BINF 702 – Research Methods

Saleet Jafri

Program in Bioinformatics and Computational Biology

George Mason University

Lecture Outline

- Introductory Remarks
- Syllabus
- Critical Analysis of a Manuscript and the Scientific Method

Introductory Remarks

- Class list with email
- Do you have a BINF computer account?
- Textbook will be available in bookstore in a week or two.
- Will not need the text until then

Syllabus – Contact Information

- Location: Prince William II, Room 185, Prince William Campus
- Course Time: Tuesdays 1:30 pm – 4:10 pm
- Instructors:

Saleet Jafri	Jennifer Weller
PW I Room 443	PW I Room 441
703-993-8420	703-993-8329
sjafri@gmu.edu	jweller@gmu.edu

Syllabus – Course Description

- Textbook: Fundamentals of Biostatistics by Bernard Rosner, 5th Edition
- Course Description: The student will learn concepts that will enable them to better understand and plan scientific experiments. The course will discuss the components of a scientific experiment with an emphasis on experimental design and data analysis. The course will also discuss the components of a scientific paper and how to critically analyze scientific work.

Syllabus – Grading Policy

- Grading Policy:

		Grade Scale:	
Homework	20%	90-100%	A
Exam 1	20%	80-89.9%	B
Exam 2	20%	70-79.9%	C
Final Exam	40%	0-69.0%	F

Syllabus – Grading Policy

- Homework assignments will be assigned several times during the semester. They will be due two weeks after they are assigned. Late homeworks will not be accepted.
- All students are expected to take the exams at the announced time. If there is a conflict and you cannot make the exam, let me know beforehand so we can agree upon a procedure to make up the exam. In order for you not to get a zero on the exam we must have agreed upon a make up procedure.

Syllabus – Academic Honesty Policy

- Academic Honesty Policy : Academic dishonesty will not be tolerated. This includes cheating, plagiarism, and falsification of academic records. That being said, you can help each other out on the homework (this does not mean that you can copy each other's homework).
- Calculators: You may use calculators in this class.

Syllabus – Important Dates

- Important Dates:
- Tuesday January 21, 2003 - First Day of Class
- Tuesday, February 18, 2003 - Exam 1
- Marcy 9-16, 2003, - Spring Break
- Tuesday, April 8, 2003 - Exam 2
  Tuesday, May 13, 2003 1:30 pm – 4:10 pm - Final Exam

Syllabus – Sage Advice

- Sage Advice: If you want to do well in course: 1) Do all the homework. 2) Ask questions in class and office hours. 3) If you are having difficulty doing the homework, be sure to see the instructor for additional help.

Syllabus – Course Schedule

Week 1	**Chapter 1 - General Overview**
	Anatomy of a Scientific Experiment and Paper
Week 2	**Chapter 2 - Descriptive Statistics**
	Chapter 3 – Probability
Week 3	**Chapter 4 - Discrete Probability Distributions**
Week 4	**Chapter 5 - Continuous Probability Distributions**
Week 5	**Chapter 6 – Estimation**
	Exam 1
Week 6	**Chapter 7 - Hypothesis Testing: One-Sample Inference**
Week 7	**Chapter 7 - Hypothesis Testing: One-Sample Inference**
Week 8	**Chapter 8 - Hypothesis Testing: Two-Sample Inference**
Week 9	**Chapter 9 - Multisample Inference**
Week 10	**Chapter 10 - Hypothesis Testing: Categorical Data**
Week 11	**Chapter 10 - Hypothesis Testing: Categorical Data**
	Exam2
Week 12	**Chapter 11 - Regression and Correlation Methods**
Week 13	**Chapter 12 - Nonparametric Methods**
Week 14	**Chapter 12 - Nonparametric Methods**
Week 15	**Chapter 13 - Hypothesis Testing: Person-Time Data**

Critical Analysis of a Manuscript and the Scientific Method

Critical Analysis

- In order to analyze a paper, one must first understand the scientific method.

Scientific Method

1. Make an observation
2. Form a hypothesis to explain the observation
3. Make predictions from the hypothesis
4. Test the hypothesis with experiments and additional observations
5. Revise the hypothesis, make new predictions, and test again.

Hypotheses

- A hypothesis provides a conceptual framework to explain existing observations and predict new ones
- We do not prove hypothesis.
- We either reject them or fail to reject them.
- If we do not reject a hypothesis is becomes an accepted theory
- An accepted theory is not a fact.

Choosing a Hypothesis: Occam's (Ockham's) Razor

- Ockham's Razor is the principle proposed by William of Ockham in the fourteenth century: "*Pluralitas non est ponenda sine neccesitate*", which translates as "entities should not be multiplied unnecessarily".
- The hypothesis which makes the least number of assumptions that explains the observation is best.
- Translation: Keep things simple.

Occam's (Ockham's) Razor

Example

- The planets move around the sun in ellipses because there is a force between any of them and the sun which decreases as the square of the distance.
- The planets move around the sun in ellipses because there is a force between any of them and the sun which decreases as the square of the distance. This force is generated by the will of some powerful aliens.

Formulation of a Hypothesis

- When formulating our hypothesis, we typically formulate a null and alternative hypothesis.
- The alternative hypothesis is typically explains the observation.
- The null hypothesis is the opposite.
- Experiments are done to reject the null hypothesis in favor of the alternative hypothesis.

Evaluation of a Hypothesis

When doing experiments to test a hypothesis, the results must be:

- Reproducible
- Statistically significant

Discovery Based Science

- Discovery based science is based on exploration of the data for patterns to yield new hypothesis.
- It does not replace the scientific method.
- It yields a way of finding new hypotheses from complex and abundant data.
- Engineering approaches that build tools or inventions can be considered to be discovery based.

Parts of a Research Article

- Abstract or Summary
- Introduction
- Methods
- Results
- Discussion
- Conclusions

Critical Analysis

- Paper Goals
- Methods
- Results
- Discussion

Paper Goals

- Do the authors describe the goals of the paper?
- What scientific question are they trying to answer?
- This is usually described in the introduction.

Methods

- Do the authors adequately describe the methods used?
- Are the methods appropriate to achieve their goals?

Results

- Are the results meaningful?
- Are the results statistically significant?
- Do the results make sense given the methods used?
- Are their other experiments that would clarify the results?

Reproducibility of Results

- Result shown in an experiment should be able to be duplicated.
- The results should have multiple trials, ie. need meaningful statistics.
- The results should be reproducible by others.

Discussion

- Do the authors put the results in context?
- Do the results support the conclusions drawn?
- Do the authors address alternative interpretations of the results?
- What should be done next?

Web Links for Scientific Method

http://teacher.nsrl.rochester.edu/phy_labs/AppendixE/AppendixE.html

http://www.selu.edu/Academics/Education/EDF600/Mod3/

http://phyun5.ucr.edu/~wudka/Physics7/Notes_www/node5.html

Why Biostatistics?

- When data is gathered, it must be evaluated to see if the results are meaningful – tests of statistical significance.
- When designing an experiment, how much data should be gathered – experimental design.
- Statistics used in bioinformatics, for example, microarray analysis