Biostatistics and Experimental Design

COURSE DESCRIPTION

This is an introductory course in experimental design and data analysis designed to encourage an understanding and appreciation of the role of experimentation, hypothesis testing, and data analysis in biology. It will emphasize principles of experimental design, methods of data collection, exploratory data analysis, and the use of graphical and statistical tools commonly used by biologists to analyze data.

INSTRUCTOR

Dr. Ann Throckmorton, Professor of Biology

Office:	311 Hoyt Science Center
Phone:	724-946-7209
e-mail:	athrock@westminster.edu
Home Page:	www.westminster.edu/staff/athrock
Office hours:	11:30-12:30 Monday 12:50- 1:50 Wednesday 10:30-11:30 Friday *or by appointment*

LECTURE

9:20 - 10:50, Tuesday/Thursday Patterson Computer Lab (TC 214)

Attendance in lecture is expected. You will not be graded on attendance except indirectly through your grades on exams and assignments. Because your success in this course is strongly dependent on your presence in class and your participation, you should make an effort to be present at all class sessions. If you know ahead of time that you will miss class, let me know so we can make arrangements for you to attend the other section. Absence may be excused for personal emergencies or health-related problems. If you miss class, it is your responsibility to contact me and to obtain lecture notes and assignments that were given during your absence.

PURPOSE OF THIS COURSE

Statistics is a science that involves collecting, organizing, summarizing, analyzing and presenting data to draw conclusions and make predictions. This course is designed to demonstrate how statistical reasoning is used in the biological sciences and how appropriate experimental design and data analysis can be used to discern patterns in living systems and to predict how those systems will react in different situations.

COURSE OUTCOMES

Outcomes related to experimental design

- to encourage an understanding and appreciation of the role of experimentation, hypothesis testing, and data analysis in biology
- to help students gain facility in designing scientific experiments and studies incorporating treatments, controls, replication, randomization, experimental units, and blocking
- to help students become aware of common pitfalls in experimental design including lack of power, poor controls, confounded effects, and pseudoreplication
- to demonstrate why researchers must include statistical planning in research design and to present statistical principles important to the design and interpretation of experiments

Outcomes related to analysis

- to present basic information about probability and probability distributions
- to provide students with the background needed to select appropriate methods of statistical analysis commonly used in biology, including parametric and nonparametric tests
- to teach students to correctly analyze experimental data and assess the statistical significance of their results in light of the calculated statistics
- to teach the students to used common statistical packages to enter data, graph and analyze data, test hypotheses, and interpret results from simple samples and experiments
- to encourage students to critique data analyses conducted by others and assess the validity of statistical analyses in the scientific literature

Outcomes related to presentation

- to demonstrate how scientific data are summarized graphically and in tables
- to give students experience in using common graphing programs to produce effective graphs
- to show how statistical results are presented in the scientific literature and to give students experience in interpreting and evaluating scientific graphs

The ultimate goal of this course, along with the rest of your education at Westminster College, is to make you a better scientist, someone who is capable of making well-informed decisions about current and future scientific discoveries and how they relate to human culture and other living things.

METHODS OF INSTRUCTION

1. **Lectures and discussion**: These will follow the schedule printed below. I expect you to attend class, pay attention, and participate actively in discussions by answering questions, asking questions, and making comments. Always bring your textbook to lecture.
2. **Reading**: The textbook provides an excellent overview of experimental design and data analysis for biologists. You will be responsible for reading the material before class so that lecture time can be spent covering essential information from the book that warrants repetition and close reading. After material has been presented in lecture you will use that information in a problem-solving session involving analysis and/or presentation of data. There will be time in each lecture to ask questions about the reading and problems that have been assigned.
3. **Exams**: Periodic lecture exams will allow you to assess your progress in learning the information presented in the class. There will be four exams in this course. The final exam will be comprehensive. You should expect multiple choice, short answer, and essay questions on the exams, as well as practical exercises in which you will demonstrate your ability to analyze, graph, and present data.
4. **Quizzes:** We will occasionally have quizzes designed to evaluate your understanding of material that we have recently covered and your ability to use techniques that have been introduced. If necessary, we will have quizzes at the beginning of class on material to be covered that day.
5. **Assignments**: Each lecture topic will be accompanied by an assignment designed to give you experience in gathering, analyzing, and/or presenting data. Many of those assignments will give students experience applying ideas from lecture and using statistical software to graph data and perform statistical tests. Data for the exercises in the textbook can be found on the CD that came with the book.

GRADING

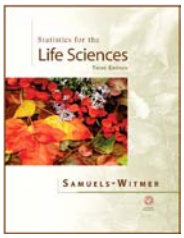
Grades will be based on exams, quizzes, and assignments, weighted as follows:

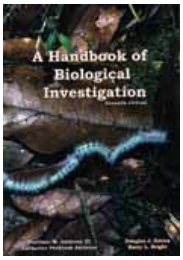
Exams (three)	= 30% of final grade
Final comprehensive exam	= 20% of final grade
Quizzes	= 20% of final grade
Assignments	= 30% of final grade

Your final grade in the course will be based on the following scale:

Above 93%: A	87% - 90%: B+	77% - 80%: C+	67% - 70%: D+	below 60%: F
90% - 93%: A-	83% - 87%: B	73% - 77%: C	63% - 67%: D	
	80% - 83%: B-	70% - 73%: C-	60% - 63%: D-	

REQUIRED MATERIALS





Textbooks: Statistics for the life sciences by M. L. Samuels and J. A. Witmer. 3rd edition, © 2003. Pearson Education, Inc.

A handbook of biological investigation, 7/e, by H. W. Ambrose III, et al., © 2007, Hunter Textbooks, Inc. (an older edition will be fine)

You may also want to have a calculator and a USB flash drive that you can use to store, organize, and transport the many computer files you will create and use in the course.

POLICY ON EXAMS AND ASSIGNMENTS

All assignments must be turned in on time unless you are absent the day that the assignment is due and have a valid excuse. Valid excuses include such things as serious illness or injury and personal or family emergencies. Points will be subtracted from assignments that are turned in late. Similarly, you must take all quizzes and exams when they are given in class, unless you have a valid excuse and have spoken to me before the quiz or exam is given.

You may turn in assignments in three ways:

1. **hard copy**: only for assignments that cannot be handed in electronically. Hand the paper to me, slide it under my office door, or give it to someone to deliver. Do not use campus mail.
2. **in the Assignments folder on the course r-drive**: if you save a file to the r-drive, the name of the file must contain your name and a description of what it contains (e.g., "Smith, Assignment 5"). You must save the file to another drive, then save it to the r-drive. If you try to save directly to the r-drive, the network will only write a blank temporary file and you will lose all of your work. Once you have saved something to the Assignments folder you will be unable to retrieve it, open it, or delete it.
3. **as an e-mail attachment:** Again, the name of the file must contain your name and a description of what it contains. You can find out if I have received your messages by looking in the Sent Items folder in your mailbox.

All electronic assignments will be returned to you in your personal folder inside the Assignments folder of the r-drive. Only you and I have access to that folder, so your work and the grades that you receive will be absolutely confidential.

ACADEMIC INTEGRITY

Academic integrity is central to the purpose and pursuit of any academic community. In this class, I expect you to adhere to the principles of academic integrity stated in the Westminster College handbook and to maintain the highest standards of academic honesty and integrity, in keeping with the philosophy and purposes of the College.

> "Academic dishonesty is a profound violation of this expected code of behavior. It can take several forms, including, but not limited to, plagiarism, cheating, purposely altering the work of another (without that person's permission), misrepresentation of attendance in class or at a College event, misrepresentation of work, facts or experimental results, unauthorized use of or intentional intrusion into another's computer files and/or programs, intentional damage to a computer system, unauthorized use of library materials and privileges, or engaging in any activity which attempts to alter or harm another's academic standing."

I encourage you to work together and discuss your assignments with other students, but all material that you turn in must be your own work. If you violate the Academic Integrity Policy, you will receive a score of zero for that assignment and a written report will be sent to the Dean of Academic Affairs. More than one violation may result in an F for the course.

TENTATIVE SCHEDULE OF COURSE TOPICS AND READINGS

All readings are in Samuels & Witmer, 2003

DATE	Topic	Reading
01/21	Introduction to the course	
01/26	Statistics and the life sciences	Chapter 1, pp. 1-8
	Description of populations and samples	Chapter 2, pp. 9-24
	Frequency distributions	Chapter 2, pp. 26-30
	Descriptive statistics	
01/28	Description of populations and samples	
	Box plots	Chapter 2, pp. 32-38
	Measures of dispersion	Chapter 2, pp. 40-48
	Statistical inference on samples and populations	Chapter 2, pp. 57-64
02/02	Randomization, probability, binomial distributions	
	Random sampling and probability	Chapter 3, pp. 71-92
02/04	Randomization, probability, binomial distributions	
	Continuous variable and density curves	Chapter 3, pp. 93-95
	Random variables	Chapter 3, pp. 96-101
	The binomial distribution	Chapter 3, pp. 102-114
02/09	The normal distribution	Chapter 4, pp. 119-145
02/11	Sampling distributions	Chapter 5, pp. 149-166
	Confidence intervals	Chapter 6, pp. 179-214
02/16	**Exam #1**	
02/18	Comparison of two independent samples	
	Standard error	Chapter 7, pp. 219-225
	Confidence intervals	Chapter 7, pp. 227-231
	The t-test	Chapter 7, pp. 234-263
02/23	Comparison of two independent samples	
	Interpretation of statistical significance	Chapter 7, pp. 266-271
	Statistical power	Chapter 7, pp. 273-277
	Assumptions of the t-test	Chapter 7, pp. 280-288
02/25	Comparison of two independent samples	
	The Wilcoxon-Mann-Whitney Test	Chapter 7, pp. 288-300
03/02	Principles of experimental design	
	Observational studies and experiments	Chapter 8, pp. 309-324
	Blocked and stratified designs	Chapter 8, pp. 326-332
03/04	Principles of experimental design	
	Replication and other sampling concerns	Chapter 8, pp. 334-343
03/06 – 03/16	**Spring break**	

DATE	Topic	Reading
03/18	TBA	
03/23	Comparison of paired samples	
	Paired sample t-tests and confidence intervals	Chapter 9, pp. 347-355
	The sign test	Chapter 9, pp. 358-369
03/25	Comparison of paired samples	
	The Wilcoxon signed-rank test	Chapter 9, pp. 372-375
	Further considerations in paired experiments	Chapter 9, pp. 377-384
03/30	Analysis of categorical data	
	Chi-square tests and contingency tables	Chapter 10, pp. 391-432
04/01	Analysis of categorical data	
	Confidence intervals for probabilities	Chapter 10, pp. 439-440
	Paired data and 2 X 2 tables	Chapter 10, pp. 441-443
	Relative risk and the odds ratio	Chapter 10, pp. 444-452
	Further considerations of chi-square tests	Chapter 10, pp. 454-455
04/06	**Exam #2**	
04/08	Comparing means of many independent samples	
	One-way analysis of variance and the F-test	Chapter 11, pp. 463-486
04/13	Comparing means of many independent samples	
	Two-way analysis of variance	Chapter 11, pp. 487-505
	Multiple comparisons	Chapter 11, pp. 507-515
	Further considerations of ANOVA	Chapter 11, pp. 516-518
04/15	Non-parametric alternatives to ANOVA	
04/20	Comparing means of many non-independent samples	
04/22	Linear regression and correlation	
	Basic concepts in regression analysis	Chapter 12, pp. 525-547
04/27	Linear regression and correlation	
	Statistical inference regarding the slope	Chapter 12, pp. 548-552
	The correlation coefficient	Chapter 12, pp. 553-563
04/29	Linear regression and correlation	
	Interpreting correlation and regression	Chapter 12, pp. 565-574
	Further considerations	Chapter 12, pp. 576-586
05/04	Transformation of variables	Chapter 2, pp. 50-55
		Chapter 4, pp. 138-139
	A summary of inference methods	Chapter 13, pp. 595-607
05/06	**Exam #3**	
05/11	**Reading Day**	
05/13, 11:30 – 2:00	**Comprehensive final exam**	