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.

strongest = RANK & SAT = -.42
weakest = GPA & SAT = .24

(b) Please construct the variance-covariance matrix for the three variables listed above.

$$\begin{array}{ccc} \sigma^2_{GPA} & & \\ \text{cov}_{GPA,RANK} & \sigma^2_{RANK} & \\ \text{cov}_{GPA,SAT} & \text{cov}_{RANK,SAT} & \sigma^2_{SAT} \end{array} = \begin{array}{ccc} .49^2 & & \\ -.1827 & 1.13^2 & \\ 25.872 & -104.412 & 220^2 \end{array}$$

(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?

(i) P(SAT > 1400) need Z to find probability so Z = $\frac{1400-1291}{220} = .4955$ which is .309

P(X > 2) = $1-\left[\binom{9}{0}(.309)^0(.691)^9+\binom{9}{1}(.309)^1(.691)^8\right]$ = 1-(.0359 + .1446) = 1-.1805= about 82%

(ii) P(890 < SAT < 1380) need 2 Z scores to find probability so $Z_{890} = \frac{890-1291}{220} = -1.82$ and

$Z_{1380} = \frac{1380-1291}{220} = +.40$ answer (1- (.345+.034)) = .621 or 62.1%

P(X > 2) = $1-\left[\binom{9}{0}(.621)^0(.379)^9+\binom{9}{1}(.621)^1(.379)^8\right]$ = 1-(.0002 + .0024) = 1-.0026= about 99.74%

(iii) P(SAT < 840) need Z again so Z = $\frac{840-1291}{220} = -2.05$ which is .020

2. Lawyers frequently receive a "year-end bonus" because law firms are partnerships and the money earned is shared among partners. There are approximately 1,000,000 lawyers in the U.S. and year 2002's "year-end bonus", when calculated as a percentage change of the 2001 "year-end bonus", is normally distributed with a mean year-end bonus of –9% (a decrease, 2002 was a bad year compared to 2001) and a standard deviation of 16%. The average age for the lawyers was 37.1 years with a standard deviation of 3.1 years. 93% of the lawyers commuted to work by automobile. SHOW YOUR WORK FOR FULL CREDIT.

a) What proportion of lawyers received year-end 2002 bonuses that were as larger as or larger their year-end 2001 bonuses?

P(BONUS ≥ 0) need Z to find probability so Z = $\frac{0 - -9}{16} = .56$ which is .288

b) A simple random sample of 100 lawyers has an average year end bonus at the 15th percentile, what is the actual value of that sample average?

The Z score at the 15th percentile for a normal variable is –1.04

So –1.04 = $\frac{X - -9}{\frac{16}{\sqrt{100}}} = \frac{9}{1.6}$ solve for X, X = -10.664 or about -10.7% (a loss)

c) A simple random sample of 100 lawyers has an average year end bonus of –7.5% (actually, a loss), what is the chance of getting a sample average of –7.5% or higher?

Z= $\frac{-7.5 - -9}{\frac{16}{\sqrt{100}}} = \frac{1.5}{1.6} = .94$ area to the right and under the curve is .174

d) What is the chance that four lawyers, selected at random with replacement, will each have year end bonuses of -7.5% or higher? (error corrected 6/9/04 at 4:23pm)

Z= $\frac{-7.5 - -9}{16} = .09$ the area to the right and under the curve is .464 so $(.464)^4$=.0464

e) What percentage of lawyers have year end bonuses between –1% and +5%? (error corrected 6/7/04 at 11:23pm)

$Z_{-1} = \frac{-1 - -9}{16} = .5$ area to the left and under the curve is .309

$Z_{+5} = \frac{+5 - -9}{16} = .88$ area to the right and under the curve is .189

so the area between is (.309-.189) = .12 or about 12%

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 Guiness 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%)?

The lowest 5% has a Z = -1.64 or -1.65, using –1.65

$Z = -1.65 = \frac{X - 100}{15}$ solve for X gives us 75.25

The highest 10% has a Z = +1.28

$Z = 1.28 = \frac{X - 100}{15}$ solve for X gives us 119.20

(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?

$Z = \frac{101 - 100}{\frac{15}{\sqrt{256}}} = \frac{1}{.9375} = 1.07$ the area to the right is .142 or 14.2%

(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?

Since we are working with means, the confidence interval we want is $\bar{x} \pm 1po\text{int}$ to be correct 99 in 100 samples (this is 99% confidence) so we set the part after the ± of the confidence interval to be equal to 1

$\bar{x} \pm Z * \left(\frac{\sigma}{\sqrt{n}} \right)$ is the confidence interval formula, set $Z * \left(\frac{\sigma}{\sqrt{n}} \right) = 1$ for 99% confidence, Z = 2.58 (because the area between –2.58 and +2.58 is 99% of the area) and we know sigma is 15 so: $2.58 * \left(\frac{15}{\sqrt{n}} \right) = 1$ solving for n gives n=1497.69 so probably 1,498

(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.

H_0: μ=100 versus H_A: μ>100
The test is Z so

$$Z = \frac{103 - 100}{\frac{15}{\sqrt{144}}} = \frac{3}{1.25} = 2.40$$

p-value is .008 this is less than .05

Since p < .05, we reject the null hypothesis. We believe college students have a higher IQ than the average human adult.

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?

Since we know 60% is the parameter, the mean is π = .60 if we are randomly dialing 1500 numbers then, the SE is

$$SE = \sqrt{\frac{(\pi)(1-\pi)}{n}} = \sqrt{\frac{(.6)(.4)}{1500}} = .0126$$

(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?

$$P_{870} = \frac{870}{1500} = .58 \text{ so } Z = \frac{.58 - .60}{\sqrt{\frac{.6 * .4}{1500}}} = \frac{-.02}{.0126} = -1.59$$ area is 1-.056 =.944 or 94.4%

$$P_{870} = \frac{840}{1500} = .56 \text{ so } Z = \frac{.56 - .60}{\sqrt{\frac{.6 * .4}{1500}}} = \frac{-.04}{.0126} = -3.2$$ area is .000687 or nearly 0%

We are assuming that (a) the sample size is large enough and (b) the sample is a random sample

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

If the response rate is 60% then P(X = 3) = $\left[\binom{4}{3}(.6)^3(.4)^1\right]$ = .3456 about 34.6%

(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?

For Los Angeles, expect 60% of 900 or 540, for Orange County we expect 70% of 600 or 420.

$$P_{400} = \frac{390}{600} = .65 \text{ so } Z = \frac{.65 - .70}{\sqrt{\frac{.7 * .3}{600}}} = \frac{-.05}{.0187} = -2.67$$ area to the left is .004 so it's very low

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

X	+3	+1.25	-20
p(x)	.50	.40	.10

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?

E(X)= μ = (+3*.50) + (1.25*.40) + (-20*.10) = 0

and the expected value for the average score from a sample of 25 restaurants is μ or 0

C. What is the standard deviation for this distribution?

$$\sigma = \sqrt{(3-0)^2(.50) + (1.25-0)^2(.40) + (-20-0)^2(.10)} = 6.7175$$

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

$$\frac{\sigma}{\sqrt{n}} = \frac{6.7175}{\sqrt{25}} = 1.344$$

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.

$Z = \frac{1-0}{1.344} = .74$ the chance is .23 or 23%

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.

H0: $\mu=266$ versus HA: $\mu>266$
The test is Z so

$$Z = \frac{269-266}{\frac{18}{\sqrt{121}}} = \frac{3}{1.6364} = 1.83$$

Since p < .05, we reject the null hypothesis. We believe older pregnant women have longer durations that other women.

p-value is .034 this is less than .05

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

$Z_{300} = \frac{300-266}{18} = 1.89$ about .029 or 2.9% of all pregnancies last 300 days or more.

(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?

Within 1% is $\pm$ 1% . To have that with 90% confidence for the proportion of pregnancies over 300 days (.029 from part B), we could do this:

$P \pm Z * \left(\sqrt{\frac{P(1-P)}{n}}\right) \Longrightarrow 1.64 * \left(\sqrt{\frac{.029 * .971}{n}}\right)$ set this equal to .01 so

$1.64 * \left(\sqrt{\frac{.029 * .971}{n}}\right) = .01$ and solve for n, it's about 758

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

$Z_{300} = \frac{300-266}{18} = 1.89$ about .029 or 2.9% of all pregnancies last 300 days or more.

$Z_{310} = \frac{310-266}{18} = 2.44$ about .007 or 0.7% of all pregnancies last 310 days or more.

P(310|300) = .007/.029 = .2413

7. Los Angeles International Airport handles an average of 6,000 international passengers an hour. Suppose 80% pass through primary security, but the remainder get detained for interrogation by the FBI. The percentage passing through primary security is not normally distributed. And suppose the FBI can handle 1,500 passengers an hour without unreasonable delays for travelers and extra costs to the airlines (due to missed flights and connections). The total cost to the airlines is approximately $70 per detained passenger with a standard deviation of $125, the total cost is not normally distributed.

a. Over the summer, it is expected that as many as 8,000 international passengers will arrive per hour. When that occurs, what is the expected number of passengers who will be detained?

20% of 8000 or 1,600

b. Using information from part a, find the approximate chance that less than 1,500 out of 8000 international passengers will be detained? (ERROR CORRECTED 6/9/04 @ 3:40pm)

1500/8000 = .1875, so $Z = \frac{.1875 - .20}{\sqrt{\frac{.2 * .8}{8000}}} = \frac{-.0125}{.0045} = -2.78$ the area to the left is about .003 so about .3%

c. Suppose the FBI decides to randomly sample passengers in order to speed up the screening process. What is the chance that a simple random sample of 100 will have between 15 and 18 passengers detained by the FBI?

$Z = \frac{.15 - .20}{\sqrt{\frac{.2 * .8}{100}}} = \frac{-.05}{.04} = -1.25$ area is .106

the chance is .309-.106=.203

$Z = \frac{.18 - .20}{\sqrt{\frac{.2 * .8}{100}}} = \frac{-.02}{.04} = -.5$ area is .309

d. Certain ethnic/racial groups appear to be detained at much higher rates than others. Suppose a human rights organization sends 64 persons who appear to be of Middle Eastern origin through the airport and 15 are detained for interrogation. Please test the hypothesis that persons of Middle Eastern origin are detained in higher proportions than the typical traveler. 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. You may treat the 64 as if it were a simple random sample and it is of reasonable size.

15/64=.234

H_0: π=.20 versus H_A: π>.20
The test is Z so

$$Z = \frac{.234 - .20}{\sqrt{\frac{.20 * .80}{64}}} = \frac{.034}{.05} = .69$$

Since p > .05, we do not reject the null hypothesis. We do not believe that persons who appear to be of Middle Eastern origin are detained in higher proportions than the typical traveler.

p-value is .245 this is NOT less than .05

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 COMPUTER OWNERS who own a Dull. Please construct a 90% confidence interval for the proportion of adults who own an "Ache–Pee". (Corrected at 10:26PM 6/9/04)

For 90% we can use a Z=1.64 or a Z=1.65. I will use a Z=1.65 for this answer. P_{Dull} is 144/529=.2722 and $P_{Ache-Pee}$ is 115/529=.2174 so the confidence intervals are:

Dull: $.2722 \pm 1.65 * \left(\sqrt{\frac{.2722 * .7278}{529}} \right) \Rightarrow .2722 \pm .0319$

Ache-Pee: $.2174 \pm 1.65 * \left(\sqrt{\frac{.2174 * .7826}{529}} \right) \Rightarrow .2722 \pm .0296$

(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?

Since 529 is "large enough: and it's a random sample, by the Central Limit theorem, we would say that the sampling distribution for a sample of size 529 is normally distributed with mean=π=.25 with standard error $\left(\sqrt{\frac{.25 * .75}{529}} \right)$ so the chance of getting P < .22 when we are expecting .25 is $Z = \frac{.22 - .25}{\sqrt{\frac{.25 * .75}{529}}} = \frac{-.03}{.0188} = -1.59$

the area to the left of –1.59 is .056 (i.e. P(Z < -1.59)). The chance of getting P > .28 when we are expecting .25 is: $Z = \frac{.28 - .25}{\sqrt{\frac{.25 * .75}{529}}} = \frac{.03}{.0188} = +1.59$ and the area to the right of 1.59 is .056. Now that we have the two extremes, we can say P(.22 < Z < .28) = 1.0 -.056 -.056 = .888

(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.

1. You can increase the sample size
2. You can decrease your level of confidence

(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?

Conditional probability, if 144 own Dulls, then 529 – 144 = 385 do not own Dulls. The chance that among those 385 the person does not know the manufacturer is 95/385= .2468

(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 Dull?

For the first 10 selected, the probability that 6 own is $\left[\binom{10}{6} (.529)^6 (.471)^4 \right]$=210*.0219*.0492=.2263

Omit the second part of this question.

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%?**

$Z = \frac{.49 - .55}{\sqrt{\frac{.55 * .45}{81}}} = \frac{-.06}{.0552} = -1.09$ the area to the left is .138 so there is a 13.8% chance that he will draw a sample of size 81 that show he will get less than 49% of the vote.

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 a 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.

$$P \pm Z_{.01} * \sqrt{\frac{P(1-P)}{n}} = .49 \pm 2.33 * \sqrt{\frac{.49(1-.49)}{81}} = .49 \pm .056$$

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

49% ± 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?

select Z=2.33 for 98% confidence because that leaves 1% in each tail under the normal curve and 98% in the central area. Use P=.55 because that's his true percentage, we want to be within 4% of that.

$P \pm Z * \left(\sqrt{\frac{P(1-P)}{n}}\right)$ ==> $2.33 * \left(\sqrt{\frac{.55 * .45}{n}}\right)$ set this equal to .04 so

$2.33 * \left(\sqrt{\frac{.55 * .45}{n}}\right) = .04$ and solve for n, it's about 839.78 or 840 voters

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.

a. Determine a 99% confidence interval for the proportion (or percentage) of deaths in California in which the deceased is cremated.

99/256=.3867 or 38.67% of the deaths in the sample involved cremation. The confidence interval is:

$$P \pm Z * \sqrt{\frac{P(1-P)}{n}} \Rightarrow .39 \pm 2.58 * \sqrt{\frac{.39 * .61}{256}} = .39 \pm .079$$

b. Suppose the confidence interval is too narrow, identify 2 things you can do to make the interval wider.

You could increase the confidence or decrease the sample size.

c. 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.

The 99% part of the confidence interval refers to the long-run percentage or probability of drawing a sample such that a confidence interval constructed from the sample information would generate an interval that contains the true population parameter. In other words, in 99 out of 100 of your possible samples of the same size, 99 of them would generate good information about the value of the parameter, one of the would not.

Parameters are fixed values and one does not speak of a parameter having a “chance” or probability. A sample statistic is a varying value and one can speak of the chance of a statistic taking on different values.

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

11. You know that every UCLA student will definitely get a job after graduation. The only uncertainty is the salary. Suppose this is what you know about the job prospects of UCLA students after graduation:

There is a 35% chance that the salary will be $20,000 per year; a 45% chance that it will be $90,000 per year; and a 20% chance that it will be $40,000 per year.

a. Construct the probability distribution for the salary of UCLA students

X	20000	40000	90000
p(x)	.35	.20	.45

b. Find the expected value and the standard deviation.

E(X)=55,500 = (20000*.35)+(40000*.20)+(90000*.45)

$$\sigma = \sqrt{(20000-55500)^2(.35)+(40000-55500)^2(.20)+(90000-55500)^2(.45)} = 32{,}011.72$$

c. Suppose a random sample of 36 UCLA students who have graduated is drawn. What is their expected average salary? What is the standard error of the salary? What is the chance that the sample average will exceed $65,000?

Their expected average is the same as the population average of $55,500. The standard error of that average salary for a sample of size 36 is $\frac{\sigma}{\sqrt{n}} = \frac{32011.72}{\sqrt{36}}$ =5,335.29

The chance that the sample average exceeds 65,000 is

$Z = \frac{65000-55500}{\frac{32011.72}{\sqrt{36}}} = \frac{9500}{5335.29} = 1.78$ so P(Z > 1.78) is .038 or there is a 3.8% chance of getting a sample average of 65,000 when you were expecting an average of 55,500

d. Two UCLA graduates meet and get married. Upon marriage, the husband agrees to deposit 65% of his salary in their joint bank account, the wife will deposit 35% of her salary in their joint account. What is expected value of their joint bank account?

Let X= husband's income and Y=wife's income. Let Z the joint bank account or Z = .65X + .35Y
The expected value of their joint account is E(Z) = .65(55,500) + .35 (55,500) = 55,500

Extra: To calculate their join variance, you need to know whether their incomes are (i) independent or (ii) correlated. Let's do both

(i) if independent then the covariance = 0 so VAR(Z) = $.65^2$(1,024,750) + $.35^2$(1,024,750) = 588,489

(ii) if correlated, suppose the correlation between a husband's and wife's incomes is .40. If so, then the covariance is Cov(X,Y) = .40*(32,011.717)*(32,011.717) = 409,900. Then VAR(Z) = $.65^2$(1,024,750) + $.35^2$(1,024,750) + 2(.65)(.35)(409900) = 744,993.25

This problem is too easy.

12. A study was conducted on the sleep patterns of infants in the United States. A sample of 25 infants was drawn at random with 54% sleeping at least 12 hours a night.

a. **Please compute a 95% confidence interval for the percentage of all U.S. infants who sleep at least 12 hours per night.**

$$P \pm Z * \sqrt{\frac{P(1-P)}{n}} \Rightarrow .54 \pm 1.96 * \sqrt{\frac{.54 * .46}{25}} = .54 \pm .195$$

b. **Suppose the sample size was increased to 100 infants, what effect would this have on the confidence interval? Assume that 54% of them slept at least 12 hours per night.**

The confidence interval would become narrower. Just plug it in.

$$P \pm Z * \sqrt{\frac{P(1-P)}{n}} \Rightarrow .54 \pm 1.96 * \sqrt{\frac{.54 * .46}{100}} \; .54 \pm .0977$$

notice specifically that the interval shrank by 50%, this is because 100 = 25*4 so increasing the sample size by 4 times results in a reduction of 1/2

c. **Please calculate a 90% confidence interval for the percentage of all U.S. Infants who sleep at least 12 hours per night. Again, please use the sample of size 25 and assume the percentage in the sample was 54% sleep at least 12 hours per night.**

We can use Z=1.64 or Z =1.65, I'll use 1.64 here

$$P \pm Z * \sqrt{\frac{P(1-P)}{n}} \Rightarrow .54 \pm 1.64 * \sqrt{\frac{.54 * .46}{25}} = .54 \pm .1634$$

This question 12 represents an attempt to test you on what can happen to confidence interval if you (a) change confidence and (b) change sample size. The other things you may want to review regarding confidence intervals is how to interpret one and how choose a sample size that will give you the exact size of a confidence interval for a given level confidence, question 9c asks just that.

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.

X	-4	0	+10
p(x)	.7	.2	.1

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

E(X)=(-4*.7)+(0*.2)+(10*.1)= -1.80 so on average, you expect –1.80 (a loss) if you played 25 times, we would expect you to lose 1.80 per play or lose $45 (it would have cost you $100 to play 25 times and you would wind up walking away with $45 less than what you started with) the total would be like E(X1 + X2 + X3 + … + X25) (see Chapter 6-2) whereas the average takes that total and divides by 25

c. Find the standard error of this same game.

Need the SD first so

$$\sigma = \sqrt{(-4 - -1.80)^2(.70) + (0 - -1.80)^2(.20) + (10 - -1.8)^2(.1)} = 4.2379$$

Now, depends on how you want to work this as a total or as an average. For an average it's $\frac{\sigma}{\sqrt{n}} = .8476$ and just like part B above, for a total of playing 25 times it's 25*.8476 or 21.19

So, you can look at this problem 2 ways, each time you play, you expect –1.80 (a loss) but it has an SD=4.2379 (which means sometimes you are in the positive). If you play 25 times, you expect an average over those 25 plays of –1.80 with an SE of .8476 so you expect to lose $1.80 on average give or take about 85 cents. It's a little easier to think of it as a total after 25 plays, in other words, if you played 25 times, we would expect you to lose $45 with a standard error of $21.19.

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%)?

$Z = \frac{-5 - -45}{21.19} = \frac{40}{21.19} = 1.89$ area under the curve to the right of 1.89 is .029 or 2.9%. Actually, I don't believe in luck ☺ especially on final exam. Luck favors the prepared.

Interpretation: The chance of someone playing 25 times and walking away with $95 dollars (out of their original $100) would be about 2.9% or .029. Typically, a person playing this game 25 times should walk away with 45 dollars less than what they began. If you did this problem as an average, the same person only lost 20 cents on average (that's $5 divided by 25)

$Z = \frac{-.20 - -1.80}{.8476} = \frac{1.60}{.8476} = 1.89$ you will get the same answer, you just need to be consistent (all totals or all averages, don't mix and match the two)

FINAL WILL BE HELD IN FRANZ 1178FINAL WILL BE HELD IN FRANZ 1178 FINAL WILL BE HELD IN FRANZ 1178

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?

8/40 = .2

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

2/10 = .2

c) Are **QUALITY AND TIMELINESS** independent? First write “yes” or “no” and then mathematically justify your choice. (Error corrected 6/10/2004 8:50am)

NO. If it’s independent then P(X,Y) = P(X)*P(Y) for all combinations of X and Y. So suppose you are looking at P(average, On Time) = P(average)*P(on time) = 13/40 * 30/40 = .2438 but P(average, on time) = .25

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

12/30

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?

This problem isn’t worded quite correctly because I need to take into account the fact that on-time homeworks are only 75% of all the possible homeworks. So if the question read “how many homeworks would I need to select to have at least a 50% chance of seeing an excellent homework” it would be

.75[1 - $(1- 6/30)^{n}$]=.50 solving for n, about 5 homework assignments

(it is a modification of the BMW problem on the second exam). Given the way the problem was worded, we would need to see about 4 on-time homeworks. Or

[1 - $(1- 6/30)^{n}$]=.50 solving for n, it’s a little over 3, since homeworks are discrete, it rounds to 4

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.

Pr(s="smash hit"|d="thumbs up")= 10/30
Pr(s="medium grosser"|d="thumbs up")= 15/30
Pr(s="flop"|d="thumbs up")= 5/30

Pr(s="smash hit"|d="thumbs down")= 0/10
Pr(s="medium grosser"|d="thumbs down")= 5/10
Pr(s="flop"|d="thumbs down")= 5/10

	Thumbs Up	Thumbs Down
Smash Hit	10	0
Medium Grosser	15	5
Flop	5	5

(b) Are success and sneak preview rating independent? Are they mutually exclusive? Please show your work for full credit.

No and No. Independent implies P(X,Y)= P(X)*P(Y) we can see for P(smash hit, thumbs down) is not equal to .25*.25 P(smash hit)*P(thumbs down)

Mutually Exclusive implies that P(X intersect Y)=0 for all combinations of X and Y

(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.

Need two tables
Thumbs Up

X	-24	36	60
p(x)	.1667	.50	.3333

E(X) = about 34 million (-24*.1667)+(36*.5) + (60*.3333)

Thumbs Down

X	-24	36	60
p(x)	.50	.50	0

E(X) = about 6 million (-24*.5)+(36*.5)