

Statistics Final Exam Review Sheet

Hint/Warning: The proportion of problems on this review sheet about probability may be *smaller* than the proportion of probability problems on the final exam itself.

1. According to the US Census bureau, 13.3% of King County residents are of Asian ethnicity. Suppose that we select 50 King County residents at random, using some brilliant and non-biased sampling strategy, and suppose we record the number of people in our sample who are of Asian ethnicity. (You can see a breakdown by specific Asian country at <http://www.cityofseattle.net/oir/datasheet/demographics.htm>)
 - a. What is the expected number of Asian people in our sample?
 - b. What is the most likely number of Asian people in our sample? With what probability will we get that number?
 - c. What is the probability of getting 5, 6, 7, 8, 9, or 10 Asians in our sample?
 - d. What is the probability of getting 15 or more Asians?
2. [Sadly, we have not covered enough material this quarter for me to give a problem on this topic on the final exam, but it is included here because it is interesting.] Story Motors has to crash-test its cars to determine the average repair cost of a 10 mile-per-hour head-on collision. This is expensive, so they decide to try it on just five of their cars. The damage costs are \$150, \$400, \$720, \$500, and \$930.
 - a. What assumptions must we make in order to find a 95% confidence interval for the 'true' average cost of repairs?
 - b. Make those assumptions, and find a 95% confidence interval.
3. In a 1997 Marist Institute survey of 950 randomly selected Americans, 54% of the sample answered "yes" to the question, "Do you think there is intelligent life on other planets?" Use this information to calculate a 95% confidence interval for the proportion of all Americans who believe there is intelligent life on other planets. Explain exactly what you are doing.
4. Suppose we roll a fair die 12 times. Which is more likely: getting exactly two 2's, or the mean die roll being 3 or less? (Hint(?): Assume a Normal distribution is an appropriate model for the sampling distribution of the mean die roll.)
5. The Wechsler Adult Intelligence Scale (WAIS) is the most common "IQ test." The scale of scores is set separately for each age group and is approximately Normal with mean 100 and standard deviation 15.
 - a. Kim Dae-jung, a young Korean statistics student, takes the test and is told that his score is 137. What proportion of all people have WAIS scores above 137?
 - b. Kim has 11 of his Korean friends at college take the test, and all together their average WAIS score is 126.3. Kim claims that this provides strong evidence that "Koreans are smarter than average." Do you agree? Explain.

6. Experimenters at Harvard wanted to test the effectiveness of aspirin in reducing heart attacks. They tested 22,071 subjects (all volunteer doctors) who were randomly assigned to two groups. Group two received one aspirin a day, and group one received a placebo once a day. The participants were then followed for five years, and the investigators recorded for each participant whether or not they had a heart attack during that time. 239 of the 10,795 people in group one had a heart attack, and 139 of the 10,898 people in group two had a heart attack. Does this experiment as described provide strong evidence that aspirin does reduce the risk of heart attack? Explain. In your explanation be sure to be clear about your hypotheses, conditions you need to check, what kind of test you are using, and what your conclusion is.

7. To test an herbal treatment for depression, 100 volunteers who suffered from mild depression were randomly divided into two groups. Each person was given a month's supply of tea bags. For one group, the tea contained the herb mixed with a spice tea, and for the other group, the bags contained only the spice tea. Participants were not told which type of tea they had. They were asked to drink one cup of the tea per day for a month. At the end of the month, a psychologist evaluated them to determine if their mood had improved. The psychologist did not know who had the tea with the herbal ingredient added.

- a. Was this experiment or an observational study? Explain why you think so.
- b. Was a placebo treatment used? Explain.
- c. Was blinding used? Explain.

8. Jon rolls a fair die 50 times and records the average of all his die rolls. Heather rolls a fair die 200 times and records the average of all *her* die rolls.

a. Which of these two events is likelier: that Jon's recorded average is less than 3, or that Heather's recorded average is less than 3? (It is not necessary to actually compute the probabilities of these events to answer this question correctly – in fact, I recommend *not* trying to calculate the probabilities.)

- b. What *are* the probabilities of the two events in question?

9. [Sadly, we have not covered enough material this quarter for me to give a problem on this topic on the final exam, but it is included here because it is interesting.] New Age Granola, Inc., claims that the average weight of its cereal boxes is at least 16 ounces. The TruLife Grocery company will send back a shipment if the average weight is any less. But there are a lot of boxes in each shipment, so the grocery company doesn't want to weigh every single box. So from the next shipment of 10,000 boxes, they take a sample of 49 and weigh each of those. Of those 49, the average weight was 15.90 ounces, and the standard deviation of the sample was .35 ounces. Should the TruLife Grocery company send back the shipment? Explain. In your explanation be sure to be clear about

your hypotheses, conditions you need to check, what kind of test you are using, and what your conclusion is.

10. A hundred participants in a school fundraiser lottery pay \$3 and are assigned a number from 1 to 100 randomly. 10 winners are chosen randomly, and one grand-prize winner is chosen from that group of 10. The grand prize is \$75, and the other prizes are \$10 each.

- a. You enter the lottery and are assigned the number 47. What is the probability that you win the \$75?
- b. What is the probability that you win any money at all?
- c. What is the expected value and standard deviation of your profit?

11. On the planet Zorblot, there are millions of shrovlians. Extensive testing has shown that when a shrovlian is exposed to argonium, there is a 28.5% chance that the shrovlian will artumplate. Also, the average weight of a shrovlian is 13.2 grumplats, and the standard deviation of the weights of all shrovlians is 1.2 grumplats. Suppose we randomly select 100 shrovlians and expose them all to argonium separately. (Assume all the shrovlians are independent.) (Also assume that whether or not a shrovlian artumplates has no effect on its weight.) (You do not need to know anything about the planet Zorblot to do this problem.)

- a. What is the expected number of shrovlians that will artumplate?
- b. What is the probability that the number of shrovlians that artumplate is twenty or less?
- c. What is the probability that the number of shrovlians that artumplate is either twenty or less OR 40 or more?
- d. What is the probability that the average weight of the shrovlians in our sample is less than 13 grumplats?
- e. Your friend independently collected a random sample of 200 shrovlians. What is the probability that the average weight of your sample is more than the average weight of your friend's sample?
- f. What is the probability that the *sum* of the average weight of your sample, and the average weight of your friend's sample, is more than 26.7 grumplats?

12. Can we use probability models based on Bernoulli trials to investigate the following situations? Explain.

- a. Forty percent of all milk chocolate M&M's are either red or yellow. What is the probability of getting 15 red or yellow candies in a package of 40?
- b. You choose M&M candies from a package until you get a red or a yellow, and record the number of candies you have to choose before stopping.
- c. A student studied Bernoulli trials using computer-assisted instruction. After the lesson, the computer presents 10 problems. The student solves each problem and enters his answer. The computer gives additional instruction between problems if the answer is wrong. Is it likely that the student will get at least 80% correct?

d. Opinion polls find that 14% of Americans “never have time to relax.” If you take a simple random sample of 500 adults, what is the probability that 70 people agree with that statement?

13. In the classic (and incredibly nerdy) tabletop role-playing game “Dungeons and Dragons,” a Fireball spell cast by a 20th level wizard will do damage equal to the sum of the rolls of twenty normal six-sided dice. (The unit for damage is a “hit point.”)

- a. What is the maximum amount of damage this spell can do?
- b. What is the expected value of damage, and what is the standard deviation?
- c. (math 211) Suppose an evil dragon has 85 hit points. What is the probability that the fireball will do 85 or more hit points of damage and kill the dragon?

14. [Sadly, we have not covered enough material this quarter for me to give a problem on this topic on the final exam, but it is included here because it is interesting.] A nutrition laboratory tests 40 “reduced sodium” hot dogs, finding that the mean sodium content is 310 mg, with a standard deviation of 36 mg.

- a. Find a 95% confidence interval for the mean sodium content of this brand of hot dog.
- b. What assumptions did you make to do part a? Are the conditions satisfied?
- c. Explain clearly what your interval means.
- d. Explain clearly what the phrase “95% confidence” means in this context.

15. [Sadly, we have not covered enough material this quarter for me to give a problem on this topic on the final exam, but it is included here because it is interesting.] The nutrition lab in the previous problem tests the hot dogs again, this time using a sample of 60 “reduced sodium” hot dogs. The new sample produces a mean of 318 mg of sodium, and the standard deviation is 32 mg.

- a. Should this larger sample of hot dogs produce a more accurate estimate of the mean sodium content?
- b. Find and interpret a 95% confidence interval.
- c. Food labeling regulations require that any food identified as “reduced sodium” must have at least 30% less sodium than its regular counterpart. If regular hot dogs average 465 mg of sodium, should this brand be labeled “reduced”? Explain, using your confidence interval.

16. [This problem is maybe *way* too long to include on a final exam, but is included here as an example of what you can do with the relatively meager tools we developed for linear regression this quarter.] The following table summarizes certain aspects of the 2007 Major League Baseball regular season:

Wins (out of 162)	RS	RA
89	892	821
88	804	750
84	810	733
73	673	783
71	790	891
85	752	690
83	802	776
78	725	829
73	723	813
72	783	853
68	724	846
90	712	732
89	851	750
89	733	657
82	735	727
71	683	720
96	867	657
94	968	777
83	753	699
69	756	868
66	782	944
96	811	704
88	887	797
79	718	725
72	693	839
69	706	778
94	822	731
88	794	813
76	741	758
75	816	844

Each row represents one of the 30 teams. The first column records the number of games they won that season (out of 162). The second column records the total number of runs they scored that season. (For those unfamiliar with baseball: the team with most number of runs at the end of the game wins that game. There are no ties allowed; a game must continue until someone wins.) The third column records the total number of runs their *opponents* scored that season, also referred to as “runs allowed.”

a. Find the mean number of games won for all teams. Also find the median, and the standard deviation.

- b. Find the mean number of runs scored, and the median.
- c. Find the mean number of runs allowed, and the median.
- d. Some people would argue that the key to winning a baseball game is to score lots of runs. Treating runs scored as the explanatory variable and games won as the response variable, determine if there is evidence of a strong relationship between those two variables. (Math 108 people: I would probably say something extra like “make a scatterplot” and “give a numeric measure of the strength of the linear relationship.” Math 211 people wouldn’t get such prompts.) Whether or not it is appropriate to do so, find the least-squares regression line and give its equation.
- e. If you made a scatterplot of wins vs. runs allowed, with runs allowed on the horizontal axis, would you expect the association to be positive or negative? Explain. Then make a scatterplot and see if you’re correct. Find the value of R , the correlation, in this case. Does it look like there is a linear relationship between these two variables?
- f. What if we consider a new list of data: for each team, calculate RS-RA. That is, for each team, subtract the number of runs allowed from the number of runs scored. Some teams will have negative values for this, because they allowed more runs in the season than they scored; we would generally expect those teams to win less games, right? So make a scatterplot and see if there is a strong linear relationship between RS-RA and wins.
- g. Find the least-squares regression line for predicting wins based on RS-RA.
- h. Interpret the slope and y -intercept of the regression line in words. Use appropriate units, if you can.