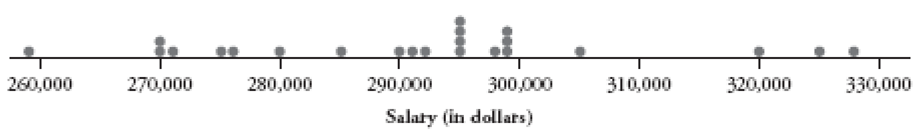
Math 393 Fall 2009 Exam 2 Review Sheet

1. Exam 2 is an in-class exam given on Monday,
2. The exam will cover Topics 9-14 and the probability and counting methods discussed in class (including Binomial variables)
3. You may have the whole class period to do the exam.
4. You may have two 3 x 5 notecards with notes on both sides for the exam, you may also use the Normal table in the back of your book.

5. Study Ideas:
   - Review all in-class and homework activities. Try homework activities that were not assigned.
   - Make a list of vocabulary words;definitions; important Theorems and facts
   - Do the practice problems below (note – answers are at the end)

Practice Problems

1. The following dotplot displays the salaries of the 23 presidents of California State University campuses for the 2007_08 academic year:



These salaries, arranged in order, are listed here:

258,680	270,000	270,315	270,568	275,000	276,055	279,500	285,000
290,000	291,179	292,000	295,000	295,000	295,000	295,000	297,870
298,749	299,000	299,435	305,008	320,329	325,000	328,209	

a. Determine the five-number summary of these salaries. Consider the following computer output:

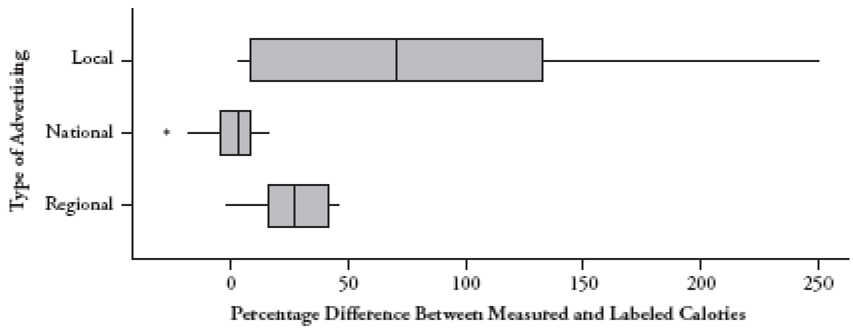
Variable	N	Mean	StDev
salary	23	291,822	17,669

b. The salary for Cal Poly's President Baker is $328,209. Calculate the *z*-score for his salary, and decide whether his salary is more than two standard deviations above the mean.

c. Check whether President Baker's salary is an outlier, according to the 1.5 x IQR rule.

2. Suppose Mary records the ages of people entering a McDonald's restaurant in a suburban shopping plaza tomorrow, while Abby records the ages of people entering a popular food court on a nearby college campus. Who would you expect to have the higher standard deviation of these ages: Mary (suburban McDonald's) or Abby (campus food court)? Explain briefly.

3. In a recent study, researchers purchased 40 food items in New York City and determined the actual calorie content of each through a laboratory analysis. They then calculated the percentage difference between the actual calorie content and the calorie count listed on the item's label. (A positive percentage difference corresponds to a food item whose actual calorie content was higher than what the label claimed.) Each food item was also classified according to whether it was marketed locally, nationally, or regionally. The boxplots below were constructed to compare the distributions:



Write a paragraph summarizing what these boxplots reveal about the percentage differences between measured and labeled calorie content among the three groups of food items.

4. Suppose that two employees have to be fired from a group of seven:
   - Amy, female, age 35
   - Bob, male, age 41
   - Chad, male, age 43
   - Diana, female, age 44
   - Elvis, male, age 49
   - Frank, male, age 61
   - Ginger, female, age 62

   Suppose for now that the company selects the two employees to be fired at random, so all possible pairs of people have the same probability of being fired.

a. List all outcomes in the sample space. In other words, list all possible *pairs* of people that can be chosen from this group of seven people. (To save time and space, just use initials. For example, write AB to represent the (Amy, Bob) pair.)

b. Determine the probability (both as a fraction and a decimal) that both employees selected are female.

c. Determine the probability (both as a fraction and a decimal) that the average age of the employees selected is older than 50.

d. Determine the probability that both employees selected are older than 60.

Suppose that Frank and Ginger are the employees selected to be fired. They proceed to file a lawsuit, alleging that the company is guilty of age discrimination for choosing the two oldest employees to fire. The company responds that it did not discriminate; it simply happened to choose these two by random chance.

e. Does your probability analysis provide evidence to support the allegation of age discrimination? Explain (with 1–3 sentences) the reasoning process behind your answer.

5. Suppose that the drying time for a certain type of paint under specified test conditions is known to be normally distributed with mean 75 minutes and standard deviation 5 minutes. Suppose also that chemists have devised a new additive that they hope will reduce the mean drying time (without changing the standard deviation). A test is then conducted to measure the drying time for a test specimen, and the company executives decide that they will be convinced that the additive is effective only if the drying time on this specimen is less than 70 minutes.

   a. If the additive actually has no effect at all on the drying time, what is the probability the company executives will mistakenly conclude that it is effective? Include a sketch with your calculation.

   b. If you want to alter the cutoff value from 70 in order to reduce the error probability in part a to .05, what cutoff value should you choose?

   c. Now suppose that the standard deviation of the drying times to 65 minutes is 2 rather than 5 minutes. Without doing any new calculations, describe how this change would affect your answers to parts a and b. Give an intuitive explanation for your reasoning in both cases.

6. Suppose 20% of all heart transplant patients do not survive the operation.

   a. Think about taking repeated random samples of 371 patients from this population. Describe how the sample proportion who die would vary from sample to sample. (*Hint:* Be sure to refer to the shape, center, and spread of its sampling distribution.) Also include a well-labeled sketch to represent this distribution.

   b. Suppose you take a random sample of 371 heart transplant patients. Determine the probability that the sample proportion who die would be .213 or higher.

7. Suppose I tell you that I flipped a coin multiple times and got 75% heads. Would you be reasonably convinced that this was not a fair coin (where “fair” means that the coin has a 50/50 chance of landing heads or tails)? If so, explain why. If not, describe what additional information you would ask for and explain why that information is necessary.

8. The distribution of house prices is skewed to the right because most houses cost a modest amount but a few cost a very large amount. If you take a random sample of 1000 houses, can you reasonably expect the distribution of the house prices to be approximately normal? Explain your answer.

9. Suppose you flip a fair coin until it lands heads up for the first time. It can be shown (do not try to calculate this) that the expected value of the number of flips required is 2. Explain (with a sentence or two) what this expected value means in this context.

10. How many distinct ways are there to arrange the letters in the word MATHEMATICS?

11. How many distinct ways are there to arrange the letters in the word PEPPER?

12. In a study group there are 3 boys and 3 girls.
    a. How many different ways can you line the students up so that the boys and girls alternate?
    b. How many ways if all the girls want to stand next to each other (e.g. GGG is somewhere in the line up)

13. What is the probability of being randomly dealt a 5-card hand from a standard 52 card deck so that:
    a. The hand contained two 5s and two 7s and one other card (not a 5 or a 7) (e.g. 5577X)
    b. The hand contained a pair (XXYZW)
    c. The hand contained a full house (XXYYY)
    d. The hand contained only clubs

14. Suppose you play a game where you roll a die and you win the amount showing in dollars if the number is even and lose the amount showing in dollars if the number is odd. What is the expected value of the winnings for this game? What does the expected value signify here?

15. Suppose you play a game where roll a fair 6-sided die and you win if the number showing is a factor of 6 and loose otherwise.
    a. What is the probability of winning the game?
    b. If you play the game 7 times, what is the probability that you will win exactly 5 times?
    c. If you play the game 10 times, how many games do you expect to win?

Solutions to Practice Problems

1.

a. Minimum: $258,680 Maximum: $328,209

(n + 1)/2 = 12, so the median is the 12th ordered value: $295,000

The lower quartile is the median of the 11 values below the (overall) median, so it is the 6th value, which is $276,055. The upper quartile is similarly the 6th value from the top: $299,000.

b. The z-score for President Baker's salary is (328,209 - 291,822)/17,669 or 2.06. This means that his salary is slightly more than two standard deviations above the average salary among CSU presidents.

c. The IQR is 299,000 - 276,055 or 22,945, so 1.5 x IQR = 34,417.5. To be a high outlier, a salary must be greater than 299,000 + 34,417.5 or 333,417.5. Baker's salary is less than this, so it is not an outlier.

2. Mary would have the greater standard deviation of ages, because McDonald's would see much more variability in its customers' ages than the campus food court would. McDonald's would have many children and elderly people as customers, as well as college-aged and middle-aged people. On the other hand, the campus food court would have mostly college-aged customers, with some older faculty and staff members and visitors.

3. The most striking aspect of these data is that locally marketed food items tend to have many more calories than advertised. The median discrepancy in this group is over 50%. There is also tremendous variability in these discrepancy percentages for the local items, ranging from close to 0 to almost 250%. On the other extreme, the nationally marketed items tend to have calorie amounts very close to what is advertised, with very little variation. The median discrepancy percentage in this group is close to 0, and there is little variability except for an outlier that actually has fewer calories than advertised. The regionally marketed items fall in between local and nationally marketed, both in terms of center and spread. The regionally marketed items do tend to have more calories than stated, but nowhere near as many as the local ones, and the variability is less than with the local items as well.

4. **a.** The sample space consists of the following pairs: AB, AC, AD, AE, AF, AG, BC, BD, BE, BF, BG, CD, CE, CF, CG, DE, DF, DG, EF, EG, FG (21 equally likely outcomes)

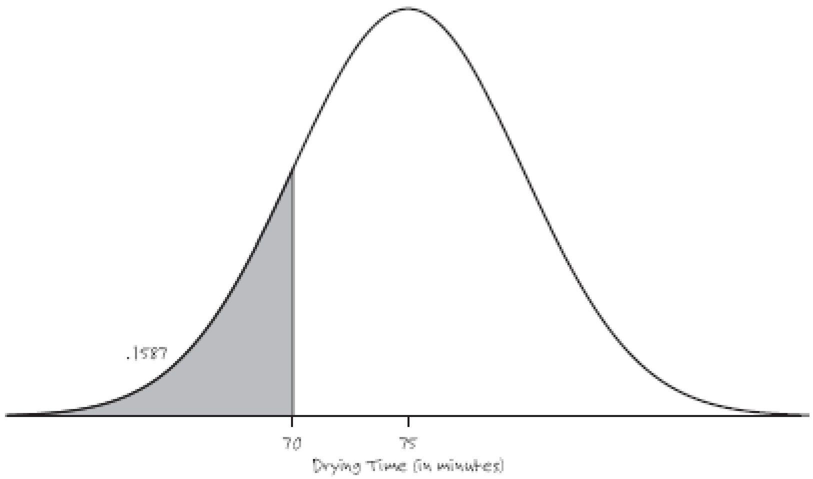
b. The outcomes for which both are female are AD, AG, DG. This probability is 3/21, which is about .143.

c. The outcomes for which the average age exceeds 50 are BF, BG, CF, CG, DF,DG, EF, EG, FG. This probability is 9/21, which is about .429.

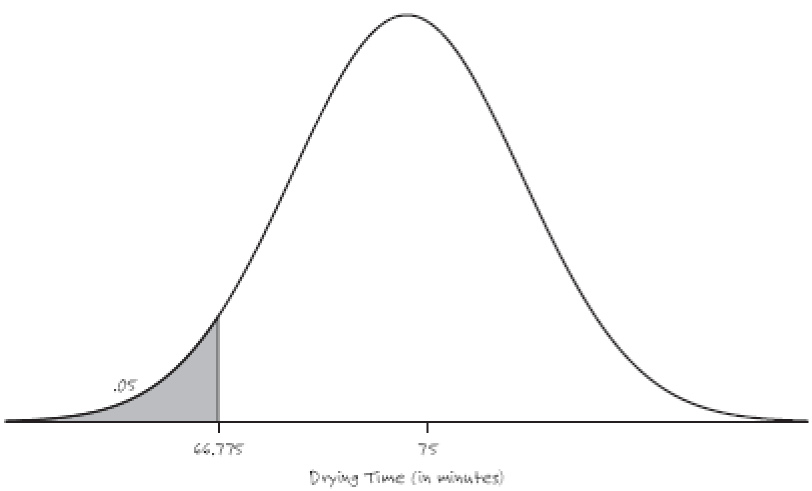
d. The only such outcome is FG, so the probability is 1/21, which is about .048.

e. If the company really selected the two employees by random chance, there would be less than a 5% chance of choosing the two oldest employees. This is certainly possible, but this probability is small enough to cast considerable doubt on the claim that the selection of the two oldest employees was by random chance. Therefore, there is at least moderate evidence to support the allegation of age discrimination.

5. The *z*-score is (70 - 75)/5 = -1.00. The area to the left of -1.00 under the standard normal curve is .1587. A sketch follows:



b. To reduce this probability to .05 requires a *z*-score of -1.645. If you let *k* denote the new cutoff value, you need (*k* - 75)/5 or -1.645, so *k* = 75 - 1.645(5) or 66.775 minutes. A sketch follows:



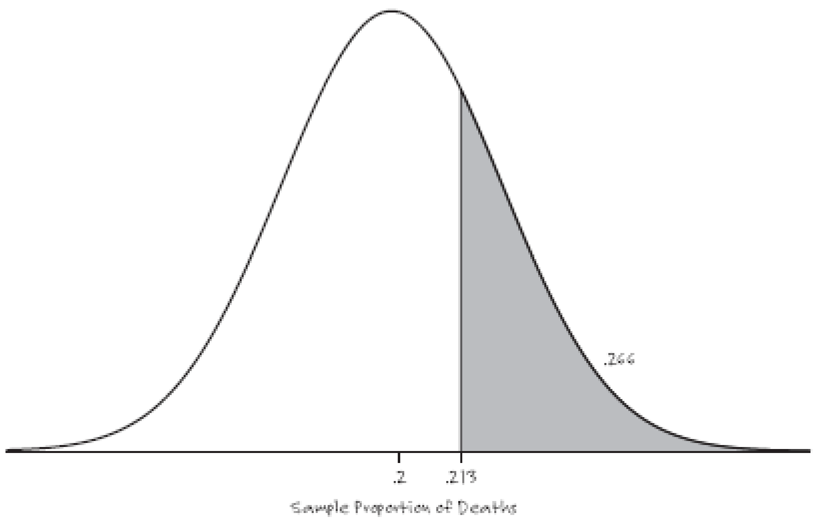
c. With a smaller standard deviation, the error probability in part a would be much smaller. The normal curve would be less spread out (taller and skinnier), so there would be less area to the left of 70.

With a smaller standard deviation, the revised cutoff value in part b would be larger. The normal curve would be less spread out (taller and skinnier), so the point at which 5% of the area is to the left would be closer to the mean and, therefore, a larger cutoff value.

6. According to the Central Limit Theorem, the sample proportion who die in a sample of 371 patients will vary according to a normal distribution, with mean equal to the population proportion of deaths (.20) and with standard deviation equal to

$$\sqrt{\frac{.20(1-.20)}{371}} \approx .0208$$

A sketch follows:



b. The *z*-score is (.213 -.20)/.0208 or 0.63. The area to the left of 0.63 under the standard normal curve is .7357, so the probability is 1 - .7357 or .2643 that the sample proportion of deaths would be .213 or higher.

7. You cannot draw a reasonable conclusion about whether this is a fair coin without knowing the sample size. If the 75% heads was based on a sample of only four flips, then there's no reason to suspect that the coin is not fair. But if the 75% heads was based on a large number of flips, that would suggest that the probability of heads is close to .75 and so the coin is not fair.

8. No, the distribution of house prices would still be skewed to the right. The Central Limit Theorem tells us that the sampling distribution of the sample *mean* house price would be very close to normal, but that result does not apply to the distribution of *individual* house prices.

9. If you were to flip a fair coin until getting the first head a very large number of times, the average number of flips required would be very close to two.

10. $\frac{11!}{2!\cdot 2!} = 9{,}979{,}200$

11. $\frac{6!}{3!\cdot 2!} = 60$

12. (Note: order matters)
    - **a.** 2*(36)=72
    - **b.** 4*(36)=144

13.

a. $\dfrac{{}_4C_2 \cdot {}_4C_2 \cdot {}_{44}C_1}{{}_{52}C_5} = 0.00061$

b. $\dfrac{{}_{13}C_1 \cdot {}_4C_2 \cdot {}_{48}C_3}{{}_{52}C_5} = 0.519$

c. $\dfrac{{}_{13}C_1 \cdot {}_4C_2 \cdot {}_{12}C_1 \cdot {}_4C_3}{{}_{52}C_5} = 0.00144$

d. $\dfrac{{}_{13}C_5}{{}_{52}C_5} = 0.000495$

14. $(-1)\left(\frac{1}{6}\right)+(2)\left(\frac{1}{6}\right)+(-3)\left(\frac{1}{6}\right)+(4)\left(\frac{1}{6}\right)+(-5)\left(\frac{1}{6}\right)+(6)\left(\frac{1}{6}\right)=0.50$

The expected winnings is 50¢. This means if the game were played over and over again then on average you would win 50¢.

15. Note: you win if you roll a 1,2,3, or 6.

a. P(Win) = P(1)+P(2)+P(3)+P(6) = 4/6 = 2/3

b. This is binomial b(7,2/3). $P(X = 5) = {}_7C_5 \cdot \left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^2 = 0.307$

c. 10*2/3 = 20/3 = 6 2/3 wins