Final for Statistics 114
Elements of Probability and Statistics - Spring 1998
Material Covered: entire course
5th May

This is a 2 hour final, worth 27% and marked out of 27 points. The total possible points awarded for each question is given in square brackets at the beginning of each question. Anything that can fit on two sides of an $8\frac{1}{2}$ by 11 inch piece of paper may be used as a reference during this quiz. A calculator and appropriate statistical tables may also be used. No other aids are permitted.

Name (please print): ______________________________ . **ID Number:**
.

last first

1[1]**.** A drug treatment program is known to be 80% effective in curing a certain disease. Twelve patients are placed in this program.

(a) [2] This problem could be described in terms of the four conditions which make up a binomial experiment. Give the interpretation of these 4 conditions which make up a binomial experiment *for this problem*[2].

Condition 1. ____.

Condition 2. ____.

Condition 3. ____.

Condition 4. ____.

(b) [1] The probability that at least 10 of the 12 patients in the drug treatment program is cured is closest to:

(i) 0.44 (ii) 0.24 (iii) 0.56 (iv) 0.28 (v) 0.73

---

2. [2] Suppose one individual is chosen at random from 500 individuals who are classified in the following way.

[1] Grosnick, Final, Math 124, May 1996.
[2] It is *not* enough just to state the four general conditions which make up a binomial experiment.

	huffer puffer	even breather
smoker	97	53
nonsmoker	99	251

(a) Given that the individual is a smoker, the chance s/he is a "huffer puffer" is: _.

(b) The chance the individual is a smoker *or* is an "even breather" is: ____________.

3. Consider the following three questions on the correlation coefficient, r. (Each question, (a), (b) and (c), is *separate* from the rest; in other words, they are *not* related to one another.)

(a) [1] Select the most likely value for the linear correlation coefficient, r, for the two variables from among those given.

- x = age (in years) of an individual
- y = number of times a year the individual visits a doctor

(i) 0.98   (ii) 0.82   (iii) 0.07   (iv) -0.65   (v) -1.28

(b) [1] Suppose we used the regression equation $y = -3x + 4$ to generate eight ordered pairs (x, y) and then used these ordered pairs to determine r. The value of r will be,

(i) -3   (ii) -1   (iii) 0   (iv) 1   (v) 2

(c) [1] **True** / **False** Two variables are not (or, at most, very minimally) related if $r = -0.02$.

---

4. [3] Suppose that samples of size 2 are picked, with a sampling with replacement where order matters procedure, from a box of five tickets, numbered 10, 20, 30, 40 and 50, where each ticket has an equal chance of being chosen.

(a) The average of the *box* is given by: ______________________.

(b) The SD of the box is given by: ______________________

(c) The SE of the average is given by: ______________________.

(d) **True** / **False** The SE of the average is *always* equal to the SD of the box divided by the square root of the size of the sample. ______________________.

(e) Determine the sampling distribution for the average by completing the following table:

ave		15							
probability of ave								$\frac{2}{25}$	

(f) A simulation procedure involves imagining (using the random numbers table to) drawing tickets (with replacement, where order matters) from the box of five above. The average of every two tickets drawn is noted and used in constructing an approximate sampling distribution. Suggest an appropriate procedure for choosing numbers from the random numbers table which simulates choosing tickets from the box by completing the following table,

if random number					$9, 0$
then "choose" ticket	10				

5. Consider a nonstandard normal with average of 47.5 and SD of 4.

(a) [1] The chance of getting between 44 and 52 is closest to,

(i) 17% (ii) 34% (iii) 52% (iv) 68% (v) 84%

(b) [1] The probability of getting less than the 52nd percentile is closest to,

(i) 17% (ii) 34% (iii) 52% (iv) 68% (v) 84%

(c) [1] The percentile of 42 is closest to,

(i) 8th (ii) -8th (iii) 92nd (iv) 68th (v) 37th

---

6. [3] The program director for an accounting program wishes to test, at 5%, the hypothesis that her students score higher than the national average of 615 on the final exam. She randomly selects 11 recent graduates of the two–year program and discovers that the average is 630 and the SD is 23.

statement of test of hypotheses: ____________________________________.

test statistic, $t =$ ____.

p–value, $P =$ ____.
conclusion: (circle one) **accept null** / **reject null** because (circle one) $P < 5\%$ / $P < 1\%$ / $P > 5\%$ /

Does it appear the students scored higher than the national average? __________.

---

7. Consider the following data which is being used to determine if the average number of magazines sold is the same or different for three different locations in fifteen stores. For example, 26 magazines (of 100 chosen magazines), located at the front window, were sold in one store; whereas, in another store, 32 magazines (of the 100) were sold at the checkout.

front window	24	26	25	25	30
back of store	26	30	35	40	45
checkout	24	24	32	33	43

(a) [1] An example of a treatment is ____________________________________.

(b) [1] Explain how “store” could be considered an extraneous confounding variable:
.

(c) [1] Randomization is used to rid the experiment of extraneous confounding variables. In this case, the randomization would be used

to assign ____

to ____.

8. [2] At Purdue University North Central, 120 students are randomly selected and asked the distance of their commute to campus[3]. From this group, an average of 9.8 miles is computed. Match the items in Column II with the statistical terms in Column I. (*All* of the items in Column I will be used up in the matching procedure; however, two items in Column II will be left unmatched.)

Column I	Column II
(a) parameter	**(a)** Process used to select 120 students and determine their distance
(b) population	**(b)** 9.8 miles
(c) sample	**(c)** All students at PU/NC
(d) statistic	**(d)** Commute distance for one student
	(e) Average commute distance for all students
	(f) The 120 students
	(g) The 120 commute distances
	(h) 8 mile commute distance for a particular student

---

9. Consider the following results for about 1,000 men. Assume all the assumptions necessary to solve this problem using the normal tables hold in this case.

average height, x = 68 inches SD $\approx$ 2.7 inches
average weight, y = 160 pounds SD $\approx$ 10 pounds $r \approx 0.45$

(a) [1] Use a linear regression to predict the average weight of men who have a height of 69 inches. This prediction is closest to:

(i) 159 (ii) 162 (iii) 163 (iv) 165 (v) 167

(b) [1] The r.m.s. error of the regression line of the weight on the height born is closest to:

(i) 8.5 (ii) 8.9 (iii) 9.1 (iv) 9.5 (v) 10.1

(c) [1] Of the men who have a height of 69 inches, the percentage are under 150 pounds is closest to:

(i) 20% (ii) 90% (iii) 10% (iv) 50% (v) 80%

[3]Based on Zarcone, A.J. and Johnson, R. "Test Bank For Johnson's Elementary Statistics", Duxbury Press, 1996.

10. Consider the following questions on confidence intervals and tests.

(a) [1] What value is always located at the center of a confidence interval for the average?

(i) expected ave (ii) population ave (iii) sample ave (iv) population SD (v) sample SD

(b) [1] **True / False** If our decision in a hypothesis test is to reject the null hypothesis, then we are certain that the null hypothesis is false.

1. (a) 12 patients,cured or not,independence of curing,80% chance cured; (b) 0.56
2. (a) 0.65; (b) 0.802
3. (a) ii; (b) ii; (c) False
4. (a) 30; (b) 14.1; (c) 10; (d) True;
   (e) 10,15,20,25,30,35,40,45,50 and $\frac{1}{25}, \frac{2}{25}, \frac{3}{25}, \frac{4}{25}, \frac{5}{25}, \frac{4}{25}, \frac{3}{25}, \frac{2}{25}, \frac{1}{25}$
   (f) 1,2–10;3,4–20;5,6–30;7,8–40;9,0–50
5. (a) iv; (b) iii; (c) i
6. $H_o : \mu = 615$ vs $H_a : \mu > 615$; $t = 2.05$; between 2.5% and 5%; reject null because $P < 5\%$; scored higher
7. (a) front window; (b) more magazines may be sold in a corner store than grocery store; (c) magazines to location
8. e,c,f,b,g,d
9. (a) ii; (b) ii; (c) iii; (d) iii
10. (a) iii; (b) False