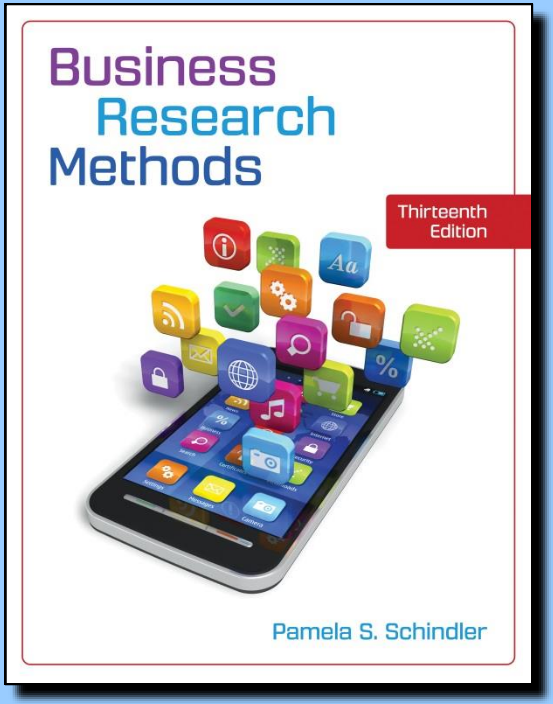
Chapter 14

STAGE 4: HYPOTHESIS TESTING

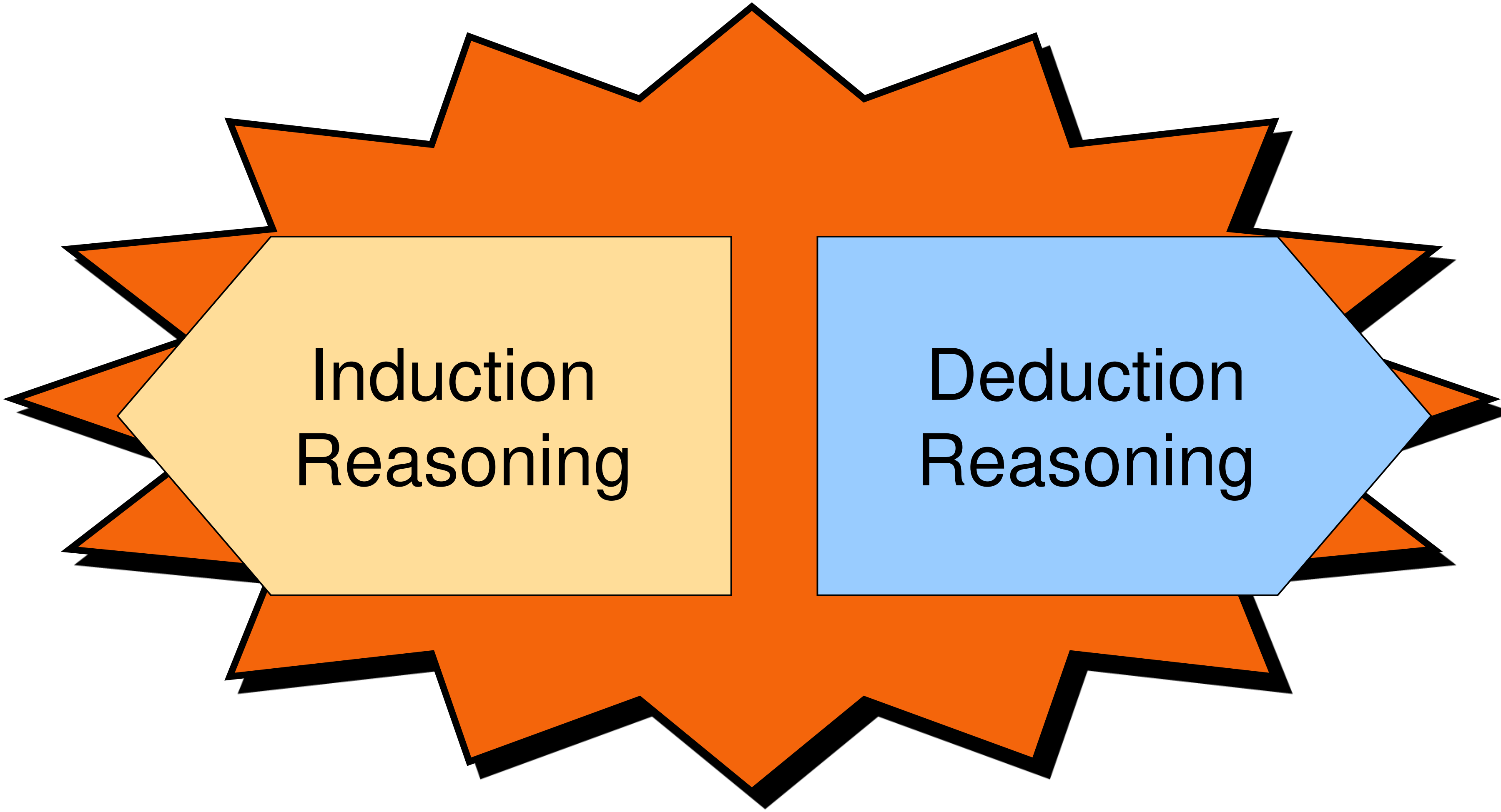


Learning Objectives

Understand . . .

- The nature and logic of hypothesis testing.
- A statistically significant difference
- The six-step hypothesis testing procedure.
- The differences between parametric and nonparametric tests and when to use each.
- The factors that influence the selection of an appropriate test of statistical significance.
- How to interpret the various test statistics.

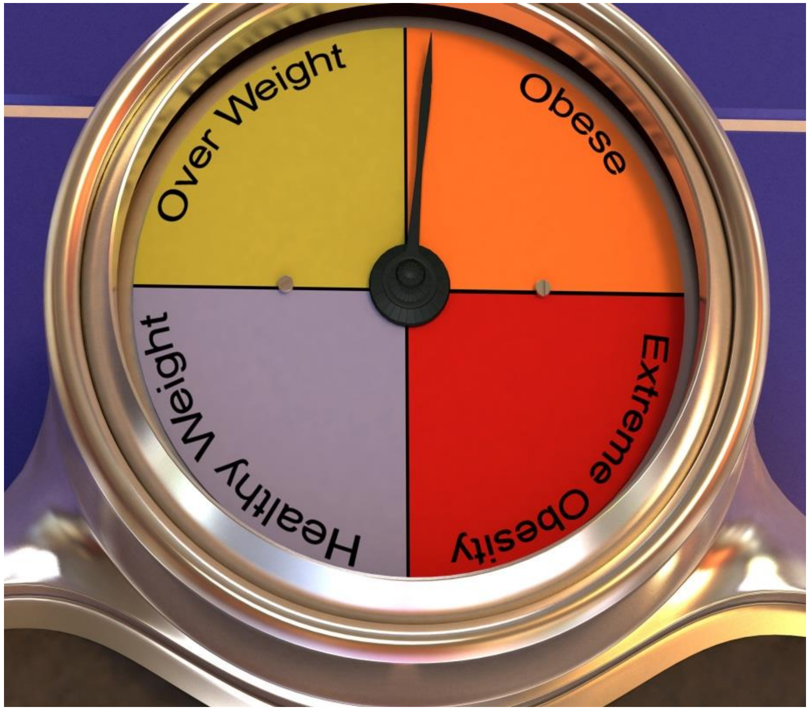
Hypothesis Testing



Reasoning and Hypotheses

Inductions are an inferential leap from the evidence presented.

=

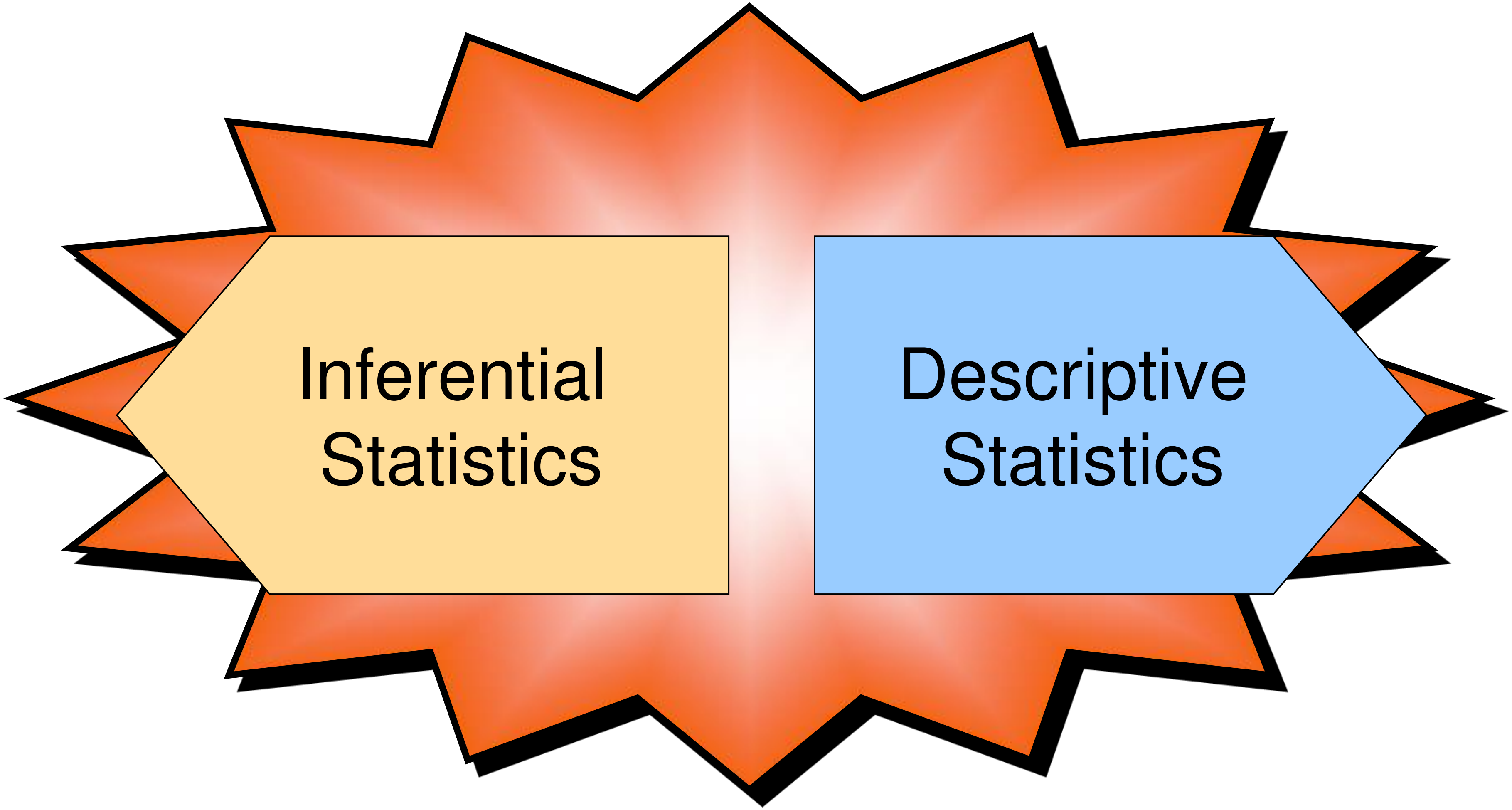


Reasoning and Hypotheses

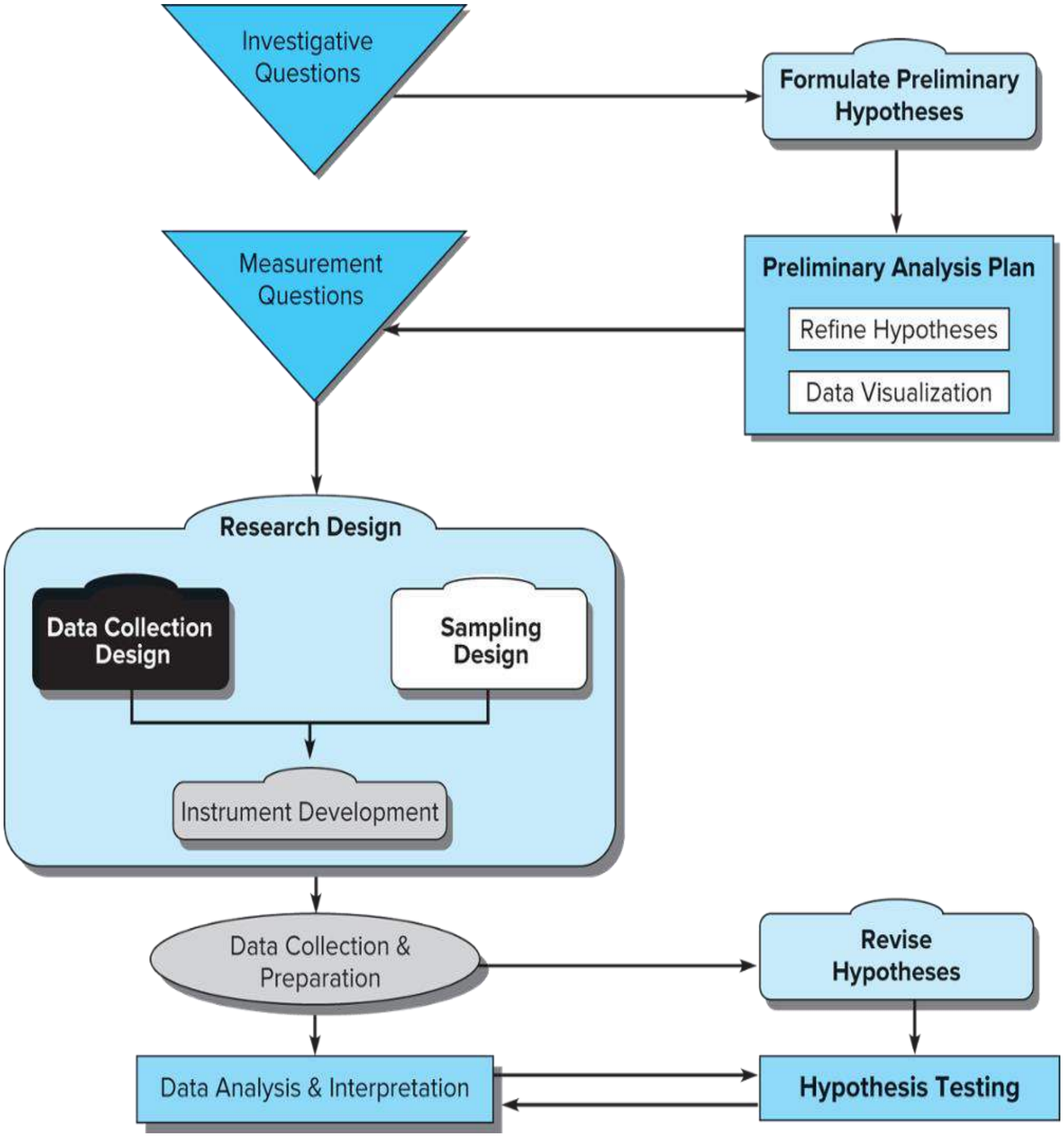
Deductions are only as good as
the premises on which they are based.

≠

Statistical Procedures



Hypothesis Testing and the Research Process



The "Ah-Ha" Moment

When researchers sift through the chaos and find what matters.

Approaches to Hypothesis Testing

Classical statistics

- Objective view of probability
- Established hypothesis is rejected or fails to be rejected
- Analysis based on sample data

Bayesian statistics

- Extension of classical approach
- Analysis based on sample data
- Also considers established subjective probability estimates

Significance & Hypotheses

Statistical Significance

Practical Significance

Null vs. Alternative Hypotheses

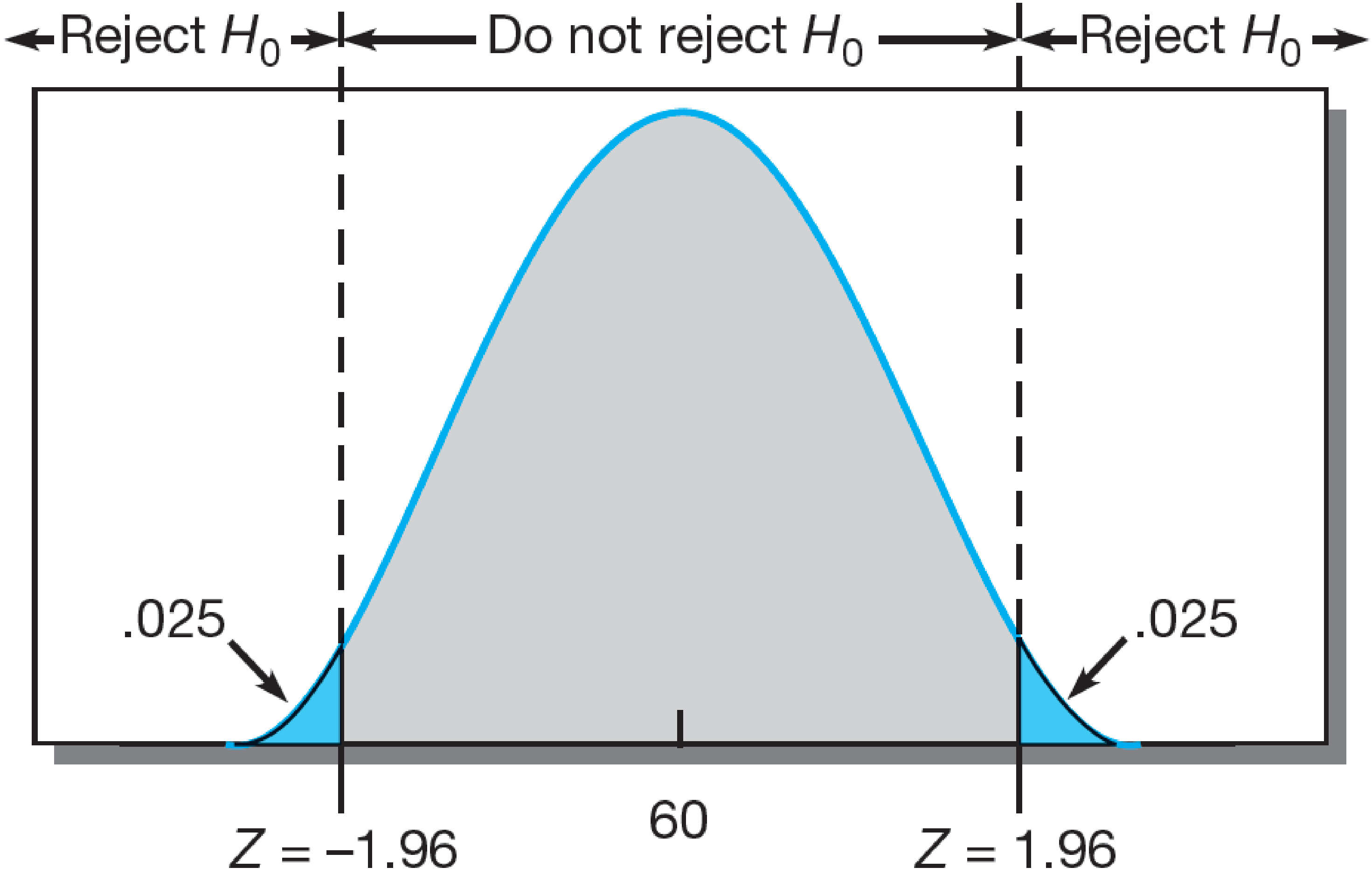
Null

- $H_0: \mu = 60$ mpg
- $H_0: \mu \leq 60$ mpg
- $H_0: \mu \geq 60$ mpg

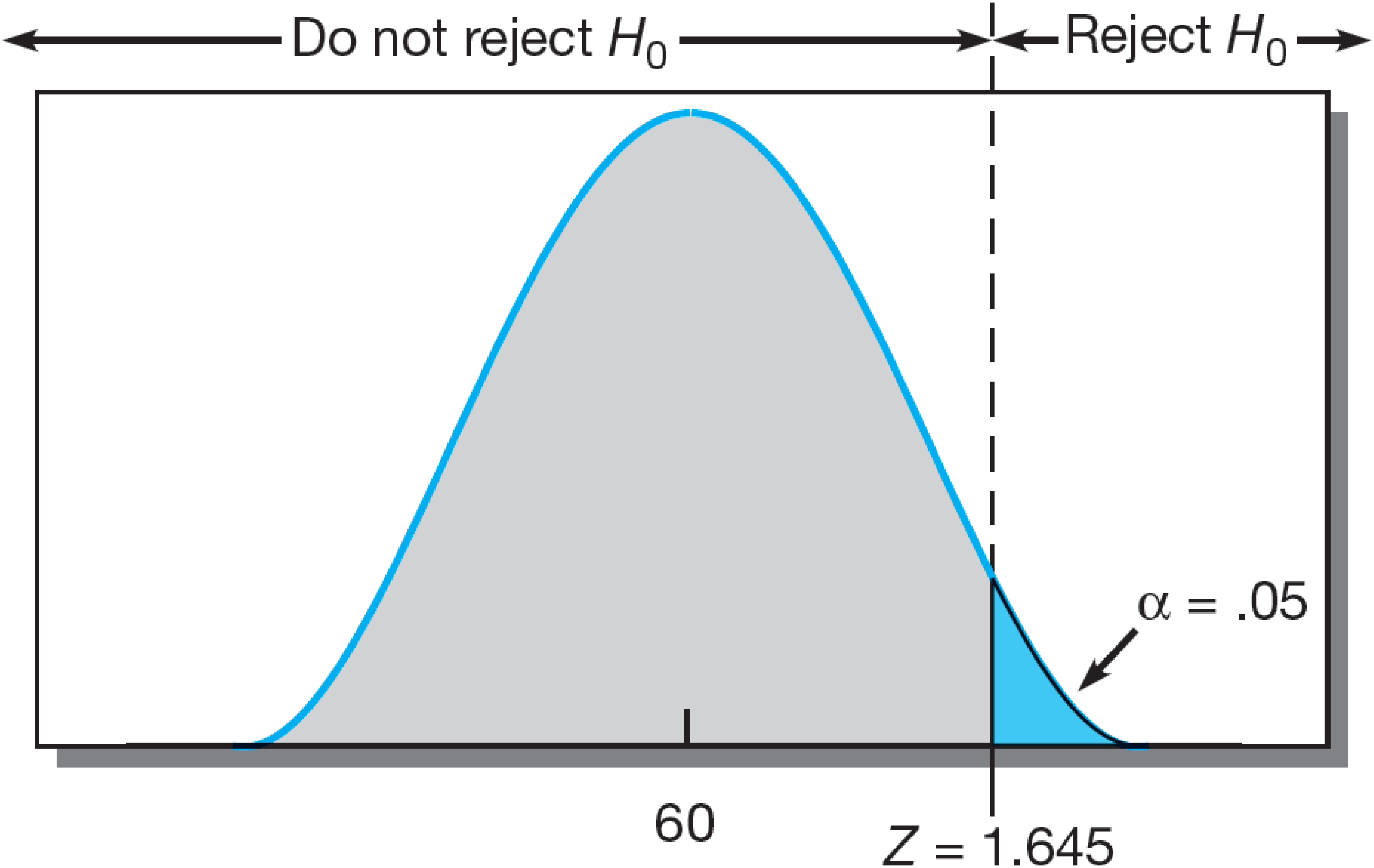
Alternative

- $H_A: \mu \neq 60$ mpg
- $H_A: \mu > 60$ mpg
- $H_A: \mu < 60$ mpg

Two-Tailed Test of Significance



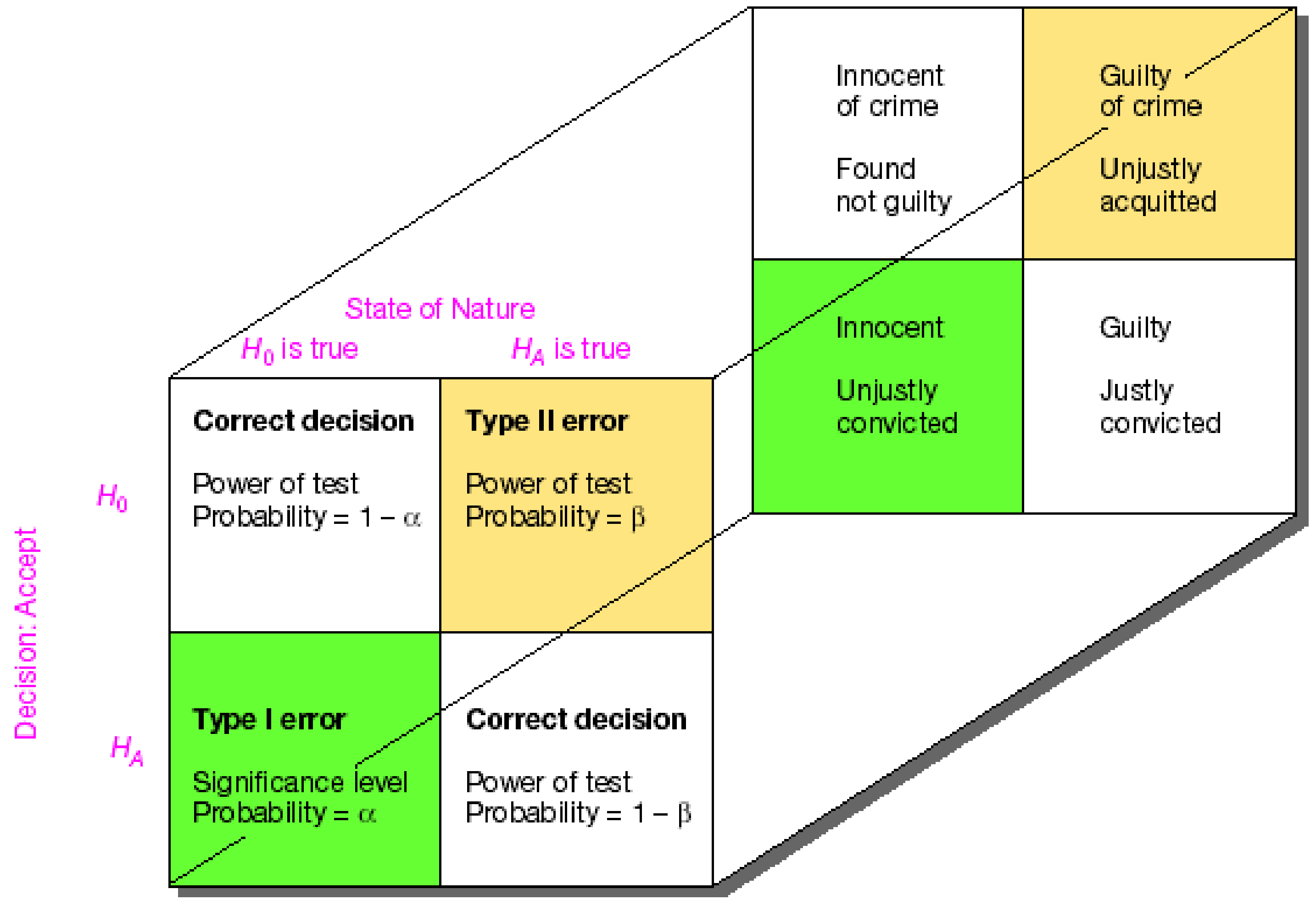
One-Tailed Test of Significance



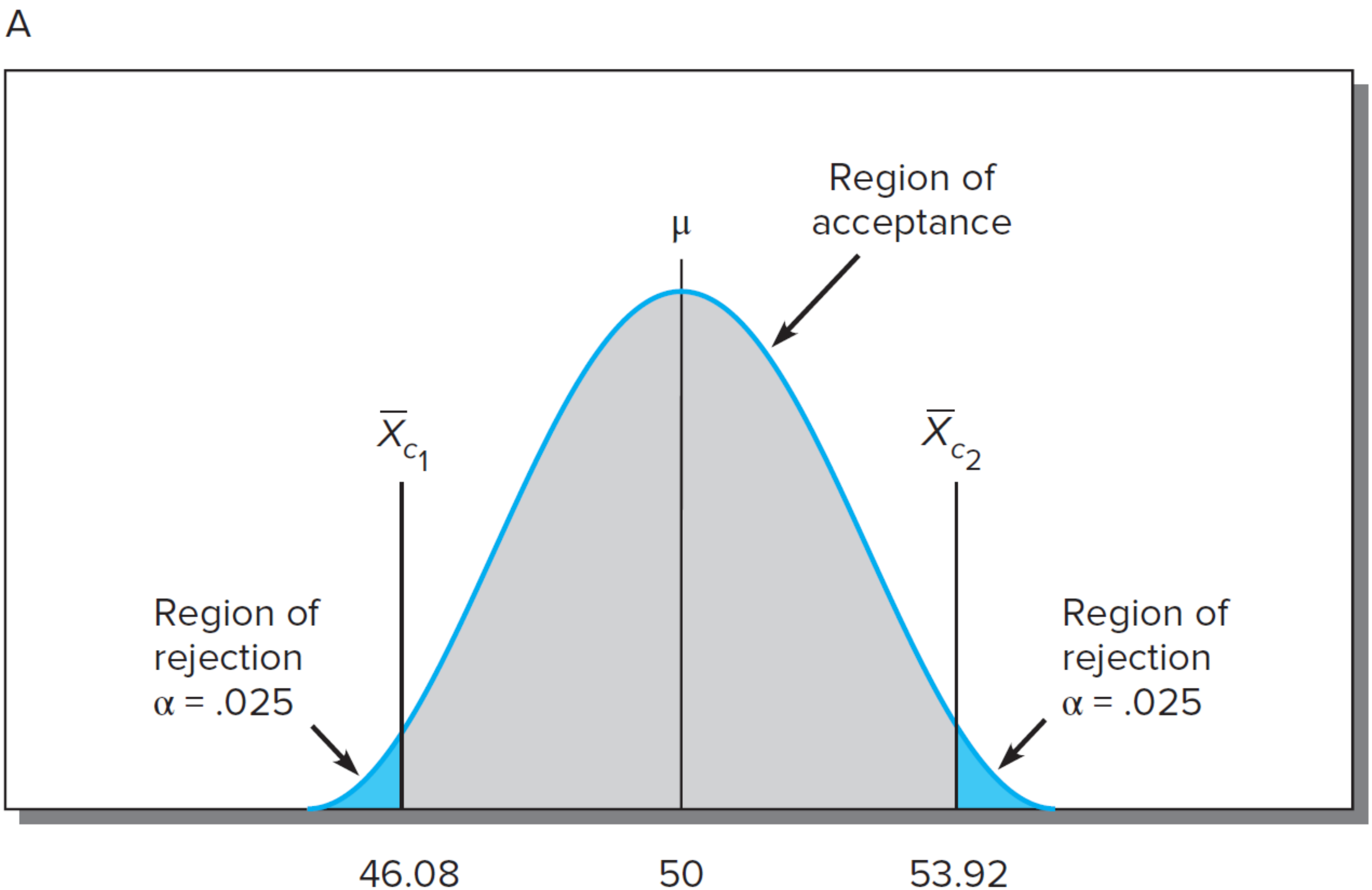
Decision Rule

*Take no corrective action if the analysis shows that one **cannot reject** the null hypothesis.*

Statistical Decisions



Probability of Making a Type I Error



Critical Values

Z = 1.96 (significance level = .05)

$\bar{X}_c$ = the critical value of the sample mean

μ = the population value stated in $H_0 = 50$

$\sigma_{\bar{X}}$ = the standard error of a distribution of means of samples of 25

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

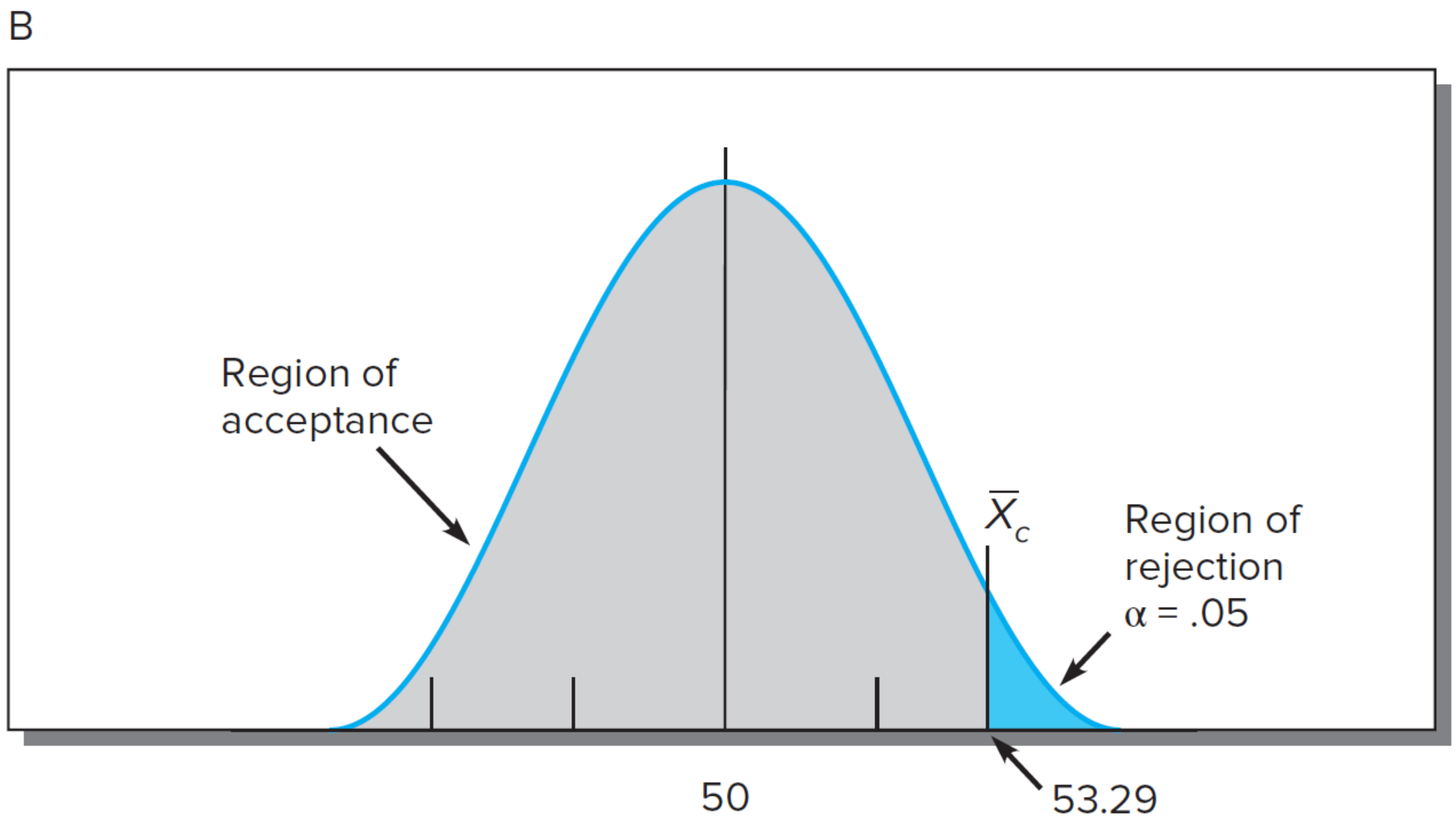
$$-1.96 = \frac{\bar{X}_c - 50}{2}$$

$$\bar{X}_c = 46.08$$

$$1.96 = \frac{\bar{X}_c - 50}{2}$$

$$\bar{X}_c = 53.92$$

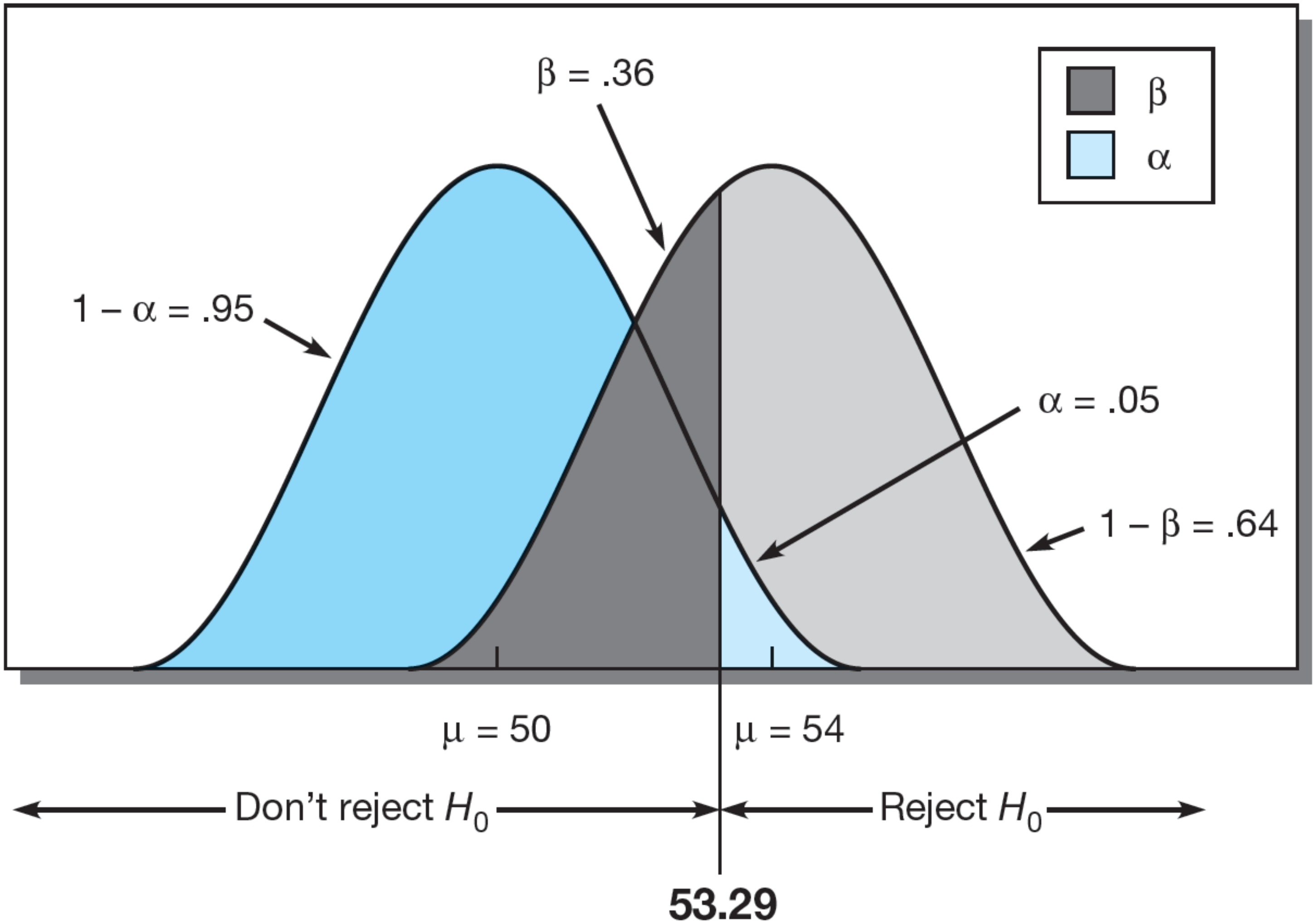
Probability of Making A Type I Error



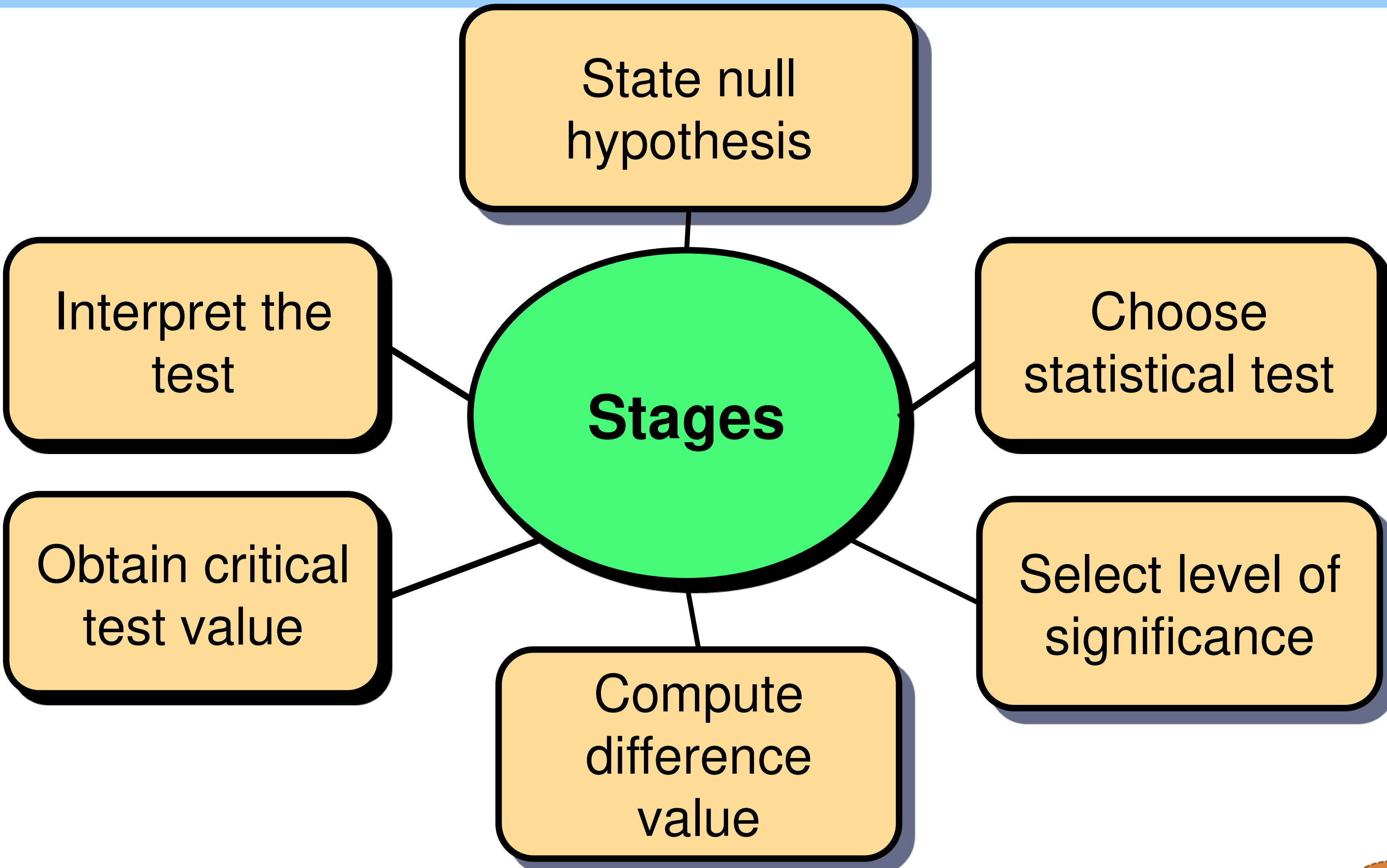
Factors Affecting Probability of Committing a β Error

- True value of parameter
- Alpha level selected
- One or two-tailed test used
- Sample standard deviation
- Sample size

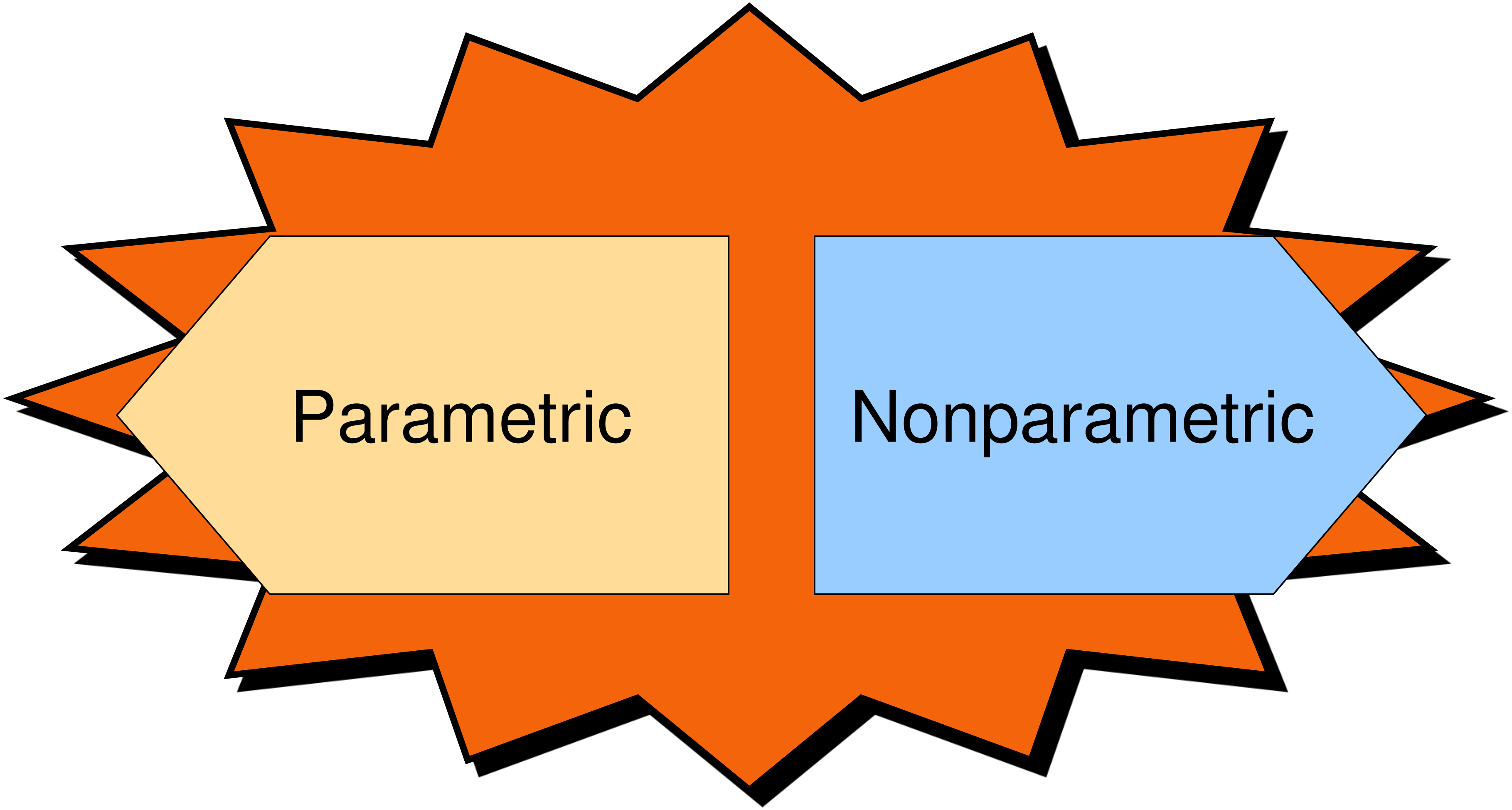
Probability of Making A Type II Error



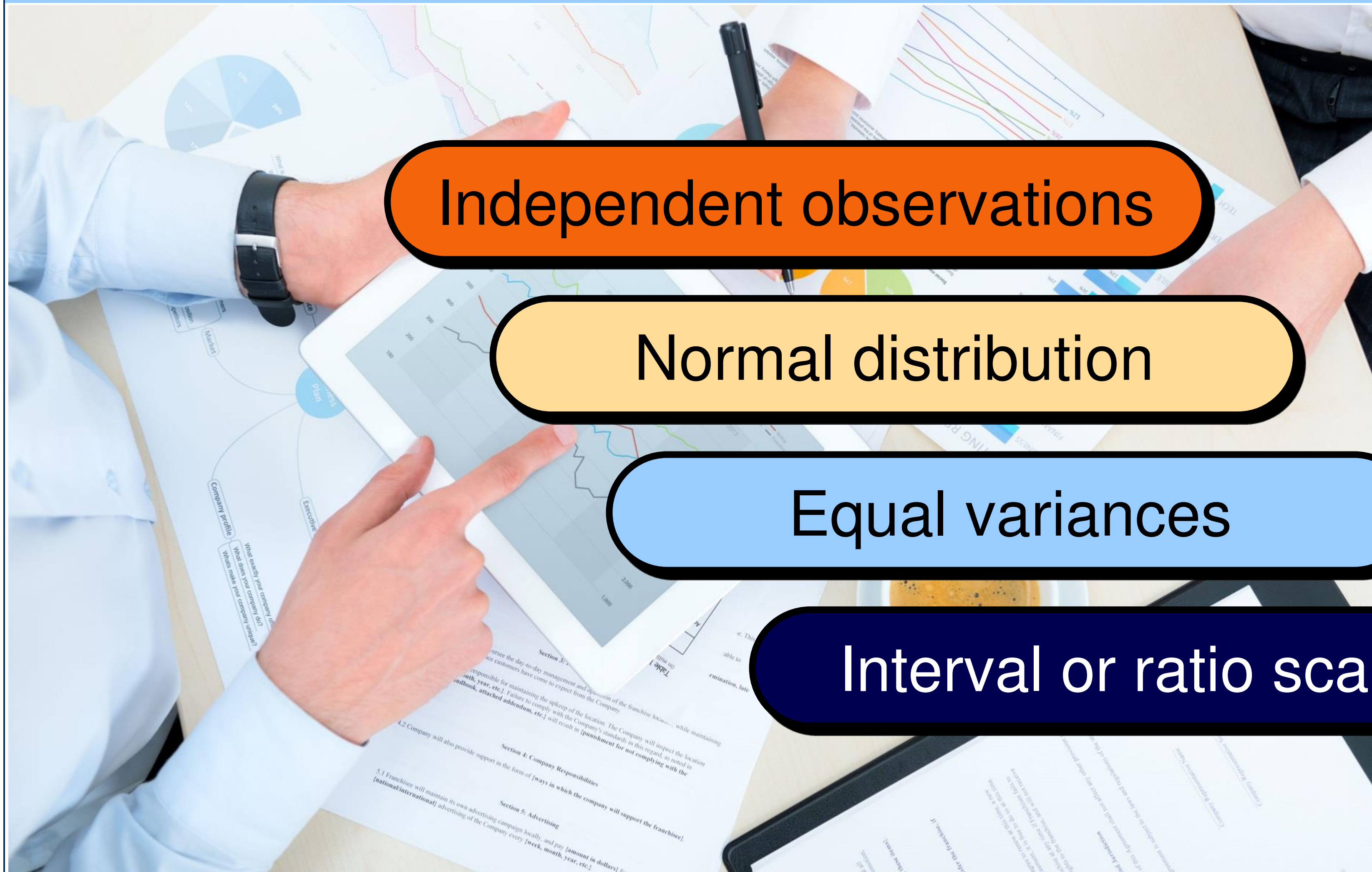
Statistical Testing Procedures



Tests of Significance



Assumptions for Using Parametric Tests



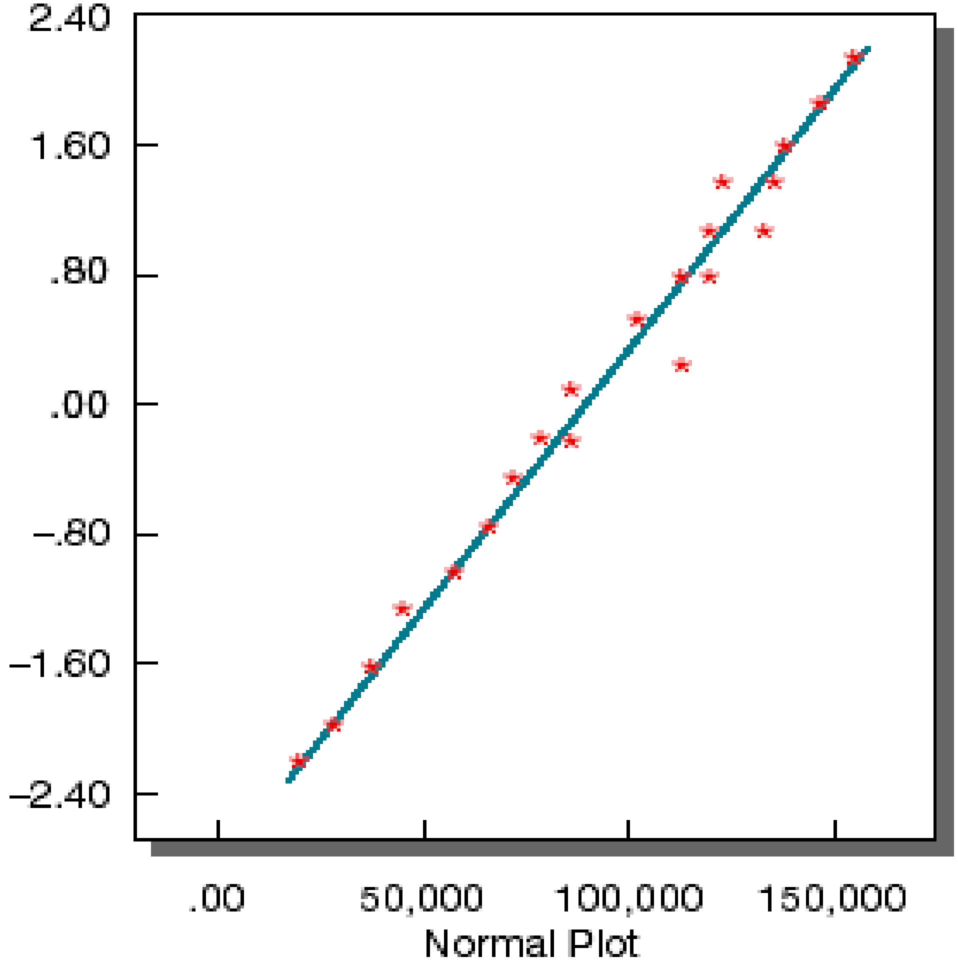
Independent observations

Normal distribution

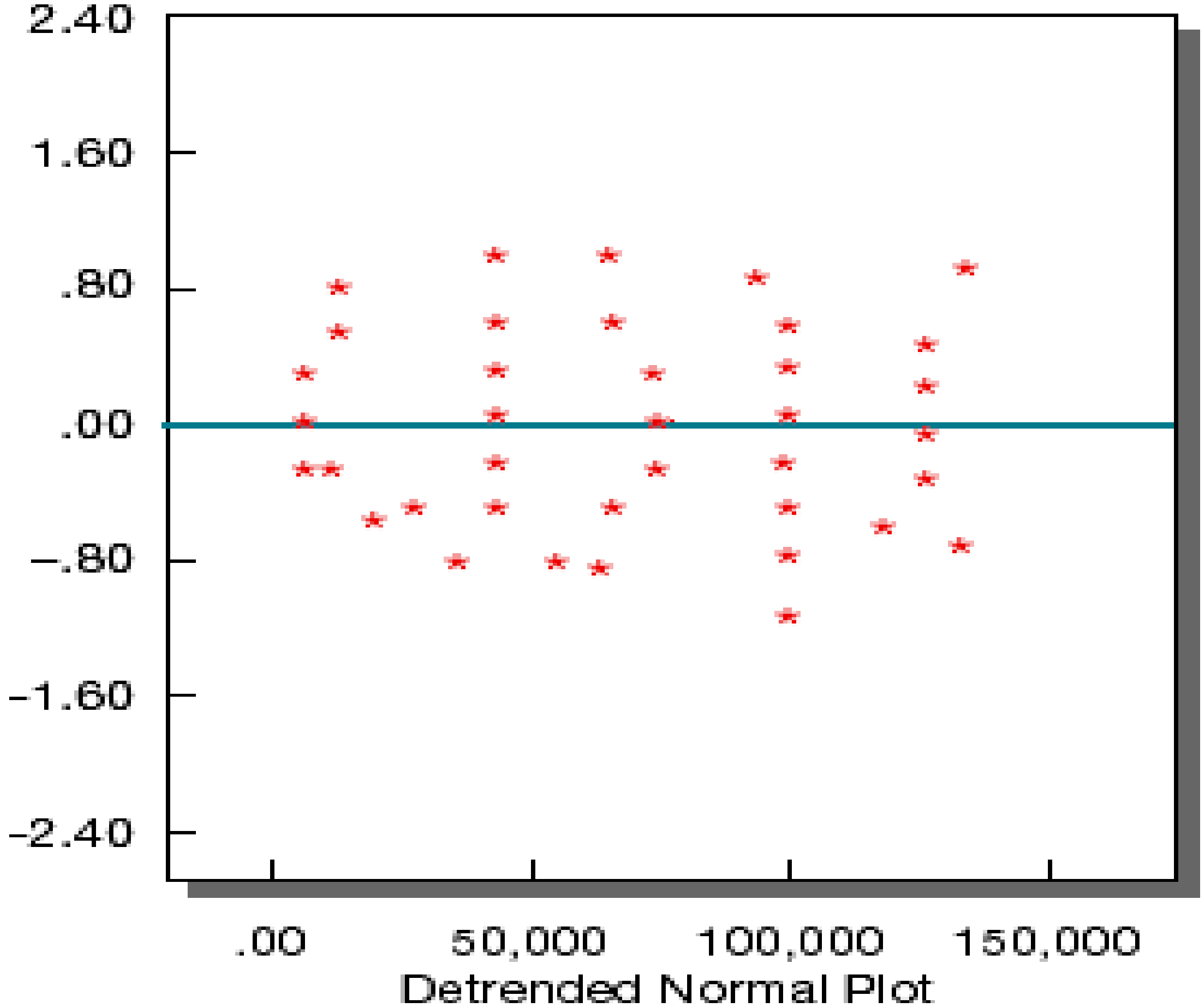
Equal variances

Interval or ratio scales

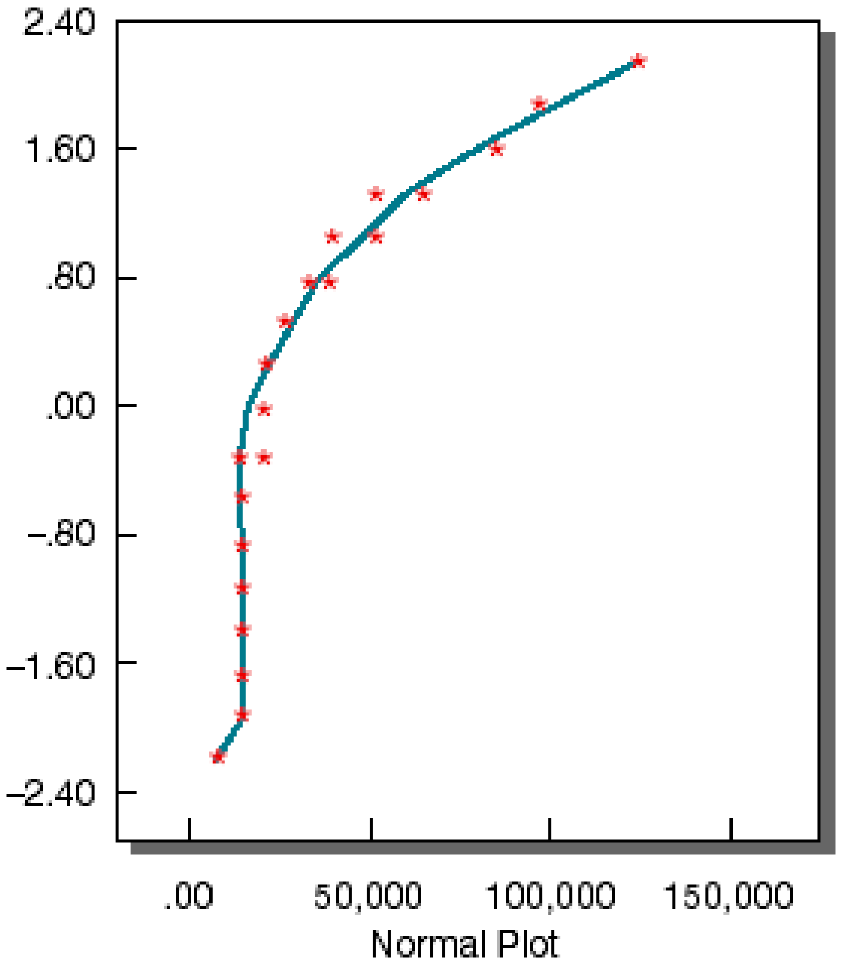
Probability Plot

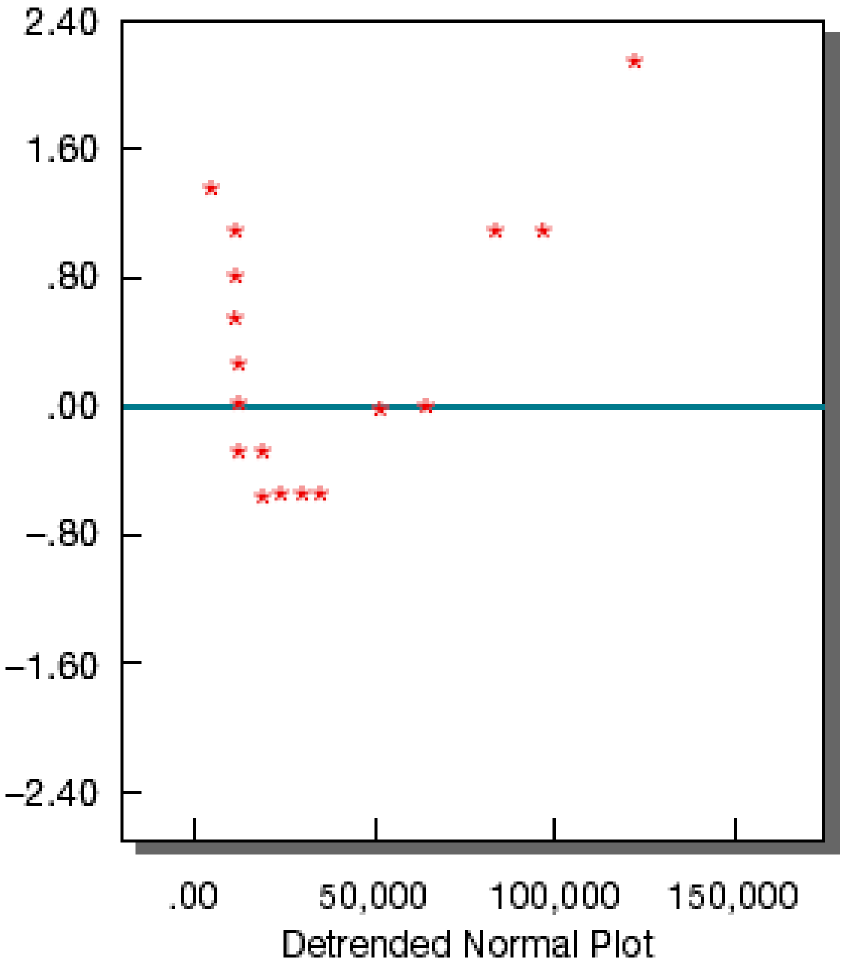


Probability Plot



Probability Plot

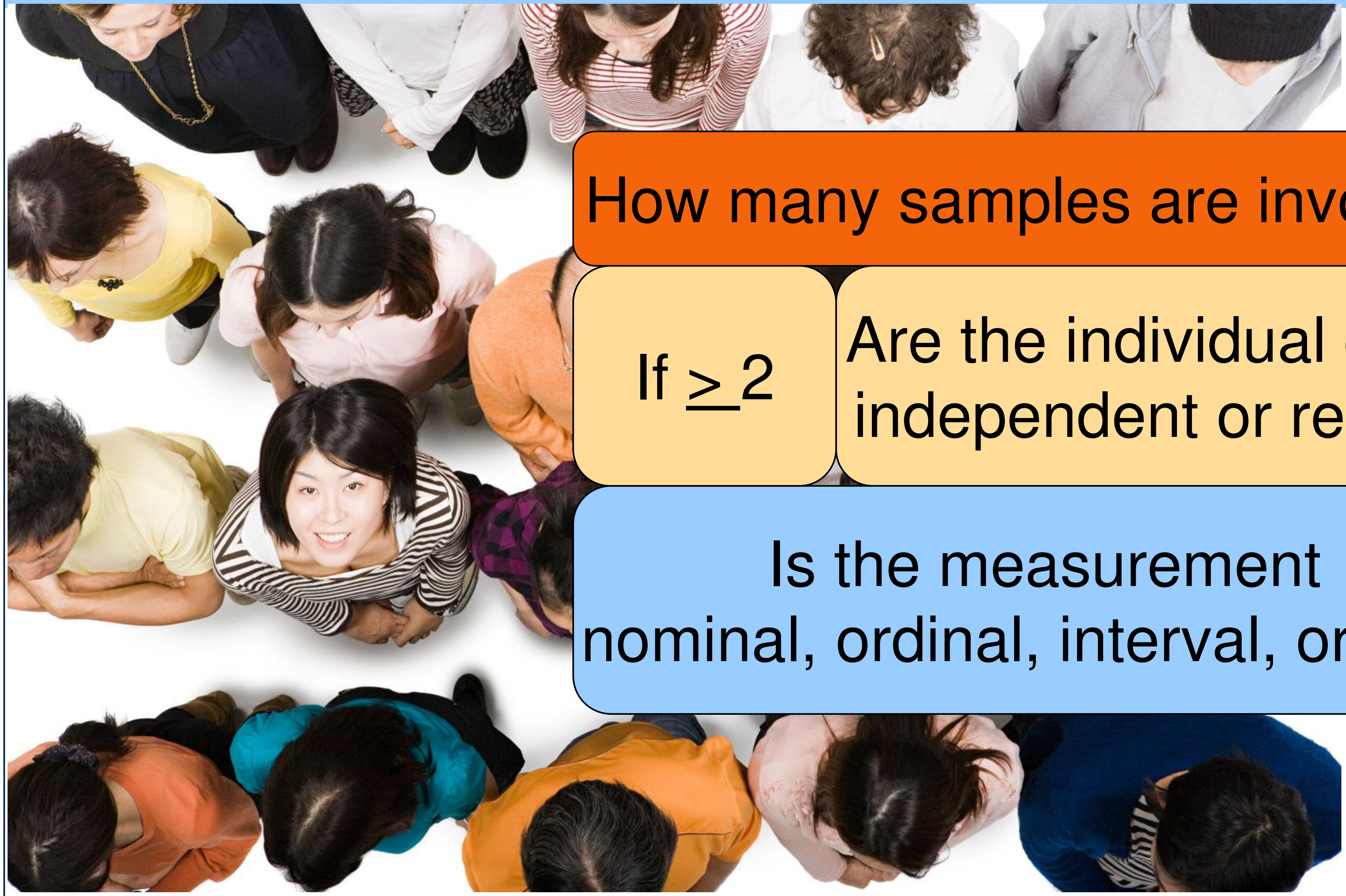




Advantages of Nonparametric Tests

- Easy to understand and use
- Usable with nominal data
- Appropriate for ordinal data
- Appropriate for non-normal population distributions

How to Select a Test



How many samples are involved?

If ≥ 2

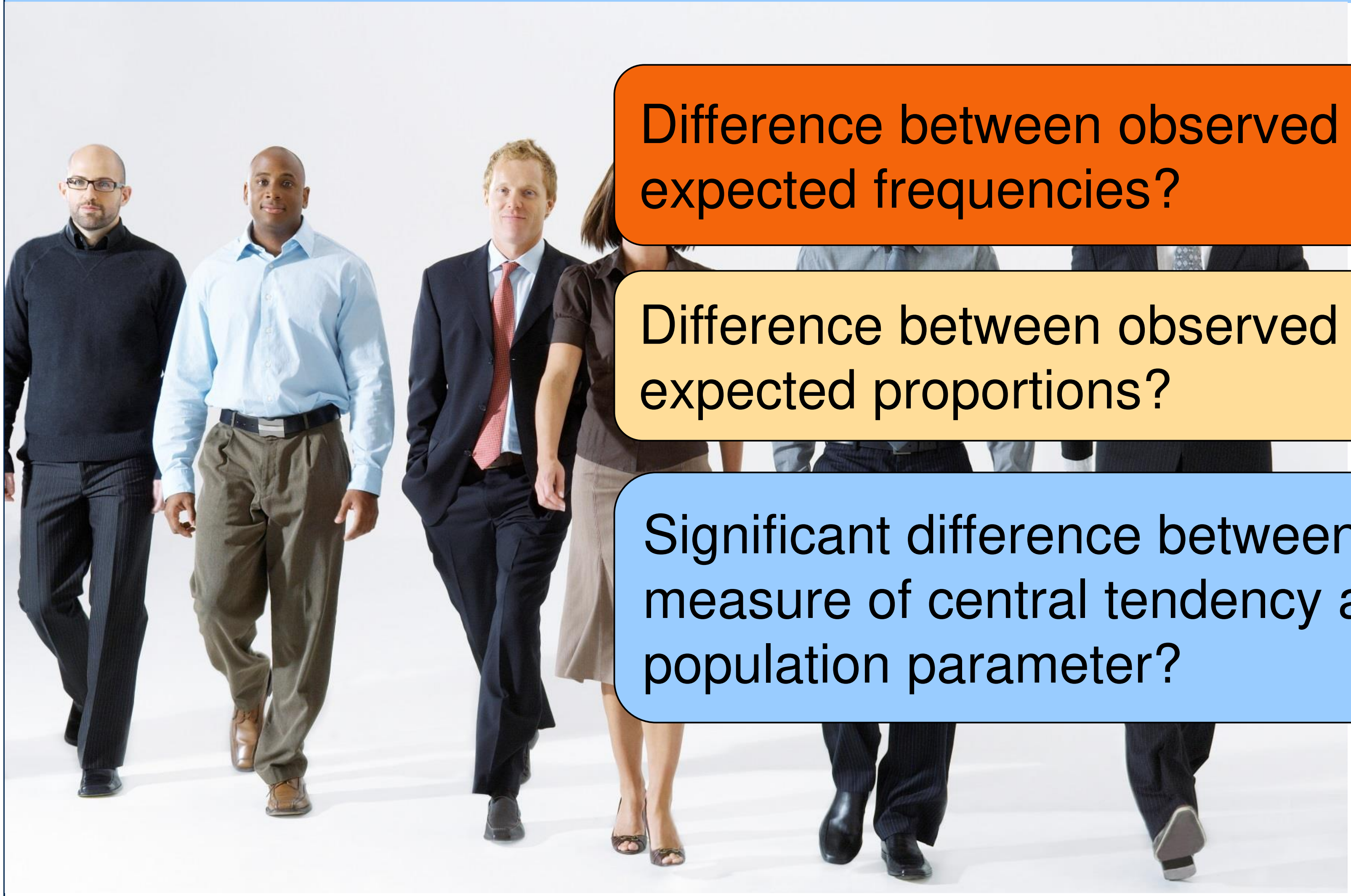
Are the individual cases independent or related?

Is the measurement nominal, ordinal, interval, or ratio?

Recommended Statistical Techniques

		Two-Sample Tests		k-Sample Tests	
Measurement Scale	One-Sample Case	Related Samples	Independent Samples	Related Samples	Independent Samples
Nominal	• Binomial • x^2 one-sample test	• McNemar	• Fisher exact test • x^2 two-samples test	• Cochran Q	• x^2 for k samples
Ordinal	• Kolmogorov-Smirnov one-sample test • Runs test	• Sign test • Wilcoxon matched-pairs test	• Median test • Mann-Whitney U • Kolmogorov-Smirnov • Wald-Wolfowitz	• Friedman two-way ANOVA	• Median extension • Kruskal-Wallis one-way ANOVA
Interval and Ratio	• t-test • Z test	• t-test for paired samples	• t-test • Z test	• Repeated-measures ANOVA	• One-way ANOVA • n-way ANOVA

Questions Answered by One-Sample Tests

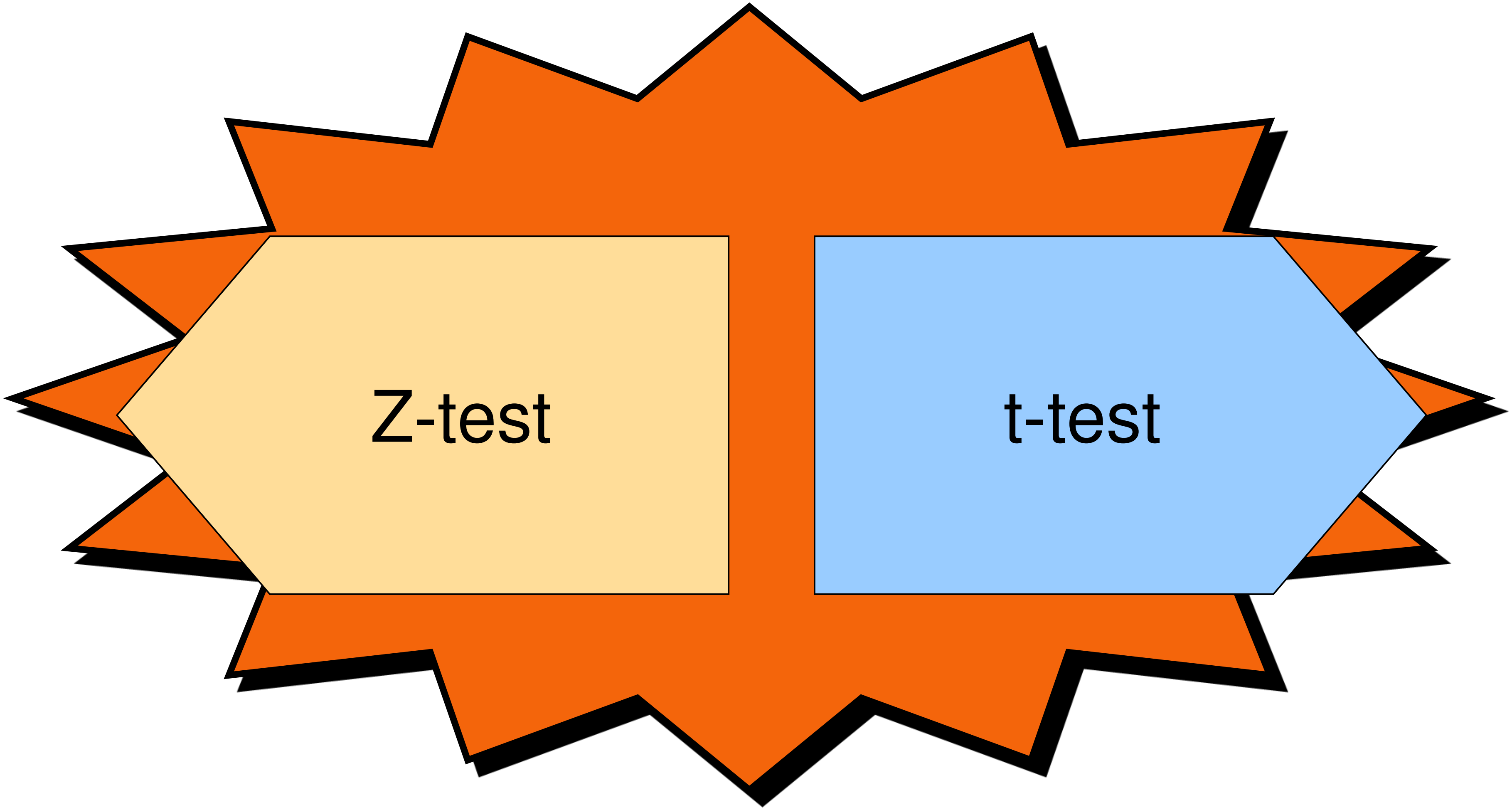


Difference between observed and expected frequencies?

Difference between observed and expected proportions?

Significant difference between some measure of central tendency and the population parameter?

Parametric Tests



One-Sample *t*-Test Example

Null	*Ho: = 50 mpg*
Statistical test	*t-test*
Significance level	*.05, n=100*
Calculated value	*1.786*
Critical test value	*1.66 (from Appendix C, Exhibit C-2)*

One Sample Chi-Square Test Example

Living Arrangement	Intend to Join	Number Interviewed	Percent (no. interviewed/200)	Expected Frequencies (percent x 60)
Dorm/fraternity	16	90	45	27
Apartment/rooming house, nearby	13	40	20	12
Apartment/rooming house, distant	16	40	20	12
Live at home	15	30	15	9
Total	60	200	100	60

One-Sample Chi-Square Example

Null	*Ho: 0 = E*
Statistical test	*One-sample chi-square*
Significance level	*.05*
Calculated value	*9.89*
Critical test value	*7.82* (from Appendix C, Exhibit C-3)

Two-Sample Parametric Tests

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Two-Sample t-Test Example

	A Group	B Group
Average hourly sales	X_1 = \$1,500	X2 = \$1,300
Standard deviation	s_1 = 225	s2 = 251

Two-Sample t-Test Example

Null	*Ho: A sales = B sales*
Statistical test	t-*test*
Significance level	*.05 (one-tailed)*
Calculated value	*1.97, d.f. = 20*
Critical test value	*1.725* (from Appendix C, Exhibit C-2)

Two-Sample Nonparametric Tests: Chi-Square

		On-the-Job-Accident		
	Cell Designation Count Expected Values	Yes	No	Row Total
Smoker	Heavy Smoker	1,1 12, 8.24	1,2 4 7.75	16
	Moderate	2,1 9 7.73	2,2 6 7.27	15
	Nonsmoker	3,1 13 18.03	3,2 22 16.97	35
	Column Total	34	32	66

Two-Sample Chi-Square Example

Null	*There is no difference in distribution channel for age categories.*
Statistical test	*Chi-square*
Significance level	*.05*
Calculated value	*6.86, d.f. = 2*
Critical test value	*5.99* (from Appendix C, Exhibit C-3)

SPSS Cross-Tabulation Procedure

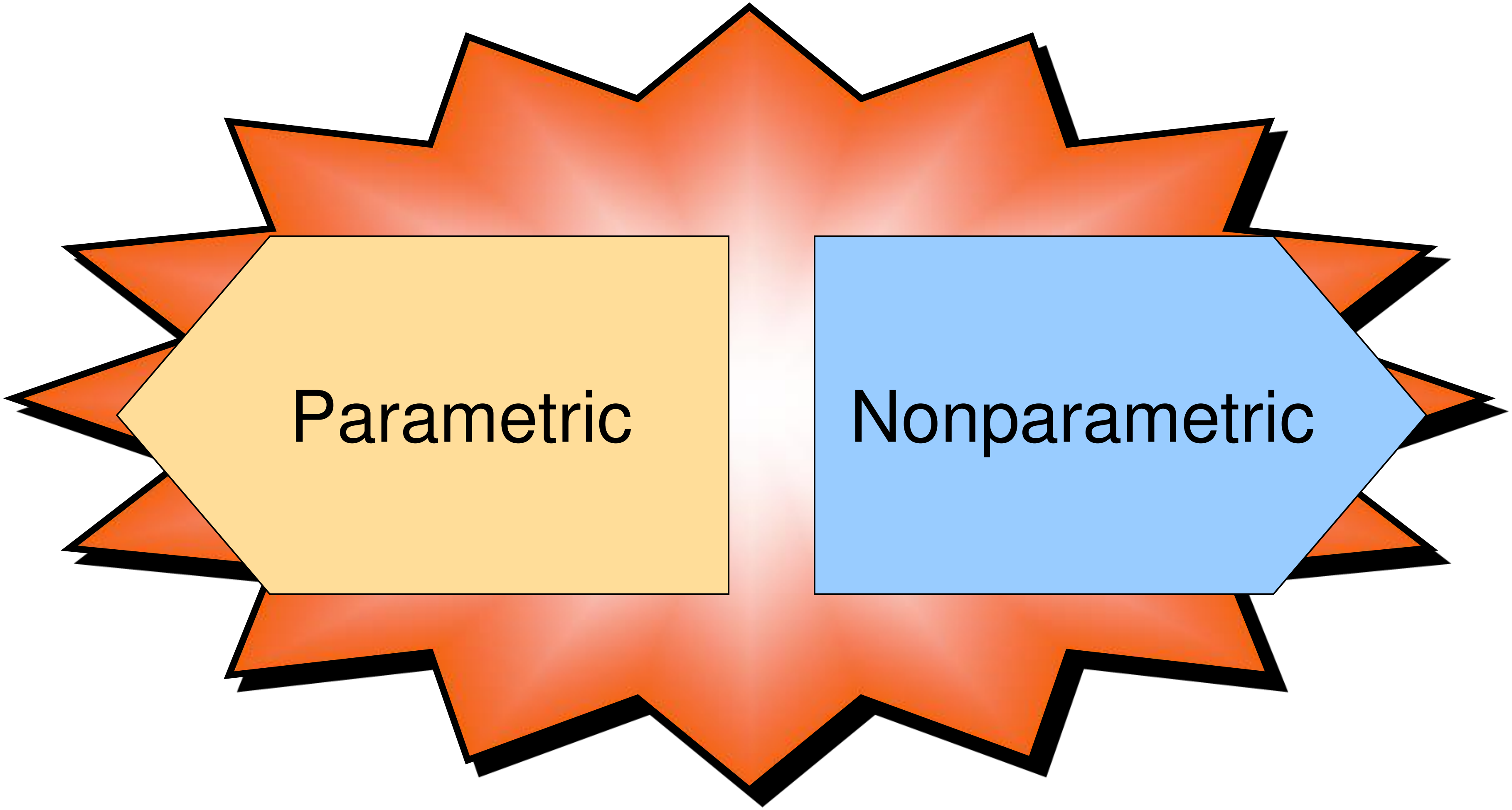
INCOME BY POSSESSION OF MBA

INCOME	Count	MBA Yes 1	MBA No 2	Row Total
High	1	30	30	60 60.0
Low	2	10	30	40 40.0
	Column Total	40 40.0	60 60.0	100 100.0

Chi-Square	Value	D.F.	Significance
Pearson	6.25000	1	.01242
Continuity Correction	5.25174	1	.02192
Likelihood Ratio	6.43786	1	.01117
Mantel-Haenszel	6.18750	1	.01287

Minimum Expected Frequency: 16.000

Two-Related-Samples Tests



Sales Data for Paired-Samples t-Test

Company	Sales Year 2	Sales Year 1	Difference D	D^2
GM	126932	123505	3427	11744329
GE	54574	49662	4912	24127744
Exxon	86656	78944	7712	59474944
IBM	62710	59512	3192	10227204
Ford	96146	92300	3846	14971716
AT&T	36112	35173	939	881721
Mobil	50220	48111	2109	4447881
DuPont	35099	32427	2632	6927424
Sears	53794	49975	3819	14584761
Amoco	23966	20779	3187	10156969
Total			$\Sigma D = 35781$	$\Sigma D = 157364693$

Paired-Samples t-Test Example

Null	*Year 1 sales = Year 2 sales*
Statistical test	*Paired sample t-test*
Significance level	*.01*
Calculated value	*6.28, d.f. = 9*
Critical test value	*3.25* (from Appendix C, Exhibit C-2)

SPSS Output for Paired-Samples *t*-Test

---*t*-tests for paired samples---

Variable	Number of Cases	Mean	Standard Deviation	Standard Error
Year 2 Sales	10	62620.9	31777.649	10048.975
Year 1 Sales	10	59038.8	31072.871	9836.104

(Difference Mean)	Standard Deviation	Standard Error	Corr.	2-tail Prob.	*t* Value	Degrees of Freedom	2-tail Prob.
3582.1000	1803.159	570.209	.999	.000	6.28	9	.000

Related Samples Nonparametric Tests: McNemar Test

Before	After Do Not Favor	After Favor
Favor	A	B
Do Not Favor	C	D

Related Samples Nonparametric Tests: McNemar Test

Before	After Do Not Favor	After Favor
Favor	*A=10*	*B=90*
Do Not Favor	*C=60*	*D=40*

k-Independent-Samples Tests: ANOVA

Tests the null hypothesis that the means of three or more populations are equal.

One-way: Uses a single-factor, fixed-effects model to compare the effects of a treatment or factor on a continuous dependent variable.

ANOVA Example

Model Summary					
Source	***d.f.***	***Sum of Squares***	***Mean Square***	***F Value***	***p Value***
Model (airline)	*2*	*11644.033*	*5822.017*	*28.304*	*0.0001*
Residual (error)	*57*	*11724.550*	*205.694*		
Total	*59*	*23368.583*			

	Means Table			
	Count	***Mean***	***Std. Dev.***	***Std. Error***
Lufthansa	*20*	*38.950*	*14.006*	*3.132*
Malaysia Airlines	*20*	*58.900*	*15.089*	*3.374*
Cathay Pacific	*20*	*72.900*	*13.902*	*3.108*

All data are hypothetical

ANOVA Example Continued

Null	$\mu A1 = \mu A2 = \mu A3$
Statistical test	*ANOVA and F ratio*
Significance level	*.05*
Calculated value	*28.304, d.f. = 2, 57*
Critical test value	*3.16* (from Appendix C, Exhibit C-9)

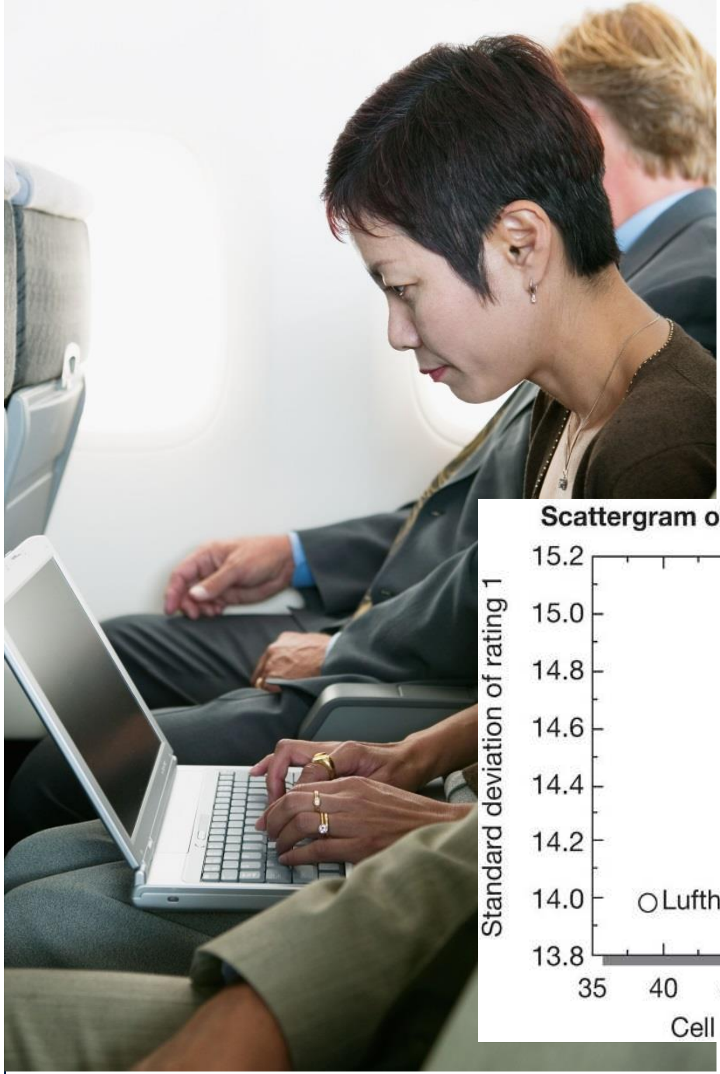
Post Hoc: Scheffe's S Multiple Comparison Procedure

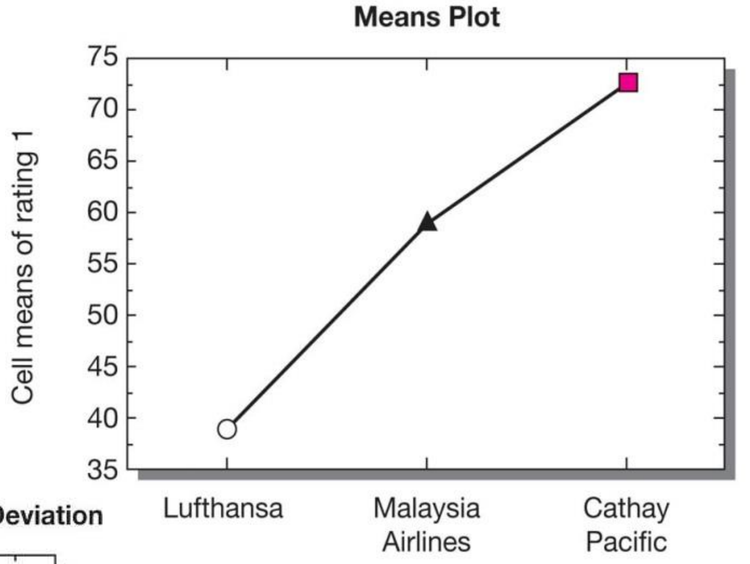
	Verses	Diff	Crit. Diff.	p Value
Lufthansa	Malaysia Airlines	19,950	11.400	.0002
	Cathay Pacific	33.950	11.400	.0001
Malaysia Airlines	Cathay Pacific	14.000	11.400	.0122

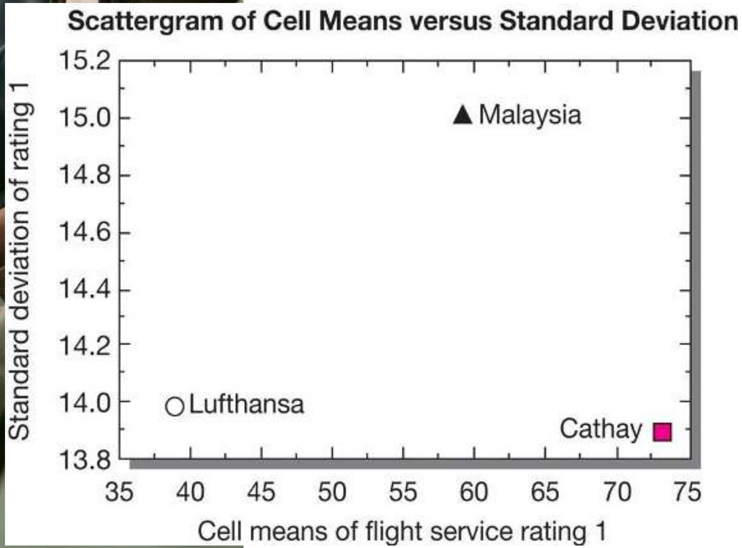
Multiple Comparison Procedures

Test	Complex Comparisons	Pairwise Comparisons	Equal n's Only	Unequal n's	Equal Variances Assumed	Unequal Variances Not Assumed
Fisher LSD	X			X	X	
Bonferroni	X		X	X		
Tukey HSD	X		X		X	
Tukey-Kramer	X			X	X	
Games-Howell	X			X		X
Tamhane T2	X			X		X
Scheffé S		X	X	X	X	
Brown-Forsythe		X	X	X		X
Newman-Keuls	X				X	
Duncan	X				X	
Dunnet's T3						X
Dunnet's C						X

ANOVA Plots




Means Plot
Cell means of rating 1
75
70
65
60
55
50
45
40
35
Lufthansa
Malaysia Airlines
Cathay Pacific


Scattergram of Cell Means versus Standard Deviation
Standard deviation of rating 1
15.2
15.0
14.8
14.6
14.4
14.2
14.0
13.8
Malaysia
Lufthansa
Cathay
35 40 45 50 55 60 65 70 75
Cell means of flight service rating 1

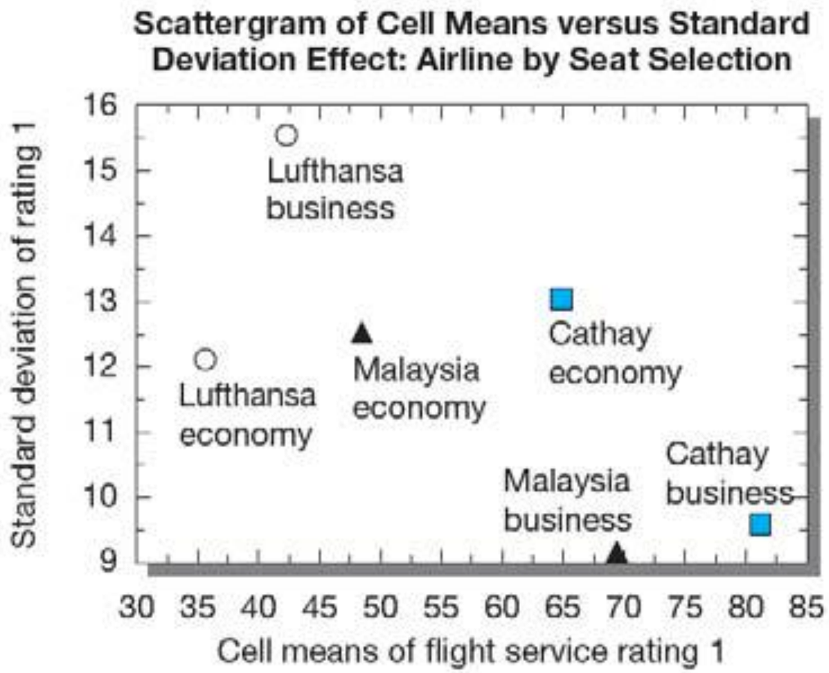
Two-Way ANOVA Example

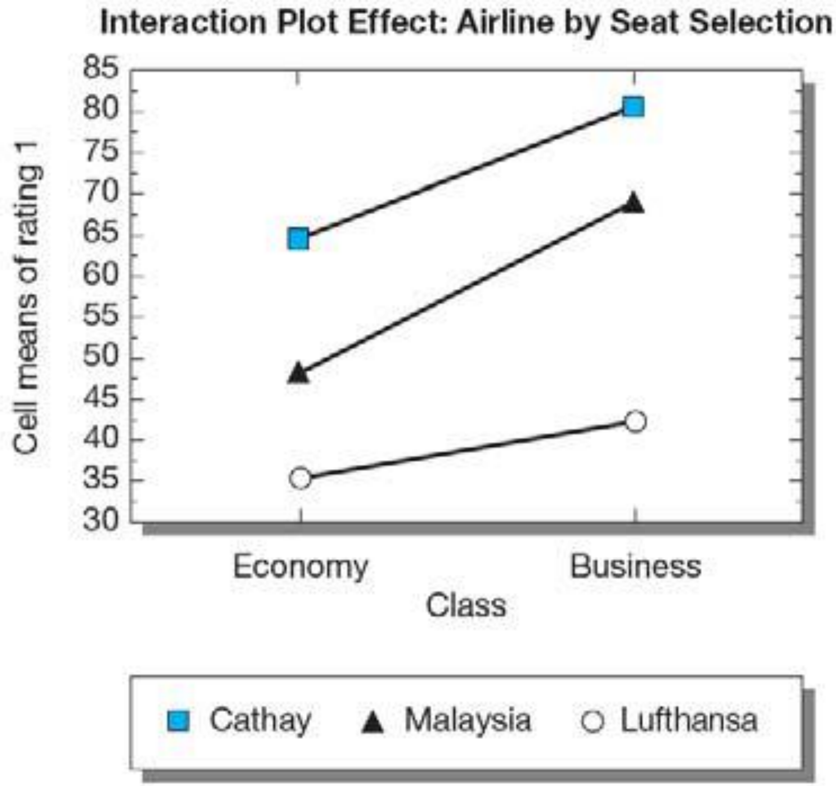
Model Summary					
Source	d.f.	Sum of Squares	Mean Square	F Value	*p* Value
Airline	*2*	*11644.033*	*5822.017*	*39.178*	*0.0001*
Seat selection	*1*	*3182.817*	*3182.817*	*21.418*	*0.0001*
Airline by seat selection	*2*	*517.033*	*258.517*	*1.740*	*0.1853*
Residual	*54*	*8024.700*	*148.606*		

Means Table Effect: Airline by Seat Selection				
	Count	***Mean***	***Std. Dev.***	***Std. Error***
Lufthansa economy	*10*	*35.600*	*12.140*	*3.839*
Lufthansa business	*10*	*42.300*	*15.550*	*4.917*
Malaysia Airlines economy	*10*	*48.500*	*12.501*	*3.953*
Malaysia Airlines business	*10*	*69.300*	*9.166*	*2.898*
Cathay Pacific economy	*10*	*64.800*	*13.037*	*4.123*
Cathay Pacific business	*10*	*81.000*	*9.603*	*3.037*

All data are hypothetical

Two-way Analysis of Variance Plots





k-Related-Samples Tests

More than two levels in grouping factor

Observations are matched

Data are interval or ratio

Summary Tables for Repeated-Measures ANOVA

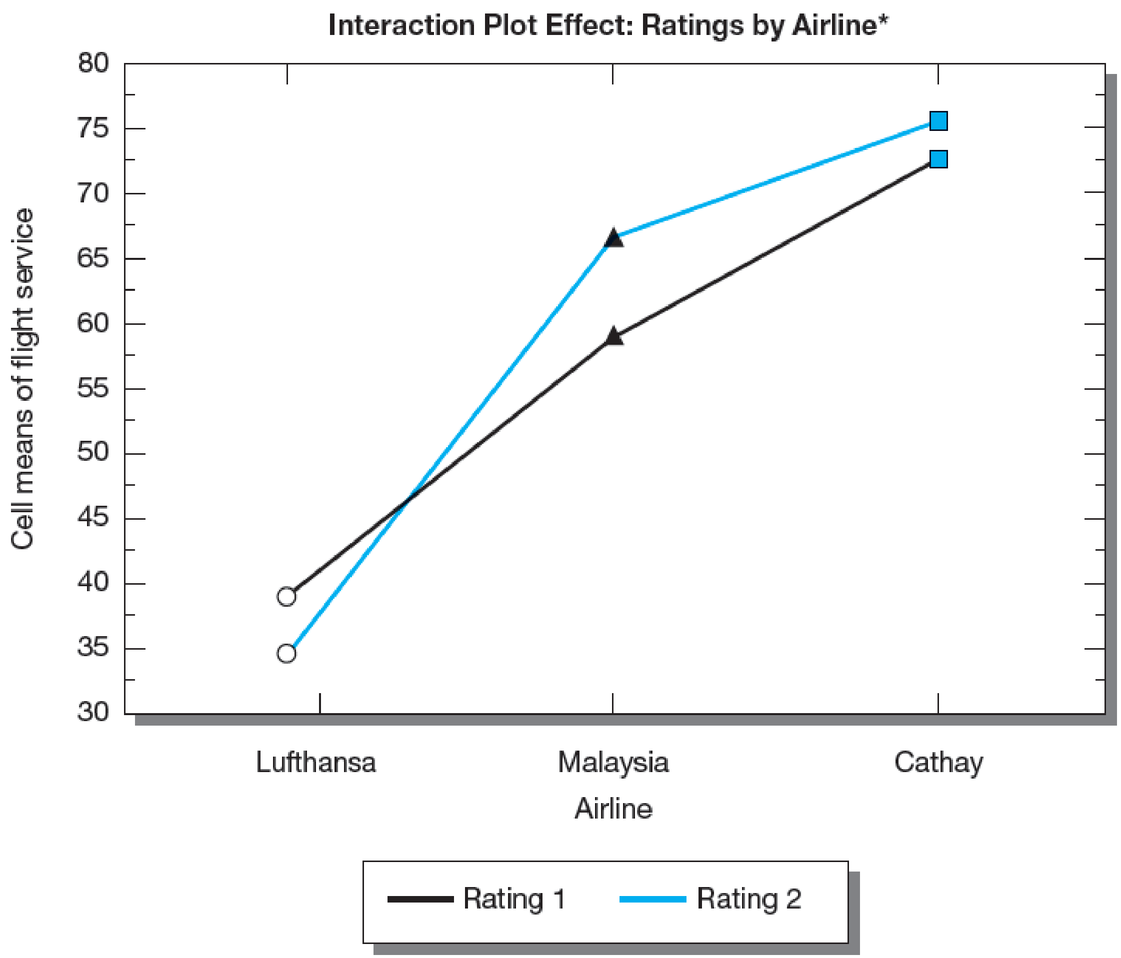
Model Summary					
Source	d.f.	Sum of Squares	Mean Square	F Value	p Value
Airline	2	3552735.50	17763.775	67.199	0.0001
Subject (group)	57	15067.650	264.345		
Ratings	1	625.633	625.633	14.318	0.0004
Ratings by air.......	2	2061.717	1030.858	23.592	0.0001
Ratings by subj.....	57	2490.650	43.696		

Means Table by Airline				
	Count	Mean	Std. Dev.	Std. Error
Rating 1, Lufthansa	20	38.950	14.006	3.132
Rating 1, Malaysia Airlines	20	58.900	15.089	3.374
Rating 1, Cathay Pacific	20	72.900	13.902	3.108
Rating 2, Lufthansa	20	32.400	8.268	1.849
Rating 2, Malaysia Airlines	20	72.250	10.572	2.364
Rating 2, Cathay Pacific	20	79.800	11.265	2.519

Means Table Effect: Ratings				
	Count	Mean	Std. Dev.	Std. Error
Rating 1	60	56.917	19.902	2.569
Rating 2	60	61.483	23.208	2.996

All data are hypothetical.

Repeated Measures ANOVA Plot



*All data are hypothetical.

Key Terms

- *a priori* contrasts
- Alternative hypothesis
- Analysis of variance (ANOVA)
- Bayesian statistics
- Chi-square test
- Classical statistics
- Critical value
- F ratio
- Inferential statistics
- K-independent-samples tests
- K-related-samples tests
- Level of significance
- Mean square
- Multiple comparison tests (range tests)
- Nonparametric tests
- Normal probability plot
- Null hypothesis
- Observed significance level
- One-sample tests
- One-tailed test

Key Terms

- *p* value
- Parametric tests
- Power of the test
- Practical significance
- Region of acceptance
- Region of rejection
- Statistical significance
- t distribution
- Trials
- t-test
- Two-independent-samples tests
- Two-related-samples tests
- Two-tailed test
- Type I error
- Type II error
- Z distribution
- Z test