Math 466 Probability & Statistics Fall 2006 Prof. Janeba Final Exam

Name_____ Score_____

Instructions: This exam is open-book and open-notes. However, consultation is allowed ONLY with the instructor. If electronic aids (Maple, calculator) are used to skip large amounts of hand computation, please note their use at that point, and tell what commands were given to the computer/calculator.

Exam is due at or before the university-appointed final exam period, Tuesday, December 12, from 8-11 am (I will accept papers until noon). We will not otherwise meet for the final exam period.

Part I: Explaining Statistics to the general public:

1. From Freedman's Statistics[1]:
   "In his novel Bomber, Len Deighton argues that a World War II pilot had a 2% chance of being shot down on each mission. So in 50 missions he is 'mathematically certain' to be shot down: $50 \times 2\% = 100\%$. Is this a good argument?"

   Your problem: Write a brief essay, not longer than one side of a page, to answer and explain the above problem for the general public. Your essay will include a different value for the probability of being shot down in 50 missions, an explanation of that different value, and a discussion of the simplifying assumptions that were necessary to find that value, including whether they are likely to be true. Keep in mind, though, that the primary focus of this problem is explaining the issues, as if for a sidebar in a newspaper or magazine article. You may find it useful to make reference to a diagram, for example a diagram of all possible outcomes of a roll of two dice.

2. Freedman's Statistics[2] quotes Jimmy the Greek:
   "The smart professional gambler, when heads comes up four times in a row, will bet that it comes up again. A team that's won six in a row will win seven. He believes in the percentages. The amateur bettor will figure that heads can't come up again, that tails is 'due.' He'll bet that a team on a losing streak is 'due' to win. The amateur believes in the law of averages."
   -Jimmy the Greek, San Francisco Chronicle, July 2, 1975

   Write a brief essay, not longer than one side of a page, discussing this issue. Your essay will include a situation in which the behavior described in the first sentence quoted above is appropriate, and one where it isn't. For this essay, assume the reader is familiar with the concept of independent events, and find a way to fit that concept into your essay.

Part II: Interpreting Statistics in the news:

3. Examine the news item: http://www.cnn.com/2006/US/10/16/iraq.poll/index.html[3], particularly the last three paragraphs.
   (it's an October poll on public approval of Bush's handling of the war in Iraq). Sort out from the article the data you need to give a 95% confidence interval for the percentage of all "adult Americans" who would "approve of how Bush is handling the war." Point out how your answer helps you interpret the last paragraph (we can't be certain about the interpretation, but we can make an educated guess; make it).

(OVER)

[1] Freedman, Pisani & Purves, *Statistics*, 4th ed., W.W.Norton & Co., 2007

[2] Ibid.

[3] [This web page is still on line as of April 29, 2013. Read it to do the problem.]

Part III: Some computational problems

4. A (hypothetical) random sample of 11 Willamette students is asked about their sleeping habits. The sample group averaged 7.1 hours of sleep on the Thursday before classes ended, with a standard deviation $S = 1.35$. If possible, give a 90%-confidence interval for the average number of hours of sleep all Willamette students got on that night. If not possible, tell why not.

5. Suppose our statistics class averaged 703 on the math SAT exam (not the actual numbers) with a standard deviation $S = 63$. For our class, $n = 18$. If possible, use this "data" to give a 90%-confidence interval for the average math SAT score of all current Willamette students. If not possible, explain why not.

6. Whether or not the study in (4) is valid, suppose *another* study finds that a 95% confidence interval for the average number of hours of sleep of Willamette students per night is 6.8 ± 0.3 hours. Write about two paragraphs for the Collegian explaining what this confidence interval does and doesn't mean. Be sure to address whether students should believe the conclusions (based on the process, not on whether the actual numbers are credible from your experience).

Part IV: All the rest

7. An urn contains 2 black balls and 3 white balls. Two balls are drawn sequentially (one is drawn, then the other) **without** replacement.

   (a) List the sample space for this experiment, and the probability of each sample point.

   (b) What is the probability that the second ball is white?

   (c) What is the probability that the second ball is white given that the first is black?

   (d) Show how the answers in (b) and (c) can be reconciled into one calculation that includes them both.

8. Now a more mysterious urn. It contains b black balls and w white balls, where b and w are unknown (but surely positive integers). We iterate the following procedure: Draw a single ball at random. If a white ball is drawn, return *two* white balls to the urn, and if a black ball is drawn, return *two* black balls to the urn. Thus after every iteration, the urn contains one more ball than it originally did.
   Let B_n be the event that the ball drawn on the n^{th} iteration is black.

   (a) Find $P(B_1)$, in terms of b and w.

   (b) Find $P(B_2)$.

   (c) Use what you have learned in (b) to find $P(B_n)$ for arbitrary values of n in terms of b, w, and n.

9. Seven taxpayers are suspected of cheating by the IRS. In fact, all seven have underpaid their taxes by different amounts, and would be caught if audited, but the IRS only has resources to audit three of them. Suppose that three of these taxpayers are randomly selected for auditing. What is the chance that the the IRS catches the third worst offender (but misses the two worst offenders)?

10. Harold has a lot of experience fishing, and he knows that at a certain location and time of day, he averages 7 catches (of a fish) per hour. Fishing in the same location at the same time of day, what is the chance that Harold catches only 4 fish in an hour?

11. Two random variables Y_1 and Y_2 have a joint probability density function given by
$$f(y_1, y_2) = \begin{cases} 24y_2(1-y_1) & 0 \le y_2 \le y_1 \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

    (a) Are these variables independent?

    (b) If $U = Y_1 - Y_2$, what is the pdf for U? (First figure out the possible values of U.)

    (c) Give $E(U)$.

Part IVb: Some Hypothesis testing problems:

12. Discuss the appropriateness of problem 10.8[4]:
"The hourly wages in a particular industry are normally distributed with mean \$13.20 and standard deviation \$2.50. A company in this industry employs 40 workers, paying them an average of \$12.20 per hour. Can this company be accused of paying substandard wages? Use an $\alpha = .01$ level test."
Don't do the problem, just discuss whether a hypothesis test makes sense. Specifically: What assumptions must be made to justify a hypothesis test? Can you think of a way to justify those assumptions?

13. Problem 10.10[5]:
"The Rockwell hardness index for steel is determined by pressing a diamond point into the steel and measuring the depth of penetration. For 50 specimens of a certain type of steel, the Rockwell hardness index averaged 62 with a standard deviation of 8. The manufacturer claims that this type of steel has an average hardness index of at least 64. Is there sufficient evidence to refute the manufacturer's claim at the 1% significance level?"
You may assume that the specimens were randomly selected from all those produced of this type.

14. Problem 10.18a[6]:
"For several years, aspirin-based products lost market share to substitute analgesics such as ibuprofen and acetaminophen. As a result, large companies such as Bayer, which had previously produced only aspirin-based products, began marketing acetaminophen-based pain relievers (*Los Angeles Times*, March 27, 1993). The following table gives the proportions of individuals who preferred each of the three major pain relievers for the years 1986 and 1991.

Pain Reliever	1986	1991
Aspirin	45%	34%
Acetaminophen	41%	41%
Ibuprofen	14%	26%

Assume that these results were obtained on the basis of two independent samples, each of size 1000.

    (a) Does statistically significant evidence exist to suggest that the proportions of aspirin users differ for 1986 and 1991? Use a .05 level test.

    (b) Is there sufficient evidence to indicate that ibuprofen has increased its market share between 1986 and 1991? Use $\alpha = .05$.

    (c) Are the tests performed in (a) and (b) related? Explain."
    [Note for 2013: What is the most appropriate test to determine if the overall market for pain relievers has changed? Do it.]

15. Problem 10.52[7]:
"A coin-operated soft drink machine was designed to discharge, on the average, 7 ounces of beverage per cup. In a test of the machine, ten cupfuls of beverage were drawn from the machine and measured. The mean and standard deviation of the ten measurements were 7.1 ounces and .12 ounce, respectively. Do these data present sufficient evidence to indicate that the mean discharge differs from 7 ounces? What can be said about the attained significance level for this test? [that's a P value. -MJ] What is the appropriate decision if $\alpha = .10$?"
You may assume that the volume of a dispensed cupful is a *random* variable (and the one we are interested in), and that these ten cupfuls were not cherry-picked from a much larger group.

[4] From Wackerly, Mendenhall, & Scheaffer, *Mathematical Statistics with Applications*, Duxbury Advanced Series, 2002.
[5] Ibid
[6] Ibid
[7] Ibid