

JEPSON FELLOW PROPOSAL COVER SHEET

Academic Year 2013 — 2014 Application for Funding in Fiscal Year 2014 — 2015

Due Monday, September 9, 2013

Project Title: Segmentation and Factor Models for Enhancing Social Media Impact

Name of Applicant: Christopher Garcia

Department: College of Business

Project Summary (50-75 words, insert in the box below):

The proposed project aims to develop quantitative models that can help organizations enhance the impact of their social media outreach campaigns. Social media outreach will be modeled as a series of interactions between an organization and different market segments. New analytical methods will be developed to enable organizations to identify how different interaction factors impact each segment, improving outreach effectiveness. The project will also result in substantial new teaching materials for multiple courses.

Required Attachments:

1. Description of a Research/Creative Project (following required format, no longer than five pages)
2. Copy of the Applicant's *Curriculum Vitae*
3. One-page Teaching Statement (describing oneself as a teacher and relating the teaching statement to the proposed project)
4. One syllabus from a course taught within the last two years; if desired, one example of supplementary material from that course (such as an assignment description) that is relevant to the project being proposed may be included

Applicant Signature: _____ Date: _____

Chair Signature: _____ Date: _____

Segmentation and Factor Models for Enhancing Social Media Impact



2014-2015 Jepson Fellowship Application

Submitted by Christopher Garcia
Assistant Professor of Quantitative Methods
College of Business
University of Mary Washington

Table of Contents

Project Proposal	4
Curriculum Vitae	8
Teaching Statement.....	12
Sample Syllabus.....	13
Sample Assignment	17

Segmentation and Factor Models for Enhancing Social Media Impact

Project Proposal

Submitted by: Christopher Garcia, Assistant Professor, College of Business

I. Project Description

It is hard to overstate the impact that social media has had in recent years. Social media outlets such as Twitter, Instagram, Facebook, blogs, and many others have ushered in a remarkable new degree of real-time connectedness. Along with these innovations have come many new and exciting opportunities to connect people to organizations in mutually beneficial ways, ranging from identifying and recruiting volunteers for humanitarian causes to connecting customers to businesses that offer products they actually want.

Amidst the constant bombardment of information one experiences when using social media, it is nearly impossible for organizations to reach their intended audience and achieve their desired outreach objectives without a carefully designed interaction experience. Compounding this difficulty is the fact that two different individuals may view the same interaction very differently. A message that one person finds compelling may be much less persuasive to another. A major reason for this is based in the differences in background, stage of life, values, and interests of these two individuals. The challenge faced by the organization is first, to correctly perceive the major differences between these two individuals, and then to tailor an appropriate social media interaction for each of them that emphasizes the alignment of the organization to their individual values, interests, and experiences. Accordingly, the basic underlying problem has two interrelated components. The first component lies in identifying the distinct groups of interested people that may be persuaded to respond to the organization's outreach. This process is called *segmentation*, and the distinct groups of people are referred to as different *segments*. An important aspect of segmentation involves determining the features that unite each distinct segment, and distinguishing these from superficial features. The second major component involves tailoring appropriate messages and interactions to appeal to the interests of each segment. One segment may be interested in environmental issues while another in economic issues. One segment might find a celebrity's Tweets to be persuasive, while another might be more convinced by reading an expert's blog. The message content, the influencer who sends or promotes the interaction, the specific type of social media, and many other aspects are all different types of *factors* that comprise the interaction. A message or interaction on social media may thus be thought of as a mixture of different factors, and the objective is to deliver the right mix of factors to each segment to maximize the likelihood of achieving the desired response.

The proposed project aims to develop quantitative models and supporting experiment designs that can help organizations enhance the impact of their social media outreach campaigns. Specifically, an approach for segmentation modeling will be developed to enable distinct outreach segments to be identified and characterized. Factor models will be developed alongside segmentation that will enable the impact of distinct factors and factor combinations on different segment groups to be quantified. This will enable organizations to make better use of limited resources and be more effective in their outreach. A major goal of this research is also to develop new, relevant, and interesting teaching materials to be used in undergraduate courses in Applied Statistics & Business Research, Foundations and Applications of Data Analytics, and Analytics Application Development.

II. Goals and Specific Objectives

Objective	Resulting Artifacts	Expected Completion
Development of core segmentation and factor models together with supporting experiment designs	Journal article (basic research)	April 15, 2015
Development of a software prototype and source code	Source code (to eventually be made publicly available)	February 15, 2015
Development of new teaching material that puts applied statistics into an interesting and highly relevant context	Teaching cases, Journal article (educational research)	May 31, 2015

III. Potential Impact and Significance

The proposed project will establish a framework through which many different types of organizations can improve the effectiveness of their social media outreach efforts. Organizations will be able to better understand who their key segments are at finer levels of granularity than were previously possible, and also understand how to design interactions with each segment to maximize the chances of obtaining the desired response. This will in turn enable organizations to make better use of their limited resources.

The methodology of this project will entail significant use of statistical modeling, design of experiments, and algorithm development. Students today relate to the world in large part through social media, and most of them have responded to social media outreach campaigns at some point in their lives. One of the major outcomes of this project will be a set of new teaching materials for use in several of the courses taught by the proposing faculty member. These materials will teach students how to apply advanced statistical methods in a context that involves real social media and real-time data. The result will be projects that are challenging, engaging, and also highly relevant. Along the way students will also gain many skills valuable for employment, including proficiency in a statistical programming language and also in several social media tools.

Finally, the project will contribute to the new and growing body of research in social media analytics. Segmentation models have existed for some time within marketing research. In the past these have largely focused on broad demographic attributes, resulting in a relatively small number of segments characterized by a small number of uniting features within each segment. By contrast, this research will combine interactive statistical analysis and experimentation with the aim of *micro-segmentation*, the identification of larger numbers of segments which are smaller in size but are more homogeneous internally. This increases the likelihood that a single type of interaction will achieve the same response among the members of a given segment, leading to improved outreach response. One of the main anticipated scholarly contributions of this research will be a new framework for interactive experimentation and analysis. The segmentation models depend to some extent on the interaction factors, and factor models similarly depend to some extent on the segments. The quantities of data involved in social media coupled with the large number of factors and segments will generally preclude the use of complete experimental designs. As a result, new incomplete designs and experimental sequences that are highly efficient must be developed to enable final factor weightings and segment parameters to be determined. These new generalized designs will represent an important step forward within the body of existing research.

IV. Procedures and Methods

The proposed project will proceed as a sequence of five phases. The final validation of the newly developed methods will be conducted experimentally by comparing them to more conventional broadcasting techniques. These five phases are described below as follows:

Phase 1 (Context Identification):

An appropriate context in which to develop and test the segmentation and factor models must be identified. The ideal context will be an organization planning to use social media for an outreach campaign who is willing to work with the proposing faculty member. An agreement will be sought after to allow the new methods to be tested on the organization's outreach. Should such an organization not be found, a simulation-based context will be developed during phase 3 instead. The simulation will involve generating randomized collections of simulated segments, each with varying degrees of receptiveness to different outreach interaction factors. The segmentation and factor models will then be applied inside a simulation to identify the segments and corresponding factor effects. The models will interact with the simulated segments in the same manner as they would with real-world segments, ensuring a degree of generalizability for the results. Finally, an important output for this phase will be a set of appropriate metrics by which to judge the favorable response rates.

Phase 2 (Development of mathematical models and experiment designs):

This phase will entail development of the core segmentation and factor models together with supporting experiment designs.

Phase 3 (Development of prototype software):

This phase will consist of developing the necessary prototype software. This will entail coding the models themselves as well as the software infrastructure necessary to connect them into their testing context (per phase 1). If the context involves a simulated environment then this will be developed during this phase. In either case, this phase will entail building connection pieces into at least one popular social media technology. This prototype software will be utilized both in the testing as well as in the future teaching materials, with the intent of eventually releasing source code to the public after the research has been published.

Phase 4 (Validation experiment):

A crucial step is the validation of the new models developed. A validating experiment will be performed that compares the response rates of the new model-based approach to that of a control based on a conventional broadcast-like approach. The validating experiment will be designed to generate large data samples for comparison. Standard mean and proportion tests will be performed to test for statistically significant differences between the two resulting response rates, and calculated confidence intervals will be obtained for each. The response rate metrics developed in phase 1 will be used as the basis of comparison in this experiment.

Phase 5 (Development of final artifacts):

This phase will entail the writing of journal papers for disseminating the research as well as the development of resulting teaching materials.

A. Timeline of Phases

Phase	Description	Expected Completion Date
1	Context Identification	10/1/2014
2	Development of mathematical models and experiment designs	11/15/2014
3	Development of prototype software	2/15/2015
4	Validation experiment	4/1/2015
5	Development of final artifacts	5/31/2015

B. Experience and Capabilities of Faculty Member

The proposing faculty member has had significant academic and industrial experience in all areas required for this project. Since 2011 he has had ten research publications that develop quantitative and computational approaches to solve challenging business problems, including five articles in refereed journals with acceptance rates of 22% or below. Additionally, this faculty member is currently serving on the editorial board of the *International Journal of Business Analytics*. Over the past five years he has also provided extensive consulting to clients in diverse industries that include defense, petroleum, finance, and nonprofit. His work as a consultant has involved developing predictive models for companies, designing and analyzing large-scale multinational experiments, developing statistical software, and developing complex optimization models. His relevant experience and capabilities may be seen in detail in the attached curriculum vitae.

C. Facilities and Essential Logistical Requirements

There are no special facilities required for this project. All software developed will be accomplished using freely-available open source programming languages (primarily R and Python). The sole essential logistical requirement will be the identification and acquisition of a testing context in which to conduct this research (per phase 1). Ideally, this will materialize in the form of applying the models developed by this research to an organization's social media outreach campaign. If such an organization cannot be found, a simulated context will be used instead.

V. Anticipated Results

The primary criteria for project success will be the presence of improved response rates obtained by the new modeling approach over the control approach (with statistical significance). Additionally, the following material artifacts are anticipated:

1. A basic research article on the new models and experiment designs developed, to be submitted to a refereed journal
2. Source code for the models, to be released to the public sometime after research is published
3. A set of teaching materials to be used in the proposing faculty member's courses
4. An educational research article on the use of social media analytics as an engaging context in which to teach applied statistics, to be submitted to a refereed educational journal or conference

VI. Dissemination of Results

The results will be disseminated in the form of one journal publication (basic research), one journal or conference proceedings publication (educational research), and a body of model source code made available on the Internet. The teaching materials developed will be used in three of the courses taught by the proposing faculty member (BUAD 259: Applied Statistics & Business Research Methods, BUAD 400: Analytics Application Development, and BUAD 403: Foundations & Applications of Data Analytics). Finally, a public lecture will be arranged on campus during the first weeks of the fall 2015 semester to showcase the results of this research to fellow colleagues at the University of Mary Washington and to the local community.

Christopher J. Garcia
Assistant Professor of Quantitative Methods
College of Business, University of Mary Washington
Curriculum Vitae

1301 College Avenue
Fredericksburg, VA 22401

cgarcia@umw.edu
Telephone: (540) 654-1456

Updated 04-September-2013

Areas of Expertise

- Applied Statistics, Data Mining, and Predictive Analytics
- Optimization
- Design of Experiments
- Complex Scheduling, Planning, and Resource Allocation Problems
- Supply Chain and Logistics Management

Education

Ph.D. Engineering Management, 2010
Old Dominion University
Dissertation Title: *Optimization Models and Algorithms for Spatial Scheduling*

M.S. Operations Research, 2008
Florida Institute of Technology

M.S. Computer Science, 2004
Nova Southeastern University

B.S. Computer Science/Mathematics, 2001
Old Dominion University

Academic Appointments

August 2011 – Present *Assistant Professor* *University of Mary Washington*

January – May 2011 *Adjunct Assistant Professor* *Old Dominion University*

Awards and Honors

- Faculty Award in Engineering Management & Systems Engineering, Old Dominion University, 2011
- National Defense Industrial Association (NDIA) Scholar, 2009
- SREB Doctoral Scholar (Virginia), 2009
- Engineering Doctoral Fellowship, Old Dominion University, 2009
- GAANN Fellowship, Old Dominion University 2001
- Phi Kappa Phi Honor Society
- Golden Key International Honour Society

Publications

Refereed Journal Articles:

Koksalmis, E., **Garcia, C.**, and Rabadi, G. (*Forthcoming*), “The optimal exam experience: A timetabling approach to prevent student cheating and fatigue”, to appear in the *International Journal of Operational Research*.

Garcia, C. (2013), “Reverse logistics sourcing models with multiproduct reliability constraints”, *International Journal of Modelling in Operations Management*, Vol. 3 No. 2, pp. 134-148.

Garcia, C. and Rabadi, G. (2013), “Exact and approximate methods for parallel multiple-area spatial scheduling with release times”, *OR Spectrum*, Vol. 35 No. 3 pp. 639-657.

Garcia, C. and Rabadi, G. (2011), “Optimising coordinated multi-team volunteer scheduling: a combined constraint programming and goal programming approach”, *International Journal of Industrial and Systems Engineering*, Vol. 9, No. 2, pp. 183-194.

Garcia, C. and Rabadi, G. (2011), “A Meta-RaPS Algorithm for spatial scheduling with release times”, *International Journal of Planning and Scheduling*, Vol. 1, No. 1 / 2, pp. 19-31.

Professional Journal Articles:

Pande, J., **Garcia, C.**, Pant, D. (2013), “Optimal component selection for component based software development using pliability metric”, *Software Engineering Notes*, Vol. 38, No. 1.

Refereed Conference Proceedings:

Garcia, C. (*Forthcoming*) “Demystifying MapReduce”, to appear in proceedings of the *Complex Adaptive Systems Conference 2013*, November 13-15, 2013, Baltimore, MD.

Henderson, D., Lapke, M., and **Garcia, C.** (2013), “SQL Injection: A Demonstration and Implications for Accounting Students”, *American Accounting Association Information Systems Section Midyear Meeting and AIS New Scholar Consortium*, Ft. Lauderdale, FL – Nominated for Best Paper of Conference

Garcia, C. (2012), “Real Estate Development Planning Using 0-1 Integer Programming”, in proceedings of the *33rd International Annual Conference of the American Society for Engineering Management*, October 17-20, 2012, Virginia Beach, VA

Garcia, C. and Rabadi, G. (2012), “Robust models for volunteer scheduling”, in proceedings of the annual *Industrial and Systems Engineering Research Conference (2012)*.

Garcia, C. and Goodrich, M (2010), “Potential Effects of Health Care Policy Decisions on Physician Availability”, proceedings of *Modeling and Simulation (MODSIM) World 2010 Conference*, October 13-15 2010, Hampton, VA.

Garcia, C. (2009), “A Decision Model for Minimizing Critical Component Costs While Achieving a Target Quality of Service”, proceedings of *INCOSE Conference on Decision Analysis and Its Applications to Systems Engineering*, November 17-18 2009, Newport News, VA

Garcia, C. and Rabadi, G. (2009), “An Optimization Model for Scheduling Problems with Two-Dimensional Spatial Resource Constraints”, proceedings of *Modeling and Simulation (MODSIM) World 2009 Conference*, October 14-16 2009, Virginia Beach, VA

Garcia, C. and Griffith, T. (2005), “A Composable Behavior Modeling System for Rapidly Constructing Human Behaviors”, proceedings of *I/ITSEC Annual Conference*, Nov. 28 – Dec. 1 2005, Orlando, FL

Papers Currently Under Review:

Garcia, C. “Effective methods for combined project selection and production planning with due-windows under limited inventory capacity”, currently under review by the *Journal of Combinatorial Optimization*.

Consulting Experience

2013 *Garden Club of Virginia*

- Supervised a team of 8 undergraduates to design and administer surveys and analyze results in support of GCV’s strategic planning
- Pro-bono service project; resulted in financial gift made to UMW foundation.

2012-2013 *Quarles Petroleum*

- Developed a statistical model to predict sales performance of fueling stations
- Model to be incorporated into a software package to assess future sites for acquisition

2012 *Capital One*

- Validation of mortgage loss and severity models for Capital One’s ARM products
- Independently worked out mathematics of loss models and implemented a model emulator
- Emulator results used to benchmark current models

2011 *US Joint Forces Command – Comprehensive Approach to Building Partnerships*

- Developed conceptual model to support multi-agency collaborative planning. This model provided an infrastructure for organizations to collectively formulate goals and appropriately align efforts based on capabilities and resources.
- Developed a mathematical optimization model to enable the optimal allocation of multi-agency efforts to objectives based on agency capabilities, capacities, strategic interests, and resources.
- Developed the Data Collection and Analysis Plan

2011 *US Joint Forces Command – Adaptive Logistics Network*

- Provided consulting for the design and analysis of a large-scale multinational defense experiment
- Developed the analytical methodology for the experiments
- Developed custom statistical software for integration and analysis of experimental data

2010 *South Hampton Roads PCA Network (Nonprofit Organization)*

- Developed software to solve a complex volunteer scheduling problem. This problem involved scheduling large numbers of volunteers across multiple teams while respecting numerous constraints including volunteer availability and preferences, work frequency preferences, and fairness maximization.

2009 Industrial/Organizational Psychologist (DOD)

- Developed a predictive model of virtual (distributed) team interactions. The result was a statistical model that predicts peer-to-peer and peer-to-leader interactions based on a set of attributes known for each team member.

Professional Service

- Editorial board member on the *International Journal of Business Analytics* (2012-2015)
- President of Richmond-Tidewater INFORMS Chapter, (2011-2012)
- Reviewer for the *International Journal of Production Research* (2 papers)
- Reviewer for the *Journal of Combinatorial Optimization* (1 paper)
- Reviewer for the *Information Processing Letters* (1 paper)
- Reviewer for the *International Journal of Global Management Studies* (3 papers)

Industrial Experience

<i>2/2010 – Present</i>	<i>Owner/ Principal</i>	<i>Blueshift Analytics LLC.</i>
<i>3/2010 – 8/2011</i>	<i>Operations Research Analyst</i>	<i>MYMIC, LLC.</i>
<i>3/2008 – 2/2010</i>	<i>Senior Software Engineer</i>	<i>Open Roads Consulting, Inc.</i>
<i>5/2005 – 1/2008</i>	<i>Senior Software Developer</i>	<i>IntraNexus, Inc.</i>
<i>1/2002 – 5/2005</i>	<i>Software Engineer</i>	<i>Alion Science and Technology</i>

Security Clearance

Active Top Secret Clearance (current through 2016)

Segmentation and Factor Models for Enhancing Social Media Impact

Teaching Statement

Submitted by: Christopher Garcia, Assistant Professor, College of Business

It is my belief that of all the roles and responsibilities of a university professor, teaching is the most valuable and important. Effective teaching and student learning is the primary mission of the university. In today's technological and highly competitive world, it is not sufficient to merely impart knowledge or mastery of a specific discipline. Information is abundantly available, and fewer students are entering careers in the disciplines in which they major in college. Today change is constant, and the skills that students need most are the ability to quickly adapt to new paradigms, the ability to think critically and handle significant amounts of complexity, and the ability to communicate professionally and effectively. The primary guiding principle for all the courses I teach is thus to foster these skills to the maximal degree possible. Accordingly, all of my teaching material and graded work are intentionally designed to be heavily loaded in these factors. To this end, all of my exams consist entirely of problem-solving exercises, and I require significant amounts of written work in all of my courses. One of the most important ways I try to impart these skills, however, is through projects that require solving problems that are realistic in nature and large in scope.

As a teacher of quantitative methods in today's high-tech world, there are abundant opportunities to illustrate how material students are learning in the classroom can be used in the real world. I have found that applying these approaches to real-world problems takes significant additional problem-solving and creativity beyond the straightforward application of methods generally required in textbook problems. Real-world applications are also inherently more interesting to students and result in a much more engaging learning experience. Accordingly, I draw heavily on my research and consulting experience to help students see the broader context of what they are learning. I am also very intentional about incorporating real-world projects and examples into my teaching. During the 2011-2012 year I proposed a change to the BUAD 259 course (Applied Statistics & Business Research Methods) that would enable students to gain experiential learning credit from selecting a suitable research project on behalf of an off-campus corporate sponsor. Since this change was approved the majority of students have selected this option for their semester projects. This has resulted in projects that are much more impactful and substantive. During the spring 2013 semester I selected and supervised a team of eight undergraduate students to design, administer, and analyze the results of a large-scale survey project for the Garden Club of Virginia to support their long-range strategic planning. These serve as examples of how I aim to incorporate larger and more realistic projects into the students' learning.

Social media analytics provides another interesting and fertile ground in which to illustrate the applicability of advanced statistical methods. Last fall I proposed an honors section of BUAD 259 to be offered with a significantly enhanced section on computational statistics. A major part involves using the R statistical programming language to perform text mining and sentiment analysis on Twitter. The new honors section is currently being taught for the first time, and students are expressing enthusiastic interest in applying statistical methods to social media. The proposed Jepson Fellowship project will result in teaching materials that will take this type of social media analytics in the classroom to a new level. While some of these materials will be designed for BUAD 259, more advanced materials will also be developed for courses required in UMW's new Data Science minor including BUAD 400 (Analytics Application Development) and BUAD 403 (Foundations & Applications of Data Analytics).

Sample Syllabus



BUAD 259: Applied Statistics and Business Research Spring 2013

Instructor:	Christopher J. Garcia, Ph.D.
Office:	George Washington Hall 215-B
Telephone:	(540) 654-1456
Email:	cgarcia@umw.edu
Office hours:	MW 3:00 pm-5:00 p.m., F 4:00-5:00 p.m. and by appointment
Prerequisite:	Math 200 or similar statistics course
NOTE:	This is a writing-intensive course

Course Description:

Most businesses will at some point encounter situations where current strategies or policies are no longer effective. This may be caused by changing economic conditions, changes in consumer bases and preferences, changes in employee needs, and many others. The underlying causes of business problems are often not readily observable and instead manifest through symptoms such as decreasing revenues or unacceptable employee turnover. In such situations, an understanding of the underlying phenomena is critical to making future decisions. To address this, managers need to have both an understanding of how to conduct rigorous research and the ability to understand the research of others. This course will provide students with an understanding of empirical business research methods along with associated quantitative and computational approaches, as well as opportunities to apply these methods to real business problems.

Learning Outcomes:

- 1) Students are able to investigate and decompose an ambiguous business problem into a clearly articulated problem statement, and from this develop a corresponding set of testable hypotheses
- 2) Students are able to design a program of research to definitively answer a non-trivial business question
- 3) Students are able to design scientifically valid surveys for use in business research
- 4) Students are able to design experiments appropriate to confirming or rejecting a set of given hypotheses
- 5) Students gain experience in applying research methods to investigate a real world management problem
- 6) Students are able to determine and apply appropriate statistical methods for hypothesis testing and prediction
- 7) Students are able to use a modern statistical computing environment and write their own programs using a statistical programming language

Additionally, this course is a writing-intensive course and has the following writing-intensive learning outcomes:

- 8) Students are able to write a business research plan and research report
- 9) Students will have completed multiple writing assignments centered on business research
- 10) Students will have revised and improved their written work based on instructor critique

Text: Zikmund, W. G. et. al (2013) Business Research Methods (9th edition), Thomson/Southwestern, Cincinnati

Required Software:

Microsoft Word and Excel

R Statistical Package (freely downloadable at <http://www.r-project.org/>)

Grading and Assigned Work:

Exam 1 (20%), Exam 2 (30%), Computational Assignment (25%), Research Project (25%)

Various homework assignments will be given to provide the opportunity to solve problems and enable students to assess their mastery of course material. Homework will be ungraded; however, it is essential that students complete assigned homework to gain the necessary problem-solving skills. There is a strong correlation between homework completion and exam performance.

For final grades a total grade of 93.34 and above receives an “A”, 90.00 to 93.33 is an “A-“, 87.34 to 89.99 is a “B+”, 83.34 to 87.33 is a “B”, 80.00 to 83.33 is a “B-“, 77.34 to 79.99 is a “C+”, 73.34 to 77.33 is a “C”, 70.00 to 73.33 is a “C-“, 67.34 to 69.99 is a “D+”, 60.00 to 67.33 is a “D”, Below 60 is an “F”. For Pass/Fail grading, a grade equivalent to a C- or higher is required to pass.

Attendance and Late Work:

Class attendance is necessary to succeed in this course. However, there is no formal attendance policy. It is the student’s responsibility to stay current on class material. Late work will be penalized by one-half letter grade per day late unless prior arrangements are made with the instructor. Apart from exceptional circumstances or prior arrangements made with the instructor, missed exams will receive a grade of 0.

Academic Integrity:

The University of Mary Washington does not tolerate academic dishonesty in any form. Penalties for cheating on exams or any other assignments in this course may include course failure and suspension or expulsion from the university. It is understood that all material submitted will be pledged in accordance with the Honor Code of UMW. Some assignments will call for collaborative effort. Details regarding appropriate “mutual assistance” will be described when the assignment is given. Otherwise the tenets of individual effort per the honor pledge are in effect.

Overview of Assignments:

- 1) **Research Project.** -- You will work as a member of a team to conduct a semester-long research project. This project will culminate in a final report, and you will submit parts of the report throughout the semester. Instructor feedback will be provided, and groups are expected to improve their research based on the feedback given. The proposal, instrument design, data collection and analysis are all group level tasks. While these drafts will not receive a grade, it is a requirement that you turn in a draft of each section. Each section will be submitted by your group. The project will investigate some facet of the business world and employ the methods discussed in the class or in your text.
 - a. **Research Proposal** -- The research proposal will include an initial draft of the introduction and methods sections. One proposal per research team will be submitted for approval and comment only. No grade will be given on the proposal. However, it must be “approved” before the research can be done. My comments should be used to improve the quality of the section and integrated into the final paper. I will schedule time to meet with each team to discuss their proposal and research methods.
 - b. **Research Report** -- Specifications for each section will be discussed in class and described on the class web site. Drafts of the introduction, methods, results and conclusions sections will be turned in for editing and comments prior to final submission. No grade will be given on these drafts, but turning them in is mandatory. The final paper should not exceed 30 pages with tables, graphs and references. Below is an outline of a typical report:

- A) Title page (1 page)
- B) Executive Summary – (1 page)
- C) Table of contents (1 page)
- D) Introduction, which includes (6 to 12 pages):
 - 1) a general overview of the problem requiring the research,
 - 2) a brief literature review
 - 3) a statement of the specific research questions or hypotheses
- E) Methods used to answer the question (2 to 4 pages)
- F) Results (2 to 4 pages)
 - 1) Statistics regarding the demographics of your sample
 - 2) Brief presentations of all analyses you conducted
- G) Conclusions (2 to 4 pages)
 - 1) summary of results in terms of your hypotheses
 - 2) overall evaluation of your research: strengths and limitations
 - 3) recommendations for future research
- H) References (1 to 2 pages)
- I) Appendices

You will find it helpful to review Chapter 25 of your book, and to review the following website when preparing your reports: <http://www.library.cornell.edu/resrch/citmanage/apa>

- 2) **Computational Assignment** -- A significant aspect of this course will involve the application of sophisticated statistical methods. The computational assignment will consist of several problems to be solved using Excel and R.

Experiential Learning Component: Students wishing to obtain experiential learning credit may do so in this course by undertaking an appropriate project as their semester research project. In order to qualify for credit, proposed research projects must immerse the students in research on a specific business problem originating outside of UMW. In general, this requires undertaking a research project on behalf of an outside company sponsor. Students wishing to obtain experiential learning credit must gain approval from instructor based on their research proposal submission during the fourth week and must sign a contract outlining the requirements for experiential learning for their proposed project. If approval is granted students will be permitted to register for a 0-credit experiential learning course.

Extra Credit:

This course involves problem solving and may at times require extensive analytical and out-of-the-box thinking. Short in-class extra credit quizzes will be given on occasion and will add a few points to the student's final grade. Quiz content will consist of puzzles - topics that do not directly involve course material, but provide opportunities to exercise problem solving skills. Study or prior preparation will never be required for these quizzes.

Disability Accommodations:

Accommodations will be made as needed as coordinated through the Office of Disability Resources.

Tentative Schedule:

Week	Start Date	Segment	Topics	Reading	Graded Assignments Due
1	1/14/2013	Business Research Fundamentals	Introduction to Business Research, Business Research Process, Ethics in Research	C1, C4, C5	
2	1/21/2013		Theory Building, Problem Formulation	C3, C6	
3	1/28/2013		Report Writing & Library Resources, Measurement & Scaling, Reliability & Validity	C25, C13	
4	2/4/2013		Attitude Measurement, Questionnaire Design, Experimental Research	C14, C15, C12	
5	2/11/2013		Experimental Research (cont.), Exam 1 Review	C12	Exam 1
6	2/18/2013	Descriptive & Inferential Statistics	Review of Elementary Statistics, Descriptive Statistics	C12	
7	2/25/2013		Univariate Hypothesis Testing & Statistical Tests	C17, C20	
8	3/4/2013		Spring Break		
9	3/11/2013		Bivariate Hypothesis Testing & Statistical Tests	C21	
10	3/18/2013		Bivariate Hypothesis Testing & Statistical Tests (cont.), Covariance & Correlation	C22	
11	3/25/2013	Computational Statistics	Exam 2 Review, Introduction to R		Exam 2
12	4/1/2013		R Programming		
13	4/8/2013		R Programming, ANOVA,	C23	
14	4/15/2013		Linear Regression, Discriminant Analysis		
15	4/22/2013		Clustering, Factor Analysis		Final Research Paper
16	4/29/2013		Final Exam Week		Computational Assignment

Sample Assignment

BUAD 259 Computational Assignment

Spring 2013

Instructor: Chris Garcia

Instructions:

1. This assignment is to be completed in groups. You will work with your research team to complete this assignment.
2. One assignment should be submitted per group by the team lead.
3. You **MUST** list each group member's name inside **EVERY** submitted file. In R code, make sure there are appropriate comment lines with **ALL** group members' names.
4. You **MUST** make sure **ALL** files are properly named. R programs will be graded automatically by another R program, so it is imperative to follow this rule to a tee.
5. All R functions that you are asked to write **MUST** be named exactly as specified and **MUST** take the arguments exactly as specified and in the specified order.
6. Please comment your R code – explain your reasoning and why you are doing what you are doing within comments (comments come after the # sign in R).
7. This is due on Wednesday **May 1** at midnight. Please treat this as an **INFLEXIBLE DEADLINE**. Submit all files in Canvas.
8. **YOU MAY NOT RECEIVE HELP FROM ANY STUDENTS OUTSIDE YOUR GROUP.**

I. Programming Exercises in R

Please submit an R file containing functions to solve the following programming exercises. The file should be named according to the following convention: <team lead's last name>.r. For example, if I a team lead my R file would be named "garcia.r". You may define as many auxiliary procedures as needed, but your file must contain the requested function definitions.

1. Non-Parametric Percentiles

Unlike the type of statistics we have worked with in class (which typically assumes a normal distribution), non-parametric statistics makes no assumption about the underlying distribution of data. Define a function called `percentileFinder` that takes two arguments 1) a vector `V`, and 2) a percentile `P` (as a decimal from 0 to 1) and returns the number at the `P` percentile based on the numbers in `V`. For example, suppose we specify `P = 0.9`. Then we need to find the number in `V` that is greater than or equal to 90% of all numbers in `V`. Your method should not assume any specific statistical distribution for `V` (it should work for any distribution). This function should have the following form: `percentileFinder(V, P)`. Here are some examples of how the function will be called:

```
> vec <- c(5,4,3,2,1,10,9,8,7,6) # Notice this contains the numbers 1 to 10
> percentileFinder(vec, 0.9) # Should return 9
> percentileFinder(vec, 0.2) # Should return 2
```

Hints: 1) Common sense will serve you well on this problem, and 2) consider the `sort` function and how it might help. You might want to try `sort(c(4,1,3,2))` and `sort(c(8,2,6,10,4))` in R and see what happens.

2. Estimating a Proportion Error

In the chapter 17 homework problems (problem 13, specifically) you were asked to estimate a proportion for a population similar to the way we did for estimating a population mean. Write a function "estimateProportionError" that takes 3 elements: 1) `n` – the number of sample elements, 2) `p` – the probability of a "success", and 3) `c` – the required confidence level. The function should have the following form: `estimateProportionError(n, p, c)`. This function will calculate and return the error E within the specified confidence c based on the probability of success p so that the following relationship holds:

$$P_i = p \pm E \text{ with } c \text{ confidence}$$

Here is an example of how the function should be called:

```
> estimateProportionError(500, 0.6, 0.95)
# Should return this number: 0.042941448508405
# In other words,  $P_i = 0.6 \pm 0.042941448508405$  with 0.95 confidence
```

Hints: 1) Use the formulas on pages 435-436, 2) look at the `estimatePopulationMean` function and consider using the `Zc` function in the "basic_functions.r" that we made.

3. Number Sequences

Consider the following sequence: 1,2,3,5,8,13,21, ... Can you guess which one comes next? It is 34, which is 13 + 21. Here, the n th number in the sequence = sequence number $(n - 1)$ + sequence number $(n - 2)$. Write a function called "nseq" that takes three numbers: 1) first – the first number of the sequence, 2) second – the second number of the sequence, and 3) `n` – the sequence number of the desired term to be returned. The function should then return the n th term according to this rule. The function should have the following form: `nseq(first, second, n)`. Here is an example of how the function should be called:

```
> nseq(1, 2, 5)
```

```
# Should return 8 - the fifth term in the sequence, where the first and second
# terms are 1 and 2, respectively
```

Hint: This is very easy to do using recursion and also easy using iteration ;)

INSTRUCTIONS FOR PARTS II and III: Please put answers to parts II and III below in a MS Word file called <team lead's last name>.doc/docx. For example, if I am a team lead I would name my file "garcia.docx".

II. Analysis of Variance

Assume you own a company that sells lawn equipment. You have multiple small locations and want to find a way to motivate your employees to sell. In an effort to accomplish this you set up a small experiment by offering different incentive packages to different locations to see which (if any) work. Prior to this experiment, monthly sales per employee across all locations were roughly the same. Please find the included spreadsheet "incentives.csv" – open it and look through it.

1. *Problem Understanding:* From looking through the data file, please identify: 1) the dependent variable, 2) each independent variable, and 3) the different levels for each independent variable
2. *ANOVA in R:* Use R to run an ANOVA on this data. Please copy-and-paste the ANOVA summary output into your MS Word file.
3. *ANOVA Interpretation:* Please answer the following questions:
 - a. Which main effects are significant?
 - b. Which interactions are significant?
 - c. It costs money to implement incentives beyond the basics, so you don't just want to offer incentives unless they are proven to work. Which incentive combination(s) would you recommend to maximize sales performance and why? (Answer in **NO MORE** than 3 sentences)

III. Multiple Regression

Assume you own a software company. When you bring on new employees your hiring process is 1) to give them an technical aptitude test, 2) bring them in for an interview (conducted by a team of three hiring managers), and if preliminarily hired, 3) put them through a 6-week training course. At the end of the course you have the option to let the employee go with no further costs or keep them on board. You keep a record of all your employees, including their interview score (an average score from 1-5 based on three hiring manager's input from the interview), their aptitude test score, the number of training classes they missed, and their annual performance score. These records are found in the included file "employee-perf.csv". You are hiring and want to identify good candidates before having to keep them on the payroll for a full year (in order to get a performance review).

1. *Conceptual Model Building:* Show a multiple regression model (conceptual) that may be used to predict employee performance based on their hiring and training data. If you use variable names (like X, Y, Z, etc.) please be sure to state what each variable name denotes.
2. *Model Implementation in R:* Implement your model in R using the "employee-perf.csv" data file. Answer the following questions:

- a. Does your model have good validity? Why or why not?
- b. How much variance of an employee's annual performance does your model account for?

3. *Prediction:* Suppose you have the following candidates that have just made it through the training course and you can either retain them or let them go. Their information is as follows:

Candidate	Aptitude Test Score	Interview Score	Missed Training Classes
Steve Jobs	84	3.55	6
Bill Gates	37	4.72	1
Sergey Brin	86	4.61	3

- a. What is the predicted annual performance score for each of these candidates?
 - b. What is the predicted range of annual performance scores for each of these candidates with 99% confidence?
 - c. Who (if anyone) would you hire? Who (if anyone) would you let go? Why?
4. *Insight:* By examining different aspects of the model, what (if any) cost-cutting policies or changes should you consider? Why or why not?