

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

Review Materials for the FINAL

Exam coverage: Chapters 3-1 through 3-5, 4-1 through 4-6, 5-1 through 5-4, 6-1 through 6-4, 8-1, 8-5, 9-3 (a little from 9-4 & 9-6)

OFFICE HOURS FINALS WEEK

Monday 6/14 5pm-7pm in Math Sciences 7608

Tuesday 6/15 11:30am-1:00pm in Boelter 9413 (the statistics computer lab)

e-mail will be guaranteed answered if sent before 6pm on Tuesday 6/15

If you need to schedule a private appointment before the final, please send e-mail to vlew@stat.ucla.edu with possible dates & times.

Suggested Extra Problems From Your Textbook:

Final Details

The final is worth 120 points spread over 5 questions. The weighting is approximate since I have not written the final yet, but this is what I am thinking about as I assemble the review materials:

Chapter 9: between 15-20 points

Chapter 8: between 20-30 points

Chapters 4& 6: between 20-45 points

Chapter 5: between 10-20 points

Chapter 3: between 10-20 points

WHAT IS ALLOWED

This exam is open note (class handouts and lecture notes and your personal notes) and open book (textbook only however). Calculators, but not cellphone calculators or PDA or laptops. Any type of writing instrument (pen, pencil, crayon, highlighter, colored pens etc). Identification (mandatory). Non-alcoholic beverages.

WHAT IS FORBIDDEN

Food, you are not allowed to eat during the final unless you have a hand written medical excuse from a U.S. board certified medical doctor (MD). Using textbooks other than the course textbook or the course handout chapters is forbidden. Notes from a different course are forbidden. Laptop computers are forbidden. PDAs (e.g. Palm Pilots) are forbidden. Cellphones forbidden. If you have these items out in plain sight or easy reach during the exam, they will be confiscated and returned to you at the end of the final.

I am not allowed to reveal final grades via e-mail or phone. If you want to know yours as soon as possible, leave a grade card with me. Otherwise, the grades will be posted on URSA in a timely manner.

What follows are practice problems, answers will be posted during 10th week, the final is not this long, it's just extra practice. Best wishes.

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

1. A study of all first year college students following their first full year of college, gives the following results for high school ranking (RANK) first year GPA (grade point average) and high school SAT score:

Average GPA = 3.19; Standard deviation GPA = 0.49
Average RANK = 2.04; Standard deviation RANK = 1.13
Average SAT = 1291; Standard deviation SAT = 220
Correlation coefficient for GPA and RANK = -.33
Correlation coefficient for GPA and SAT = .24
Correlation coefficient for RANK and SAT = -.42

Assume the SAT scores had a minimum of 910 and a maximum of 1580 and are normally distributed. RANK has a minimum of 0 and a maximum of 6 and is not normally distributed. GPA has a minimum of 1.56 and a maximum of 3.98 and it is not normally distributed.

- (a) Of the 3 correlation coefficients given to you above, please identify which pair has the strongest correlation and which has the weakest correlation.
- (b) Please construct the variance-covariance matrix for the three variables listed above.
- (c) Suppose 9 first year students are selected at random from the population of first year college students. (i) What is the chance that at least 2 of them possess SAT scores over 1400? (ii) What is the chance that at least 2 of them possess SAT scores between 890 and 1380? (iii) What percentage of all first year students have SAT scores below 840?

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

3. The IQ scores of adult humans (age 18 and over) are approximately normally distributed with a mean of 100 and a standard deviation of 15. The highest IQ of a currently living adult as reported by the Guinness Book of World Records belongs to Marilyn vos Savant who scored a 186 (nearly six standard deviations above average). The maximum IQ score is 200, values estimated above it are deemed unreliable. The lowest score on record is 40.

**(a) How low is the lowest 5% of all IQ scores (that is, at or below what IQ score is the lowest 5%)
How high is the highest 10% of IQ scores (that is, at or above what IQ Score is the highest 10%)?**

(b) A simple random sample of size 256 is drawn from the adult human population. What is the chance that the sample average will exceed 101?

(c) How large of a sample would a researcher need to select to insure that he or she is within plus or minus 1 IQ point of the population mean IQ with 99% confidence?

(d) A simple random sample of 144 college students is drawn from the adult human population. For the sample the average IQ is 103 and the standard deviation is 30. Please test the hypothesis that college students have higher IQ scores than the average human. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule.

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

4. You got a job working for a marketing company and your supervisor is planning a random sample which will survey some of households in Los Angeles County. Your supervisor instructs you to contact households by random-digit dialing phone numbers. Your supervisor knows from past experience that about 60% of the households you contact in this manner will respond.

(a) If you randomly dial 1500 telephone numbers, what are the mean and standard error of the proportion of households who respond?

(b) Find the probability that you will get at least 870 responses if you randomly dial 1500 phone numbers. What is the probability that you will get less than 840 responses? What two assumptions are you making to find these probabilities?

(c) Calculate the chance that exactly three of the first four people contacted will respond.

(d) Past studies of random digit dialing households in Orange County show that 70% of households contacted in this manner will respond. If your supervisor instructs you to random digit dial 900 households in Los Angeles County and 600 in Orange County, what is the expected number of households who will respond? What is the chance that you will get less than 390 Orange County Responses?

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

5. There are 20,000 restaurants in the County of Los Angeles, 50% of them received a letter grade of "A" during inspections, 40% received either a B or a C grade and 10% failed their inspections. Restaurant grades are not normally distributed. My financial adviser, the Oracle, has hired you as a temporary personal assistant. Your job is to schedule his next 16 dinners (Oracle never eats at home). Unfortunately, you didn't know about the rating system and you never eat out because you don't have the money. So you listened to your best friend and picked 25 restaurants at random with replacement from an internet database of the 20,000 restaurants in Los Angeles. The Oracle will give you +3 points if you choose "A" restaurants, +1.25 points if you choose "B" or "C" restaurants, and -20 points if you choose a restaurant with a failing grade. Treat your restaurant selections as if they were a simple random sample of restaurants.

A. Construct the probability distribution for this problem

B. What is the expected value for this distribution? What is the expected value for the mean restaurant scores for a sample of 25 restaurants selected at random with replacement?

C. What is the standard deviation for this distribution?

D. What is the standard error for the average score of a sample of 25 restaurants?

E. To convert your temporary job into a permanent job, you must have accumulated an average of at least +1 points from the Oracle after picking 25 restaurants for him. What's your chance of getting an average of at least +1 points after picking 25 restaurants? If it is not possible to calculate the chance, please write "not possible" below and explain why.

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

6. The pregnancy duration of human females (age 18 and over) is approximately normal with a mean of 266 days and a standard deviation of 18 days. It is believed that older pregnant women have longer pregnancy durations. A simple random sample of 121 older pregnant women is drawn from the population of all pregnant women. The average pregnancy duration for the sample is 269 days and the sample standard deviation is 35.

(a) Please test the hypothesis that older women have longer pregnancy durations than the average woman. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule.

(b) What proportion of pregnancies have durations as long as or longer than 300 days?

(c) Suppose a researcher is only interested in studying the proportion of pregnancies that have durations as long as or longer than 300 days. How large of a sample would he or she have to select in order to be able to construct a 90% confidence interval that was within 1% of the true proportion over 300 days?

(d) Given that we know a pregnancy lasts at least 300 days, what is the probability that it lasted longer than 310 days?

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

8. A marketing survey interviewed 1000 adults selected at random from the population of all U.S. adults. Of the adults, 529 said they currently own a personal computer. When asked about the manufacturer of their computer, 144 of them said "Dull", 115 of them said "Ache-Pee", 175 of them said "some other company" and the rest of them said "I don't know".

(a) Please construct a 90% confidence interval for the proportion of adults who own a Dull. Please construct a 90% confidence interval for the proportion of adults who own an "Ache-Pee".

(b) Suppose Ache-Pee's true market share is truly 25%. What is the chance that among 529 computer owners you would get less than 22% of them saying they owned an Ache-Pee? What is the chance that you would get between 22% and 28% saying they owned an Ache-Pee? What is the chance that you would get at least 28% saying they owned an Ache-Pee?

(c) Suppose the confidence interval constructed in part (a) above is too wide, please identify two things you can do to decrease the size (width) of the confidence interval.

(d) Given that a computer owner does not own a Dull, what is the chance that he or she does not know the manufacturer of his or her computer?

(e) What is the chance that 6 out of the first 10 adults interviewed own a computer? Given that someone owns a computer, what is the chance that 6 out of the first 10 owners own a Dell?

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

9. Suppose we are psychics and we know that former police chief Bernard Parks will be the next Mayor of Los Angeles with a final winning percentage of 55%. Unfortunately, we don't know Parks and he doesn't return our phone calls or e-mails so he doesn't know he will get 55% of the vote after the next election. In fact, he is spending a lot of money right now on random samples of size 81 which are supposed to help him make decisions about the upcoming election.

- a. What is the chance that one of his surveys will give a result showing that he will get 49% or less of the vote if it is true that he really has 55%?

- b. Suppose again that Parks does not know that he will get 55% of the vote and suppose he takes another random survey of size 81 and it shows that 49% will vote for him. Can you construct an 98% confidence interval for the population percentage of votes for Parks?

Circle one: Yes No

If you circled yes, please construct a confidence interval. If you circled no, please explain why you cannot construct a confidence interval.

- c. Suppose Parks is handed a confidence interval that looks like this:

$49\% \pm 6\%$

and he says to his consultant "this isn't very useful. I need to know if I'm going to win this election! What you are telling me is that I might have between 43% and 55% of the vote!" Suppose Parks is really going to win with 55% of the vote and all he needs is 51%. What sample size would the consultant need to use to be 98% confident that Parks is within 4% of the true percentage?

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

10. There were a total of 226,324 deaths in California in 1999, 71% involved non-Hispanic white persons of European descent. The average age at death was 63.4 years. 51% of the deaths involved heart disease. A random sample of 256 deaths was selected for a funeral industry study. Detailed research determined that the deceased was cremated in 99 of the deaths.
- Determine a 99% confidence interval for the proportion (or percentage) of deaths in California in which the deceased is cremated.
 - Suppose the confidence interval is too narrow, identify 2 things you can do to make the interval wider.
 - A classmate comes up to you and says, this is the interpretation of a 99% confidence interval:
"There is a 99% chance that the true parameter is within the interval you gave in part (a)"
Is your classmate's interpretation correct? (circle one) YES NO
And justify your choice in the space below.

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

13. It's a family tradition: your professor goes to Las Vegas every year for Thanksgiving. A new casino has opened and they are playing a modified roulette game that has 40 possible numbers that can be spun on a wheel. To play you bet \$4 and you get to choose 4 numbers. If the wheel lands on any number that you chose, you win \$10. If the wheel does not land on a number you choose but on one of 8 "special numbers" you don't win or lose anything. If the wheel lands on any of the remaining numbers (not the ones you choose or the special numbers), you lose your bet of \$4. Suppose the typical person plays 25 times.

a. This game of modified roulette can be represented by probability distribution, please construct a reasonable one in the space below.

b. The 25 plays can be treated like a random sample of size 25. Find the expected value of this game.

c. Find the standard error of this same game.

d. Suppose the professor decides to spend \$100 playing a total of 25 times and she lost \$5. Calculate the chance that your professor could lose 5 or fewer dollars playing this game. Show all of your work and answer this question – based on your calculations is she lucky? (let's suppose lucky means the chance of losing 5 or fewer dollars is less than 5%)?

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

14. The table below shows homework handed in by students of a class classified according to its timeliness and quality.

	Quality				
Timeliness	excellent	good	average	below average	Total
On Time	6	12	10	2	30
Late	2	4	3	1	10
Total	8	16	13	3	40

a) Suppose one homework is selected at random, what is the probability that the homework that is selected is excellent?

b) What is the probability that the one homework that is selected is excellent given that it is late?

c) Are late and excellent independent? First write "yes" or "no" and then mathematically justify your choice.

d) Given that the one homework selected is on time, what is the probability that it is average or below average?

e) How many on-time homeworks would I need to select (at random, with replacement) to have at least a 50% chance of seeing a excellent homework?

FINAL WILL BE HELD IN FRANZ 1178, not the usual lecture hall

15. The movie studios sometime test a movie's appeal by showing it in so-called "sneak previews" before deciding whether to open it in limited engagements or in widespread release. Unfortunately, sneak preview audiences often rate even eventual flops quite high, so the information obtained from a sneak preview is not very reliable. And since it will cost two million dollars to conduct a sneak preview campaign, some of the studio executives are against doing so. Recent experience with sneak previews is as follows: 40 movies have been sneak previewed, of which 30 were given "thumbs up" ratings by the audience (10 of these turned out to be smash hits, 15 turned out to be medium grossers, and 5 were flops) and 10 movies were given "thumbs down" by the preview audiences (five of these turned out to be medium grossers, and five were flops).

- (a) Determine all six conditional probabilities $\Pr(s|d)$, where s denotes what actually happened after the release and d denotes the studio's information it got from the reaction ("thumbs up" or "thumbs down" of the sneak preview audience.
- (b) Are success and sneak preview rating independent? Are they mutually exclusive? Please show your work for full credit.
- (c) Suppose Smash Hits earn \$60 million for a studio, medium grosser earn \$36 million and flops cost a studio \$24 million. What is the expected value for movies which get a "thumbs up" rating and what is the expected value for movies which get a "thumbs down" rating.