**Comm. 516 — Introduction to Data Analysis in Comm.**                      **R 6-9**
Dr. Mike Schmierbach                                   Office hours: WF 1:30-3 & by appt.
mgs15@psu.edu                                           Office: 217 Carnegie
Mobile #: 814-933-9934                                 Office #: 814-865-9582

**Readings:**

Bryman & Cramer, Quantitative Data Analysis with IBM SPSS 17, 18 & 19 (ISBN: 978-0-415-57919-3)

Collier, Using SPSS Syntax: A Beginner's Guide (ISBN: 978-1-4129-2218-0)

**Course overview:** This class serves as an introduction to data analysis techniques commonly employed in the field of communications and in related disciplines. The course will employ a commonly-used statistical package to illustrate concepts (e.g., Statistical Package for the Social Sciences, SPSS), and instruction will be provided on how to employ statistical software to conduct a variety of specific analysis techniques. These techniques will include descriptive statistics, analysis of variance, correlation and regression, and exploratory factor analysis.

**Objectives:** Students who successfully complete assigned coursework should develop the following skills:

- Familiarity with the basic SPSS (briefly called PASW) interface and its core functions, including the ability to effectively use documentation and create syntax
- Ability to identify appropriate statistical procedures for various types of data and analytical goals, and to properly conduct and interpret those procedures
- Increased skill with evaluating presentation of statistical data and with presenting data in written and spoken contexts

**Policies:** My guiding standard is that you are adults, with the rights and responsibilities that go along with that status. In particular, please note the following:

- You are grad students. Act like it. Come to class, ask reasoned questions, contribute to discussion as you are able and allow those around you to learn. In addition, there are some specific considerations for this course:
  - This is a hands-on learning course. While the basic notes are available online, you should expect each session to cover a wide range of material and to present multiple examples and in-class activities. It is very difficult to succeed in this course if you don't come to class, and you should not expect the instructor to provide private tutoring if you are absent. Instead, you are responsible for getting notes and asking specific questions as needed. The class has no attendance policy or points for participation; I expect you to exercise reasonable judgment as to whether you need to miss class. A heads-up regarding both planned and unexpected absences is helpful and polite, but not required.
  - We meet in a computer lab, which obviously offers a number of distractions. Both research and prior experience suggest that students who choose to use the

computers for personal activities (such as checking e-mail or Facebook) not only miss important material, they also distract those near them. As such, non-course activities are not allowed during class (aside from breaks), and students engaging in them may be asked (nicely, I hope) to leave. In addition, College rules on food and drink in the labs are very strict — you are not allowed either. We will have sufficient breaks for eating and drinking if you need to do those things.

- Note to students with disabilities: Penn State welcomes students with disabilities into the University's educational programs. If you have a disability-related need for reasonable academic adjustments in this course, contact the Office for Disability Services, ODS located in room 116 Boucke Building at 814-863-1807(V/TTY). For further information regarding ODS, please visit its website at www.equity.psu.edu/ods/. Instructors should be notified as early in the semester as possible regarding the need for reasonable academic adjustments.
- This syllabus and any other course materials, including grading policy, are subject to change if necessary. I will give you advance notice of any changes.

**Academic integrity:** Academic integrity is the pursuit of scholarly and creative activity in an open, honest and responsible manner, free from fraud and deception, and is an educational objective of the College of Communications and the university. Cheating, including plagiarism, falsification of research data, using the same assignment for more than one class, turning in someone else's work, or passively allowing others to copy your work, will result in academic penalties at the discretion of the instructor, and may result in the grade of "XF' (failed for academic dishonesty) being put on your permanent transcript. In serious cases it could also result in suspension or dismissal from the university. As students studying communication, you should understand and avoid plagiarism (presenting the work of others as your own). A discussion of plagiarism, with examples, can be found at: http://tlt.its.psu.edu/plagiarism/tutorial/. The rules and policies regarding academic integrity should be reviewed by every student, and can be found online at: www.psu.edu/ufs/policies/47-00.html#49-20, and in the College of Communications document, "Academic Integrity Policy and Procedures." Any student with a question about academic integrity or plagiarism is strongly encouraged to discuss it with his or her instructor.

**Special thanks to…:** This course has been taught for many years by Mary Beth Oliver. Her hard work in developing the materials and pedagogical approach has been critical in its prior success. I hope to live up to those prior accomplishments. If so, it will be thanks in part to her generosity in sharing those materials with me; while the bulk of course material I use is my own or drawn from other sources, I have integrated many of her ideas as well. In addition, I want to thank all of you, the students. While I am confident in the material and have put significant time into developing the course, I know that there will be some need for flexibility and experimentation. I thank you all for enrolling in the course and helping me develop it further.

**Grades:** Grades are based on the average of the exams and optional project, weighted as follows:

Exam 1 ................................................. 30%
Exam 2 ................................................. 30%
Take-home exam *or* project................... 30%
In-class final......................................... 10%

*Grading scale:* Final grades are assigned based on point totals using this scale:

$$
\begin{array}{ll}
\text{93-100+} & \text{A} \\
\text{90-92.9} & \text{A-} \\
\text{87-89.9} & \text{B+} \\
\text{83-86.9} & \text{B} \\
\text{80-82.9} & \text{B-} \\
\text{77-79.9} & \text{C+} \\
\text{70-76.9} & \text{C} \\
\text{60-69.9} & \text{D} \\
\text{0-59.9} & \text{F}
\end{array}
$$

*Exams 1&2:* The first two exams will be in-class, open-book exams using SPSS as well as including some questions that involve interpreting and writing about output, results and quantitative research. Both exams are cumulative to that point, but exam 2 will primarily focus on material from the topics after exam 1. You must complete your own work; no consultation with others (other than questions of clarification for the instructor) is allowed. Students should study and review their notes, homework and any review questions. Even though the exam is open book, you won't have enough time to complete it successfully if you are not reasonably familiar with the material.

*Take-home exam/Final project:* For the third exam, students have the option of completing a take-home exam, which requires completing problems based upon a shared data set, or carrying out an independent project.

Exam option — For the exam option, students must complete all work on their own, and will have one week from receiving the data and questions until the work must be submitted. Scoring will be similar to that for any exam, evaluating responses entirely based on correctness. This is a comparatively "high-risk" option, as — like any exam — low scores are a definite possibility, but if you are confident in the material and don't have any outside data you wish to analyze, this is a viable option, and the time required could vary greatly.

Project option — For the project option, students must *individually* identify a relevant set of hypotheses and test those hypotheses using either existing data (including data you personally collected for another course, provided you have written permission from all those involved) or data collected during the semester. Your grade will be based upon your in-class presentation of your findings, as well as a more careful review of your presentation slides and the corresponding data and syntax used to generate your results. More details are provided in a separate direction sheet. Since you have greater control over the process, this is comparatively lower risk compared with the exam, but high scores will require good work. This is a great option if you already have data you want to analyze, know of an existing source that would work for secondary analysis, or even already have a planned project for which you could easily collect data. Most students seriously thinking about academic research in communications should choose this option, as it provides a much more useful outcome. However, since this does require some theory development (and corresponding outside reading) it potentially will involve more time than the

take-home option. Once you have committed to the project option, you cannot shift to the exam option or vice versa.

***In-class final:*** Regardless of the above option chosen, all students will complete an in-class final exam during finals week. This exam is cumulative, closed-book and does not involve any computer work. While our primary focus this semester is on hands-on work with SPSS, it is important that you have sufficient familiarity with the conceptual ideas as well as with reading tables and output that you can answer basic questions about those topics without having to reference notes. I will provide a review sheet to highlight the key concepts and tasks you should expect to be tested on; the exam will be multiple choice.

***Extra credit questions:*** Prior to each exam, you may submit up to three sample practice questions for extra credit toward the relevant exam. Each question should be similar in format to the homework questions (more below), and must include a) data (or draw from data provided in the course) and b) a description of the answer as well as a copy of the output or syntax used to generate that answer. You may submit no more than one question on any given topic. For each appropriate question, you will receive a 2-point bonus on your exam score (out of 100). Questions must be submitted via e-mail no later than midnight the Sunday prior to the exam, allowing me to review and format questions to post for helping the rest of the class prepare for the exam. (Questions for the final are due by the Sunday prior to presentations.) Questions submitted for the final exam will count toward the 30% take-home/presentation aspect, not the in-class component.

**Basic course format and schedule:** For several reasons, this class is organized differently from most. The class is divided into a series of topics, each of which is posted as a self-contained PowerPoint lecture on ANGEL. (Early lectures are already up; later ones will be filled in as we go and as I learn from prior lectures.) The order of topics is set, as numbered below, but the exact date of topics is not determined. We'll cover as many topics each day as we comfortably can; this might be three or even four early on, but some complex days may just address one or two. Exam dates *are* provisionally set, as noted. To give us the necessarily flexibility, some topics, noted with a B after the number, are "bonus" topics. We will cover these as time permits, but we can comfortably drop them without missing the core material of the course.

Each topic includes a series of lecture-oriented slides, which will sometimes be supplemented by direct demonstrations of content in SPSS. At the end of each topic, we will have a set of review questions, which will always include SPSS practice and may also include a few quick self-assessment quiz questions. The gap between topics will also serve as a break, so use that time to work the problems but also to take a couple minutes to stretch, get a snack, or whatever else you may need to do. You should feel free to ask questions for clarification of material *at any point* during a topic lecture; you do not have to wait until the end and indeed I would very much appreciate questions right away to help me recognize unclear or confusing material.

**Homework:** In conjunction with the lectures, I have also posted practice problem sets for each topic. Depending on the complexity of the topic, these may be a few simple questions or just one or two more involved problems. Working these problems is not required and can be done at your own pace — the answers are included in each problem set (after a couple blank pages to avoid

"spoilers"). I strongly encourage you to work as much of the homework as possible, and indeed to look ahead at problems after reviewing the book. Completing these activities along with the in-class questions is the best way to focus your studying and master the material.

**List of topics:** The following list indicates the basic content of each topic. As noted above, topic numbers that include a "B" are "bonus" topics and may not be addressed; you are not responsible for the material if we don't cover it in class, but the notes will be up if you're curious. The position of exams within the topic sequence, and the estimated dates for exams 1 and 2, are indicated. See the notes at the end regarding readings.

1. SPSS basics: Navigating the file types and basic menus, including syntax; data entry and data view management (Ch. 2)

2. Basic descriptive statistics & frequencies (Ch. 2, Ch. 5)

3. Recoding and computing variables (Ch. 3)

4. Basic variable construction and reliability (Ch. 4)

5B. Principle components analysis (Ch. 11)

6. Visual representations of data (Ch. 5)

7. Assessing for normal distributions (Ch. 2, Ch. 5)

8. Chi-square and crosstabs (Ch. 7, Ch. 8)

EXAM — Week 5 (Feb. 9)

9. t-tests (Ch. 7)

10. Basics of one-way ANOVA, including post-hoc tests (Ch. 7)

11. ANOVA with multiple IVs; interpreting interactions (Ch. 9)

12. ANCOVA and other model variations (Ch. 7, Ch. 9)

13. Repeated measures in ANOVA (Ch. 7, Ch. 9)

14B. MANOVA (Ch. 9)

15B. Other variations on ANOVA (Ch. 7, Ch. 9)

EXAM — Week 10 (March 22)

16. Basics of correlation (Ch. 8)

17B. Partial correlation (Ch. 10)

18B. Alternative correlation statistics (Ch. 8)

19. Principles of regression (Ch. 8)

20. Multivariate regression (Ch. 10)

21. Hierarchical regression and interpreting F (Ch. 10)

22B. Alternatives to OLS (Ch. 10)

23. Dummy coding and regression as unifying linear model (Ch. 10).

24B. Testing interactions in regression (Ch. 10).

25. Logistic and other regression variants (Ch. 10).

*Presentations the last scheduled week of class. See separate assignment sheet for details.*

**A note on the readings:** Each topic touches on material from one or more chapters in the Bryman and Cramer book. In general, that book is most useful as a guide *while* you are completing homework and grappling with in-class activities. Unfortunately, its distribution of material and organization do not always match our class plan. The listed chapters are not meant as requirements for a particular unit, but as a reminder of where the textbook tackles that topic. However, I do feel that a preliminary reading of the book, generally in its intended order, would be a helpful first orientation to each topic *prior* to class. By reading ahead, you can see the upcoming principles sketched out and have a sense of how the individual activities develop into a more coherent approach to statistical analysis. Here is my suggested calendar of reading:

Week 2: Chapters 1-5 (This covers all the basics of SPSS for description and reiterates a number of important statistical concepts.)

Week 3: Chapter 11 (This will get you ready for "factor analysis" by the time we cover it.)

Week 4: Chapters 6 and just the sections on chi-square in chapters 7 and 8 (This covers the logic of hypothesis testing, in case you forgot from your methods class, and chi-square – the latter are pretty small sections of those chapters, and this is where the book really breaks from our schedule.)

Week 6: Chapters 7, 9 (This will give you a preview of everything we do with comparing means between groups, including all aspects of ANOVA.)

Week 11: Chapters 8, 10 (And this covers everything about correlation and regression.)

In addition to the Bryman and Cramer book, we also have the Collier book that summarizes key aspects of SPSS syntax. This is *purely* a reference, but having a reference for syntax is very important. For many of the activities and even exams in this class, you will be expected to

include your syntax as part of your work. Good practice in SPSS is always to either work in syntax or at least generate the syntax that was used to produce your results, so anyone could reproduce them later. Careful commenting is also extremely helpful. While we will cover syntax in class, remembering its details can be tricky and exhausting. The Collier book is there to help. I would refer to it whenever you have a question about syntax, using the index to find the topic of interest.