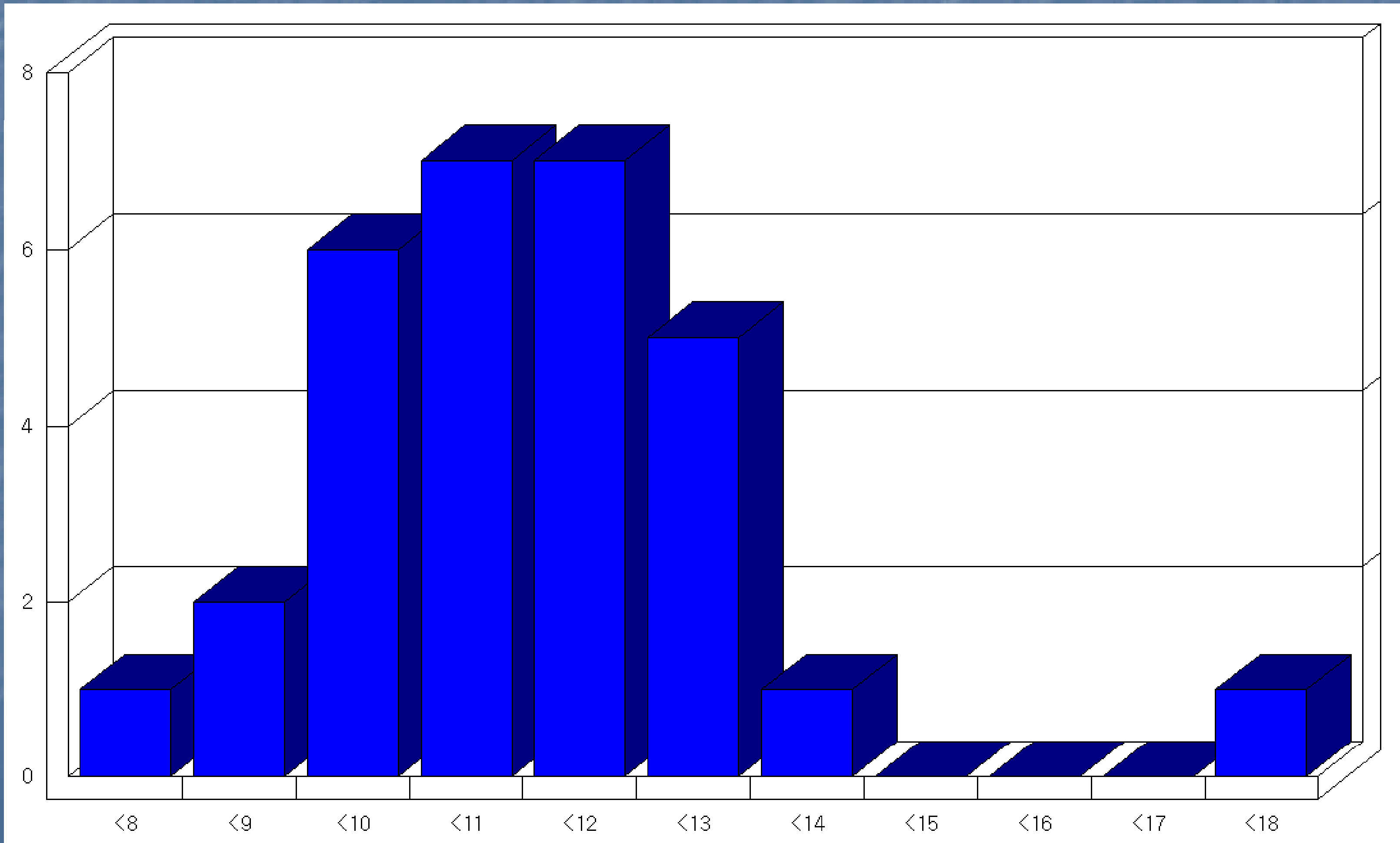
ID	Cortisol	ID	Cortisol	ID	Cortisol
001	13.1	011	9.7	021	10.6
002	8.6	012	11.3	022	11.4
003	11.7	013	11.8	023	10.8
004	10.4	014	8.9	024	9.7
005	9.0	015	12.5	025	10.1
006	12.9	016	7.7	026	9.4
007	9.5	017	10.0	027	17.3
008	12.0	018	9.6	028	11.5
009	10.2	019	12.2	029	10.9
010	11.6	020	11.8	030	12.0

Descriptives

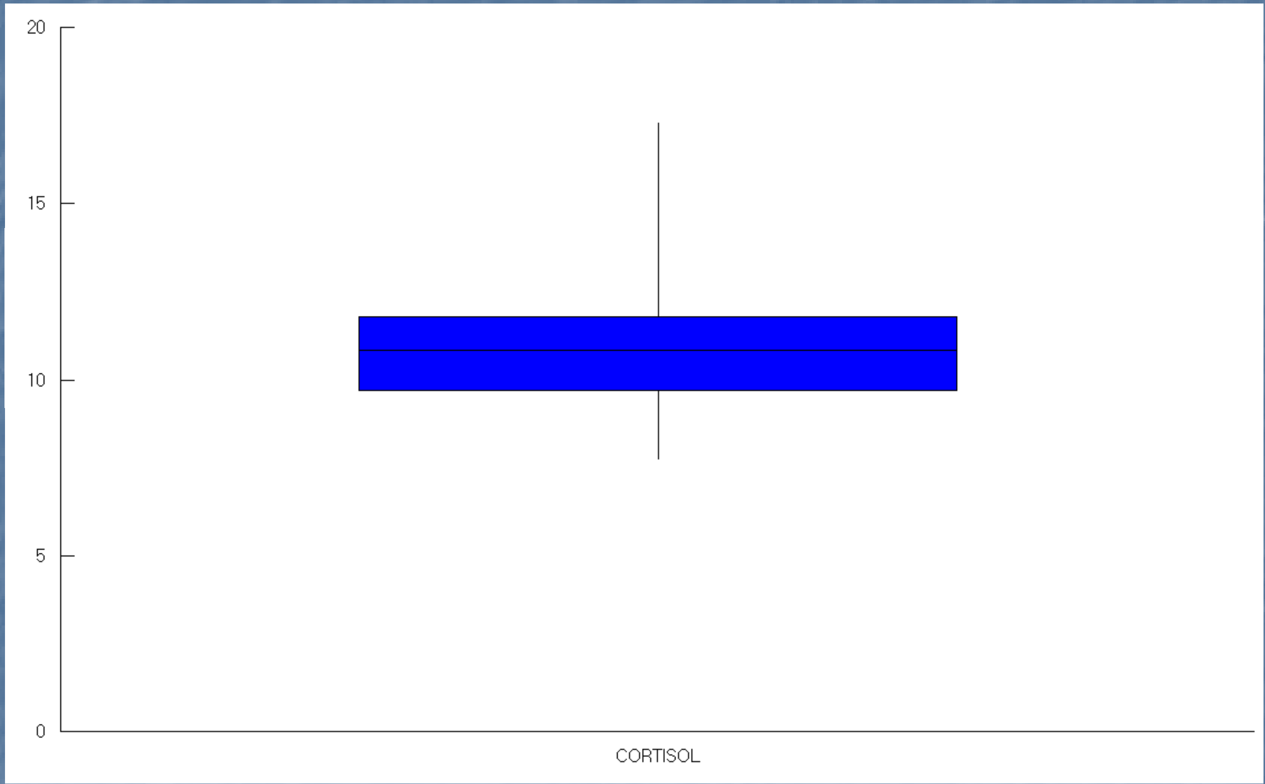
Sample size	Center	Dispersion	Shape
30	Mean 10.94	SD 1.80	Skewness 1.3
	Median 10.85	IQ Range 2.10	Kurtosis 4.2

Variable	N	Mean (SD)	Median (IQ range)
Cortisol	30	10.94 (1.80)	10.85 (2.10)

Histogram

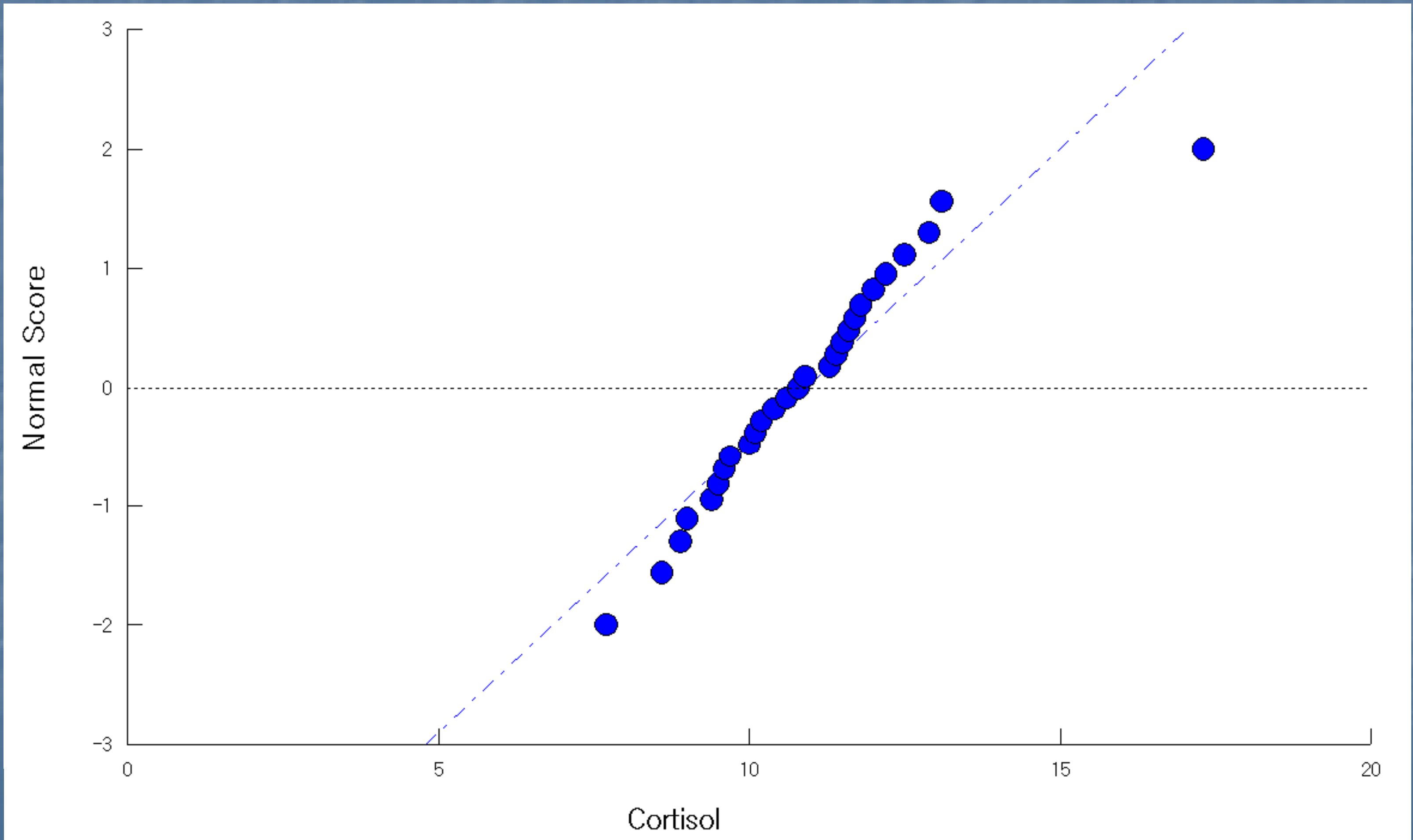


Box Plot and Outlier



	Value	Z score	Outlier
Minimum	7.7	1.80	No
Maximum	17.3	3.53	Yes

Normal Q-Q Plot



Normality Test

[Kolmogorov-Smirnov Test]

검정하고자 하는 분포 - 정규분포

평 균:	10.9400	표준편차:	1.8018
표본수:	30	D MAX :	0.1115
유의수준 (α = 0.05): T = 0.242		확 률: P = 0.8837	

결 론: 정규분포를 이룹니다.

[Lilliefors Test]

Max Difference D = 0.1115

확 률: P = 0.100* (P > 0.10)

결 론: 정규분포를 이룹니다.

[Shapiro-Wilk Test]

Shapiro-Wilk W Statistic = 0.9138

확 률: P = 0.0205

결 론: 정규분포를 이루지 않습니다.

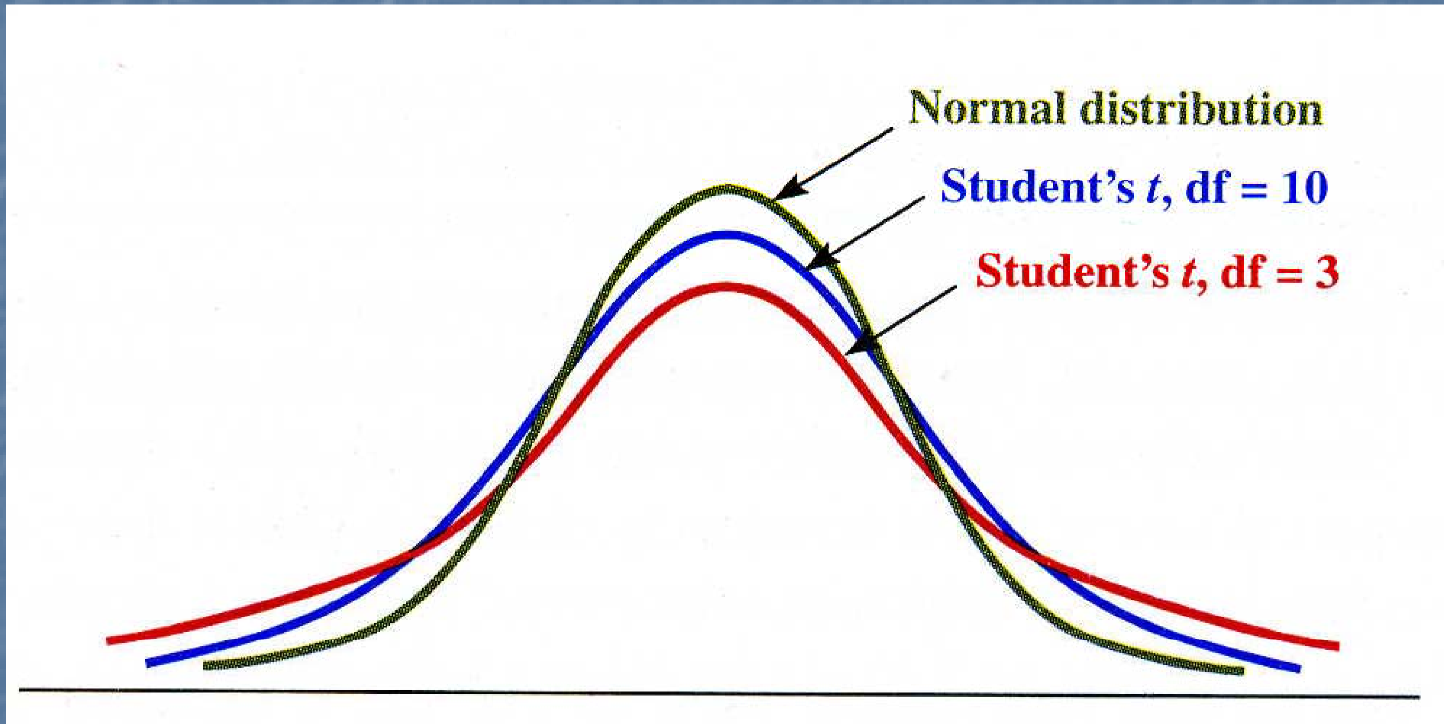
Central Limit Theorem

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

$N > 30$

t Distribution

- Sampling distribution of sample means



Degrees of Freedom, df = n – 1

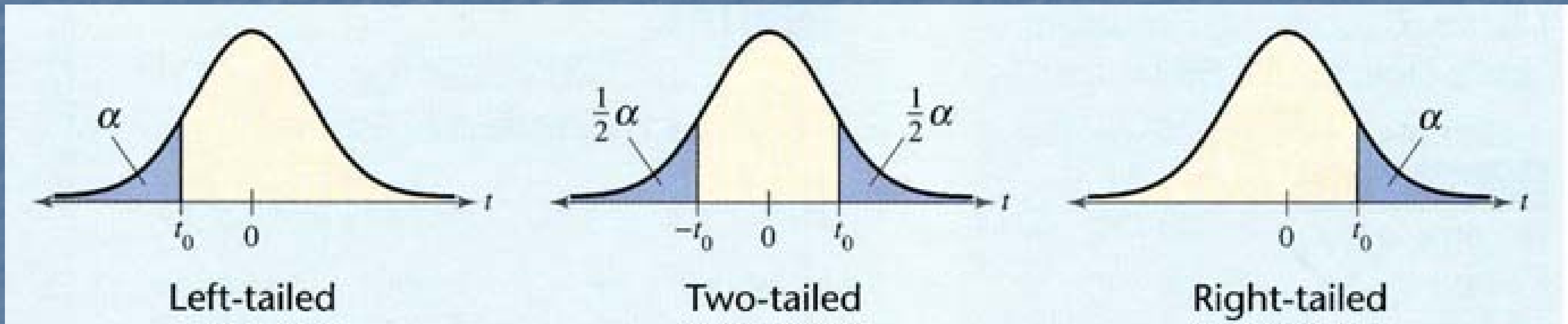
A parameter that identifies each different distribution of the Student's t-distribution.

Statistical Hypothesis

1. Null Hypothesis
2. Alternate Hypothesis

Rejection Region

One-tailed | Two-tailed | One-tailed



*H*a: $\mu < k$ | *H*a: $\mu \neq k$ | *H*a: $\mu > k$

Type I and Type II Errors

Null Hypothesis (*Ho*)

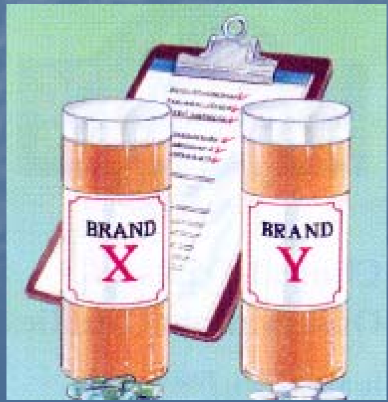
	True	False
Reject	**Type I**	Correct
Accept	Correct	**Type II**

- Type I errors – False positive result, no clinically important difference
- Type II errors – False negative result, small sample size

Level of Significance

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

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

1. Parametric Statistics
2. Nonparametric Statistics

Parametric Test

- Z test
- t test
- F test
- ANOVA

When to Use Parametric Test

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

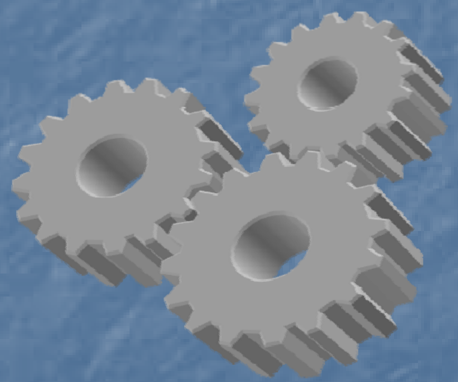
Nonparametric Test

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

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?

Example: Paired Data

Two tests were given to 7 individuals.

Subject	*First Test*	*Second Test*
001	50	69
002	66	85
003	73	88
004	84	70
005	57	84
006	83	78
007	76	90

Example: Independent Data

Heart rate was assessed among newborns.

ID Number	*Gender*	*Heart Rate*
001	1	125
002	1	110
003	1	130
004	1	120
005	2	133
006	2	136
007	2	124
008	2	118

Gender 1 = male, 2 = female

Independent vs. Dependent

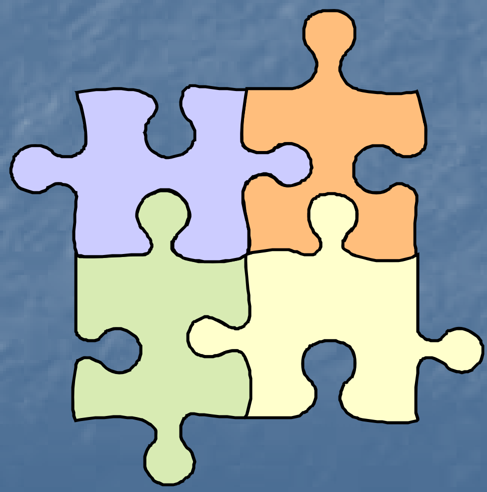
- The classification of variables depends on the study design.
- In most observational and experimental studies, the exposure will be the explanatory (independent) variable and the disease will be the outcome (dependent) variable.

Example

R x C Contingency Table

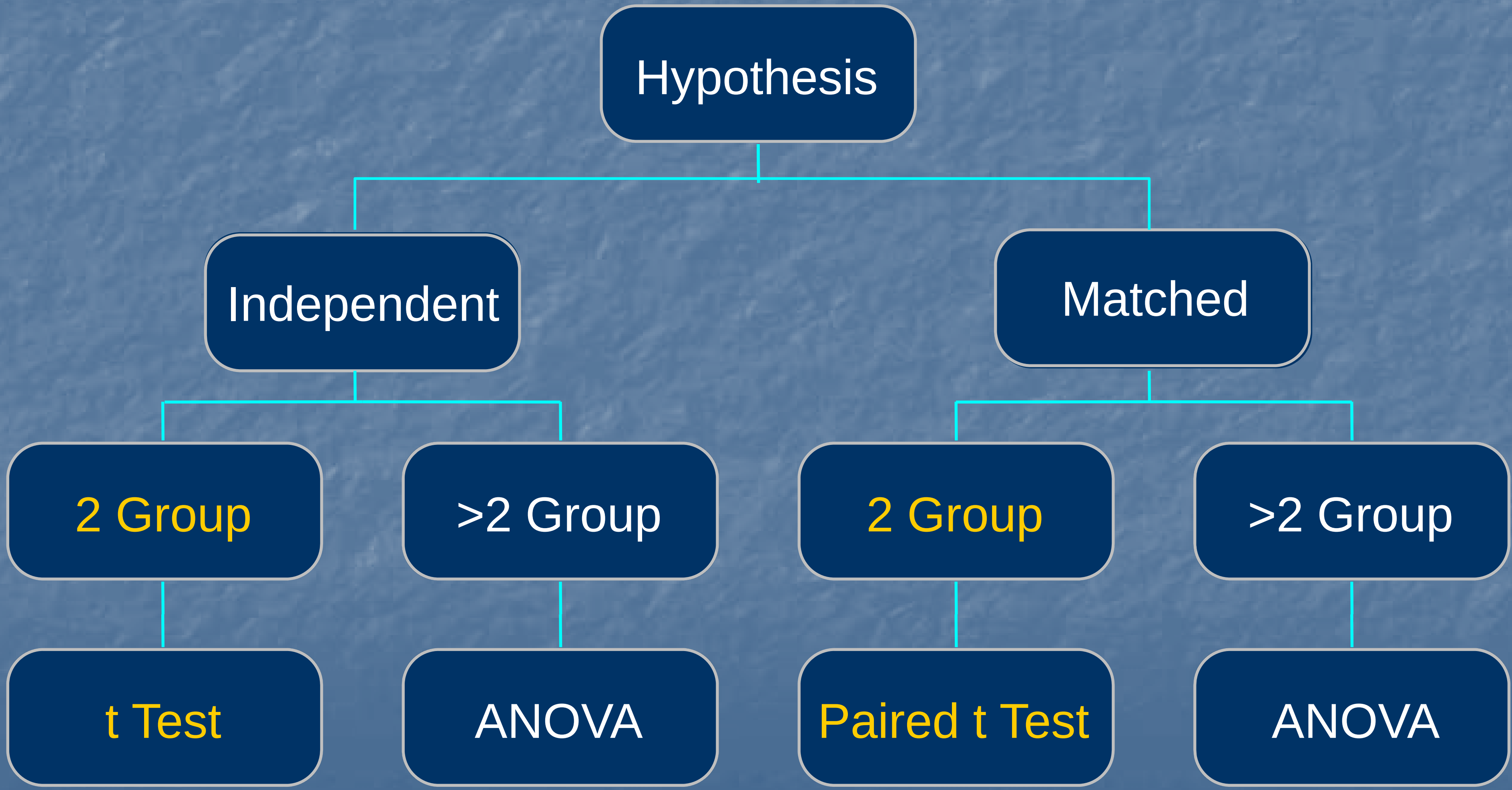
Independent \ Dependent	Asthma (-)	Asthma (+)
Smoking (-)	300	200
Smoking (+)	100	400

You Will Need to Know Which Statistical Test to Use



Two-Sample Methods

Statistical Test
Interval Data



Example: Student's t-test

24 hour total energy expenditure in groups of lean and obese women

Patient No.	Lean (n = 13)	Obese (n = 9)
1	6.13	8.79
2	7.05	9.19
3	7.48	9.21
4	7.48	9.68
5	7.53	9.97
6	7.58	11.51
7	7.90	11.85
8	8.08	12.79
9	8.09	
10	8.11	
11	8.40	
12	10.15	
13	10.88	

Practical Statistics for Medical Research, 1991 by Altman

When to Use Student's t Test

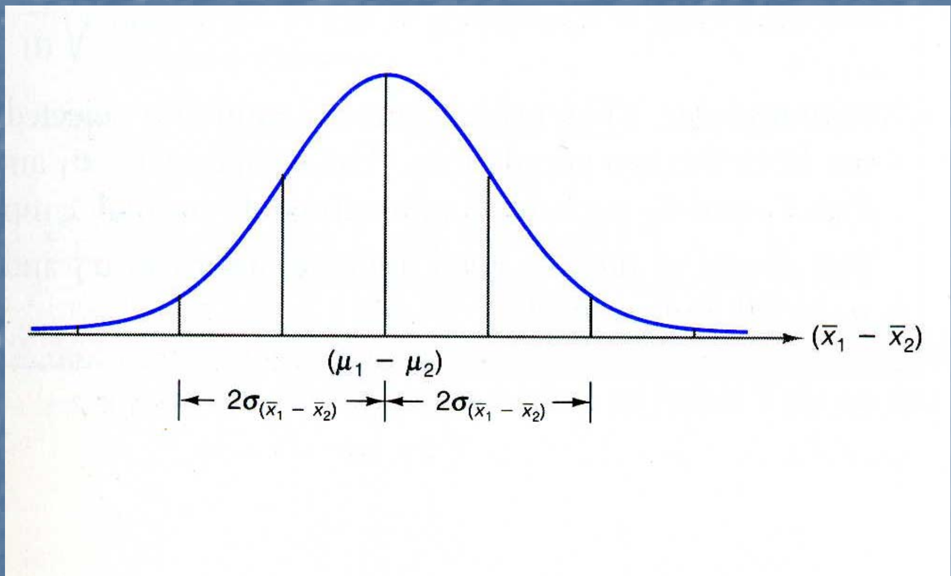
Difference - Means of 2 Samples

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

*Satterthwaite t-test – Unequal Variance

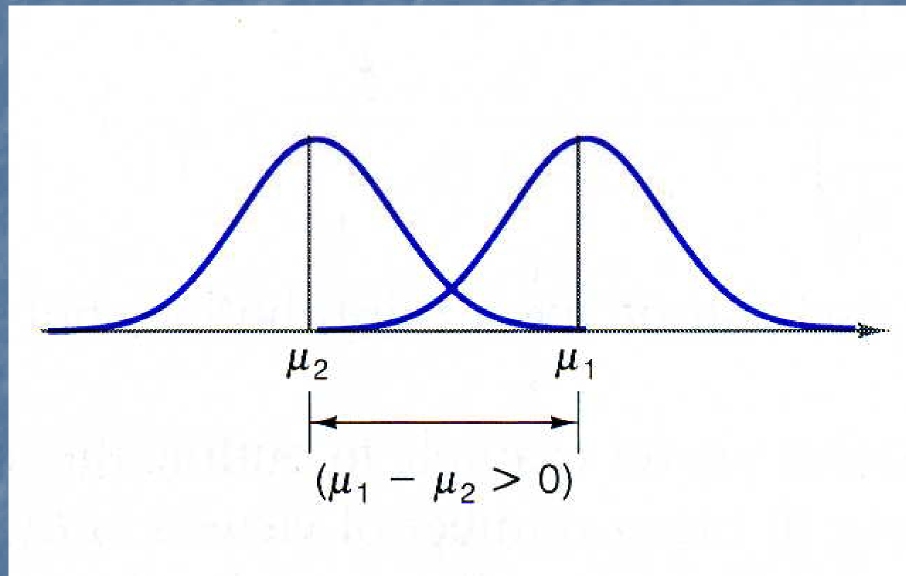
Assumptions for two-sample t

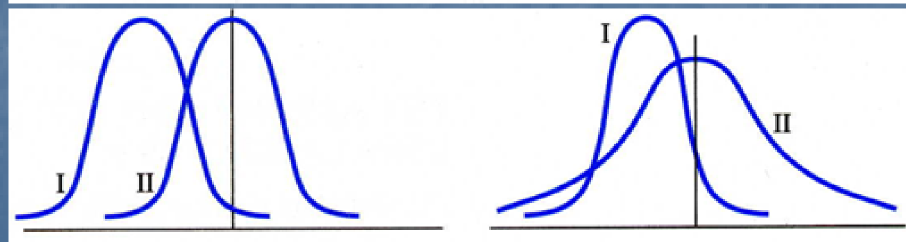
- Normal Distribution



Sampling Distribution of $X_1 – X_2$

- Equal Variance





Student's t test

WILLIAM GOSSET: The "Student" in Student's t-Distribution

William Sealy Gosset was born in Canterbury, England, on June 13, 1876, the eldest son of Colonel Frederic Gosset and Agnes Sealy. He studied mathematics and chemistry at Winchester College and New College, Oxford, receiving a first-class degree in natural sciences in 1899.

After graduation Gosset began work with Arthur Guinness and Sons, a brewery in Dublin, Ireland. He saw the need for accurate statistical analyses of various brewing processes ranging from barley production to yeast fermentation, and pressed the firm to solicit mathematical advice. In 1906, the brewery sent him to work under Karl Pearson (see Biography in Chapter 13) at University College in London.

During the next few years, Gosset developed what has come to be known as Student's *t*-distribution. This distribution has proved to be fundamental in statistical analyses involving normal distributions. In particular, Student's *t*-distribution is used in performing inferences for a population mean when the population being sampled is (approximately) normally distributed and the population standard deviation is unknown. Although the statistical theory for large samples had been completed in the early 1800s, no small-sample theory was available before Gosset's work.

Because Guinness's brewery prohibited its employees from publishing any of their research, Gosset published his contributions to statistical theory under the pseudonym "Student"—thus the name "Student" in Student's *t*-distribution.

Gosset remained with Guinness his entire working life. In 1935, he moved to London to take charge of a new brewery. But his tenure there was short lived; he died in Beaconsfield, England, on October 16, 1937.

Example: Paired t-test

A study was done to see how blood cholesterol level changed following a heart attack.

Patient No.	Cholesterol Level 2 Days After	Cholesterol Level 4 Days After
1	270	218
2	236	234
3	210	214
4	142	116
5	280	200
6	272	276
7	160	146
8	220	182
9	226	238
10	242	288
11	186	190
12	266	236

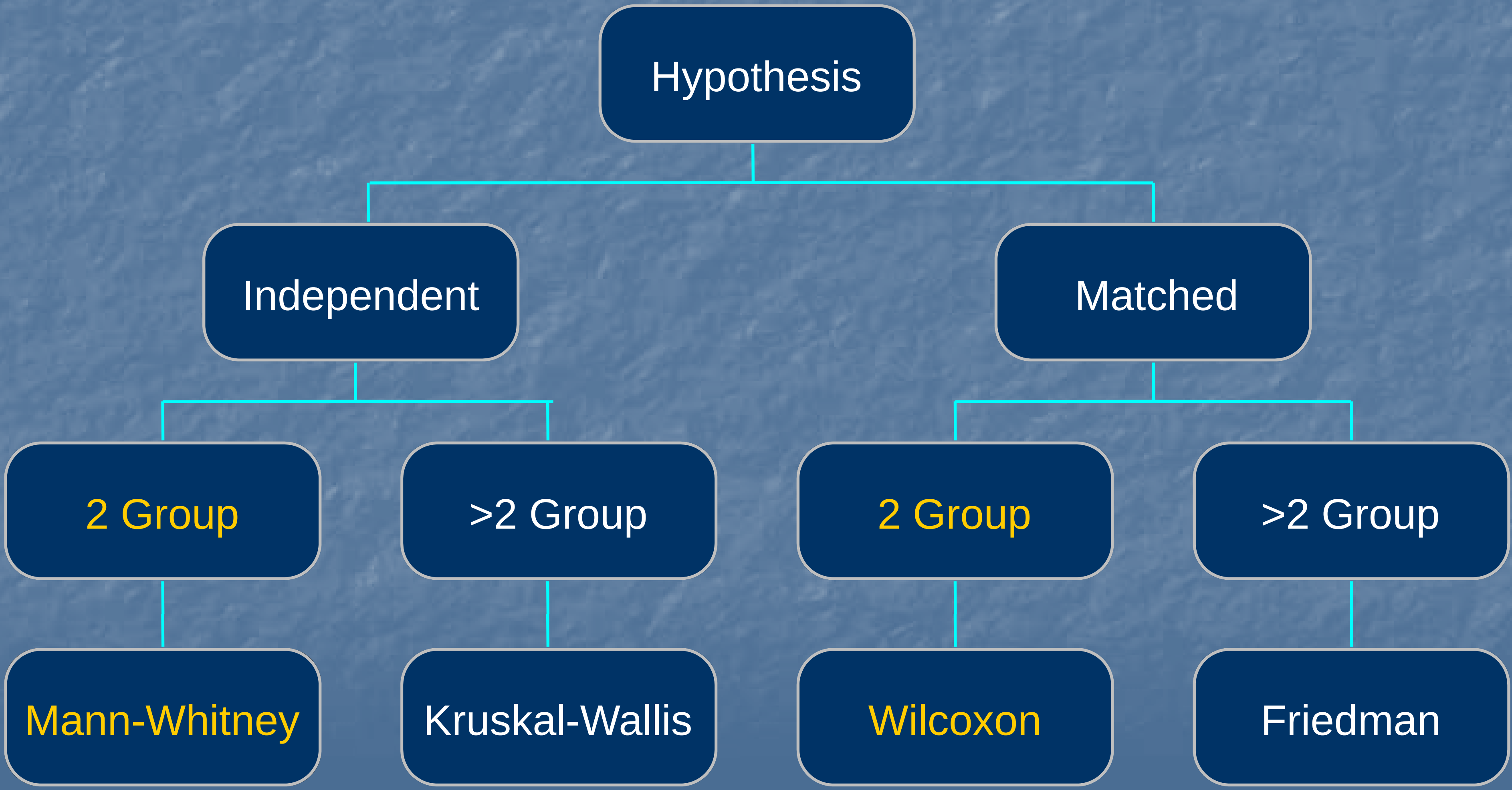
Minitab Handbook , 2^{nd} edition

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 Test

An optometrist is studying the length of time patients can wear two types of contact lens.

Patient ID	Type A Lens	Patient ID	Type B Lens
001	8	011	7
002	11	012	4
003	10	013	10
004	15	014	8
005	12	015	22
006	8	016	6
007	12	017	7
008	9	018	6
009	11	019	6
010	16	020	10

Each patient wears the lens for a week and reports the maximum number of hours they can wear without irritations.

When to Use Mann-Whitney Test

Difference - Distribution of 2 Samples

- Ordinal Scale
- Two Independent Samples > 3
- Same-shape populations
- Wilcoxon Rank Sum Test - Median

Practical Example

이 환자들에서 수술 후 부갑상선 호르몬의 변화를 관찰한 결과 부갑상선 호르몬이 6개월이 지나도 정상 이하로 떨어지지 않고 2차례 이상 정상범위를 중심으로 증가와 감소를 반복한 그룹(그룹 1, 12예)과 수술 후 일시적으로 증가되었다가 수술 후 6개월이 지난 후 추적 관찰 기간 동안 정상화된 그룹(그룹 2, 6예)으로 나눌 수 있었다. 이들을 대상으로 수술 후 병리조직검사 결과, 골밀도 검사를 조사하고, 총 칼슘, 이온화 칼슘, 알칼라인 포스파타제, 부갑상선 호르몬, 오스테오칼신, 비타민 D 등의 혈청치에 대해서 알아보고 비교하였다.

통계적 검정은 SPSS 10.0을 사용하여 Kruskal-Wallis test 및 Chi-square test에 의해 분석하였으며 유의 수준은 P< 0.05로 하였다.

Practical Example

Two Group: N1 = 10, N2 = 5

능항진증 환자 15명을 부갑상선 전절제술 및 전완부 자가 이식을 시행하면서 수술 중 부갑상선 호르몬 측정을 시행하였다. 전완부 부갑상선 자가 이식은 1~2 mm의 부갑상선

시간을 구분하였고, 성공적 부갑상선 절제술을 시행한 환자와 실패한 부갑상선 절제술 환자간의 데이터 비교는 Student t-test를 시행하였다.

Example: Wilcoxon Signed Rank Test

A double-blind randomized crossover study was done to see the effect of a new drug for headache.

Patient No.	Aspirin	New Drug
1	2	4
2	3	6
3	1	3
4	5	6
5	4	4
6	5	3
7	6	7
8	5	7
9	3	6
10	3	7

Each patient ranks their relief from excellent (1) to poor (10).

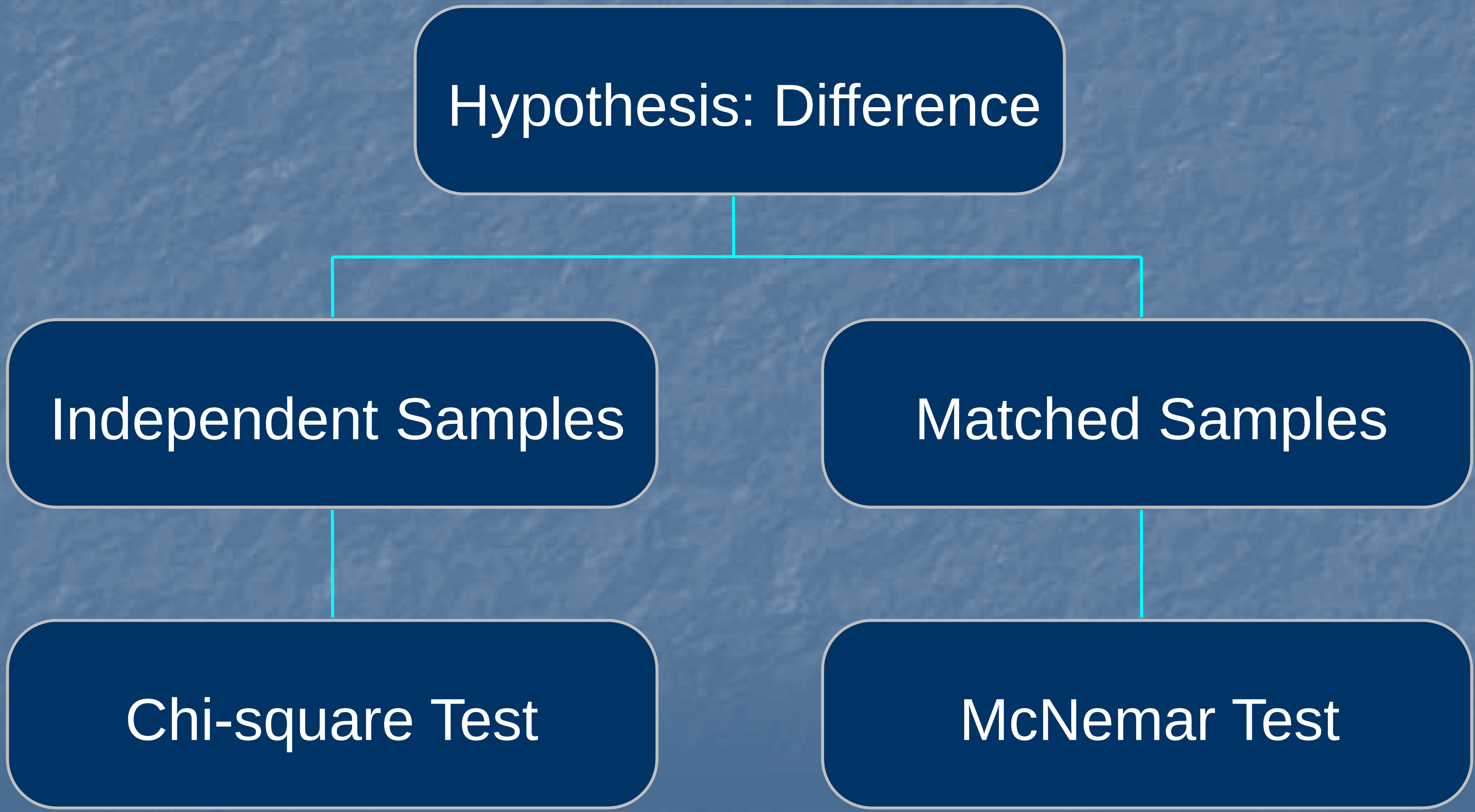
When to Use Wilcoxon Signed Rank Test

Difference - Distribution of 2 Samples

- Interval Scale
- Two Paired Samples > 3
- Symmetric Distribution of Differences
- Differences are Independent

Statistical Test

Nominal Data



Example: Chi-Square Test

A group of 480 patients with pneumonia were treated with either amoxicillin or erythromycin.

	Improvement	
	Yes	No
Amoxicillin	144 (60%)	96 (40%)
Erythromycin	160 (67%)	80 (33%)

Example: Chi-Square Test

Comparison of effects of treatment of pneumonia with amoxicillin or erythromycin.

ID	Antibiotics	Improvement	ID	Antibiotics	Improvement
001	**Amoxicillin**	**Yes**	**241**	**Erythromycin**	**Yes**
002	**Amoxicillin**	**Yes**	**242**	**Erythromycin**	**Yes**
003	**Amoxicillin**	**Yes**	**243**	**Erythromycin**	**Yes**
004	**Amoxicillin**	**Yes**	**244**	**Erythromycin**	**Yes**
…	**Amoxicillin**	**Yes**	**245**	**Erythromycin**	**Yes**
144	**Amoxicillin**	**Yes**	**…**	**Erythromycin**	**Yes**
145	**Amoxicillin**	**No**	**400**	**Erythromycin**	**Yes**
146	**Amoxicillin**	**No**	**401**	**Erythromycin**	**No**
…	**Amoxicillin**	**No**	**---**	**Erythromycin**	**No**
240	**Amoxicillin**	**No**	**480**	**Erythromycin**	**No**

When to Use Chi-square Test

Association – Proportion of 2 Samples

- Nominal Scale
- $\leq$ 20% Cells, Expected Frequency < 5
- No Cell, Expected Frequency < 1
- A Random Sample

Table with Small Expected Counts

A pathologist is studying the relationship between histologic type and tumor stage in epithelial ovarian cancer.

	Stage 1	Stage 2	Stage 3	Stage 4
Serous	4	16	16	6
Mucinous	1	7	7	3
Endometrioid	3	4	12	5

3 x 4 table

Table with Small Expected Counts

Table with Observed and Expected Frequencies for Example.

	Stage 1		Stage 2		Stage 3		Stage 4	
Serous	4	4.5	16	15.3	16	14.2	6	7.9
Mucinous	1	1.9	7	6.6	7	6.1	3	3.4
Endometrioid	3	1.5	4	5.1	12	4.7	5	2.6

Expected frequency = Row total x Column total / Grand total

Table with Small Expected Counts

A pathologist is studying the relationship between histologic type and tumor stage in epithelial ovarian cancer.

	Early Stage	Advanced Stage
Serous	20	22
Mucinous	8	10
Endometrioid	7	17

3 x 2 table

Table with Small Expected Counts

Table with Observed and Expected Frequencies for Example.

	Early Stage		Advanced Stage	
Serous	20	17.5	22	24.5
Mucinous	8	7.5	10	10.5
Endometrioid	7	10.0	17	14.0

Expected frequency = Row total x Column total / Grand total

Example: Fisher Exact Test

Comparison of effects of treatment of pneumonia with amoxicillin or erythromycin.

ID	Antibiotics	Improvement	ID	Antibiotics	Improvement
001	**Amoxicillin**	**Yes**	**011**	**Erythromycin**	**Yes**
002	**Amoxicillin**	**Yes**	**012**	**Erythromycin**	**Yes**
003	**Amoxicillin**	**Yes**	**013**	**Erythromycin**	**Yes**
004	**Amoxicillin**	**Yes**	**014**	**Erythromycin**	**Yes**
005	**Amoxicillin**	**Yes**	**015**	**Erythromycin**	**Yes**
006	**Amoxicillin**	**Yes**	**016**	**Erythromycin**	**Yes**
007	**Amoxicillin**	**No**	**017**	**Erythromycin**	**Yes**
008	**Amoxicillin**	**No**	**018**	**Erythromycin**	**No**
009	**Amoxicillin**	**No**	**019**	**Erythromycin**	**No**
010	**Amoxicillin**	**No**	**020**	**Erythromycin**	**No**

Example: Fisher Exact Test

A group of 20 patients with pneumonia were treated with either amoxicillin or erythromycin.

	Improvement	
	Yes	No
Amoxicillin	6 (60%)	4 (40%)
Erythromycin	7 (70%)	3 (30%)

When to Use Fisher Exact Test

Association – 2 x 2 Contingency Table

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

Practical Example

Table 5. Incidence of hematoma in thyroidectomy patients

Group	Incidence	P-value*
Surgery		1.00
Total thyroidectomy	2.8% (13/465)	
Less than total thyroidectomy	2.6% (5/190)	
Disease		0.024
Malignancy	2.5% (11/442)	
Benign	2.1% (4/188)	
Graves' disease	11.1% (3/27)	

*P-value was calculated by fisher extract test.

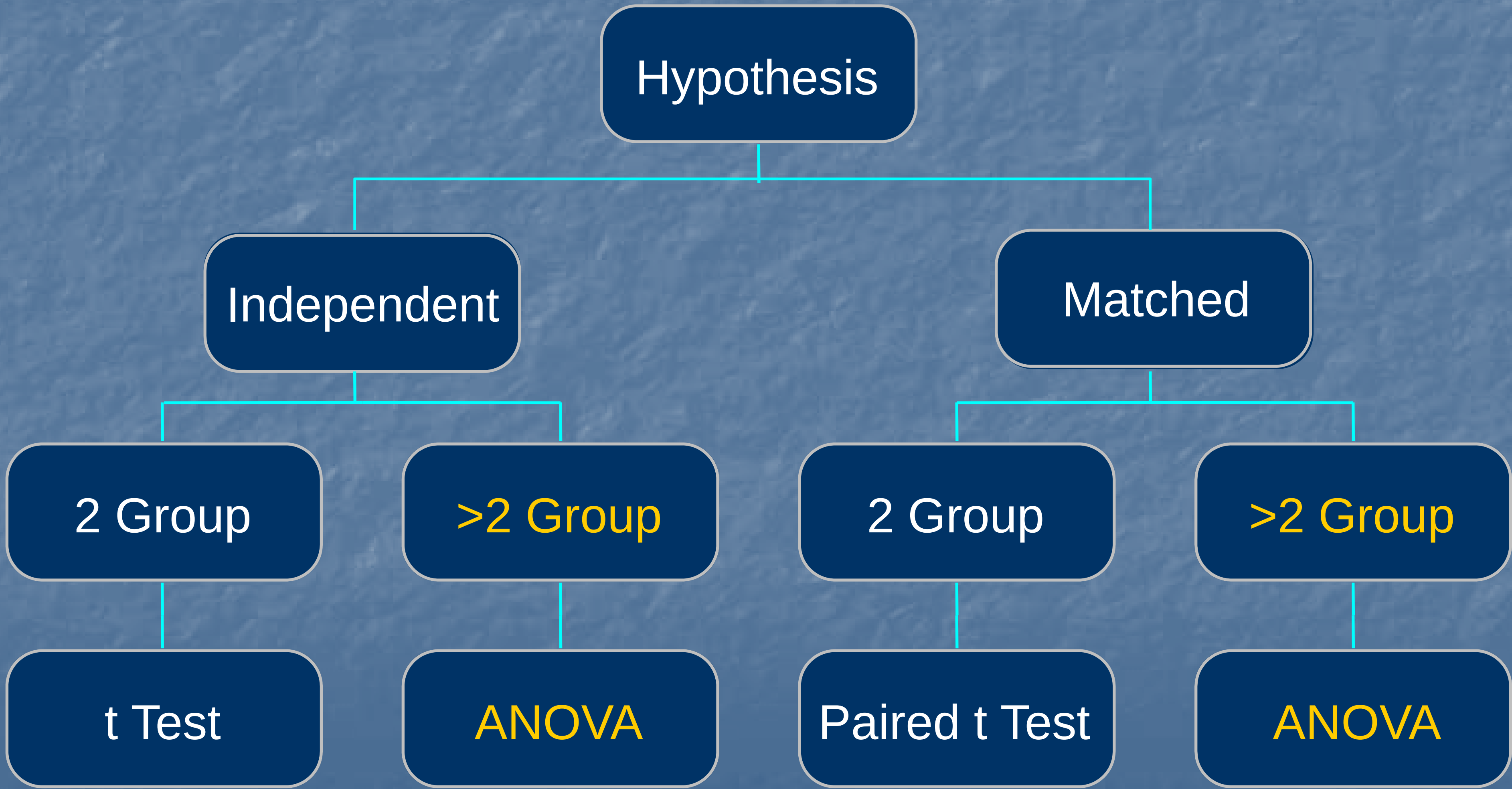
Practical Example

		AMES		P value
		High risk (%)	Low risk (%)	
P53				0.936
	Positive	4 (30.8)	5 (29.4)	
	Negative	9 (69.2)	12 (70.6)	
Bcl-2				0.042
	Positive	5 (38.5)	13 (76.5)	
	Negative	8 (61.5)	4 (23.5)	

성별, 나이 및 피막 침범 유무와 TNM 병기, AMES 체계 등에 따라 분석하였으며, SPSS 11.0을 이용하여 chi-square test와 student's t-test를 사용하였다. P value 0.05 이하인 경우 통계적으로 유의한 것으로 판정하였다.

Multi-Sample Methods

Statistical Test
Interval Data



Example: One-Way ANOVA Test

Red cell folate levels in three group of cardiac bypass patients given different levels of nitrous ventilation.

Group I (n = 8)	Group II (n = 9)	Group III (n = 5)
243	206	241
251	210	258
275	226	270
291	249	293
347	255	328
354	273	
380	285	
392	295	
	309	

Practical Statistics for Medical Research by Altman, 1999

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

Test for Variance

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

Multiple Comparison After ANOVA

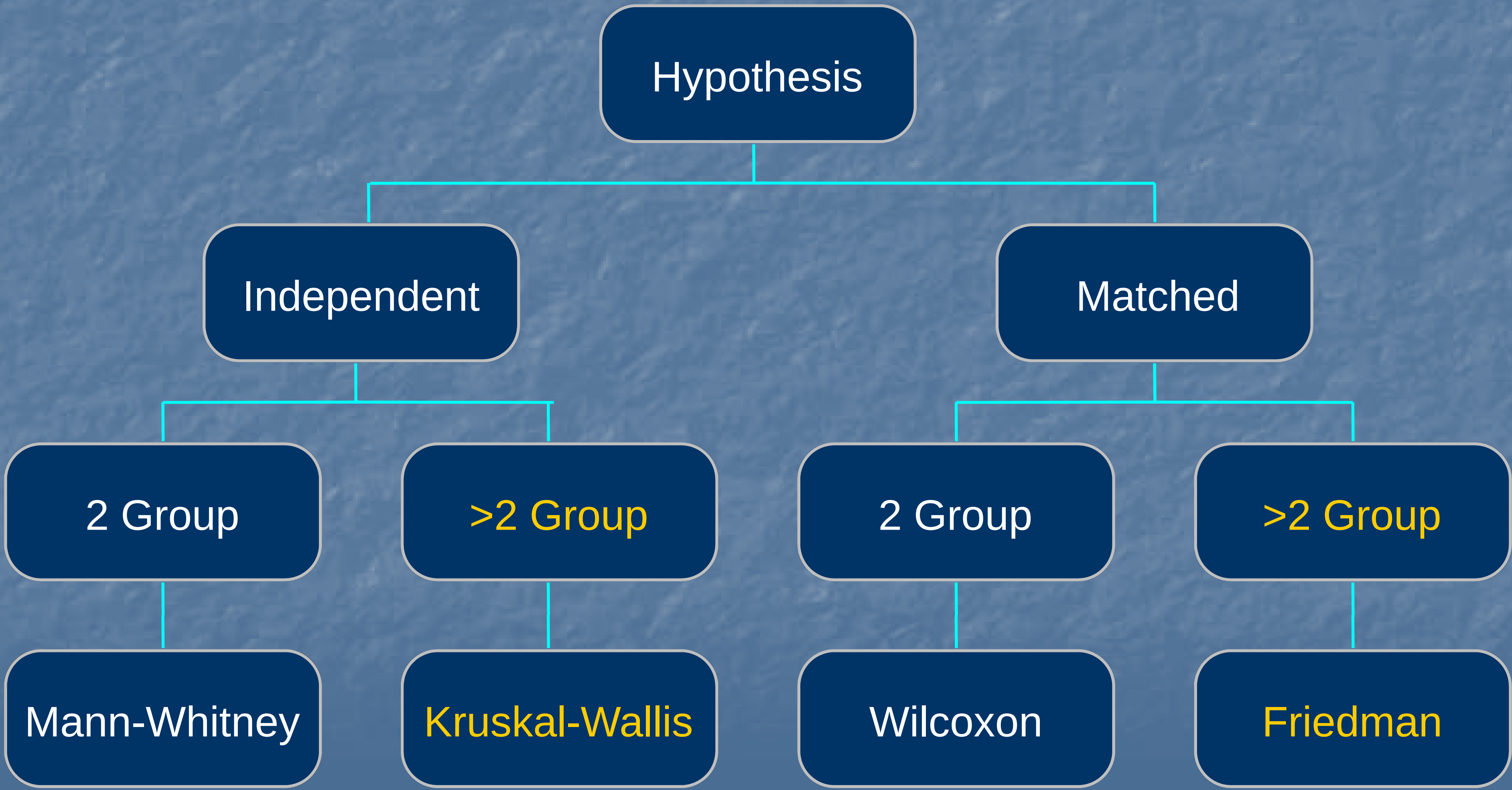
Do not use multiple t test

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

post-hoc tests

Post-hoc test	Equal size	Group-wise	Pairwise
Scheffe	-	+	+
Bonferroni	-	-	+
Tukey's HSD	+	+	+
SNK	+	+	-
Duncan	+	+	-
LSD	+	-	+
Games Howell	-	-	+
Dunnet's C	-	-	+

Statistical Test
Ordinal Data



Example: Kruskal-Wallis Test

Ocular anti-inflmmatory effects of four drug on lid closure after administration of arachidonic acid

Rabbit Number	Indomethicin	Aspirin	Piroxicam	BW755C
1	+2	+1	+3	+1
2	+3	+3	+1	0
3	+3	+1	+2	0
4	+3	+2	+1	0
5	+3	+2	+3	0
6	0	+3	+3	-1

Fundamentals of Biostatistics, 6th edition

When to Use Kruskal-Wallis Test

Difference - Medians of ≥ 3 Samples

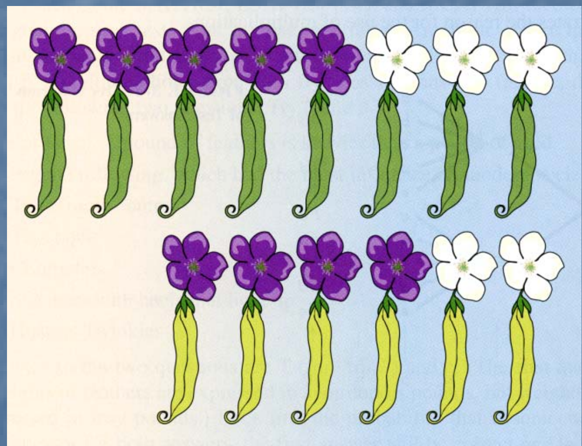
- Ordinal Scale
- k Independent Samples
- One-Way ANOVA by Ranks
- All Sample Sizes ≥ 5

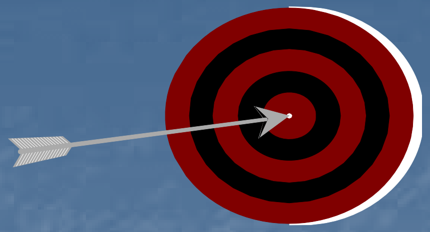
Multiple Comparison After KW

Do not use multiple Mann-Whitney test

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

Summary of Inferential Tests

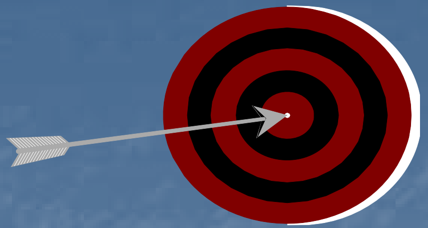




Statistical Test

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

Comparative Statistics

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

Chi-square	Mann-Whitney Kruskal-Wallis	Student t-test One-Way ANOVA
McNemar	Wilcoxon Friedman	Paired t-test Repeated ANOVA

Classification of Statistical Methods

- Z-test
- Student's t-test
- paired t-test
- ANOVA

  (1) one way / two-way / multi-way

  (2) repeated measures
- multiple comparisons

  (1) Bonferroni (2) Scheffe (3) Duncan (4) other
- Pearson correlation analysis
- simple linear regression

Classification of Statistical Methods

- contingency tables
  (1) chi-square test (2) Fisher's exact test
  (3) McNemar's test (4) other
- nonparametric test
  (1) Mann-Whitney U test (Wilcoxon rank sum test)
  (2) Wilcoxon signed rank test
  (3) Kruskal-Wallis test
  (4) test for trend
  (5) Spearman rank correlation
  (6) other

Classification of Statistical Methods

- survival analysis
  (1) life table method (2) Kaplan-Meier method
  (3) log-rank test (4) Cox regression
  (5) other
- epidemiological statistics
  (1) sensitivity/specificity, odds ratio, relative risk
  (2) other
- multivariate analysis
  (1) multiple regression
  (2) logistic regression (3) other
- others

Check Lists in Assessing the Statistical Contents

- 연구종류
  1) 원저 2) 증례 3) 종설 4) 기타

- 통계기법 종류
  1) 추측통계 2) 기술통계 3) 사용안함 4) 기타

- 통계기법 적용

1. 기술통계
1) 대표값 (1) 적절 (2) 부적절 (3) 사용안함
2) 산포도 (1) 적절 (2) 부적절 (3) 사용안함

Check Lists in Assessing the Statistical Contents

- 통계기법 적용

2. 추측통계
1) 방법
2) 적정성 (1) 적절 (2) 부적절
3) 통계적 가정
   (1) 정규성 (2) 등분산성 (3) 독립성
   (4) 표본크기 (5) 기대도수 (6) 기타
4) 통계적 절차
   (1) 사후검정 (2) 기타

Check Lists in
Assessing the Statistical Contents

- 방법론 (기법의 서술)
  (1) 상세 기술 (2) 유의수준 (신뢰구간) (3) 통계프로그램
  (4) 양측/단측 검정 (5) 기타
- 분석결과
  (1) p-값 (2) 신뢰구간 (3) 통계량, 자유도 (4) 해석 오류
  (5) 계산 오류 (6) 비교군 언급 (7) 연구가설과 일치
  (8) 한계점 (9) 인용문헌의 통계적 서술 (10) 기타
- 기타
  (1) 통계용어
  (2) 보조적 분석기법 (Table, Chart,…)
  (3) 통계학적 방법론에 관한 참고문헌

의학논문 1

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.

통계학적 방법론

- 통계프로그램

  통계분석은 dBSTAT(ver 4.0)을 이용하여 실시하였다.

- 통계적 기법

  각 측정치는 평균±표준편차로 요약하였으며 스테로이드 투여 전후간의 PC_{20}의 차이는 paired t-test를 적용하여 유의성을 검정하였다.

- 유의수준

  양측검정을 시행하였으며 p 값이 0.05 미만인 경우 통계적으로 유의하다고 판정하였다. (유의수준은 0.05로 하였다.)

의학논문 1

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, † $P<0.01$

통계학적 방법론

- 통계분석 결과 해석

  PC_{20}의 평균값은 치료전 1.89 mg/mL에서 치료후 4.56 mg/mL로 통계적으로 유의한 증가를 보였다($p = 0.028$).

Table 1. Comparison of PC_{20} (n=40)

	Before Tx	After Tx	p-value*
PC_{20}(mg/mL)	1.89±0.72	4.56±1.26	0.021

value are expressed as mean±SD

* calculated by paired t-test

Obstetric and Gynecology
Vol. 102, No 1, July 2003

METHODS

Data were analyzed with the Student *t* test if data were determined to be normallv distributed or with the Wilcoxon rank-sum test for nonparametric data. All comparisons were two tailed, with a *P* value less than .05 considered significant. All analyses were completed using SAS statistical software (SAS Institute).

Obstetric and Gynecology
Vol. 102, No 1, July 2003

RESULTS

Table 1. Patient Characteristics by Hemostasis Modality (Mean ± Standard Deviation, Median, Range)

	EBVS (n = 30)	Suture (n = 30)	Statistical significance (P)
Age (y)	46.2 ± 8.6 45.7 (31.6–69.8)	47.5 ± 9.4 46.1 (25.8–68.0)	.60
BMI (kg/m^2)	26.2 ± 4.6 25.0 (20.1–39.3)	26.1 ± 6.0 24.4 (18.4–42.2)	.54
Uterine mass (g)	247.2 ± 206.3 185 (46.0–865)	250.8 ± 221.7 200.0 (50.0–1065.0)	.95
Parity	1.7 ± 1.1 2.0 (0–4)	1.6 ± 1.0 2.0 (0–3)	.71
Length of stay (d)	0.10 ± 0.3 0.0 (0–1)	0.33 ± 0.5 0.0 (0–1)	.03

EBVS = electrosurgical bipolar vessel sealer; BMI = body mass index.
Age data normally distributed and analyzed with the Student t test. All other data analyzed with the nonparametric Wilcoxon rank-sum test.

Amygdala Volume and Depressive Symptoms in Patients With Borderline Personality Disorder

Thomas Zetzsche, Thomas Frodl, Ulrich W. Preuss, Gisela Schmitt, Doerthe Seifert, Gerda Leinsinger, Christine Born, Maximilian Reiser, Hans-Jürgen Möller, and Eva M. Meisenzahl

Statistical Analyses

All morphometric data were tested for normal distribution with Kolmogorov-Smirnov nonparametrical tests. Amygdala volumes refer to the gray matter volume that was calculated by the

possible effects. Post hoc analyses were performed with univariate ANCOVAs with age and total cranial volume as cofactors (see Table 2).

Intergroup differences of total brain volume and demographic variables (e.g., age, weight, and height) were tested with Student's *t* test. Potential associations between amygdala volumes and the HAMD score (for determination of depressive symptoms) were evaluated by calculating Spearman correlation coefficients. Relationships between aggressive/impulsive symptoms and amygdala volumes were analyzed in the same way.

For all statistical analyses a p value of $< .05$ (two-sided) was considered as statistically significant. A p value $\geq .05$ and $\leq .07$ was defined as a statistical trend.

Higher Fasting Serum Insulin Is Associated with Increased Resting Energy Expenditure in Nondiabetic Schizophrenia Patients

Xiaoduo Fan, Ellen J. Anderson, Paul M. Copeland, Christina P. Borba, Dana D. Nguyen, Oliver Freudenreich, Donald C. Goff, and David C. Henderson

Statistical Analysis

The data were analyzed by using SPSS (version 13.0; SPSS Inc., Chicago, IL). Descriptive statistics were used to describe demographics and anthropometric and laboratory measurements. Fasting serum insulin was not distributed normally and was, therefore, log transformed before analysis. Analysis of variance or independent *t* test was used for group comparison as appropriate. Pearson correlation coefficients were used to quantify relations between fasting serum insulin and other variables. Further, stepwise multiple regression was used to examine whether fasting serum insulin is an independent predictor of REE when other potential confounding variables also considered. The criteria of $p = .05$ for a variable to enter and $p = .10$ for a variable to be removed were used in the multiple regression analysis. For all statistical analyses, a p value of less than .05 (2-tailed) was used to test for statistical significance.