Final for Statistics 503
Statistical Methods In Biology - Fall 1997
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[1]**.** The following data to 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] Fill in the following ANOVA table.

Source	Sum of Square	Degrees of Freedom	Mean Squares
Between Groups			
Within Groups			
Total			

(b) [1] The statement of the test is given by: ____________________________.

(c) [1] The p–value is given by: ____________________________________.

(d) [1] We will (circle one) **accept** / **reject** the null hypothesis

because: ____.

[1] Samuels, 12.10, p 407, 1989.

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

(a) [2] Fill in the following table with fictitious data that agrees with the statement: "Headache and rash occur independently of each other" and verify your answer by calculating and comparing two appropriate proportions from your data.

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

one proportion (give both in a fraction and in a decimal form): ________________.

another proportion (give both in a fraction and in a decimal form): ________________.

(b) [2] Fill in the following table with fictitious data that agrees with the statement: "A person who experiences headache is more likely (than a person without headache) to experience a rash" and verify your answer by calculating and comparing two appropriate proportions from your data.

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

one proportion (give both in a fraction and in a decimal form): ________________.

another proportion (give both in a fraction and in a decimal form): ________________.

[2]Samuels, 11.40, p 385, 1989

3[3]**.** 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] Calculate a pooled 95% confidence interval, assuming the data is *not* paired: .

(b) [2] The 95% confidence interval, when the data is assumed to be paired, is (circle one) **larger than** / **smaller than** / **the same width as** the 95% confidence interval, when the data is assumed to be not paired, because

(i) the standard deviation in the paired case is smaller than in the unpaired case.
(ii) the t test statistic in the paired case is larger than in the unpaired case.
(iii) the sample sizes for the two sites are equal.
(iv) the pooled standard deviation falls inbetween the standard deviations for each site
(v) of reasons (i) and (ii) above.

(c) [1] A nondirectional t test will give the same results as an F test if the data is (circle one) **paired** / **not paired**.

4[4]**.** A *portacaval shunt* operation is sometimes used to deal with patients with cirrhosis of the liver who start to hemorrhage. The results of 51 different kinds of study are given in the table below.

	patients improvement →	marked	moderate	none
design	no controls	24	7	1
of	controls, but not randomized	10	3	2
study	randomized controlled	0	1	3

(a) [1] Explain what it means, in this case, to say “no controls”. What is a possible control here?

(b) [1] How could the studies in this case not be randomized, if there was a control? What else could have been done?

[3]Samuels, 9.28, p 323, 1989
[4]Freedman et. al, p 7, 1993.

(c) [1] Randomization deals with extraneous, possibly confounding, variables. What is a possible extraneous confounding variable in this case?

(d) [1] Is it possible to double–blind here? Explain.

5[5]**.** 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 $\mu \neq 106$" where $\sigma = 4$ and normality is assumed.

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

(b) [1] The probability of a type I error, assuming the null is rejected if $\bar{X} < 103$ or if $\bar{X} > 109$, is:

(c) [1] In this example, as the sample size increases, the

(i) probability of a type I error decreases, probability of a type II error increases
(ii) probability of a type I error increases, probability of type II error increases
(iii) probability of a type I error decreases, probability of type II error decreases
(iv) probability of a type I error increases, probability of type II error decreases
(iv) not enough information to say

---

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 $n = 3$.

(a) [1] The probability distribution is given by (circle one):

(i) $\begin{pmatrix} 4 \\ x \end{pmatrix} (0.4)^x (0.6)^{10-x} \quad x = 0, 1, 2, \ldots 10$

(ii)

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

(iii) $P(X = 3) = 0.064$
(iv) 40%
(v) not enough information

(b) [1] Construct the sampling distribution for $\bar{X}$, based on the *random* sample of size 2, by filling in the following table.

$\bar{x}$							
$P(\bar{X} = \bar{x})$							

[5]Samuels, 7.46, p 235, 1989

(c) [1] Determine the following.

$E(X) =$ ____.

$V(X) =$ ____.

7. Flip a coin 2500 times. Assume the probability of getting a head in one flip of a coin is 0.3 and that each flip is independent of each other flip.

(a) [1] A version of the central limit theorem tells us that if np and $n(1-p)$ are both larger than 5, the normal distribution can be used to approximate the binomial distribution where the normal is given by:

(i) N(750, 0.3)
(ii) N(750, 525)
(iii) N(750, 22.9)
(iv) N(2500, 0.3)
(v) N(5, 0.3)

and correction factors are used.

(b) [1] The probability of getting at most 1545 heads is closest to:

(i) 0.999 (ii) 0.099 (iii) 0.325 (iv) 0.088 (v) 0.473

(c) [1] The empirical rule tells us: ______________________________.

____

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] Two cards are drawn without replacement, where order matters and at random. The chance of getting a jack and a queen is closest to:

(i) $\frac{16}{2652}$ (ii) $\frac{51}{2652}$ (iii) $\frac{32}{2652}$ (iv) $\frac{52}{2652}$ (v) $\frac{4}{51}$

(b) [1] Two cards are drawn without replacement, where order matters and at random. The chance of getting a jack *followed by* a queen, *in that order*, is closest to:

(i) $\frac{16}{2652}$ (ii) $\frac{51}{2652}$ (iii) $\frac{32}{2652}$ (iv) $\frac{52}{2652}$ (v) $\frac{4}{51}$

(c) [1] Two cards are drawn without replacement, where order matters and at random. The chance of getting a jack *given that* (or conditional on) a queen has already been drawn, is closest to:

(i) $\frac{16}{2652}$ (ii) $\frac{51}{2652}$ (iii) $\frac{32}{2652}$ (iv) $\frac{52}{2652}$ (v) $\frac{4}{51}$

9[6]**.** 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:

	Maples $\rightarrow$	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 hickories were absent is closest to:

(*i*) 47% (*ii*) 38% (*iii*) 82% (*iv*) 24% (*v*) 18%

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

(*i*) 47% (*ii*) 38% (*iii*) 82% (*iv*) 24% (*v*) 18%

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

(*i*) 47% (*ii*) 38% (*iii*) 82% (*iv*) 24% (*v*) 18%

10[7]**.** 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] Describe the scatter plot for this data: ______________________________.

(b) [1] The least squares line is: ______________________________.

(c) [1] The predicted percentage of damaged squares when the age is 20 days is: __.

[6]Samuels, 11.19, p 139, 1989

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

(1) (a) 4.55, 3, 1.52; 0.474, 12, 0.0395; 5.02, 15; (b) means same vs means different; (c) 0; (d) reject, p–value smaller than α

(2) (a) $40/60 = 0.67$ and $160/240 = 0.67$ (b) $60/40 = 1.5$ and $140/260 = 0.54$.

(3) (a) (-4.578, 19.245); (b) smaller than, i; (c) not paired

(4) (a) portacaval shunt operation was not compared against other different kinds of heart operations that accomplish what the shunt operation accomplishes; (b) patients could have been assigned by "clinical judgement"; (c) health of patient; (d) not really possible (unethical).

(5) (a) alternative accidentally rejected; (b) 0.45; (c) iii.

(6) (a) ii; (b) 0, 0.5, 1, 1.5, 2, 2.5, 3; 0.047, 0.186, 0.186, 0.056, 0.036, 0.004; (c) 1.2, 0.24

(7) (a) ii; (b) i; (c) 67% of data within one SD of average

(8) (a) iii; (b) i; (c) v.

(9) (a) ii; (b) iii; (c) i.

(10) (a) positive moderate linear scatter plot; (b) $\bar{y} = -19.676 + 3.285x$ (c) 46.024