Experimental Design and Statistical Analysis Seminar

Shah Golshan
sgolshan@Health.ucsd.edu
858-642-1264

1

Shahrokh 'Shah' Golshan, Ph.D.

- I am a Project Scientist at UCSD/SDVAMC.
- 1985-2000: I was the principal biostatistician and director of the Data Management and Analysis Core for the Mental Health Clinical Research Center at UCSD.
- 2000-2013: I was the PI and director of the Methodology, Biostatistics and D.M. Unit for the Advanced Center for Innovation in Services and Intervention Research.
- 2013-2016: I was the director of the VMRF Vetstats Biostatistics Core, Veterans Medical Research Foundation.
- 2015-Present Director of the Data System, Krupp Center for Integrative Research, La Jolla, CA
- 2022-Present Co-Director of the Krupp Center for Integrative Research, La Jolla, CA
- I have served multiple times as a member of review committees for NIMH and I am a member of UCSD, FISMA, IT Security and IT Technical Advisory Boards.
- I have taught an annual Research and Experimental Design seminar in the department of Psychiatry since 1986.
- I have more than 35 years experience in research, Co-PI on more than 36 grants funded by the NIH, VA or DoD and co-author on 150 peer reviewed publications.

2

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR

Beginner Session (9 Weeks):

1: Introduction to Study Design (2 Weeks)
2: Basic Principles in Experimental Design (2 Weeks)
3: Definitions And Descriptive Statistics (2 Weeks)
4: Normal Curve And Derived Scores (1 Week)
5: Data Screening (2 Weeks)
6: Inferential Procedures (2 Weeks)

Advanced Session (10 Weeks):

7: Correlation (1 Week)
8: Regression (1 Week)
9: T-test (1 Week)
10: Completely Randomized: Single Factor Design (2 Weeks)
11: Completely Randomized: Two Factors Design (1 Week)
12: Single Factor With Repeated Measures (1 Week)
13: Split-plot Factorial Design (Mixed Design) (1 Week)
14: Frequency Data Analysis (1 Week)
15: Order Data Analysis (1 Week)

3

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR

Beginner Session (9-12 Weeks):

1: Introduction to Study Design (2 Weeks)
2: Basic Principles in Experimental Design (2 Weeks)
3: Definitions And Descriptive Statistics (2 Week)
4: Normal Curve And Derived Scores (1 Week)
5: Data Screening (2 Weeks)
6: Inferential Procedures (2 Weeks)

4

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR: 1: STUDY DESIGNS (2 Weeks)

1. Classification of Study Designs
2. Activities By Stage of Clinical Trial
3. Essential Design Features of a Controlled Clinical Trial
4. Quality Assurance
5. Computer Facility
6. Data Security Precautions
7. Preparation of Analysis Files

5

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR: 2: Experimental Design (2 Weeks)

1. **Main Principles of Experimental Design**
2. **Selecting an appropriate Design**
3. **Criteria for Evaluating An Experimental Design**
4. **Threats to Valid Inference-Making**
5. **Overview of Experimental Designs Types:**
   i. Completely Randomized design (CR)
   ii. Randomized Block Design (RB)
   iii. Latin Square Design (LS-p)
   iv. Completely Randomized Factorial Design
   v. Split-Plot Factorial Design

6

3: Definitions and Descriptive Statistics (2 Weeks)

I: Preliminary Concepts

1. Descriptive statistics
2. Inferential statistics
3. Population
4. Sample
5. Random Sampling
6. Research
7. Data Type
8. Independent and Dependent Variables
9. Statistical Inference And Hypothesis Testing
10. Errors in Hypothesis Testing

II: Descriptive Statistics

1. Frequency Distributions
2. Graphic Representation
3. Measures of Central Tendency
4. Measures of variability

7

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR: 4: Normal Curve and Derived Scores (1 Week)

1. The Normal Curve
2. Derived Scores
3. Linear Transformation: Z Scores
4. Non-Linear Transformation: Centile, T and Stanines Scores
5. Comparability of Scores
6. Combining Measures From Different Distributions

8

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR: 5: Data Screening (2 Weeks)

1. Accuracy of data file
2. True correlations among variables
3. Missing data
4. Outliers
5. Normality, Linearity, Homoscedasticity
6. Multicollinearity and Singularity
7. Common Data Transformations

9

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR: 6: Inferential Procedures (2 Weeks)

- Hypothesis Testing
  1. Probability
  2. General Procedure For Hypothesis Testing
  3. Factors Affecting Type II Error
  4. Calculating Type II Error
  5. Estimating Sample Size
  6. Steps In Testing A Hypothesis
  7. Hypothesis Testing About Single Mean and Two Means
- Estimation
- The Relation Between Estimation and Hypothesis Testing

10

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR

Beginner Session (9 Weeks):

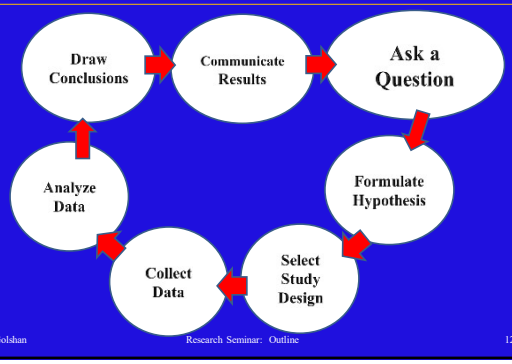
1: Introduction to Study Design (2 Weeks)
2: Basic Principles in Experimental Design (2 Weeks)
3: Definitions And Descriptive Statistics (2 Weeks)
4: Normal Curve And Derived Scores (1 Week)
5: Data Screening (2 Weeks)
6: Inferential Procedures (2 Weeks)

Advanced Session (10 Weeks):

7: Correlation (1 Week)
8: Regression (1 Week)
9: T-test (1 Week)
10: Completely Randomized: Single Factor Design (2 Weeks)
11: Completely Randomized: Two Factors Design (1 Week)
12: Single Factor With Repeated Measures (1 Week)
13: Split-plot Factorial Design (Mixed Design) (1 Week)
14: Frequency Data Analysis (1 Week)
15: Order Data Analysis (1 Week)

11

Research Process



12

Why Learn about statistics and why these selected topics?

- **Everyday life** (fitbit, Smart-Watch)
- **Research :** Research is the empirical investigation of the relationship between or among several variables.
- There are three basic goals in any research:
  1. To collect data that are free of bias.
  2. to draw valid conclusions concerning the effects of an independent variable.
  3. to make valid generalizations to populations and setting of interest.

13

Why Learn about statistics and why these selected topics?

- If you want to do your own research, you need to have a good understanding of how to conduct research properly.
- If you are involved in research at any level, you must have a basic knowledge of research.
- When you read a research article, the authors describe how the data were collected and analyzed.
- To understand what the authors are trying to say, you need to understand their lingo.

14

Why Learn about statistics and why these selected topics?

- The research world is highly competitive.
- Less than 7% of grant applications receive funding.
- Few pages are allowed for grant proposals.
- It is easier for just one thing to make or break a chance of getting funded.
- The experimental design and analysis sections of grants are subject to more scrutiny than ever.
- A spectrum of statistical methodologies is needed to address different research questions.
- It is critical that experimental design and analysis sections are written accurately and
- that statistical analysis plans are appropriate and tailored to each specific hypothesis.

15

Statistical Myth

1. Knowing a statistical software is the main requirement to analyze your own data: **False**
2. Statistics doesn't make sense: **True**
3. The best way to learn how to analyze data is to analyze some: **True**
4. Most researchers feel comfortable dealing with statistical/design issues: **False**
5. It's often difficult to find someone knowledgeable enough who is willing and able explain statistical issues: **True**
6. You can fix almost any design problems with a good statistics: **False**

16

Statistics, good, bad and ugly: Good Statistics

1. Statistics can force us to look at data and facts, rather than relying on opinions.
2. Statistics helps us test our beliefs and to learn from it.
3. Even when statistical analysis fails to confirm our initial hypothesis, we are forced to reevaluate the hypothesis, which leads to improved understanding.
4. Statistics are the bridge between raw data and knowledge and understanding.
5. Statistical thinking uses data to separate variation from special and common causes.
6. Statistical analysis is especially useful when we are faced with complex situations that challenge human understanding.

17

Statistics, good, bad and ugly: Bad Statistics

1. They are an abstraction of reality, and, because of this, they are not the reality itself.
2. Sometimes, statistics are worse than raw numbers. A mean is a single number that may represent thousands of individual measurements.
3. Statistics traps us into analysis paralysis. No amount of statistical analysis can ever produce certainty.
4. Statistics are confusing.
5. Statistics are often used without graphics. The first three rules of data analysis may be plot the data, plot the data, and plot the data.

18

Statistics, good, bad and ugly: Ugly Statistics

1. There are three kinds of lies: lies, damned lies, and statistics." (Benjamin Disraeli)
2. The use of statistics to support weak arguments.
3. The tendency of people to criticize statistics that do not support their positions.
4. Or to overlook or ignore statistics that contradict their own beliefs.
5. Statistics are often misused by data bullies to dismiss their opposition.

19

Why Use Statistics?

1. It can summarize and simplify large amounts of data.
2. Statistics looks at data and facts, rather than relying on opinions and can draw conclusions about data.
3. Statistics may reveal underlying patterns in data.
4. Statistical techniques are probabilistic and not definitive and can separate the probable from the possible.
5. Science is more in the question business than the answer business.
6. Statistics are the bridge between raw data and knowledge and understanding
7. **Research is about the evolution of knowledge and is the slow march of accumulating evidence.**

20

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR: OBJECTIVES

The main emphasis of this seminar is on understanding statistical procedures rather than memorization or calculation. We will discuss:

- Basic statistical concepts (i.e. randomization, & hypothesis testing)
- Basic principles in Experimental Design
- Variety of Experimental Designs, their strengths and weaknesses
- ~~A selected group of statistical test procedures~~

21

References:

1: Cohen, Jacob. Statistical Power Analysis for the Behavioral Sciences.1988.

2: Dawson-Saunders B. and R. G. Trapp. Basic and Clinical Biostatistics. 1990, Appleton & Lange.

6: Kirk, R. E. Experimental Design (2nd Edition). 1982, Brooks/Cole.

7: Keppel, Geoffrey. Design and Analysis, A Researcher's Handbook.1973.

9: Marigold, L. and P. S. Gallo, Jr. The Practical Statistician: Simplified Handbook of Statistics. 1975, Wadsworth Publishing Co.

13: Pedhazur, E. J. Multiple Regression In Behavioral Research (2nd Edition). 1982, Holt, Rinehart and Winston.

14: Shott, S. Statistics for Health Professionals. 1990.

15: Singer J.D. and J.B. Willett. Applied Longitudinal Data Analysis. 2003

16: Siegel, S. S. Nonparametric Statistics for the Behavioral Sciences. 1956

18: Tabachnick, B. and L. S. Fidell. Using Multivariate Statistics. 1989.

19: Winer, B. J. Statistical Principles in Experimental Design. 1971.

20: Witte, Robert S. Statistics. 1989. Holt, Rinehart and Winston, Inc.

22

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR

Beginner Session (9 Weeks):

- **1: Introduction to Study Design (1 Week)**
- **2: Basic Principles in Experimental Design (2 Weeks)**
- **3: Definitions And Descriptive Statistics (1 Week)**
- **4: Normal Curve And Derived Scores (1 Week)**
- **5: Data Screening (2 Weeks)**
- **6: Inferential Procedures (2 Weeks)**

Advanced Session (10 Weeks):

- **7: Correlation (1 Week)**
- **8: Regression (1 Week)**
- **9: T-test (1 Week)**
- **10: Completely Randomized: Single Factor Design (2 WEEKS)**
- **11: Completely Randomized: Two Factors Design (1 WEEK)**
- **12: Single Factor With Repeated Measures (1 WEEK)**
- **13: Split-plot Factorial Design (Mixed Design) (1 WEEK)**
- **14: Frequency Data Analysis (1 WEEK)**
- **15: Order Data Analysis (1 WEEK)**

23

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR: 7: Correlation (1 Week)

1. Characteristics of the Correlation Coefficient
2. Pearson Correlation Coefficient
3. Covariance
4. Testing Difference Between Two r Coefficients
5. Spearman Rank Correlation
6. Partial Correlation
7. SPSS Program and Output

24

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
8: Regression (1 Week)

1. Characteristics of the Regression Equation
2. Best Fitting Straight Line
3. Error of Predication
4. The Coefficients of Alienation and Determination
5. Regression Methods
6. Assessing Regression Model
7. Assumptions of the Linear Regression Model
8. Logistic Regression
9. Odd Ratio
10. SPSS Program

25

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
9: t-test (1 Week)

1. Introduction to t-test
2. Degrees of Freedom
3. Characteristics and Application of t Distribution
4. Hypothesis Testing:
   a) Single Mean
   b) Two Independent Means
   c) Two Dependent Means
5. Interval Estimates of μ and $\mu_X - \mu_Y$
6. Power calculation
7. Strength of Association
8. Sample Size estimation
9. SPSS Screens

26

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
10: Design And Analysis of Completely Randomized: Single Factor Design (2 WEEKS)

1. Definition and Applicable Designs
2. ANOVA or t-test
3. Design Issues
4. Sources of Variability
5. F Test and F Ratio
6. Design Assumptions
7. Unequal N and Nonorthogonality
8. Fixed And Random Effects
9. One-Way ANOVA
10. Power calculation for F Tests

27

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
10: Design And Analysis of Completely Randomized: Single Factor Design (2 WEEKS)

11. Specific Comparisons
12. Type I Errors
13. Planned Versus Post-hoc Comparisons
14. Planned Comparisons
15. Evaluation of Independent Comparisons
16. Post-Hoc (Multiple) Comparisons
17. Multiple Comparison Methods
18. Selection of a Multiple-Comparison Test
19. SPSS Screens

28

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
11: Design And Analysis of Completely Randomized: Two Factors Design (1 WEEK)

1. Design considerations
2. Main and Interaction effects
3. Design assumptions
4. Analysis method
5. Strength of association
6. Planned and Post-hoc Comparisons
7. Comparisons Involving Marginal Means (No Significant Interaction)
8. Orthogonal Comparisons
9. Multiple Comparisons
10. Comparisons Involving Cell Means (Significant Interaction)
11. Analysis of Simple Main Effects
12. SPSS Program and Outputs

29

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
12: Design And Analysis Of Single Factor With Repeated Measures (1 WEEK)

1. Randomized Block Design Considerations/Assumptions
2. Advantages And Disadvantages of RB Designs
3. Single Factor Design Considerations and Assumptions
4. Comparisons Involving Treatment Means
5. Direct Method
6. Weighted Sums Method
7. Post Hoc Multiple Comparisons
8. Variance And Covariance Matrix
9. Power Calculations
10. SPSS Menus, Syntax and Outputs

30

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
13: Design And Analysis Of Split-Plot Factorial Design (Mixed Design) (1 WEEK)

1. Split-Plot Factorial Design Considerations
2. Design Assumptions
3. ANOVA Table
4. Example
5. Simple Main effects of the Repeated Factor
6. Error Term for Simple Main effects of the Repeated Factor
7. Simple Main effects of the Non-repeated Factor
8. Methods 1 and 2 for Error Term for Simple Main effects of the Non-repeated Factor
9. SPSS Menu, Syntax and Output

31

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
14: Frequency Data Analysis (1 WEEK)

1. Type of Chi-Square tests
2. Chi-square Limitations and Degrees of Freedom
3. One-way Chi-square
4. Two-way Chi-square Test
5. Repeated-measures Chi-square
6. Test of Association
7. Mcnemar's Test
8. Multiple Comparisons
9. SPSS Menu, Syntax and Output

- Null and Alternative Hypotheses
- Calculating Expected Frequencies
- Data Analysis Method
- Example

32

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
15: Order Data Analysis (1 WEEK)

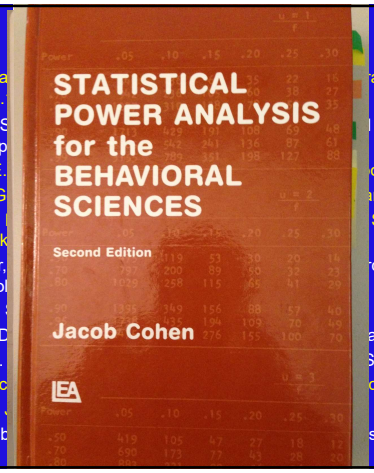
1. Design considerations
2. Between Subject Design
   a) Rank-Sum
   b) Mann-Whitney
   c) Kolmogorov-Smirnov
   d) Kruskal-Wallis
3. Within Subject Design
   a) The Sign test
   b) Wilcoxon's Signed Ranks Test
4. Two within Design: Analysis Of Variance By Ranks: The Friedman
5. Multiple Comparison Nemenyi's Test
6. SPSS Menu, Syntax and Output

33

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
16: Introduction To Multivariate (1 WEEK)

1. Multivariate Analysis of Variance (MANOVA)
2. Discriminant Analysis
3. Principal Component Analysis
4. Factor Analysis

34



1: Cohen, Ja … Sciences.
2: Dawson-S … 1990, App … Biostatistics.
6: Kirk, R. E. … oks/Cole.
7: Keppel, G … andbod.1973.
9: Marigold, … Handbook … Simplified
13: Pedhazur, … 1982, Hol … rch (2nd Edition).
14: Shott, S. …
15: Singer J.D … alysis. 2003
16: Siegel, S. … Sciences. 1956
18: Tabachnic … cs. 1989.
19: Winer, B. … 1971.
20: Witte, Rob … ston, Inc.

35

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
Tests and Requirements

- Class attendance is taken at every class and it is Mandatory.
- Any class absence = Official Letter Of Attendance Failure.
- Grades A or B for all Quizzes and Tests are required.
- Test:
  1. Weekly Quiz (10 Multiple-Choice Questions)
  2. Monthly Test (10 Open-Ended Questions)
  3. Session Test:
     I. 40 Multiple-Choice Questions
     II. 20 Open-Ended Questions
  4. Final Seminar Test:
     I. 60 Multiple-Choice Questions
     II. 40 Open-Ended Questions

36

EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS SEMINAR:
Tests and Requirements

JUST KIDING

There is NO Quiz

There is NO Test

This seminar is for you and just for you.

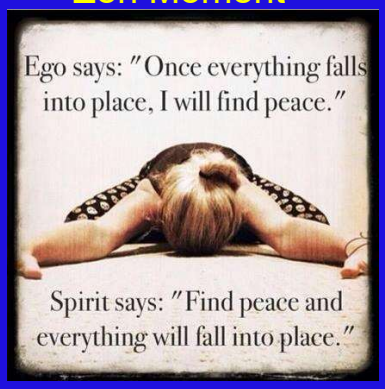
If you come, you will learn something new.

If you don't come, well, we will miss you.

The last person who enters the room, have to bring cookies for everyone.

37

Zen Moment



38

My Thanks to:

Assistant Statistician	**Marge Innovera**
Staff Psychologist	Les Moody
Staff Cat Feeder	Ken Opener
Wardrobe Assistant	Joaquin Closet
Sunscreen Supervisor	Les Brown
Staff Mediator	Sue First
Spiritual Counselor	Miss Dolly Lama
UCSD Foreman	Luke Bizzy
Russian Chauffeur	Picov Andropov
Computer Repair Consultant	Bill M. Moore
Photographer	F. Stop Fitzgerald
Parking Lot Attendant	Kent Stirwell
DNA Researcher	Gene Hackman
Cafeteria Chef	Benny Honda
Travel Manager	Will Price Randomly

39

Experimental Design and Statistical Analysis Seminar

Shah Golshan
sgolshan@Health.ucsd.edu
858-642-1264

40