Math 3200 **Fall 2014**

STATISTICS

Welcome to Math 3200! My name is Professor Edward Spitznagel. This is the successor course to Math 320. It is a calculus-based introductory course in statistics and the underlying probability theory supporting it. Since this course is now differentiated (and integrated—☺) from the effectively non-calculus-based Math 2200, a paragraph or two of explanation is warranted.

When I began teaching Math 320 in 1970, it had an enrollment of 21 students. At that time, it was a calculus-based course. Over the years, it grew until four years ago it had over 400 students. Gradually, the calculus prerequisite became a nominal one-semester dose (Math 131), which meant that the quality of the course really suffered. Perhaps that would not have been a problem, except for the fact that many of our upper level courses depended on students being prepared for them by Math 320. Without that preparation, those courses had to spend their first third in reviewing what should have been covered in Math 320, and thus themselves became watered down.

By returning Math 320 to its roots, we hope to upgrade the quality of all our statistics offerings for both mathematics majors and minors. Of course, any student, major, minor, or not, who has the calculus background is welcome in the revitalized Math 320. Although what we are doing is in fact restoring Math 320 to what it once was, it was decided that it might be more politically correct to give it a new number—thus its new designation as Math 3200.

Times and Places

Our course meets Monday, Wednesday, and Friday 10-11 in Busch 100. **Before you come to class, please preview the section of the book to be covered that day.** Naturally I don't expect you to learn all the material from that reading. What I do expect is that you will be able to ask much better questions, having done that preview.

My *official* office hours are from 11 to 12 on Monday, Wednesday, and Friday. That's right after class. My office is Room 118 in Cupples I. As far as unofficial hours are concerned, you are *welcome* to knock anytime you see the light on. However, if you are elsewhere on campus or off-campus, I recommend calling in advance to see if I'm in. My telephone number is 935-6745.

Textbook

The text is Tamhane and Dunlop's *Statistics and Data Analysis: From Elementary to Intermediate*. This is one of a very few books from which a junior level course can be taught. Most other books are either too hard (too much mathematics) or too soft (too little mathematics). Like Baby Bear's bed, Tamhane and Dunlop is just right.

Hand Held Technology

The **Texas Instruments calculators TI-83, TI-84, and TI-89 (and the new TI-Nspire series)** contain essentially every probability function and statistical program we will be using during the course. It would be foolish not to use such technology in our course, as it saves memorizing a huge number of arcane formulas. I have therefore declared the above to be the official calculators for the course. I have a computer emulation of the TI-83, with which I will frequently work problems in class, projecting an image of the calculator on the screen. These calculators also contain functions that supersede the distribution tables in the back of the book. I will not provide those tables for the examinations; you will be expected to use the calculator instead. ***Verbum sapienti!***

Manual Homework

There are six recommended homework problems per day of class. In most instances, two are odd-numbered, with answers in the back of the book. The other four are even-numbered book problems or are taken from Society of Actuaries Exam P sample questions. I will usually have time to work two even-numbered problems in class, leaving you with a net four problems per day to do on your own. These problems will not be graded. Your primary motivation for keeping up with the homework is that most of the examination problems will be homework problems with simple changes in the data.

Computer Technology

There is a wide variety of computer software for doing statistics, ranging from the relatively primitive capabilities in Microsoft Excel® to the extremely powerful SAS® package. We will use two statistics packages, SAS and R®.

We will begin with SAS, using it alone for the first three weeks, then adding R to compare-and-contrast. I will demonstrate them in class, and will assign homework problems for you to do and hand in for grading. Thus you will be able to claim at least passing familiarity with the two most popular statistics packages when the time comes to interview for jobs and internships.

Computer Homework

There are three required computer homework problems per week of class. When it is convenient, these problems are chosen from the recommended manual homework problems. These problems are due in class each Friday, with the exception of our first Friday (August 29), Fall Break (October 17), and Thanksgiving Break (November 28), That works out to a total of twelve assignments.

Examinations

As mentioned earlier, examinations are closely linked to the homework problems. If you faithfully work the problems, you should have no trouble scoring well on the examinations. Each examination will contain twenty multiple choice problems, of which approximately fifteen will be homework problems with altered numbers. You may bring one 4×6 inch notecard to each in-semester examination. For the final exam, you will be permitted to bring all your previous notecards, plus one you have prepared for the final exam. You may use both sides of each note-card.

Over the four examinations, you can achieve a maximum of 80 points. With the computer homework added in, your maximum number of points will be 100. At the end of the semester, the A range will be 90 and above, the B range will be 80 to 90, the C range will be 70 to 80, and the D range will be 60 to 70, with plus and minus grades at the tops and bottoms of each of these ranges.

Students ask if I ever grade on a "curve." Curve grading was popular about fifty years ago. It assigned six letter grades A, B, C, D, E, and F based on a Gaussian, also called a "normal" curve. The grade of A corresponded to being 2 standard deviations above the mean and was awarded to the upper 2.5% of all students. The grade of B corresponded to being one to two standard deviations above the mean and was awarded to 13.6% of all students. The most common grades were C and D, at 34.1% each. I doubt any of you would like the grades to be assigned based on that system.

Instead, I will follow the modern convention, in which the A range will be 90 to 100, the B range will be 80 to 90, the C range will be 70 to 80, and the D range will be 60 to 70, with plus and minus grades at the tops and bottoms of each of these ranges. If you are registered pass/fail, you must achieve at least 70 points to pass, which is the lowest score for a C–.)

In addition to calculating the straight sum of points, I will also average the examination scores following a weighting process, in which each in-semester examination counts 16% and the final counts 32%, giving you whichever score is higher. (The computer homework will still be counted at 20%.)

This alternative weighting system rewards students who have tended to improve over the semester.

Examination Schedule

The three in-semester examinations will be given from 7:00PM to 9:00PM the following **Wednesday evenings**: September 17th , October 22nd, and November 19th. The homework problems suggested for those three days will not be covered on the exams those nights, but rather will be covered on the following exams.

The final examination will be given on **Thursday, December 11th, 3:30PM-5:30PM.**

As always, examination room assignments are posted on the Math Dept website:

http://www.math.wustl.edu/seatlookup/

the day of the examination.

Recommended Homework

Following are the recommended homework problems. At the risk of preaching to the choir, let me say that mastering these and reading the book should give you the two hours-out-of-class-for-every-one-in-class needed for success in the typical undergraduate course.

Two schools, CalTech and MIT, award credits equal to the weekly sum of lecture hours and expected amount of hours outside of class. As a reality check, I visited their websites and found the credits for their equivalent statistics courses to be:

CalTech:	Ma112a lists 9 units of credit.
MIT:	18.443 lists 12 units of credit.

Thus, these two schools expect their students to spend between two and three hours outside of class for every hour inside class.

Aug 25	Chapter 2	6,9,11,12,A1,A2
Aug 27	Chapter 2	17,20,24,27,A3,A4
Aug 29	Chapter 2	28,29,33,34,A5,A6
Sep 1	**Labor Day**	
Sep 3	Chapter 2	35,39,40,42,A7,A8
Sep 5	Chapter 2	46,48,51,53,A9,A10
Sep 8	Chapter 2	61,62,64,69,A11,A12
Sep 10	Chapter 2	73,74,75,76,A13,A14
Sep 12	Chapter 2	80,81,82,83,A15,A16
Sep 15	Chapter 3	1,4,6,7,8,10
Sep 17	Chapter 3	12,14,15,16,17,18
Sep 17	**First Examination**	
Sep 19	Chapter 3	20,21,22,23,24,26
Sep 22	Chapter 4	2,3,4,5,7,8
Sep 24	Chapter 4	12,13,22,23,24,26
Sep 26	Chapter 4	30,31,33,34,38,40
Sep 29	Chapter 5	4,6,7,11,A17,A18
Oct 1	Chapter 5	16,18,20,21,22,23
Oct 3	Chapter 5	24,25,26,28,30,33
Oct 6	Chapter 6	2,3,4,7,8,10
Oct 8	Chapter 6	11,12,13,14,15,16
Oct 10	Chapter 6	17,18,20,22,25,30
Oct 13	Chapter 7	3,6,8,12,13,16
Oct 15	Chapter 7	17,18,19,20,21,24
Oct 17	**Fall Break**	
Oct 20	Chapter 8	2,3,4,5,6,8
Oct 22	Chapter 8	9,10,14,18,20,23
Oct 22	**Second Examination**	
Oct 24	Chapter 9	5,6,9,12,14,16

Oct 27	Chapter 9	17,20,24,27,28,30
Oct 29	Chapter 10	2,4,5,6,7,8
Oct 31	Chapter 10	11,12,15,16,20,24
Nov 3	Chapter 10	28,29,30,31,32,36
Nov 5	Chapter 11	2,5,10,11,12,17
Nov 7	Chapter 11	22,23,28,30,34,39
Nov 10	Chapter 11	40,41,42,44,45,46
Nov 12	Chapter 12	1,2,3,4,5,7
Nov 14	Chapter 12	8,9,11,12,13,16
Nov 17	Chapter 12	18,19,20,21,24,28
Nov 19	Chapter 13	2,3,6,16,17,22
Nov 19	**Third Examination**	
Nov 21	Chapter 13	25,26,28,29,30,34
Nov 24	Chapter 14	12,19a,23a,26
Nov 26-Nov30	**Thanksgiving Vacation**	
Dec 1	Chapter 14	35,36,37,38
Dec 3	Chapter 15	1,2,3,5
Dec 5	Chapter 15	14,15,16,17
Dec 11	**Final Examination**	

Required Homework

Here are the required computer homework problems. Three problems are due per week, always on Friday, at the beginning of class. The total number of assignments is equal to twelve. All assignments are to be done with SAS, and Assignments 4 through 12 also with R. Starting with SAS by itself will give you a chance to get used to it before learning about R.

Sep 5	2.6, 2.20(simulate), 2.27
Sep 12	2.29, 2.51, 2.53 (simulate all 3)
Sep 19	2.61, 2.73b, 2.83
Sep 26	3.14(n=120), 3.20, 3.23
Oct 3	4.5, 4.24, 4.33
Oct 10	5.6, 5.21, 5.28(simulate)
Oct 24	6.7, 6.12, 6.25
Oct 31	7.13, 8.10, 9.14
Nov 7	9.16, 10.4, 10.20
Nov 14	10.28, 11.23, 11.28
Nov 21	11.40, 12.3, 12.13
Dec 5	12.21, 13.2, 13.26

"A" Problems

A1. A survey of a group's viewing habits over the last year revealed the following information:
(i) 28% watched gymnastics
(ii) 29% watched baseball
(iii) 19% watched soccer
(iv) 14% watched gymnastics and baseball
(v) 12% watched baseball and soccer
(vi) 10% watched gymnastics and soccer
(vii) 8% watched all three sports.
Calculate the percentage of the group that watched none of the three sports during the last year.

A2. The probability that a visit to a primary care physician's (PCP) office results in neither lab work nor referral to a specialist is 35%. Of those coming to a PCP's office, 30% are referred to specialists and 40% require lab work. Determine the probability that a visit to a PCP's office results in both lab work and referral to a specialist.

A3. A public health researcher examines the medical records of a group of 937 men who died in 1999 and discovers that 210 of the men died from causes related to heart disease. Moreover, 312 of the 937 men had at least one parent who suffered from heart disease, and, of these 312 men, 102 died from causes related to heart disease. Determine the probability that a man randomly selected from this group died of causes related to heart disease, given that neither of his parents suffered from heart disease.

A4. An insurance company examines its pool of auto insurance customers and gathers the following information:
(i) All customers insure at least one car.
(ii) 70% of the customers insure more than one car.
(iii) 20% of the customers insure a sports car.
(iv) Of those customers who insure more than one car, 15% insure a sports car.
Calculate the probability that a randomly selected customer insures exactly one car and that car is not a sports car.

A5. An insurance company determines that N, the number of claims received in a week, is a random variable with $P(N = n) = 1/(2^{n+1})$, where $n \geq 0$. The company also determines that the number of claims received in a given week is independent of the number of claims received in any other week. Determine the probability that exactly seven claims will be received during a given two-week period.

A6. The loss due to a fire in a commercial building is modeled by a random variable X with density function $f(x) = 0.005(20-x)$ for $0<x<20$, and 0 elsewhere. Given that a fire loss exceeds 8, what is the probability that it exceeds 16?

A7. An insurance policy pays an individual \$1000 per day for up to 3 days of hospitalization and \$250 per day for each day of hospitalization thereafter. The number of days of hospitalization, X, is a discrete random variable with p.m.f. = (6–x)/15 for x=1,2,3,4,5. Calculate the expected payment for hospitalization under this policy.

A8. An actuary determines that the claim size for a certain class of accidents is a random variable X with moment generating function

$$M_X(t) = (1/(1-2500t))^4$$

Determine the standard deviation of the claim size for this class of accidents.

A9. An insurance policy pays a total medical benefit consisting of two parts for each claim. Let X represent the part of the benefit that is paid to the surgeon, and let Y represent the part that is paid to the hospital. The variance of X is 5000, the variance of Y is 10000,
and the variance of the total benefit, X +Y , is 17000. Due to increasing medical costs, the company that issues the policy decides to increase X by a flat amount of 100 per claim and to increase Y by 10% per claim. Calculate the variance of the total benefit after these revisions have been made.

A10. A company insures homes in three cities, J, K, and L . Since sufficient distance separates the cities, it is reasonable to assume that the losses occurring in these cities are independent. The moment generating functions for the loss distributions of the cities are:

$M_J(t) = (1-2t)^{-3}$

$M_K(t) = (1-2t)^{-2.5}$

$M_L(t) = (1-2t)^{-4.5}$

Let X represent the combined losses from the three cities.
Calculate $E(X^3)$.

A11. An actuary has discovered that policyholders are three times as likely to file two claims as to file four claims. If the number of claims filed has a Poisson distribution, what is the variance of the number of claims filed?

A12. A company establishes a fund of \$12,000 from which it wants to pay a bonus to any of its 20 employees who achieve a high performance level during the coming year. Each employee has a 2% chance of achieving a high performance level during the coming year, independent of any other employee. Determine the maximum value of the bonus for which the probability is less than 1% that the fund will be inadequate to cover all payments for high performance.

A13. The number of days that elapse between the beginning of a calendar year and the moment a high-risk driver is involved in an accident is exponentially distributed. An insurance company expects that 30% of high-risk drivers will be involved in an accident during the first 50 days of a calendar year. What portion of high-risk drivers are expected to be involved in an accident during the first 80 days of a calendar year?

A14. The lifetime of a printer costing 200 is exponentially distributed with mean 2 years. The manufacturer agrees to pay a full refund to a buyer if the printer fails during the first year following its purchase, and a one-half refund if it fails during the second year. If the manufacturer sells 100 printers, how much should it expect to pay in refunds?

A15. Two instruments are used to measure the height, h, of a tower. The error made by the less accurate instrument is normally distributed with mean 0 and standard deviation 0.0056h. The error made by the more accurate instrument is normally distributed with mean 0 and standard deviation 0.0044h. Assuming the two measurements are independent random variables, what is the probability that their average value is within 0.005h of the height of the tower?

A16. A company manufactures a brand of light bulb with a lifetime in months that is normally distributed with mean 3 and variance 1 . A consumer buys a number of these bulbs with the intention of replacing them successively as they burn out. The light bulbs have independent lifetimes. What is the smallest number of bulbs to be purchased so that the succession of light bulbs produces light for at least 40 months with probability at least 0.9772?

A17. Claims filed under auto insurance policies follow a normal distribution with mean 19,400 and standard deviation 5,000. What is the probability that the average of 25 randomly selected claims exceeds 20,000 ?

A18. For Company A there is a 60% chance that no claim is made during the coming year. If one or more claims are made, the total claim amount is normally distributed with mean 10,000 and standard deviation 2,000 . For Company B there is a 70% chance that no claim is made during the coming year. If one or more claims are made, the total claim amount is normally distributed with mean 9,000 and standard deviation 2,000 . Assume that the total claim amounts of the two companies are independent. What is the probability that, in the coming year, Company B’s total claim amount will exceed Company A’s total claim amount?