Psychology 420: Advanced Psychological Statistics and Design of Experiments
Spring Semester 2000

Instructor

Howard B. Lee, Office: Sierra Tower 302, Phone: 818-677-4863, Messages: 818-677-2827
Office Hours: Monday 3-4 PM Tuesday, 3-4 PM

Course Objectives.

There are multiple objectives for the course. Most important among them is to introduce students to a range of basic experimental designs and to provide them with the skills to *statistically* examine data collected with each type of design. Other important objectives are to provide students with skills: (a) in designing experiments which fit the their substantive questions, (b) in reporting the results of experiments, and (c) in reading mathematical/statistical textbooks and journal articles. Students who have not taken an introductory statistics course or have taken one “a long time ago” will be at a clear disadvantage in the course. No provisions, adjustments or exceptions will be made for these students.

Design of the Course

The course content is divided into twelve sections--except for the one section, each presents a basic model of the analysis of variance or some variation on it (e.g. regression, general linear model). The major emphasis of the course is on design of experiments and the corresponding statistical analyses. The instructor will also present material that is not covered in the main textbook.

Class time will be divided between lectures, discussions, and workshops. The distribution of time between each will depend on the flow of the class. Reading assignments parallel each section. The major text is Kirk. The Hays and Howell texts are the supplementary texts (see be1ow). Each text might be considered a different instructional treatment of the same material. Students can select the text that best fits their style of learning.

Since I assume that design and statistics can only be learned by reading and doing, there are seven assignments. Time during the lab period will be allocated to allow students to work on these assignments. Whatever can not be finished in class or in the lab must be completed outside of class and submitted by the due date. Each assignment corresponds more or less to one or more sections in the course. By completing each assignment successfully, the student will have demonstrated mastery in the use of’ basic ANOVA models.

Please note that in the schedule of classes, the lecture portion of the class is on Mondays and the Lab is on Wednesdays. This was done for administrative purposes only. Lecture and lab material will be combined throughout the semester except in those cases where the instructor states differently. On certain Wednesdays where there is a departmental faculty meeting, class will either be deferred or start at a later time.

Textbooks:

Kirk, R. E. (1995). *Experinienta1 design: Procedures for the behavioral sciences*. (3rd Ed.). Pacific Grove, CA.: Brooks/Cole (Primary Test).

Hays, W. L. (1994). *Statistics.* (5th Ed.). Fort Worth, TX: Harcourt College Publishers. (Supplementary Test)

Howell, D. C. (1997). *Statistical methods for psychology*. (4th Ed.). Belmont, CA: Duxbury

The following books are dated, but certain students may be able to benefit from their presentational style. Most of these are out-of-print but are probably available in the Library.

Dayton, C. M. (1970). *Design of educational experiments*. New York: McGraw-Hill.

Edwards, A. L. (1984). *Experimental design in psychological research*. (5th Ed.) Reading: AW

Li, J. C. R. (1957). *Introduction to statistical inference*. Ann Arbor, MI: Edwards Brothers.

Natrella, M. G. (1963). *Experimental statistics*. Washington, D.C.: National Bureau of Standards.

On the course outline below, the 12 sections of the course are enumerated along with readings from Kirk. Parallel coverage of some of this material can be found in Hays and in Howell. Handouts will be given in class for material covered in the lecture that is not in the Kirk text

Course Outline

Topic	Kirk
I. Introduction[1]	Chapter 2
II. The General Linear Model	3.1 to 3.4, Chapter 6
III. Simple ANOVA	5.1-5.3, 5.5, 5.6,5.8, 5.9, 3.5, 3.6
IV. A Significant F, Multiple Comparisons	Chapter 4
V. Randomized Block & Latin Squares	Chapters 7, 8
VI. Factorial Designs with 2 factors	Chapter 9
VII. Factorial Designs with 3 or more factors	Chapter 10
VIII. Hierarchical and Nested Designs	Chapter 11
IX. Repeated Measures ANOVA	Chapter 12
X. Fractional Factorial and Response Surface	Chapter 14
XI. Analysis of Covariance	Chapter 15
XII. Introduction to Analysis of Time Series	Handout Material

[1] See also Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally. and
Kerlinger, F. N. & Lee, H. B. (2000). *Foundations of behavioral research*. (4th Ed.). Chapter 18. Fort Worth, TX: Harcourt College Publishers.

Assignments

The six of the seven assignments are based on an experiment that is described on the pages that follow. From this experiment, data have been collected; each person in the class has a different set of data obtained from separate replications of the study. The study was designed so that these data may be used for the six different statistical analyses corresponding to each assignment. You are to treat each assignment as if you were writing an article for a journal such as Journal of Experimental Psychology or the Human Factors Journal. Of course, you will provide more information about the statistical analyses in your write-ups than would be permitted by either journal. The specific assignments are given in detail on the pages following the write-up of the experiment.

Grading

This is an advanced class; so students are expected to be highly interested in the course since the skills taught are essential to your future academic pursuits (e.g. Ph.D.) or professional careers. Thus, I am more concerned about your mastering the material to be learned than about how you compare with each other in the class. The 7 assignments will count for 80% of the total grade in the course. The final examination will count for 20%. This setup will hopefully make students less grade conscious and more concerned with the important work of learning about experimental design and their statistical analyses. In this case, I would **not** accept an assignment until all aspects of it demonstrates mastery. Upon reaching mastery, each student will receive the full number of points for that assignment. In the sole judgement of the instructor, papers not reaching mastery will be returned to the student for revisions. Points will not be given until mastery has been reached. You would be expected to turn in all *7* papers in complete form, i.e., word-processed or type written in APA style. The hand computations are not to be typewritten or word-processed. Computer outputs must be **clearly annotated** and labeled. Sloppy work in the opinion of the instructor will be returned to the student no evaluation or grade. There will be tentative deadlines that the student must follow in order to complete the course in the amount of time given. Students who are unable to complete and achieve mastery on all seven papers will be given a grade of "I" (incomplete). The final examination is cumulative over the entire course and is open book and notes.

Communications

Each student is **required** to give or send the instructor a valid and working email address. The instructor will email the data set used in all assignments to the student. Also email will be used to communicate vital information concerning the class. Students who do not provide a valid and working email address or do not check their email on a regular basis will be technically dropped from the class. The instructor's email address is hblee@email.csun.edu

The Experiment: Some Effects of Level of Feedback arid Type of Grading on Students' Learning to Write Papers on Psychological Experiments

Students' academic performance has always been a major focus of education. Information about a student's performance has been used for administrative purposes (e.g., grading, selecting, and certifying students), evaluative purposes (e.g., comparing instructional methods, comparing curricula), and instructional purposes (e.g., diagnosing weaknesses, correcting errors with feedback). In some cases, this information has been used to compare students with one another or to compare instructional methods (norm-referenced comparisons). In other cases, this information has been used to determine whether the student has mastered a subject matter or whether a particular curriculum leads students to mastery of the subject matter (criterion-referenced comparisons). While all these various uses deserve further research, this study focuses on a particular combination of uses of information about student performance. The purpose of this study, in broad terms, is to examine the effects of the level of corrective feedback and the effects of normed- and criterion-referenced grading practices on students' academic performance.

More specifically, teachers often use some information about a student's current performance to help him or her improve his or her subsequent performance. This is especially true of teachers' feedback on students' papers. At one extreme, this feedback takes the form of a letter grade with no further information while at the other extreme, the feedback takes the form of extensive corrective notes to the student at each step in the paper. There is, however, little empirical evidence to suggest which form of feedback is most effective. It may be that somewhere in between the two extremes is optimal. One purpose of this study, then, was to determine a level of feedback that optimizes student performance on writing a subsequent paper.

In addition to the amount of feedback, teachers usually assign a grade–most often a letter grade–to the students' papers. The letter grade tells the student much about where she or he stands among her/his peers but little about whether he or she has mastered the instructional material. Some recent models of instruction (Bloom, 1971; Carroll, 1963; Keller, 196?) suggest that grades should do just the reverse. They should reflect the student's mastery of a subject matter rather than his rank order among peers. A second purpose of this study was to determine the effects of these two grading practices on students' performance on writing a subsequent paper.

In this study, then, students wrote two scientific papers as part of their regular classroom assignments. After the first paper was turned in, the students were randomly assigned to treatments. When the papers were returned, the student received either a letter grade or a mastery grade and either <u>none</u>, <u>some</u> or <u>many</u> written comments on the paper, depending on the particular treatment he or she was assigned to. The effects of' these treatments on students' performance on the second paper was then examined. Students were subsequently allowed to do a third paper.

While limited in scope, this study begins to provide some empirical evidence of a. practical nature on two important instructional variables amount of corrective feedback and type of grading. If the amount of feedback is unrelated to subsequent performance, the practical implication is that teachers can save themselves considerable time and effort by just jotting a grade on the paper. The results of this study will bear on whether this feedback should be in the form of a letter grade or a mastery grade. Additionally, this attempt to deal with practical problems in education takes on theoretical importance when instructional models such as Carroll's are considered. The effect of the amount of feedback and type of grading on students' subsequent performance bear directly on two important components of models.

Method

Participants

Participants were 72 tenth-grade, high-school students enrolled in 4 sections of a required psychology course. The high school is located in a suburban area which falls within the middle to upper-middle socioeconomic strata. In general, the students are achievement oriented and most of them master the basic concepts in the course.

Tasks

As part of their course assignments, students are required to write a research paper on each of three experiments conducted in their class. The experiments are taken from the Psychological Science Curriculum Study's textbook and are described in chapters 3 and 10 of the text. For each paper, students follow an outline which parallels a form established for scientific articles but which is simplified to a level suitable for 10th grade students.

Treatments

There were 6 treatment conditions in this study. Students received a mastery or letter type grade accompanied by one of three levels of corrective feedback. Mastery was defined as a score on the performance measure that was equivalent to a grade of C+ (see below). The first level of corrective feedback, a control condition, was no corrective feedback (NCF). Participants received either a mastery type or *letter* type grade on their papers without *further comment*. The second level was minimum corrective feedback (Min CF). Students who received a grade of non-mastery or a letter grade ≤ C received the following comments on the front page of their papers: "You did a good job on the following research concepts as applied to experiment #1 (list concepts). But you did not master the following concepts (list concepts). Check your textbook--with a little studying, you will (move up to mastery or your grade will improve)." Students who received a grade of mastery or ≥ C+ received the following comments on the front page of their paper: "Good job--in general, you mastered most of the research concepts as applied to experiment #1. You might check again the following concepts since you had some difficulty with them (list concepts)."

The third level was maximum corrective feedback (Max CF). Participants in this condition received the same feedback as those in Min CF with the addition of written comments inserted in their papers. For each key concept, there was a standard written comment explaining the concept.

Instrumentation

Two raters using a concept-by-concept scoring system scored all papers. The raters, using certain criteria, rated each concept as correct or incorrect. For each paper, each rater scored a total of 65 concepts. The student's total score was the average of the two raters' scores. Thus, scores could range from 0-65**.** The interrater reliability was 0.95. The total score earned by a student on a paper was interpreted as a measure of his performance.

In addition, scores on the Wechsler Intelligence Scale for Children (WISC) were available for all participants from the school's records.

Procedures

The 72 participants were randomly assigned to one of 6 groups. These groups were randomly assigned to one of the 6 levels of the treatment variable.

During the first 6 weeks of class, students received instruction in psychology, research, and research writing. At the end of the 7th week, they turned in their first paper. Upon receipt of feedback, students were given class time to rewrite their papers based on the feedback. In addition, instruction continued from week 8 to week 12 when the second paper was turned in. A third paper was submitted in the 18^{th} week. In all sections of the class, instruction was the same.

Results

I. Assignment 1 - CR Design

A. This assignment involves the statistical analysis of the completely randomized design described in the write-up of the experiment. For this assignment, use your entire data set divided into the 6 treatment groups.

B. State the hypotheses about the effects of the various (6) treatments. The hypotheses should take two forms: (1) research hypothesis in which you rank order the treatments in terms of' differences between means and your rationale for this rank ordering and (2) null hypothesis appropriate for the statistical analysis.

C. State the linear model for the CRD ANOVA and the assumptions of the model. Write the linear model in terms of matrices. Specifically provide the appropriate coefficients in the design matrix.

D. For each assumption, present logical and/or empirical evidence which bears upon whether the assumption is tenable. Present descriptive statistics (e.g., means and standard deviations) and statistical, tests (F-tests) when appropriate. For example, if participants were randomly assigned to treatment, the means and variance; should not differ statistically from one group to the next.

E. Report the results of the ANOVA on the participants' total scores on the second paper. Be sure to include descriptive statistics, an ANOVA table, and a level of α chosen in advance. Also report (1) the strength of relation between the treatment and dependent variables (use both η^2 and ω^2 and show your calculation), (2) the power of your statistical test (show calculations), and (3) the number of participants for $1 - \beta = 0.97$ (show work).

F. Discuss your hypothesis in light of these findings.

G. What may you conclude?

Note: Make sure you have 2 copies of each assignment. One of them is submitted to the instructor and the other for your use in subsequent assignments. In this way, any delay in my getting comments back to you will not delay your progress.

II. Assignment 2 - Multiple Comparison Tests

A. This assignment uses and extends your work on the first assignment. With this assignment, you examine the results of the study more thoroughly in terms of differences between means.

B. Formulate and test a set of hypotheses about differences between treatment means using orthogonal contrasts

1. State these hypotheses in written form.
2. Form a table similar to Kirk's Table 4.2–1.
3. Provide evidence that the contrasts you have chosen are orthogonal (show work).
4. Select a level of α in advance.
5. Perform the statistical tests (show work).
6. Discuss the findings as they bear on your hypotheses.

C. Formulate and test a set of' hypotheses about differences between treatment means; you need not restrict yourself to planned orthogonal contrasts.

1. State these hypotheses in written form.
2. Test hypotheses involving pairwise comparisons using:
   a. Tukey's HSD test
   b. Scheffe's method
   c. Newman-Keul's Test
3. Test hypotheses involving complex comparisons using Scheffe's method. One comparison should be the control group (letter grade--no corrective feedback) against all other groups. For this comparison, also use Dunnett's Test.

4. Discuss the results of these tests as they bear on your hypotheses. Where 2 tests give a different result, discuss this too.

D. Explore the 3 most important & priori hypotheses using orthogonal contrasts (you did this in B above) and then test the remaining variability (sums of squares) using a post-hoc method.

III. Assignment 3 - Randomized Block Design

A. In our experiment, participants actually were blocked on level of anxiety (low, medium, high). Specifically, 3 levels of anxiety were selected at random and subjects within blocks were randomly assigned to a treatment. For each treatment, then, 4 Ss are low in anxiety, 4 Ss medium, and 4 Ss high. In order to form a randomized blocks design, select one participant at each level of anxiety in each of the *6* treatment groups. N = 3x6 = 18 with $n_j = 3$. Your data matrix may be presented schematically as:

Treatment

		1	2	3	4	5	6	
	Low	S_1	S_5	S_9	S_{13}	S_{17}	S_{21}	$n_L = 6$
Anxiety	*Medium*	S_{25}	S_{29}	S_{33}	S_{37}	S_{41}	S_{45}	$n_M = 6$
	High	S_{49}	S_{53}	S_{57}	S_{61}	S_{65}	S_{69}	$n_H = 6$
		$n_1 = 3$	$n_2 = 3$	$n_3 = 3$	$n_4 = 3$	$n_5 = 3$	$n_6 = 3$	N = 18

B. State the hypothesis about the effects of the various (6) treatments. (You have already done so in Assignment #1.) But you can change them since this is a “different” experiment.) State hypothesis about the blocking variable. The hypotheses should represent both research hypotheses and null hypotheses appropriate for the statistical analysis. Also, indicate whether you expect a treatments-by-block interaction.

C. State the linear model for the RB ANOVA and the assumptions of the model. Write the linear model in terms of matrices. Specifically provide the appropriate coefficients in the design matrix.

D. For each assumption, present logical and/or empirical evidence which bears upon whether the assumption is tenable. Present descriptive statistics (e.g., means, standard deviations, and correlations) and statistical tests when appropriate. Be sure to show your work.

E. Report the results of the RB ANOVA on the participants’ total scores on the second paper. Be sure to include (1) descriptive statistics, (2) ANOVA summary table, (3) the strength of relation, and (4) an estimate of the variance components.

F. Formulate and test a set of hypotheses about differences between treatment means using orthogonal contrasts. Follow the format used in Assignment #2.

G. Discuss your hypotheses in light of these findings.

H. Compare the results of the RB design with those of the CR design.

IV. Assignment 4 - Completely Randomized Factorial Design 2 ~

A. In this assignment, the completely randomized design of Assignment #1 is recognized as a 2×3 factorial design with 2 levels of grade and 3 levels of corrective feedback. For this assignment use the entire data set.

B. State a set of hypotheses about the effects of the two independent variables and the interaction term. These hypotheses should be stated in research form along with a rationale and in the null form appropriate for the CRF ANOVA

C. State the linear model for the CRF ANOVA and the assumptions of the model. Write the linear model in terms of matrices. Specifically provide the appropriate coefficients in the design matrix.

D. For each assumption, present logical and/or empirical evidence which bears upon whether the assumption is tenable. Present descriptive statistics and statistical tests when appropriate. Be sure to show your work.

E. Report the results of the 2×3 ANOVA. Be sure to include descriptive statistics, ANOVA table, strength of relation (eta^2 and omega^2) and data on the power of your F tests. Finally, if possible, pool the error, and interaction terms. If you do so, provide data in support of this decision.

F. Formulate and test a set of hypotheses about differences between means using either tests of simple main effects, orthogonal contrasts, or post hoc comparisons. Defend the particular method you have chosen.

G. Discuss your hypotheses in light of the findings.

H. Compare the results of the CRF design with those of the CR design.

Assignment 5 – Repeated Measures ANOVA

A. Use both the scores on the initial evaluation of the research paper and the second evaluation of the paper as the repeated measures.

B. Draw a schematic representation of the design of your study. Be sure to define the symbols you used for the various factors. List three advantages of the repeated measures design and three limitations. Evaluate the relevance of the model for psychological research in your area of specialization.

C. Write a linear model for your design and define each term in the model. Be sure to use Greek letters for fixed effects and Roman letters for random effects. Write the linear model in terms of matrices. Specifically provide the appropriate coefficients in the design matrix.

D. State the assumptions of your model. In particular focus on the variance-covariance assumptions. Present empirical data on the variance-covariance matrix assumptions. Use both the Box test and the "eyeball" approach to determine if the assumptions concerning the variance-covariance matrix has been satisfied. Present also a logical strategy for avoiding Box's test in many cases. Why does this strategy especially apply to your design?

E. State the null and alternative hypotheses for this design.

F. Explain why "participants" can be considered a nested variable in certain parts of your design (state which parts) and a crossed variable in other parts (what parts are these?). How does this nesting affect various interactions you get as compared to the completely crossed factorial design? What about the error term(s)? How does this relate to the Between Subjects and Within Subjects parts of your design?

G. Present the ANOVA table for your data analysis. Include source, sum-of-squares, df, mean square F and expected mean squares for your model. Indicate significant F's at a specific α. Indicate why two error terms are needed.

H. Indicate the steps you would follow in examining differences between means. In particular discuss variations in the Est Var (ψ) for comparison of certain means.

I. Compare the results of this analysis with the results of the 3×2 factorial ANOVA used in assignment 4. Indicate similarities and differences.

VI. Assignment 6 – Fractional Factorial, Screening Designs and Analysis of Covariance.

A. Give an explanation as to why the Latin Square covered earlier is similar to the fractional factorial design. Also explain why the fractional factorial design is closely related to confounded factorial designs.
B. What are the assumptions underlying the use of the fractional factorial design?
C. What is the major purpose of screening designs? Explain its relationship with fractional factorial designs.
D. Explain the rationale in using Analysis of Covariance. Give the assumptions involved.
E. What are some of the major problems with using the analysis of covariance?

The actual computations and result section write-up for assignment 6 will be determined at a later date. More information will be given at that time in terms of expectations. It is unlikely that you will do both fractional factorial w screening designs and the analysis of covariance. One will be assigned and the other may be made into an extra credit assignment.

VII. Assignment 7 – Time Series Analysis

A. What are the components of a time series? Give a brief definition of each. How many (which ones) are particularly relevant in psychological research?

B. What are the types of designs that would benefit from the use of time series analysis?

C. What is the major goal of smoothing techniques? Name two methods for smoothing a time series. For the data provided to you, apply both methods of smoothing to the data.

D. Plot each of the smoothed data along with the raw (unsmoothed) time series data. Were the smoothing methods effective in accomplishing what they are supposed to do?

E. Write out the autoregressive forecasting model for the data you are analyzing.

F. Using the first $n-1$ data points, fit the data to the model and get the appropriate regression coefficients. Use this equation to predict the value at time n. How good was the estimated value?

Layout of the Five Basic ANOVA Designs

(a) The completely randomized design (one-way ANOVA) AKA CR Design or AS Design

Independent Variable		
A_1	A_2	A_3
S_1	S_5	S_9
S_2	S_6	S_{10}
S_3	S_7	S_{11}
S_4	S_8	S_{12}
Group 1	Group 2	Group 3

For the table above it would be an A_3S_4 or CR–3 design.

(b) Randomized Block Design (one-way ANOVA). AKA RB Design or SA design.

Independent Variable		
A_1	A_2	A_3
S_1	S_1	S_1
S_2	S_2	S_2
S_3	S_3	S_3
S_4	S_4	S_4
S_5	S_5	S_5
Group 1	Group 1	Group 1

For the table above it would be an S_5A_3 or RB–3 design.

(c) Randomized Factorial Design (Two-way ANOVA)

	Independent Variable 1		
Independent Variable 2	A_1	A_2	A_3
B_1	S_1 S_2 S_3 Group 1	S_7 S_8 S_9 Group 3	S_{13} S_{14} S_{15} Group 5
B_2	S_4 S_5 S_6 Group 2	S_{10} S_{11} S_{12} Group 4	S_{16} S_{17} S_{18} Group 6

(d) Two-way Within-subjects Design (Two-way ANOVA)

	Independent Variable 1		
Independent Variable 2	A_1	A_2	A_3
B_1	S_1 S_2 S_3 Group 1	S_1 S_2 S_3 Group 1	S_1 S_2 S_3 Group 1
B_2	S_1 S_2 S_3 Group 1	S_1 S_2 S_3 Group 1	S_1 S_2 S_3 Group 1

(e) Split-plot, mixed factorial design (Two-way ANOVA)

	Independent Variable 1		
Independent Variable 2	A_1	A_2	A_3
B_1	S_1 S_2 S_3 Group 1	S_1 S_2 S_3 Group 1	S_1 S_2 S_3 Group 1
B_2	S_4 S_5 S_6 Group 2	S_4 S_5 S_6 Group 2	S_4 S_5 S_6 Group 2