Final for Statistics 503
Statistical Methods In Biology - Fall 1995
Material Covered: entire course
18th December

This is a 2 hour final, worth 34% and marked out of 34 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. The following data is used to test if the mean susceptibility to illness (measured in concentration of lymphocytes in peripheral blood) is the same or different for four different stress levels at $\alpha = 0.05$.

no stress	1.45	1.19	1.05	1.07
mild stress	1.70	2.04	1.49	1.91
moderate	0.21	0.58	0.11	0.27
high stress	1.34	0.99	1.17	1.30

(a) [1] An example of one treatment is: _____.

(b) [1] A t test would not be used in this experiment because: _____.

(c) [1] The response is: _____.

(d) [1] The *number* of experimental units is: _____.

2[1]**.** A study is conducted of the adverse side effects; namely, headache and rash, on 500 patients of a fictitious drug.

	headache →	yes	no	row totals
rash	yes	60	140	200
	no	40	260	300
	column totals	100	400	

(a) [1] Complete the following table if the two variables "headache" and "rash" are assumed to be independent of one another.

	headache →	yes	no	row totals
rash	yes			200
	no			300
	column totals	100	400	

(b) [1] A natural way of thinking about this data is as a (circle one)
two independent samples, one variable / one sample, two variables
contingency table.

(c) [1] The observed value of the chi–square statistic is given by: _____.

(d) [1] The p–value is given by: _____.

3[2]**.** In a study to determine the density of nerve cells at specified sites in the intestines of horses, the following data was collected.

site 1	50.6	39.2	35.2	17.0	11.2	14.2	24.2	37.4	35.2
site 2	38.0	18.6	23.2	19.0	6.6	16.4	14.4	37.6	24.4

(a) [1] This is a (circle one) **independent samples design / paired design** which involves (circle one) **blocking / repeated measures / blocking over time**.

(b) [1] Calculate a 95% confidence interval: _____.

(c) [1] The p–value for a nondirectional test at $\alpha = 0.05$ is: _____.

(d) [1] The t critical value for the test in part (c) is: _____.

[1]Samuels, 11.40, p 385, 1989
[2]Samuels, 9.28, p 323, 1989

4[3]**.** An experiment was carried out to determine the effect of providing free milk to school children in Lanarkshire, Scotland. The children were assigned at random either to receive or not to receive milk.

(a) [1] This is most likely a **controlled experiment** / **observed study** because:

(b) [1] Describe the sample.

(c) [1] What, do you think, is probably the purpose of this experiment?

(d) [1] Describe and compare both a possible statistic and parameter.

5[4]**.** Coumaric acid is a compound that may play a role in disease resistance corn. Consider a test for the concentration of coumaric acid in corn seedlings: "$H_o : \mu = 106$ versus $H_a : \mu = 102$" where $\sigma = 2$ and normality is assumed.

(a) [1] A type I error occurs, *in this case*, if:

(b) [1] The probability of a type II error, assuming the null is rejected if $\bar{X} < 103$, is:

(c) [1] In this example, the probability of the type I error increases if the probability of the type II error (circle one) **increases** / **remains the same** / **decreases**.

---

[3]Freedman et. al, 5, p 23, 1993.
[4]Samuels, 7.46, p 235, 1989

6. Assume that 40% of babies born have blue eyes and each birth is independent of one another and, in general, this problem obeys the conditions of a binomial experiment where $n = 3$. The probability of x ($x = 0, 1, 2, 3$) of the next $n = 3$ children having blue eyes is given by:

x	0	1	2	3
$P(X = x)$	0.216	0.432	0.288	0.064

(a) [2] Using the TI–83 calculator, *with seed 2*, and following the simulation procedure,

if random number	1,2	3,4,5,6	7,8,9	0
then "count" blue–eyed babies born	0	1	2	3

collect 5 *pairs* of numbers (10 numbers in all). Determine the average of these 5 pairs and display your results in the following table.

$\bar{x}$							
proportion (out of 5)							

(b) [1] Suggest a better simulation procedure (which uses three digits) than given in part (a) above, by filling in the following table,

if random number				
then "count" blue–eyed babies born	0	1	2	3

---

7. Flip a coin 25 times. Assume there is a 30% chance of getting a head in one flip of a coin and that each flip is independent of each other flip.

(a) [1] The probability of getting 2, 3 or 4 heads is closest to:

(i) 0.091 (ii) 0.099 (iii) 0.081 (iv) 0.088 (v) 0.101

(b) [1] Use the normal to approximate the exact probability determined in part (a). This probability is closest to:

(i) 0.091 (ii) 0.099 (iii) 0.081 (iv) 0.088 (v) 0.101

(c) [1] Since np and $n(1 - p)$ are both (circle one)

greater than / the same as / less than

five, this implies the approximation in part (b) should be (circle one)

close to / the same as / better than / very different than

the exact answer in part (a).

8. Consider an ordinary deck of 52 cards, with four *suits* (spades, hearts, clubs and diamonds), where each suit has 13 cards (two, three, . . . , ten, jack, queen, king and ace).

(a) [1] Three cards are drawn with*out* replacement, where order matters and at random. The chance of getting three jacks is closest to:

(i) $\frac{24}{132{,}600}$ (ii) $\frac{64}{132{,}600}$ (iii) $\frac{2{,}197}{132{,}600}$ (iv) $\frac{1{,}716}{132{,}600}$ (v) $\frac{3}{52}$

(b) [1] Three cards are drawn *with* replacement, where order matters and at random. The chance of getting three jacks is closest to:

(i) $\frac{24}{132{,}600}$ (ii) $\frac{64}{132{,}600}$ (iii) $\frac{2{,}197}{132{,}600}$ (iv) $\frac{1{,}716}{132{,}600}$ (v) $\frac{3}{52}$

(c) [1] Three cards are drawn with replacement, where order matters and at random. The chance of getting two jacks and a queen is closest to:

(i) $\frac{24}{132{,}600}$ (ii) $\frac{64}{132{,}600}$ (iii) $\frac{2{,}197}{132{,}600}$ (iv) $\frac{1{,}716}{132{,}600}$ (v) $\frac{3}{52}$

9[5]**.** From an area of 21 acres, an ecologist noted the presence or absence of maples and hickories in each of 144 randomly selected plots of 38 feet square each:

	Maples →	present	absent	subtotals
Hickories	present	26	63	89
	absent	29	26	55
	subtotals	55	89	144

(a) [1] In a randomly chosen plot, the probability maples were absent is closest to:

(i) 47% (ii) 31% (iii) 62% (iv) 24% (v) 18%

(b) [1] In a randomly chosen plot, the probability the maples were absent, given that the hickories were absent, is closest to:

(i) 47% (ii) 31% (iii) 62% (iv) 24% (v) 18%

(c) [1] In a randomly chosen plot, the probability the maples and hickories were absent is closest to:

(i) 47% (ii) 31% (iii) 62% (iv) 24% (v) 18%

[5] Samuels, 11.19, p 139, 1989

10[6]**.** Infestation of crops by insects has long been of great concern to farmers and agricultural scientists. Below, is data on the age of a cotton plant (days), x, and percent damaged squares, y.

x	9	12	12	15	18	18	21	21	27	30	30	33
y	11	12	23	30	29	52	41	65	60	72	84	93

(a) [1] Give the scatter plot below.

(b) [1] The correlation coefficient is: _____.

(c) [1] The correlation coefficient (circle one) **remains the same** / **changes** if the age of cotton plants is measured in weeks rather than days.

[6] Devore and Peck, 4.30, p 155, 1993

(1) (a) no stress; (b) more than two treatments; (c) susceptibility to illness; (d) 16.

(2) (a) 40, 160, 200; 60, 240, 300; 100, 400, 500; (b) one sample, two variables; (c) 20.8; (d) 0.

(3) (a) paired design; (b) blocking; (c) (1.3, 13.3); (d) 0.22; (e) 2.31.

(4) (a) controlled experiment; (b) children given free milk; (c) does free milk improve children's health? (d) average difference in health before and after giving children free milk for sampled children (statistic) or for children in Lanarkshire (population).

(5) (a) 106 accidentally rejected; (b) 0.31; (c) decreases.

(6) (a) 0, 0.5, 1, 1.5, 2, 2.5, 3; 0.2, 0.2, 0.2, 0.4, 0, 0, 0; (b) 001–216, 217–648, 649–936, 937–000; 0, 1, 2, 3.

(7) (a) 0.088; (b) 0.091; (c) greater than, close to

(8) (a) i; (b) ii; (c) ii.

(9) (a) 62%; (b) 47%; (c) 18%.

(10) (a) positive moderate linear scatter plot; (b) 0.949; (c) remains the same