old exams

1

Biostatistics Midterm 1 1999 50pts total

I. Answer the following pertaining to the experiment and data given below. 8pts

An investigator wished to determine whether individual that have been exposed to high altitudes have higher hemoglobin levels than the normal population, which has a mean of 15.80mgp/ml with a standard deviation of 5. She measured the hemoglobin levels on 25 randomly selected adult males from a remote Andean village with the following results: The data is given in the order in which it was collected.

17.7	17	17.5	16	17.6
17.1	17.4	17.5	17.3	16.9
16.9	17.2	15.9	16	17.1
15.4	16.3	17	16.1	16.8
17	16.4	16.6	15.5	17.2

1. What is the population about which she can make legitimate inferences based on this sample?

adult ♂ living in that one village

2. Has the data been collected to a sufficient number of decimals? Explain.

cycling ortho

17.7
− 15.4
2.3 → 23 units

not between 30 & 300
too few decimals

3. Have a sufficient number of measurements been made? Explain.

yes do cumulative running mean − it is < 5%

no? − compare with 15.80 − looks pretty different but
$\sigma_{\bar{x}} = 5/\sqrt{25} = 1$ ∴ $15.80 \pm 1.96(1) \rightarrow 17.80$
our mean is within this range − may be results are not clear cut

4. What is wrong with the way the data was collected? How would you correct this?

pseudo replication use > 1 village
to compare with normal population must use ♂s

II. Calculate the following. Assume a normally distributed population. 9pts

1. Find 95% confidence limits for a sample mean of 12.4 from a sample of size 25 with a standard deviation of 2.9. 3pts

$\bar{X} \pm t_{.05(2)[24]}(s_{\bar{X}})$

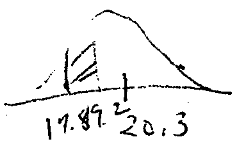
$12.4 \pm 2.064(2.9/\sqrt{25}) =$

$12.4 \pm 1.2 \quad (12.4 \pm 1.19)$

2. Find the probability of selecting a plant between 17.8 and 19.2 cm tall from a population with a mean of 20.3 and a standard deviation of 1.2. 3pts

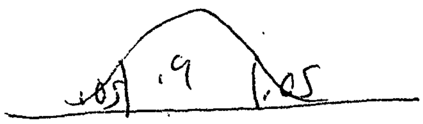
$z = \frac{17.8 - 20.3}{1.2} = -2.08 \qquad z = \frac{19.2 - 20.3}{1.2} = .92$

$p = .0188 \qquad p = .1788$



$.1788 - .0188 = .1600$

3. Which two points enclose the central 90% of all values from the unit normal distribution? 1pt



$z = \pm 1.645$

2006 must also be able to do problems with binomial & Poisson

4. Find the probability of obtaining a value of t ≥ -.685 with 23 degrees of freedom. 2 pts

from Table $t_{\alpha(1)[23]}$

.75

.75

III. Give the name of the sampling distribution, the test statistic formula, and the null and alternate hypotheses in words and symbols that should be used to solve the following. Assume a normally distributed population. 4pts

1. To test the hypothesis that mothers with low socioeconomic status (SES) deliver babies that are lower than normal the birth weights of 100 babies of low SES were measured. They were found to have a mean of 115 oz. with a

1/2 for words

standard deviation of 24 oz. The mean birth weight for the normal population is known to be 120 oz with a standard deviation of 25 oz.

Normal dist H_0 $\mu \geq 120$ not lower than normal
H_1 $\mu < 120$ lower than normal

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

2006 only z for hypothesis testing

2. A particular subspecies thought to be extinct has a mean bill length of 7.8mm with a standard deviation of 1.2 based on a sample of 47 individuals. An individual is found with a bill length of 5.9. Does this individual differ in length from the population?

t dist $t = \frac{X - \mu}{s}$

$H_0: \mu = 7.8$ individual does not differ
$H_1: \mu \neq 7.8$ individual differs

IV. Answer the following. 18pts

1/2

1. Which statistic should be used to indicate the following? Give the name and symbol where appropriate. Explain briefly. 6pts

a. Central tendency of movie quality rated on a scale of 1-5 stars

ordinal data ∴ median 1/2 for properties

b. Population dispersion of blood pressure

s measures dispersion of Xs; essentially unbiased & easier to interpret than s^2

c. Central tendency of femur lengths with values ranging from 5 to 8 cm

$\bar{X}$, mean fairly symmetrical measurement data ⇒ mean lowest sampling error includes values of all data equally

2. Write the formula for the statistical variance and explain why it is calculated the way it is. 3pt

formula 1 $\frac{\Sigma(X-\bar{X})^2}{n-1}$ unbiased otherwise would underestimate σ^2

unbiased 1

because range of sample < range of population

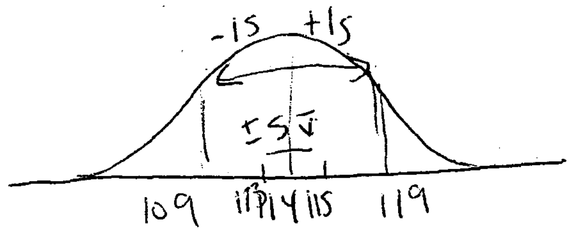
df etc 1

lose 1 df in estimating $\bar{X}$

3. For a sample of size 25 with a mean of 114 and a standard deviation of 5 sketch a graph indicating the mean , the mean ± 1 standard deviation and the mean ± 1 standard error. 2pts

no values 1/2 1 each

$s_{\bar{X}} = 1$



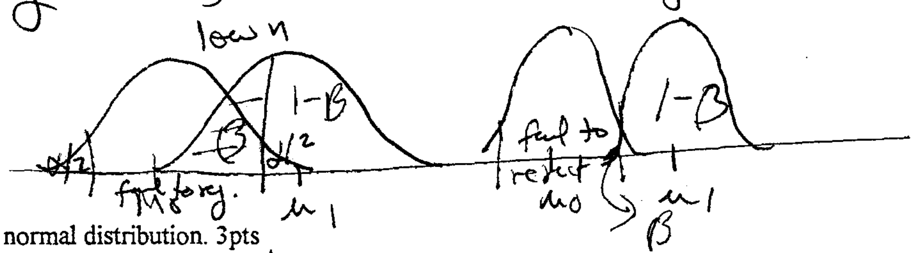
/14

4. What effect does increasing sample size have on Type II error? Illustrate using a sketch that indicates β and power for small and large n. 3pts

answer ↓ Type II 1

narrows sampling dist of $\bar{X}$ 1
or dep

graph 1



5. List the properties of the (Standard) unit normal distribution. 3pts

1 st 2 pts

-1/2 each

continuous

$-\infty$ to $+\infty$

symmetrical unimodal

$\mu = 0$

$\sigma = 1$

6. Why are random samples preferable to regular ones? 1pt

useful in further statistical analyses

5

V. Test the hypothesis that a particular species of plant should be shorter when grown at high altitudes given that the mean of the lowland members of the species is known to be 30.6 cm with a variance of 12.3 cm^2. A sample of 50 individuals from the high altitude population has a man of 32.2. 11pts

1. Write the null and alternate hypotheses in symbols. 1pt

$H_0: \mu \geq 30.6$

$H_1: \mu < 30.6$

2. State the level of significance. 1pt

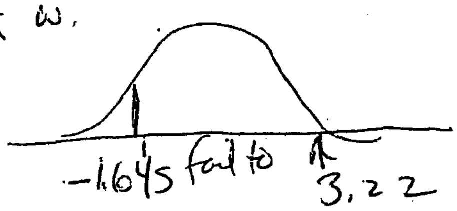
$\alpha = 0.05$

3. Calculate the appropriate test statistic. 2pts

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = 3.22$$

4. Sketch the regions of reject and fail to reject on a graph indicating the critical value and indicate the value of your calculated test statistic. 3pts

1/2 if consistent w. 2 tails



1.645 1pt
stat 1pt
reject f to 1pt

5. Find the probability of obtaining a value as deviate as your calculated value or worse. 1pt

1 - p in

$1 - .0006 = 0.9994$

6. State your statistical conclusions. 1pt

fail to reject

7. Verbalize your conclusions. 2pts

1 proper wording - consistent with #6 full

Here you can say no evidence that they are shorter.

Biostatistics Midterm 1 2001

I. Answer the following pertaining to the experiment and data given below. 62 pts

An experiment was performed to determine whether blood acetocholine levels are elevated in birds nesting within 1km. fields sprayed with an organophosphorous insecticide. Sixteen birds exposed to spray were measured with the results shown below. These are to be compared with birds not exposed to spray which are known to have an average of 18.1. The data is given in the order in which it is taken.

Acetocholine concentration in mg% per liter X 10^{-3}

21	14	17	11
13	16	19	17
16	16	15	18
15	16	17	14

21	17	$\boxed{16}$	14
19	17	16	14
18	16	15	13
17	$\boxed{16}$	15	11

1. Find : 16 pts

a. Mean 3 pts

$$\bar{X} = \frac{\Sigma X}{16} = \frac{255}{16} = 15.9375$$

b. Mode 3 pts

most frequently occuring value = 16

c. Median 4pts

$$\frac{n+1}{2} = 8.5 \Rightarrow \text{use average of } 8^{th} \text{ and } 9^{th} \text{ value}$$

$$\frac{16+16}{2} = 16$$

d. 95% confidence limits (the standard deviation of this sample =2.38) 6 pts

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{2.38}{\sqrt{16}} = 0.595$$

$$CL = \bar{X} \pm t_{0.05(2),15} S_{\bar{x}}$$

$$= \bar{X} \pm (0.595)(2.131)$$

$$= \bar{X} \pm 1.2679 = 15.9375 \pm 1.2679$$

2. What are the independent and dependent variables? On what scale is each measured? 8 pts

dependent : blood acetocholine levels (ratio scale ; continuous data)

independent : proximity to organophosphorous insecticide (nominal ; either close or not close (within 1km)

3. Test to determine whether levels are in fact elevated in birds from the sprayed area. 22pts

a. Write the null and alternate hypotheses in words and symbols 4 pts

$H_0 : \mu_x \leq \mu_0 = \mu_x \leq 18.1$ the levels are not elevated

$H_1 : \mu_x > \mu_0 = \mu_x > 18.1$ the levels are elevated

b. Central tendency for species (ex. D. pulex, D. rosea, D. dubia)

mode

– needs to be used due to nature of data (nominal)

c. Central tendency for lifespan

lifespan high + skew

~~mode~~ ~~[illegible]~~ ~~median~~ ~~Mean X~~ Median

highly skewed

~~– ratio scale data; continuous~~

~~– [illegible] values far from middle~~

(such as infant death)

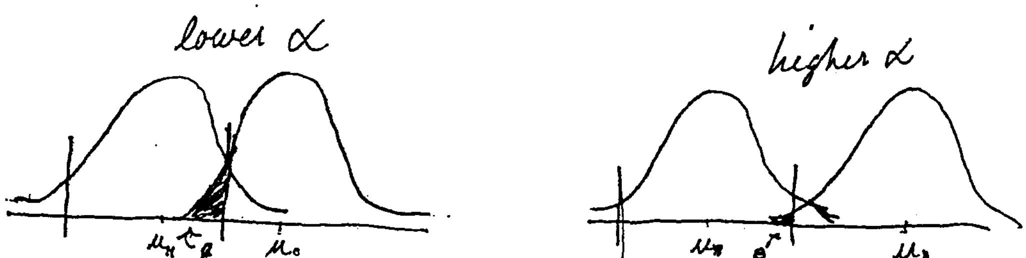
2. Write the formula for the statistical variance and explain why it is calculated the way it is. 6 pts

$$S^2 = \frac{\Sigma (X - \bar{X})^2}{n-1}$$

* the formula for statistical variance replaces the N in the denominator of parametric variance with n-1. This is to eliminate the bias which would occur if $\bar{X}$ was used to approximate μ. We are removing one degree of freedom to account for this discrepancy

3. What effect does decreasing α have on β? Power? Sketch β and power for two different levels of α and label. 6 pts

* decreasing α increases β, and therefore decreases power



4. Which properties of the t distribution differ from those of the Normal? Include the relationship between t and Normal. 6 pts

NORMAL	T-DIST
• $\mu = 0$; $\sigma = 1$	• $\mu = 0$; $\sigma > 1$ (more spread out)
	• many df

* all other properties are the sa (symmetrical, unimo $-\infty \to \infty$, mean=med=m

* as df → ∞ ; the t-distribution = normal (as sample size

b. State the level of significance 2pts $\alpha = 0.05$

c. Find the appropriate statistic 4pts

do not know σ, only s $\Rightarrow$ use t dist. - 2006 we will have only z on mt 1

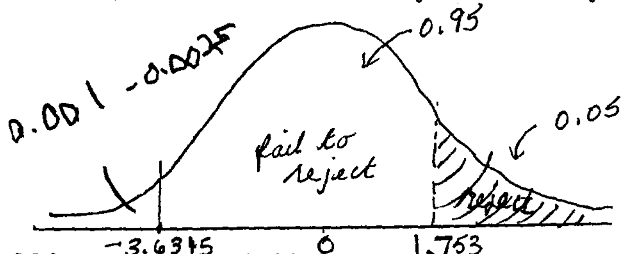
$$t = \frac{\bar{X} - \mu}{S_{\bar{x}}}$$

$P = 1 - 0.001, .0025, .999$
$1 - 0.9025$

d. Find p 2 pts

$$S_{\bar{x}} = \frac{2.38}{\sqrt{16}} = 0.595 \qquad t = \frac{15.9375 - 18.1}{0.595} = -3.6345$$

e. Find the critical value and sketch the reject and fail to reject regions; locate your statistic on the graph. 6pts



$v = 15$

f. Make your statistical decision. 2 pts

fail to reject null hypothesis (H_0)

g. Verbalize. 2pts

There is not significant evidence that blood acetocholine levels are elevated in birds nesting within 1 km of fields sprayed with the insecticide

4. Has the data been collected to a sufficient number of decimals? Explain. 4 pts

21
− 11
10 → not between 30 and 300

21.0
− 11.0
10.0 → valid

∴ the data should have been collected using 1 more decimal place

5. Have a sufficient number of measurements been made? Explain. 4 pts

do ~~exxx~~ running mean ⇒ less than 5% difference between classes

∴ yes, a sufficient amount of measurements has been made

6. List at least 4 possible confounding variables that must be controlled. 4 pts

– sex, diet, time of day, age of birds

7. What must the investigator do to avoid pseudoreplication? 4 pts

use data from more than one field

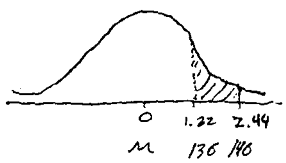
II. Answer the following. Assume a normally distributed population. 8 pts

1. Given that blood pressure has a mean of 120 and a standard deviation of 8.2, what is the probability of finding an individual with a value between 130 and 140? 4pts

$\mu = 120$

$\sigma = 8.2$

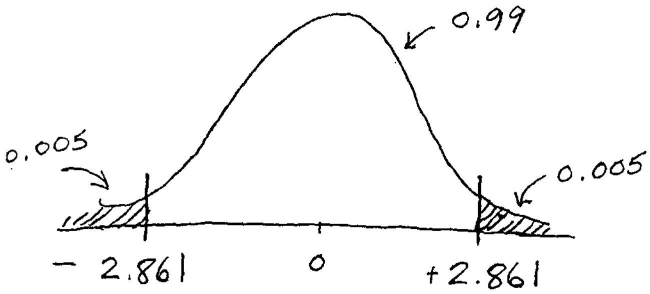
$Z_{130} = \frac{X - \mu}{\sigma} = \frac{130 - 120}{8.2} = 1.22$



$Z_{140} = \frac{140 - 120}{8.2} = 2.44$

$p = 0.1112 - 0.0073 = \underline{0.1039}$

2. What Z values enclose the central 99 percent of all values from a t distribution with 19 degrees of freedom? 4 pts



$\nu = 19$

$t_{0.01(2),19} = \pm 2.861$

III. Answer the following. 30 pts

1. Which **statistic** should be used to indicate the following? Give the name and symbol where appropriate. Explain briefly. 12 pts

a. Population dispersion for femur lengths

s = standard deviation

a measure of the average deviation from the mean ⇒ easier to interpret than s^2 (variance).

Biostatistics Midterm I: 2002 87pts

I. Answer the following pertaining to the experiment described below. 24pts

A study was performed to determine the effect of photoperiod in inducing the formation of resting eggs in the copepod *Diaptomus leptopus*. Thirty females from Stoneycroft Pond at the Morgan Arboretum were randomly divided into four treatment groups: 10, 12, and 14 hrs of light in a 24 hr period respectively. The **proportion** of resting egg clutches each female produced was recorded.

Female 1	10 hrs light	12 hrs light	14 hrs light
1	0.00	0.25	0.75
2	0.167	0.5	1.00
3	0.25	0.67	0.67
4	0.00	0.33	1.00
5	0.167	0.25	0.80
6	0.00	0.33	0.50
7	0.00	0.50	0.75
8	0.20	0.25	1.00
9	0.167	0.75	1.00
10	0.20	0.33	0.80

2 2

1. What is the dependent variable? On which type of scale is it measured? 4pts

proportion of clutches with resting eggs ratio

2 2

2. What is the independent variable? On which type of scale is it measured? 4pts

photoperiod ratio

3. List 4 confounding variables that must be controlled. 4pts 1 each

amount of light ~~food~~ age of copepod
temperature genetics of copepod density (container size)

4. What is the largest population about which the experimenter can make a legitimate inference based on this experiment? 2pts

the females of Stoneycroft pond

5. How will the experimenter avoid pseudoreplication? 2pts

each ♀ in own container
if want to infer about more than that pond
use > 1 pond

11 —

6. Is the sample size large enough? Explain. 4pts

from comparing the groups for a clearcut difference — 2 for doing it yes 2pts 2 running means must differ by less than 5%

if do running means on each group at n=5, n=10 yes

if do running mean at n=9 & n=10 for last: no

7. Which measure of central tendency would you expect to be most appropriate for this type of data? Explain briefly. 4pts

2 2

median - proportions are binomially distributed & expect to be skewed if $p > .7$ or $p < .3$

II. Answer the following pertaining to the distribution below. 7pts

Number of flowers per plant in 32 *Dubius speciosa*

Midpoint	Class limits	Frequency	Relative Frequency
7	6 - 8	1	1/32
10	9 - 11	4 5	4/32
13		5 10	
16	2pts	6 16	1pt
19		6 17	
22		4	
25	showing	5	
28	discrete (1pt)	1	

1. Calculate the class limits and relative frequencies for the first two classes. 3 pts

2. Find the median. 2pts

median : $\frac{32+1}{2} = 16.5$ 16th - 16 17th obs - 19 av = 17.5

3. Describe the shape of the distribution using the appropriate statistical jargon. 2pts

very little skew but platykurtic

III. Answer briefly. 15pts

2 2

1. Write the formula for statistical variance and explain briefly why it is calculated that way. 4pts

$$s^2 = \frac{\Sigma(X-\bar{X})^2}{n-1}$$

n-1 to eliminate bias ÷ n is underestimate because range of sample <<< of population ∴ ↓ denom

lose 1 degree of freedom in

13

2. Write the properties of the t distribution that differ from the normal. 3pts

$1\frac{1}{2}$ each $\sigma > 1$ for t

different distribution with each df

($t_{[\infty]}$ = normal)

3. Why are 95% confidence limits those most commonly calculated? 3pts

compromise between accuracy (high %) and

precision (narrow limits) $2\frac{1}{2}$ if don't explain

4. Define the sampling distribution of the mean. 3pts

Dist. of sample $\bar{X}$s from all possible random

samples of same size from same n.

5. What is the purpose of a paired design t test? 2pts

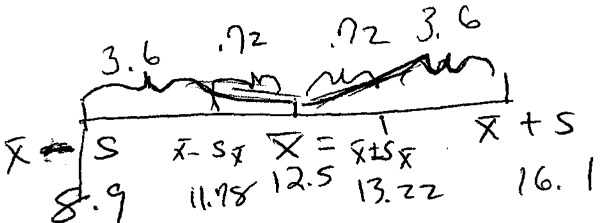
It allows you to eliminate a major confounding

variable when comparing 2 groups. and is usually more powerful

IV. Sketch and answer the following. 13pts

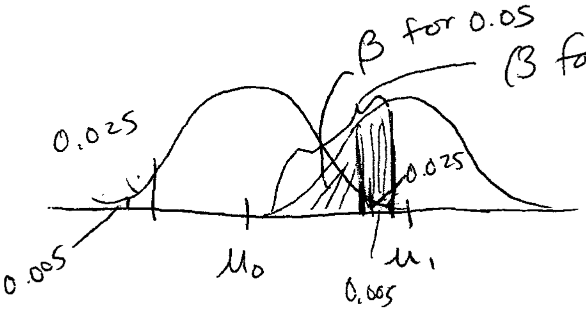
1. A population has a mean of 12.5 and a standard deviation of 3.6. Find the standard error and show the mean, the standard deviation, and the standard error on a graph. 4pts $n = 25$

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{3.6}{\sqrt{25}} = .72 \quad 2pts$$



2pts

2. Show a graph comparing power with α = 0.01 and 0.05. Which has the higher power? 4pts



↑ β for 0.01 ↓ power

2pts for 0.05

higher power

B

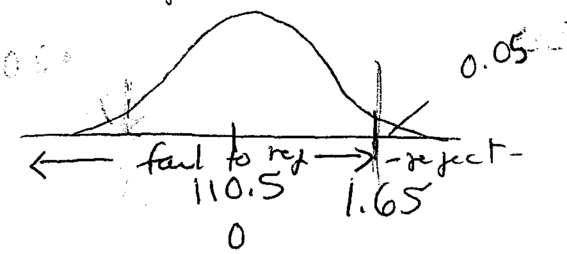
3. In a test using the normal distribution your research hypothesis is that the mean of the population you are studying is greater than 110.5. Write the null and alternate hypotheses and sketch. Indicate your level of significance and critical values and shade the rejection region. 5pts

2 pts H_0 $\mu \le 110.5$
H_1 $\mu > 110.5$

1 pt $\alpha = 0.05$

1 pt $Z_{crit}(1) = \pm 1.65$

1 pt



V. Calculate: 10pts 2006 - calc. with binomial & Poisson

1. Given a mean of 18.4 and a variance of 4.56 from a sample of size 25, find the 95% confidence limits. 4pts

1 pt $\bar{X} \pm t_{.05(2)(24)} S_{\bar{X}}$

1 pt $S_{\bar{X}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{4.56}{25}} = .427$

18.4
$\bar{X} \pm 2.064(.427)$
1 pt $\boxed{18.4 \pm .881}$

$P(17.519 < \mu < 19.282) = .9$

1 pt $t_{.05(2)[24]} = 2.064$

2. What is the probability of finding an individual of between 15.2 and 16.2 with a mean of 18.2 and a standard deviation of 1.5? Include the formula and a sketch. 6pts

$z = \frac{X - \mu}{\sigma}$

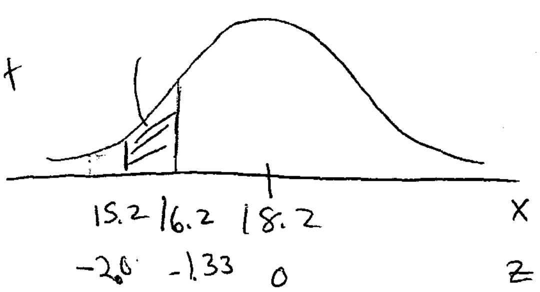
1 pt $z = \frac{15.2 - 18.2}{1.5} = -2$ 1 pt $p = 0.0228$

1 pt $z = \frac{16.2 - 18.2}{1.5} = -1.33$ 1 pt $p = 0.0918$

1 pt $p = 0.0918 - 0.0228 = 0.0690$

?

1 pt



VI. Test the following hypothesis. 18pts

The normal mean number of eggs produced in a single brood of green iguanas is known to be 44.5 with a standard deviation of 15.8. The mean of iguanas from an unusual habitat is 37.2 with a standard deviation of 16.4 based on a sample of size 20. Do the iguanas from the unusual habitat differ from the norm?

1. State the hypotheses in words and symbols. 3pts

1pt for good symbols

H_0 $\mu = 44.5$ the mean does not differ from norm

H_1 $\mu \neq 44.5$ the mean does differ

2. State your level of significance. 2pts

$\alpha = 0.05$

3. Calculate the appropriate statistics. 4pts

$$z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{37.2 - 44.5}{15.8/\sqrt{20}} =$$

use 16.4 −1

3.53

2.066 → 2.07

4. Find p. 3pts

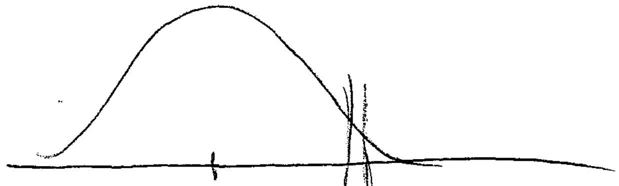
.0192 × 2 = .0384 use t −1

5. State your statistical conclusion. 2pts

reject Ho

6. Verbalize. 4pts

the iguanas differ sign. from norm



$\mu_0 = 44.5$

0

z calc = 2.07

z crit = $z_{.05(2)}$ = ±1.96

BIOLOGY 322
BIOSTATISTICS MIDTERM EXAM 2: PART I Closed **BOOK 67PTS**
NOVEMBER 12, 1998

I. For each of the following give the test statistic (formula), the name of the sampling distribution and the null and alternate hypotheses for all tests. 18 pts

1/2 credit for consistent hypotheses & credit for dist

1. You wish to determine whether two populations are equally variable.

F dist. 1 pt

$$F = \frac{s_1^2}{s_2^2}$$ 1 1/2 pts

$H_0: \sigma_1^2 = \sigma_2^2$ 2 pts

$H_1: \sigma_1^2 \neq \sigma_2^2$

2. You wish to determine whether there are equal numbers of a particular insect in two areas. You choose ten dates at random and on each date you sample each area. You count the number of insects present in each area on each date.

(paired t) t sampling dist

$$t = \frac{\bar{D} - 0}{s_{\bar{D}}}$$ prefer t

$F = \frac{MS_{blocks}}{MS_{error}}$

$H_0: \mu_D = 0$

$H_1: \mu_D \neq 0$

or random block ANOVA

$F = \frac{MS_{treat}}{MS_{error}}$ $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

H_0: blocking not effective
H_1: blocking effect

3. You wish to compare the size of territory defended by a particular species of fish in two habitats--one with much vegetation and one that has little vegetation. You choose three ponds with much vegetation and three different ponds with little vegetation. You measure the size of the territory of twenty five fish in each pond. Partial results are shown below.

	Much vegetation			**Little vegetation**		
Pond	*A*	*B*	*C*	*D*	*E*	*F*
	66	85	94	48	65	44
	72	65	80	55	61	49
		etc			etc	

F sampling dist (Nested ANOVA)

$H_0: \mu_v = \mu_{lv}$

$H_1: \mu_v \neq \mu_{lv}$

$$F = \frac{MS_{gr}}{MS_{subgr}}$$

H_0: no variability among ponds within vegetation type

H_1: variability

$$F = \frac{MS_{subgr}}{MS_{error}}$$

4. Mean cholesterol levels in children ages 2-12 is known to be 175mpc with a standard deviation of 29mpc. The mean for 40 children whose fathers have had a heart attack is 207.3 with a standard deviation of 30mpc. Is the cholesterol level elevated in children with a family history of heart attack?

Normal sampling dist

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

$H_0: \mu \leq \mu_0$ or $\mu \leq 175$

$H_1: \mu > \mu_0$ or $\mu > 175$

II. For the problem below give the name of the test, the test statistic, and the null and alternate hypotheses of all tests that we have covered that would be needed to analyze the data correctly. What else should you test that we have not had a test for? 18 pts

A study was performed to determine the effect of exposure to smoke on respiratory function as measured by forced midway expiratory flow (FEV). The authors wished to determine which groups had significantly lowered FEV as compared to the non-smokers. Given are the summary data. Assume the original data is available.

	Non-smokers	Passive	N-inhal.	Light	Heavy
$\bar{X}$	3.78	3.30	3.32	3.23	2.59
s	0.79	0.77	0.86	0.81	0.88
n	200	200	200	200	200

Levene's F or Fmax

we do Levene

4 1/2 + 1/2 homoscedasticity Fmax $= \frac{s_l^2}{s_s^2}$

Ho: $\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2 = \sigma_E^2$

H_1: not all $\sigma^2 =$

6 1/2 ANOVA - single classification F

Ho: $\mu_A = \mu_B = \mu_C = \mu_D = \mu_E$

H_1: not all $\mu =$

$F = \frac{MS_{GR}}{MS_{error}}$

$H_0 = \mu_C \leq \mu_T$

$H_1 = \mu_C > \mu_T$

Dunnett's test

4 1/2 + 1/2 $q' = \frac{\bar{X}_t - \bar{X}_{NS}}{SE}$

2 also should test for normality

18

III. Answer the following briefly. 19 pts

1. When is a randomized block design (2way without rep.) **not** more powerful than a completely randomized one? Explain briefly. 4

2 when the blocks are not different

then it won't reduce SS error enough $F = \frac{MS_{treat}}{MS_{error}}$

2 to compensate for lower df

randomized: $SS_{error} = SS_{TOTAL} - SS_{TREAT}$ RB: $SS_{error} = SS_{TOT} - SS_{TRT} - SS$

$df_{error} = df_{TOT} - df_{Treat}$ $df_{error} = df_{TOT} - df_{TRT} - df$

2. How does one calculate the SS for the subgroups within groups? Why is it calculated this way? 3

$SS_{SUBGR\,WITHIN} = SS_{ALL\,SUBGR} - SS_{GR}$

2

If the groups are different all subgroups will differ because of group effect - we want pure subgroup effect

1

3. Which design is better (if possible) a **two-way anova with replication** or a **nested anova**? Explain. 4

1 two way w. replication

1 1/2 higher power

1 1/2 can test for interaction

4. Which is more powerful a test involving the normal distribution or a t test on the same data? Explain. 4

2 Normal

2 more narrow dist. → less overlap smaller fail to reject

lower critical

5. Why does the null hypothesis in Anova imply that the parametric value of F=1? 4

if $\mu_1 = \mu_2 = \mu_3$ we can use sampling

1 distribution of mean;

3) $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

2

$n\sigma^2_{\bar{x}} = \sigma^2$

↑ MSgr estimates this

↑ MSerror estimates this

1

$F = \frac{MS\,gr}{MS\,error}$ both num & denom estimate same thing

IV. Which assumptions, if any, of Anova etc. are violated by the following? Which transformations, if any, might correct the problem? 12 pts 2 2

1. proportion of *Acnathes* with values ranging from .85-1.00

proportions > .7 violate normality
binomial
homoscadasticity
arc sin $\sqrt{p}$

2. lifespan ranging from 1-5 yrs

+ skew ∴ not normal
maybe log

3. weight ranging from 1-8 g

none probably
in humans weight is + skew

Key 21

Biostatistics Midterm II 1999

I. Give the null and alternate hypotheses and the test statistics (formula) for the following tests. Use symbols where appropriate: 25 pts

2006 could have binomial or Poisson

1. A two-tailed paired comparison t test

$H_0: \mu_D = 0$

$H_1: \mu_D \neq 0$

$$t = \frac{\bar{D}}{s_{\bar{D}}}$$

2.The subgroup effect in a nested (hierarchical) Anova

$$F = \frac{MS_{subgroup}}{MS_{error}}$$

H_0: no difference among subgroups within groups

H_1: difference among subgroups within group

3. A t test where the research hypothesis is that removal of predators increases weight gained by kangaroo rats

$H_0: \mu_{prem} \leq \mu_{pred}$

$H_1: \mu_{prem} > \mu_{pred}$

$$t = \frac{\bar{X}_{prem} - \bar{X}_{pred}}{s_{\bar{x}_{prem} - \bar{x}_{pred}}}$$

4. The interaction in a two way Anova with replication (factoral).

H_0: no interaction between factors in effect on dependent

H_1: interaction between factors in effect on dependent

$$F = \frac{MS_{A \times B}}{MS_{error}}$$

5. An F test where the hypothesis is that variability is decreased in areas of chronic pollution

$H_0: \sigma^2_{up} \leq \sigma^2_{p}$

$H_1: \sigma^2_{up} > \sigma^2_{p}$

$$F = \frac{s^2_{up}}{s^2_{p}}$$ or largest over smallest

2006 also includes Binomial + Poisson

II. Which of the following is more powerful for the same data? Explain briefly. 18pts

1. Matched pairs or Unpaired t

5 matched pairs usually 3
eliminates effect of confounding variable
1 if variable not important, less powerful
because lose df

2. Multiple t or Dunnetts

3 Multiple t

3 $\alpha > 0.05$

3. 2 Way with replication or Nested Anova

3 2 way with rep. 3 df error ↑
for comparison with groups – probably
reduces SS error better

III. Which assumptions of t and Anova tests are violated by the following? Which transformation is most likely to be effective? 10 pts

1. Number of colonies per plate with values ranging 1-11

1 counted data low mean or $\sqrt{1/2}$ $\sqrt{x+1/2}$
2 ∴ not normal
not homoscedasticity → Poisson
2

2. Percentage of banded snakes with values ranging from 16-28%

1 not normal if $p < .3$ Binomial
2 not homoscedastic
2

2 arc sin $\sqrt{p}$

23

IV. For each of the following state the type of test and the sampling distribution. Assume a parametric test is valid or that the data will be transformed. 20 pts

1. An investigator wished to determine whether larger males have a mating advantage in copepods. She exposed one female to two males simultaneously and then measured the size of the loser male and the size of the winner male. The experiment was then repeated with 49 other females.

Female	**1**	**2**	**3**	**etc**
winner	123	134	107	
loser	133	145	162	

paired t or random block 3 (or 4)

t or F 2 pts

should use paired t since one-tailed; if use Anova stae the p must be divided by 2.

2. A study was performed to study the effect of three pesticide treatments (control, synthetic pesticide, and biologically derived pesticide) on yield under field conditions of four crop species. Five fields were planted under each combination of treatment and species. Given below are the means values. Assume the original data is available for analysis.

	Control	**Synthetic Pesticide**	**Biological Derivative**
Species 1	150	175	181
Species 2	142	161	152

2 way ANOVA with replication (Factoral

F

3. A study was performed to compare the pH of fertilized and unfertilized fields. Ten fertilized and ten unfertilized fields were chosen and 5 samples were taken in each field for pH determination.

Nested (hierarchical) ANOVA

F

4. A study was performed to determine the effect of 4 different photoperiod treatments on sodium plasma levels in 40 adult male mice. Ten mice were randomly assigned to each treatment.

5 Single classification ANOV
1 way

F

24

V. Test the following. Assume the data is suitable for parametric testing or has been transformed.

A study was performed to test the effect of four treatments in inducing diapause in the copepod D. leptopus. The percentage of diapausing individuals was measured under each treatment on populations from three different ponds. Three replicates were performed.

1. Complete the following table. 9 pts

−1/2 for subsequent errors

Source of Variation	SS	DF	MS	F	Fcrit
Treatment	5878	3	1959.3	94.79	3.40
Pond	2068	2	1034	50.02	3.01
PondXTreatment	806	6	134.3	6.50	2.51
Error	496	35−11=24	20.67		
Total	9248	35			

2. State whether you reject or fail to reject and verbalize your conclusions. 9 pts

reject all 3 6

sign effect of treatment and Pond

sign interaction between pond & treatment on diapause induction

3. Describe the interaction if any. 4pts

Treatment	**Means 1**	**2**	**3**	**4**
Pond 1	31.7	33.3	31.3	32.3
Pond 2	58	67.7	75.3	45.7
Pond 3	50.7	65	69.3	42.7

treatments had little effect on pond 1

VI. List the properties of the F distribution that differ from the t distribution. 5pts

(1) + skew

(2) $0 \rightarrow +\infty$

(3) $\mu \rightarrow 1$

(4) 2 df

(5) $t^2 = F$

Key

25

Biostatistics Midterm II 2006

I. Give the null and alternate hypotheses, the test statistics (formula), and the name of the sampling distribution for the following tests. Use symbols where appropriate: 36 pts

2006 can be binomial or Poisson

1. To test whether three means differ

$H_0: \mu_1 = \mu_2 = \mu_3$ 2

H_1: not all means =

$$F = \frac{MS_{group}}{MS_{error}}$$ 3

F dist 1

2. To test the block effect in a random block Anova

H_0: no variation among blocks ∴ blocking not effective

H_1: variation among blocks ∴ blocking effective

$$F = \frac{MS_{blocks}}{MS_{error}}$$

F dist

3. To test whether a particular drug lowers blood pressure where the data is paired and you subtract after from before

A less than B

B − A positive

$H_0: \mu_D \leq 0$

$H_1: \mu_D > 0$

$$t = \frac{\bar{D} - 0}{s_{\bar{D}}}$$

t dist

4. The interaction in a two way Anova with replication (factoral)

Ho: no interaction between factors ~~ea~~

H1: interaction between factors in their effect on the dependent variable.

$$F = \frac{MS_{A \times B}}{MS_{error}}$$

F dist

5. A test for variances where you hypothesize that group 1 should have a larger variance than group 2

$H_0: \sigma_1^2 \le \sigma_2^2$

$H_1: \sigma_1^2 > \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2}$$

F dist

6. To test whether an individual has a blood pressure than is hi:her than the norm of 120 given the population variance.

$H_0: \mu \le 120$

$H_1: \mu > 120$

$$Z = \frac{X - 120}{\sigma} \quad \text{or} \quad \frac{X - \mu}{\sigma}$$

Normal

2 2

II. Which of the following is more powerful for the same data? Explain briefly. 12pts

1. Random block or single classification anova

Random block subtracts effect of major confounding variability from SS error; however df error is also smaller and it will be only more powerful if ↓ in SS error more than compensates for ↓ df error. ie when blocks are very variable

2. Tukey or Dunnetts

Dunnetts (1) ÷ α over fewer comparisons since only compare treatments to control
(2) can be 1-tailed
(3) ~~xxxxxmultiplexxxxxrangexx~~ not true anymore; test has been changed

3. 2 Way with replication or Nested Anova for the group effect assuming the data is the same

2-way tests $\frac{MSgr}{MSerror}$ rather than $\frac{MSgr}{MSsub}$

∴ denominator probably smaller ∵ df error > df subgroups

III. Which assumptions of t and Anova tests are violated by the following? Which transformation is most likely to be effective? Explain. 12 pts

1. Weights of animals on three different diets

Mean	3.58	5.26	8.91
Variance	0.97	3.65	8.89

violates additivity
homoscedastic.
normality

transf 2 pts
why -1.5
4 pts

because multiplicative effects ∴ log

2. Grades on this test if the average is 75%

$p > .7$ ∴ binomial + − skew
probably not homoscadastistic

arc sine $\sqrt{p}$

IV. For each of the following state the type of test. Assume a parametric test is valid or that the data will be transformed and that the original data is available for analysis. 12 pts

One sample Z(X)	Unpaired t	Single classification Anova
One sample Z($\bar{X}$)	Paired t	Random block Anova
One sample t(X)	Factoral Anova	Nested (hierarchical) Anova
one sample t($\bar{X}$)	2006	binomial & Poisson

1. An investigator wished to determine whether people with ~~high blood pressure~~ glaucoma have higher blood pressure than average. A group of 50 people with glaucoma had a mean systolic pressure of 38.5 with a standard deviation of 25.1. is this higher than the average of normal individuals of that age group which is 128.5.

   one sample t ($\bar{X}$)

2. A study of bacterial contamination compared the concentrations of E. coli in 5 different sandy bays. They were each sampled on the first and fifteenth days of June, July, and Aug. Five sites were sampled on each date. Given are the means for each date. Assume the original data is available for analysis.

29

	Bay 1	Bay 2	Bay 3	Bay 4	Bay 5
June1	20	121	85	132	94
June 15	313	360	279	480	220
:					
Aug 15	1200	7900	1000	6600	2300

2 - way w. replication

3. A study was performed to determine the effect of fertilizer addition on soil pH. The pH of soils at 10 sites was measured before and after fertilizer addition.

paired t (or random block ANOVA)

4. A study was performed to determine the effect of 4 different photoperiod treatments on sodium plasma levels in 40 adult male mice. Ten mice were randomly assigned to each treatment. Two measurements were made on each mouse.

nested ANOVA

V. Test the following. Assume the data is suitable for parametric testing or has been transformed. 20pts

An experiment was performed to compare the concentrations of PCBs in three different lakes that are sources of drinking water. Four sites were used in each lake and two determinations were made on each site. Analyze using all the steps in hypothesis testing: State all pairs of hypotheses, the level of significance, the test statistics (Complete the table), your statistical conclusions, and verbalize. →

Source of Variation	SS	df	MS	F
Lakes	61.17	3-1 = 2 (2)	(5) 30.585	$\frac{30.585}{.5}$ = 61.17 (8)
Sites	4.5	9	(6) 0.5	$\frac{0.5}{1.5}$ = .33 (9)
Error	18.0 (1)	12 (3)	(7) 1.5	
Total	83.67	3×4×2-1 (4) (23)		

$F_{critical}$ for lakes is 3.89; $F_{critical}$ for sites is 2.80.

Fcalc > Fcrit (10) Fcalc < Fcrit (11)

reject (12) fail to reject (13)

PCBs vary sign among lakes (14) (16) 2pts no sign variability among sites within lakes

5 (15) 2pts

2. A study was performed to compare the concentrations of mercury in two species of fish in three different types of lakes. Three lakes of each type were used as replicates. Analyze for differences among lakes, fishes, and interaction. Describe the interaction if any. 20pts

	Type I	Type II	Type III	
Trout	9 6 13	12 15 17	18 25 22	
	Σ =28	**Σ =44**	**Σ =65**	**Σ Σ =137**
Bass	4 8 3	6 3 8	4 7 10	
	Σ =15	**Σ =17**	**Σ =21**	**Σ Σ =53**
	Σ Σ =43	**Σ Σ =61**	**Σ Σ =86**	**Σ Σ Σ 190**

Assume ↘ give

SS Total	734.4444
SS Fish	392
SS Type	155.4444
SS error	106.6667

Biostatistics Midterm II 2003

2006 binomial & Poisson

31

I. Answer the following briefly. 38 pts

1. What is the parametric value of F predicted by Ho in Anova? Explain why this is the predicted value. 4pts

1pt $F = 1$

1pt $F = \frac{n s_{\bar{x}}^2}{s_p^2}$ 4pts 4 Ho: $\mu_1 = \mu_2 = \mu_3$ is true then from sampling dist. of $\bar{X}$

(3) $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \Rightarrow n\sigma_{\bar{x}}^2 = \sigma^2$

$n s_{\bar{x}}^2$ est $n\sigma_{\bar{x}}^2$ ∴ est also σ^2

s_p^2 also estimates σ^2 (pooled s_1^2, s_2^2, s_3^2) each est. σ^2; numer est σ^2 ∴ denom est. σ

2. Which is more powerful- a completely randomized design with a single classification Anova or a random block design with a two way Anova without replication? Explain. 4pts

1) if the blocks are very different the random block ANOVA is more powerful: $SS_{error} = SS_{TOT} - SS_{gr} - SS_{blx}$ thus SSerror will be smaller (so will denom.) → $\frac{MS_{groups}}{MS_{error}\downarrow}$ → ↑F↑,

$\frac{\downarrow SS_{error}}{\downarrow df_{error}} = MS_{error}\downarrow \rightarrow F = \frac{MS_{groups}}{MS_{error}\downarrow}$

df error will be smaller but ↓ SSerror will more than compensate

2) if blocks are not different then it is less powerful because of lower df

3. What can you test for in a factoral (2way with replication) design that you cannot test for with any other design? 2pts

interaction

68

4. Would you expect the following to meet the assumptions of parametric testing? If not state which assumptions they are likely to violate and which transformation, it any, would be most likely to correct the problems. 15pts

   a. height in cm. with values ranging from 2-11

   probably OK likely to be normal & additive

   b. percent infected with values ranging from 10 to 25

   1 no – 1 normality – 1 binomial – skewed + (p<.5)
   1 homoscedasticity
   1 arc sine $\sqrt{p}$

   c. counted data with values ranging from 3-9

   1 no 1 normality – 1 Poisson – skewed +
   1 homoscedasticity
   1 $\sqrt{\ }$

5. What is the expected mean square for the group factor in a nested Anova? 4pts

   groups $\sigma^2 + \sigma^2_{subgroups} + \sigma^2_{groups}$

6. List in order of decreasing power: multiple t tests, Tukey, or Dunnets? Explain. 5pts

   3½ multiple t > Dunnetts > Tukey
   1 $\alpha > 0.05$ 1 0.05
   fewer comp than Tukey
   ½ { multiple range
   can be 1-tailed

   69

33

7. List the properties of the F distribution that differ from the t. 4 pts

$0 \to +\infty$

skew +

$\mu \to 1$

2 df

$(t^2 = F)$

II. **For the following state the name of the sampling distribution, the hypotheses using symbols where appropriate, and the test statistic for the following. 5pts each (25pts)**

1. Testing two groups for homoscedasticity

1 F

2 $H_0: \sigma_1^2 = \sigma_2^2$
$H_1: \sigma_1^2 \neq \sigma_2^2$

2 $F = \dfrac{s_1^2}{s_2^2}$

2. Testing to determine whether the mean of group A is less than the mean of group B

t

2 consistent

$H_0: \mu_A \geq \mu_B$
$H_1: \mu_A < \mu_B$

$$t = \frac{(\bar{X}_A - \bar{X}_B) - 0}{s_{\bar{X}_A - \bar{X}_B}}$$

3. A posteriori (post hoc) tests comparing each mean with each other

Tukey q

$H_0: \mu_A = \mu_B$
$H_1: \mu_A \neq \mu_B$

$$q = \frac{(\bar{X}_A - \bar{X}_B) - 0}{SE}$$

70

7

omit; repeat of earlier page

b. Central tendency for species (ex. D. pulex, D. rosea, D. dubia)

mode

- needs to be used due to nature of data (nominal)

c. Central tendency for lifespan

lifespan high + skew

~~mode~~ ~~median~~ ~~Mean ($\bar{X}$)~~ median

highly skewed

~~- ratio scale data; continuous~~

- ~~values far from middle~~ (such as infant death)

2. Write the formula for the statistical variance and explain why it is calculated the way it is. 6 pts

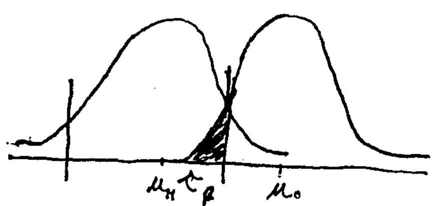
$$S^2 = \frac{\Sigma(x - \bar{x})^2}{n-1}$$

* the formula for statistical variance replaces the N in the denominator of parametric variance with $n-1$. This is to eliminate the bias which would occur if $\bar{X}$ was used to approximate μ. We are removing one degree of freedom to account for this discrepancy

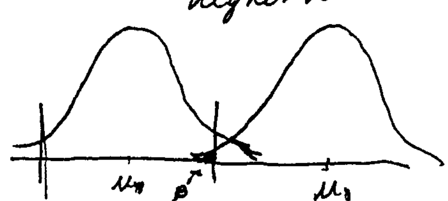
3. What effect does decreasing α have on β? Power? Sketch β and power for two different levels of α and label. 6 pts

* decreasing α increases β, and therefore decreases power

lower α



higher α



4. Which properties of the t distribution differ from those of the Normal? Include the relationship between t and Normal. 6 pts

NORMAL	T-DIST
• $\mu = 0$; $\sigma = 1$	• $\mu = 0$; $\sigma > 1$ (more spread out)
	• many df

* all other properties are the same (symmetrical, unimodal, $-\infty \to \infty$, mean=med=mode)

* as df $\to \infty$; the t-distribution = normal (as sample size approaches population

4. All tests in a nested Anova

F dist

Ho: $\mu_1 = \mu_2 = \mu_3$
H1: not all μ =

$$F = \frac{MS_{groups}}{MS_{\cancel{sub}\,groups}}$$

Ho: no variation in subgroups within groups
H1. variation.

$$F = \frac{MS_{subgroups}}{MS_{error}}$$

5. A test for skew

Normal dist.

Ho: $\gamma_1 = 0$
H1: $\gamma_1 \neq 0$

$$z = \frac{g_1 - 0}{\sigma_{g_1}}$$

III. What type of test is each of the following? 8pts

Single sample $\underline{X}$, given σ	independent 2 sample	2way Anova without replication
Single sample $\underline{X}$, given $\sigma_{\bar{X}}$	paired samples	2way Anova with replication
Single sample X, given $s_{\bar{X}}$	1 way Anova	nested Anova

2006 binomial + Poisson

IS US NS
p n=3
t

Assume the data is either suitable for these tests or will be transformed to meet the assumptions.

1. A study was performed to investigate the effects of disease and pond permanence on survival in the gray tree frog. Eighteen artificial ponds were used with the following treatments: Three permanent ponds with infected snails, three temporary ponds with infected snails, three permanent ponds with uninfected snails, three temporary ponds with uninfected snails, three permanent ponds with no snails, and three temporary ponds with no snails. The percentage survival of tadpoles was recorded in each pond.

2 way w replication

2. A particular species of mammal is known to have an average litter size of 11.2. A sample of 25 animals with supplemental feeding has an average of 13.8 with a standard deviation of 1.4. Does the supplemental feeding increase egg production?

$$t = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

Single sample give $s_{\bar{X}}$

7/

35

3. A study was performed to compare the effects of three different diets on the protein content of the dorsal root ganglion in rats. Twenty rats were placed on each diet and three readings were made on each rat.

Single Nested ANOVA

4. A study was performed to test the effect of fertilizer addition on soil pH. Three fertilizers and a control were compared. They were each applied to 6 different soil types and the pH of each soil type on each treatment was measured 2 months after treatment.

2 way w/o rep Anova (random block)

IV. Answer the following pertaining to the data below.

In a study of the effect of habitat fragmentation on dispersal in the endangered Fender's butterfly, investigators compared the "move length" of males and females as a function of whether they were inside, outside, or at the edge of a patch of their food source (also threatened). The **mean** move lengths, based on **30** individuals in each category are summarized below. 29pts

	Habitat		
sex	Edge	Inside patch	Outside patch
Male	8.4	5.2	11.5
Female	7.7	2.7	5.8

1. State the hypotheses using symbols where appropriate. 3pts

Ho: $\mu_I = \mu_O = \mu_E$ H₁: not all μ =

Ho: $\mu_f = \mu_m$ H₁: $\mu_f \neq \mu_m$

Ho: no interaction between sex & habitat on move length

2.State the level of significance and the number of tails. 2pts

$\alpha = 0.05$ 1-tail

3. Calculate the appropriate statistics and degrees of freedom by completing the table below: 13 pts

7

36

Source of Variation	SS	df	MS	F
Habitat	24.41	3-1=2	12.105	28.56
Sex	4.75	2-1= 1	4.75	11.21
Habitat X Sex	3.39	2x1 = 2	1.695	4.00
Error	73.74	6(30-1)=174	0.424	
Total	106.29	179		

4. Find the appropriate p values, state whether you reject or fail to reject, and verbalize your conclusions for each hypothesis. 9pts

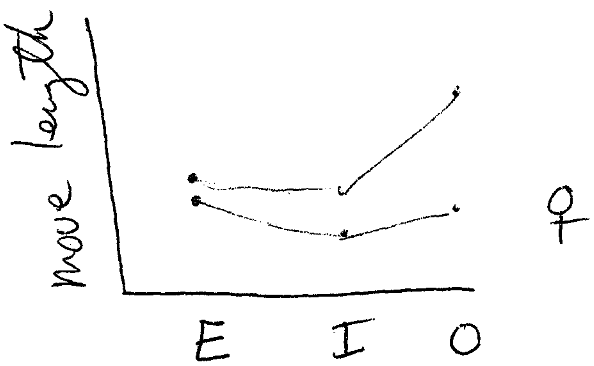
habitat $p < 0.0005$ reject move lengths differ significantly among habitats

sex $p \geq 0.001$ reject move lengths differ significantly between sexes

interaction $.025 < p < .01$ reject significant interaction between habitat & sex on move length

5. Describe the possible interaction even if you find it not significant. 2pts

males move outside patch more & females move more on edge of patch



73

37

Biostatistics Midterm II 2005

2006 includes binomial + Poisson

I. Answer the following briefly. 40pts

1. a. What is the parametric value of F predicted by Ho? 2pt

$F_P = 1$

b. What is the "intuitive" test statistic? 2pt

$$F = \frac{n s_{\bar{X}}^2}{s_P^2}$$

c. Derive this formula from the sampling distribution of the mean. 4pts

if Ho: is true $\mu_1 = \mu_2 = \mu_3$ etc & then from SDM $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

$\therefore \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \rightarrow n\sigma_{\bar{X}}^2 = \sigma^2$

s_P^2 estimates σ^2, $n s_{\bar{X}}^2$ estimates $n\sigma_{\bar{X}}^2 \therefore \sigma^2$

2. a. Why is a random block design (or paired t) likely to be more powerful than a completely randomized ~~(independent~~ samples) design? 4pts

allows you to subtract the variation due to blocks from $SS_{TOTAL} \rightarrow \downarrow SS_{error} \rightarrow \downarrow MS_{error} \rightarrow \uparrow F$

b. Why might it not be more powerful? 2pt

$\downarrow$ df error $MS_{error} = \frac{SS_{error}}{df_{error}}$ & fewer df to lookup F with

So if blocking is not effective - you lose

3. What can you test for in a factoral (2way with replication) design that you cannot test for with any other design? Write the hypotheses for this test. 4pts

2pts interaction

2pts Ho: no interaction between factors in effect on dependent

H_1: interaction between factors

4. Would you expect the following to meet the assumptions of parametric testing? If not state which assumptions they are likely to violate and which transformation, it any, would be most likely to correct the problems. 12pts

a. percent infected with values ranging from 35-65

4pts probably OK proportions .3 - .7 likely to be symmetrical

3pts if say no binomial arc sin $\sqrt{p}$ violates normality & homo

38

b. mean $\quad$ 5.2 $\overset{3}{\sim}$ 15.4

s^2 $\quad$ 0.43 $\underset{3^2=9}{\sim}$ 3.82

multiplative effects violate additivity normality ✓ homo

use log

c. Number of eggs per clutch (values ranging from 3-8)

Counted data low mean : Poisson violates normality + skew ✓ homo

5. What is the expected mean square for the group factor in a nested Anova? 2pt

$EMS = \sigma^2_{group} + \sigma^2_{subgroup} + \sigma^2$

6. List in order of decreasing power: multiple t tests, Tukey, or Dunnets. Explain. 8pt

4pts Mult t > Dunnetts > Tukey

3pts $\alpha > 0.05$ $\quad$ $\alpha = 0.05$ fewer comp than tukey $\quad$ $\alpha = 0.05$ more comparison

1pt one tail sometimes multiple range

1p $\quad$ 2pts

II. For the following state the name of the sampling distribution, the hypotheses using symbols where appropriate, and the test statistic for the following. 25pts

2pts

1. Testing to determine whether the length of the island population is **less variable** than the mainland

F

Ho: $\sigma^2_{I_2} \geq \sigma^2_m$

H1: $\sigma^2_I < \sigma^2_m$

$$F = \frac{S_m^2}{S_I^2}$$

(actually here in 1-tail I would put expected larger on top so I know I can reject)

39

2. Testing to determine whether the **mean** wing **length** of the island population is **less** than that of the mainland.

$H_0: \mu_I \geq \mu_m$

$H_1: \mu_I < \mu_m$

t dist or Normal dist

for 2 samples

$$t = \frac{\bar{X}_I - \bar{X}_m}{S_{\bar{X}_I - \bar{X}_m}}$$

if assume know σ of mainl

$$z = \frac{\bar{X}_I - \mu_m}{\sigma_{\bar{X}}}$$

3. To test for skew

Normal

$H_0: \gamma_1 = 0$

$H_1: \gamma_1 \neq 0$

$$z = \frac{g_1 - 0}{\sigma_{g_1}}$$

4. A posteriori (post hoc) tests comparing a treatment with a control

Dunnetts q' dist

$H_0: \mu_T = \mu_c$

$H_1: \mu_T \neq \mu_c$

$$q' = \frac{\bar{X}_T - \bar{X}_c}{SE}$$

40

5. All tests in a two way Anova with replication

F dist

Ho: $\mu_1 = \mu_2 = \mu_3$ $F = \frac{MS_A}{MS_{error}}$
H1: not all μ =

Ho: $\mu_A = \mu_B = \mu_C$ $F = \frac{MS_B}{MS_{error}}$
H1: not all μ =

Ho: no interaction between A & B $F = \frac{MS_{A \times B}}{MS_{error}}$
H1: interaction between A & B

III. Which type of test is each of the following? If the data is not suitable for parametric testing assume it will be transformed. 12pts

Single sample normal $\underline{X}$	independent t	2way Anova without replication (random block)
Single sample normal X	paired t	2way Anova with replication (Factorial)
Single sample t	1 way Anova	nested (hierarchical) Anova
	Poisson	binomial

1. A study was performed to compare the effects of three different diets on the protein content of the dorsal root ganglion in rats. Twenty rats were placed on each diet and the protein content for each rat was determined.

A B C
n = 20 n = 20 n = 20

Single or 1 way ANOVA

2. An experiment was performed to compare the yields of two varieties of beans and to determine which of 4 sowing rates maximizes yield. Twenty-four plots were used and each variety was sowed at each of the 4 densities for a total of 8 combinations and three plots per combination.

2 way w rep / factorial

Sow	Var 1	Var 2
1	n = 3	n = 3
2	n = 3	n = 3
3	n = 3	n = 3
4	n = 3	n = 3

3. A study was performed to compare the abundances of *D. coloradensis* in eutrophic, mesotrophic, and ultraoligotrophic ponds in Colorado. Four ponds of each type were selected and three samples were collected in each pond.

eu meso ultroo
Pond 1 2 3 4 1' 2' 3' 4' 1'' 2'' 3'' 4''
n = 3 in each

Nested ANOVA

A study was performed to determine the effects of photoperiod and temperature on the gonadosomatic index of the fish *M. terrae-sanctae*. Two photoperiod treatments were used (long and short day) and three temperatures (16, 18, and 27C) for a total of 6 distinct treatments. 60 fish were randomly assigned to one of the 6 treatments (10 per treatment). Each fish is housed in its own tank.

a. Which advanced Anova is this?

b. Write all sets of null and alternate hypotheses for this Anova.

c. Which SS add together to give you SStotal in this Anova?

d. Why is a random block design generally more powerful than a completely randomized design?

A study was performed to determine the effects of mercury on the gondaosomatic index in guppies. Three concentrations of mercury were used; four tanks of guppies were used for each treatment with 10 fish per tank.

1. Which advanced Anova is this?

2. Write all sets of null and alternate hypotheses for this Anova.

3. Which SS add together to give you SStotal in this Anova?

4. Why might a random block design be less powerful than a one way?

42

BIOLOGY 322 FINAL EXAMINATION DEC. 9, 1998
PART I: CLOSED BOOK 67pts

I. Answer the following. 18pts

1. Which statistic is best to indicate central tendency for degree of sunburn on a scale of 1-5? Explain.

median because it is ordinal data

2

2. Which test is preferred for a paired comparison on # mutations per plate with values ranging from 0 to 8? Explain.

2

Sign test   counted data low mean ⇒ Poisson
violated normality & homoscedasticity
∴ non-parametric test
highly skewed ∴ sign not Wilcoxon

3. Why are 95% the most commonly calculated? (What is the problem with 90%? With 99%?).

1

90% limits are less accurate
99% limits are less precise; they are too broad

4. What is the advantage of random sampling design over regular sampling design?

1 random samples allow statistical analysis

5. How can you have statistical significance without biological significance? Give a hypothetical example.

statistical significance means you reject Ho:
3 you can reject Ho and the difference (or the correlation) is so slight as to have little biological meaning; the independent variable has an effect but the effect is slight. this occurs with high n.

6. What relationship do you expect between the parametric variance and the parametric mean if for number of plants per quadrat if the distribution is contagious?

$s^2 > \bar{x}$ ← statistics

1 $\sigma^2 > \mu$ parameters

7. How does one test for significance of group effect in a nested Anova? Explain in terms of expected MS.

2 $$F = \frac{MS\ gr}{MS\ subgr}$$

Gr MS $= \sigma^2 + \sigma^2_{subgr} + \sigma^2_{gr}$

Sub-gr MS $= \sigma^2 + \sigma^2_{subgr}$

to isolate the effect of groups we test over subgroups

8. What properties of the Poisson and Binomial distributions differ?

Binomial	Poisson
$p \not\to 0$	$p \to 0$
$n \not\to \infty$	$n \to \infty$
skew +, - or 0	skew +
range 0 - n	range 0 - ∞
2 parameters	1 parameter
$\mu = np, \sigma^2 = npq$	$\mu = \sigma^2$

3

9. Why are all Anovas one-tailed?

3 because F_p can only be = or > 1

Gr expected MS $= \sigma^2 + \sigma^2_{gr}$

if $\sigma^2_{gr} = 0$ Ho is true $F_p = 1$

if $\sigma^2_{gr} \neq 0$ H_1 is true $F_p > 1$

a difference in means ↑ MS groups beyond that predicted by variance within groups (error)

45

II. There is a problem with each of the following experiments either eith the way the data was collected or in the analysis. Explain each problem briefly and state how to correct it. 15pts

2 or IV 3

1. A study was performed to determine whether density dependent factors control the population density of oystercatchers. If density dependent factors are important, a high density one year should result in a lower density the following year. Hence they performed a correlation between the *ratio of the population density in the following year/the population density in a particular year* (Y) vs. *population density in the particular year* (X).

Ratio	1.5	1.44	1.36	etc
Density	46	39	32	etc

not mathematically independent 3
(also non-linear)
Y = ~~X~~/X
→ do pop density in following year 2
vs. pop. density in particular year
2 points for Spearman ← won't really help
or 1 if get dependence

2. An experiment was performed to look at the effect of dredging on water quality. Thirty measurements of oxygen in mg/l were made in a harbor prior to dredging and again two weeks following dredging. The sites before are not the same as those after. Data was analyzed by an unpaired t test.

Before	10.33	10.62	11.43	etc
After	6.18	7.88	9.22	etc

3

pseudoreplication
need more than 1 dredged site 2

2 pts if say use same sites & pair

46

3. A study was performed to measure the effect of metal pollution on species richness (#species). Measurements were made on Pb concentrations and species richness at 25 randomly chosen separate localities and the data was analyzed by parametric correlation.

Pb conc.	10	16	1	25 etc
# species	7	6	24	3 etc

counted data low mean not normal 2
but Poisson
may then be non-linear as well 1½
try a $\sqrt{x}$ or Spearman if
the transform doesn't work 1½

III. For the following state the type of test (name) that would best be used to analyze the data. Where necessary assume a parametric test applies or that the data will be transformed. The list of names follows.18pts

Unpaired t	Single classification Anova	Regression (I)
Paired t	Random block Anova	Correlation
	Factoral Anova	χ^2 Goodness of fit to the (specify)
	Nested (hierarchical) Anova	χ^2 Contingency

1 pt for wrong χ^2
1 pt for ANOVA

1. An experiment was performed to determine whether children whose fathers have had myocardial infarctions (heart attacks) have elevated blood cholesterol levels. 100 male and 100 female 14 year olds whose father had had MI were compared with 100 100male and 100 female 14 year olds whose father had not had MI. Summary results are shown below. Assume the original data is available.

Fathers with MI:	males $\bar{X}$=207.3% s = 35.6%	females $\bar{X}$=198.2% s= 36.8%
Fathers without MI:	males $\bar{X}$= 174.5% s = 32.6%	females $\bar{X}$=145.5% s= 34.9%

factoral ANOVA

2. A study was performed to determine whether the occurrence of a particular abnormality in plants is at random. One hundred areas were chosen and in each ten plants were examined for the occurrence of the abnormality.

# plants/area	1	2	3	4	5	6	7	8	9	10
# areas	9	20	14	15	3	5	8	6	12	7

χ^2 G of fit to Binomial

2lr for G of fit

3. A laboratory study was performed to determine if the metabolic efficiency of egg production in Drosophila decreases as temperature increases. 20 eggs were used per temperature treatment and the mean energy required per egg was measured. A high energy requirement means a low efficiency.

Temperature	20	21	22	23	24	25	26	27	28
μl O_2/egg	9.5	9.8	10.2	10.9	11.0	10.9	11.5	14.2	18.3

Regression

3-48

4. A study was performed to test the effectiveness of three drug treatments in treating gonorrhea. With penecillin 40 of 200 individuals tested positive, with ampicillin 10 of 100 individuals tested positive, with streptomycin 15 of 100 had positive tests.

χ^2 Contingency

5. A study was performed to determine whether a particular clone of **Daphnia pulex** is evenly distributed in the water column.

Depth	0	10	20	30	40	50	60	70
# Daphnia	2	20	14	23	45	30	12	8

χ^2 GoF fit to a uniform
1:1:1:1 etc

6. A study was performed to compare the biomass of algae in streams contaminated with Cd and uncontaminated streams. Four contaminated and 4 uncontaminated streams were measured. Ten sites were chosen in each stream and the biomass of algae was measured at each site.

Nested ANOVA

2 1

IV. For the following state the test statistic, the name of the sampling distribution, and the null and alternate hypotheses.16pts

1

1. A particular species is known to have a standard deviation of 2.34 for the mean body length. A population exposed to metals form a smelter for 15 years has a standard deviation of 2.15 based on a sample size of 200. Is the population dispersion reduced in the population from the polluted site?

$H_0: \sigma^2 \geq$

$H_1: \sigma^2 < (2.34)^2$

$$\chi^2 = \frac{\nu s^2}{\sigma^2}$$

χ^2

2. A particular species is known to have a mean number of bristles on the fifth leg of 8.9 with a standard deviation of 2.1. An individual is found with a value of 6.7. Does this individual differ form the norm?

$H_0: \mu = 8.9$

$H_1: \mu \neq 8.9$

Poisson $P(x) = \dfrac{e^{-\lambda} \lambda^x}{x!}$

3. A study was performed on 4 strains of bacteria: vsr-; vsr-h-; vsr-l-; and vsr-s-. Each culture was plated onto 100 plates and the number of mutants per plate were counted. Means and standard deviations of each were obtained. Test for homoscedasticity.

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$

H_1: not all $\sigma^2 =$

$F_{max} = \dfrac{s_l^2}{s_s^2}$ or Levene's (no formula)

Fmax

4. The proportion of deaths due to lung cancer in males aged 15-64 is known to be 0.12. 5 of 20 deaths among workers in a particular chemical plant were attributed to lung cancer. Is the occurrence of cancer in the plant abnormally high?

$H_0: p \leq .12$

$H_1: p > .12$

$p(x) = \dfrac{n!}{x!(n-x)!} p^x q^{n-x}$

Binomial

54

BIOLOGY 322 FINAL EXAMINATION DEC. 20, 1999 CLOSED BOOK

I. Answer the following. 27pts

1. Why is the denominator of the sample variance n-1 rather than n?

- To eliminate bias - otherwise the estimate would be too low.
- the numerator will always be too small since the range of the sample is less than the range of the population; decreasing the denominator compensates.
- the numerator is only zero around μ. In estimating μ by $\bar{X}$ we lose one degree of freedom

2. Is a random block Anova always more powerful than a single classification Anova on the same data? Explain.

No - $MS\ error = \frac{SS\ error \downarrow}{df\ error \downarrow}$ in blocked

— SS error will be smaller in blocked ANOVA as one subtracts the effect of blocks, but df error will be smaller as well. SS error must decrease sufficiently to compensate for decrease in df error. This occurs if the blocks are very different.
- also you have fewer df in blocked ANOVA

3. Why are 95% confidence limits the most commonly calculated? (What is the problem with 90%? With 99%?).

95 compromise between percision & accuracy

90% more precise: lower range

99% more accurate; closer to 100%

4. Explain why in Anova Ho means the parametric value of F =1 .

Ho implies $F = \frac{MS\ among}{MS\ error} = 1$

MS among is $n\,S^2_{\bar{X}}$

MS error is S^2

property # 3 of sampling dist of mean:

$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$ $\therefore$ $n\,\sigma^2_{\bar{X}} = \sigma^2$

5. How can you have statistical significance without biological significance? Give a hypothetical example.

statistical significance means you reject Ho. Biological significance means the effect you are testing is important.

For example in a correlation between size of newborns and maternal consumption of coke you might find a statistical significant correlation if you use a very large sample size, but the coefficient of determination is .005.

6. How does one test for significance of group effect in a nested Anova? Explain in terms of expected MS.

$F = \frac{MS\ groups}{MS\ subgroups}$

expected MS

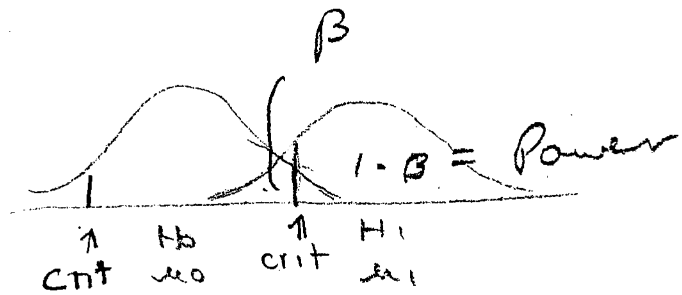
groups $\sigma^2 + \sigma^2_{sub} + \sigma^2_{grou}$

subgroups $\sigma^2 + \sigma^2_{sub}$

error σ^2

to isolate the group effect you test over sub not error

7. What is power? Include a sketch.



8. Which of the following is the most powerful? The least? Explain briefly.
Multiple t, Dunnetts, or Tukey tests?

Multiple t most α > 0.05
Dunnets 2nd 0.05 ÷ fewer comparisons
also can be 1-tailed + is multiple range
Tukey least

9. What are the degrees of freedom in a chi square goodness of fit to the normal with 10 classes? Explain.

10 - 1 - 2
lose one always because
parameters you must estimate to solve the normal dist
$\Sigma O = \Sigma E$
∴ $\Sigma|O-E|$ absolute value = 0

II. There is a problem(s) with each of the following experiments either in the way the data was collected or in the analysis. Explain each problem briefly and state how to correct it. 15pts

1. A study was performed to determine how disturbance affects species richness. Two areas were chosen- a highly disturbed area and a fairly undisturbed area. In each area 15 plots were chosen at random and the number of species in each was determined with the following results. Data was analyzed by unpaired t. (there are at least two errors.)

Number of species per plot:

Undisturbed	7	19	10	8	9	12	8	9	7
	10	9	14	8	7	11			
Disturbed	4	4	3	3	2	6	3	5	3
	4	4	10	5	4	4			

counted data low mean Poisson ✓ not homosced.
√x or use Mann Whitney U
pseudo replication use > 1 disturbed ✓ > 1 undist.

2. In a study of pollution of ponds in a particular area an investigator wished to determine how the ratio of nitrogen to phosphorous changed with increasing nutrient loading. The concentrations of nitrogen and phosphorous were measured in six randomly chosen ponds and the ratio of N:P was plotted against P. Data was analyzed for a negative regression. (There are at least 3 errors.)

Pond	**1**	**2**	**3**	**4**	**5**	**6**
N/P	55	34	22	15	13	7
P	1	5	14	41	55	98

do correlation not regression – mensurative
not linear OK give 3 pts
don't do N/P vs. P ; not mathematically
independent – do N vs P

SS

III. For the following state the type of test (name) that would best be used to analyze the data. Where necessary assume a parametric test applies or that the data will be transformed. The list of names follows. 16pts

Unpaired t	Single classification Anova	Regression (I)
Paired t	Random block Anova	Correlation
	Factoral Anova	χ^2 Goodness of fit to the (specify)
	Nested (hierarchical) Anova	χ^2 Contingency

1. An experiment was performed to determine whether children whose fathers have had myocardial infarctions (heart attacks) have elevated blood cholesterol levels. 100 six year old children whose fathers had MI were compared with 100 whose fathers had not had MI. The children were classified into three groups: normal, elevated, and slightly elevated cholesterol .

	Normal	**Slightly elevated**	**Highly elevated**
Fathers with MI:	24	35	41
Fathers without MI:	56	30	14

χ^2 contingency

2. A study was performed to determine whether the occurrence of a particular abnormality in a particular species is at random. One hundred areas were chosen and the number of abnormalities was recorded in each.

# abnormalities/area	1	2	3	4	5	6	7	8	9	10
# areas	9	20	14	15	3	5	8	6	12	7

χ^2 G of fit to Poisson

3. A laboratory study was performed to determine if the metabolic efficiency of egg production in Drosophila decreases as temperature increases. Three temperatures were used. The metabolic efficiency of twenty eggs each from two different strains was measured. *Given are the means of each. Assume the original data is available.*

	10°	**15°**	**20°**
strain 1 means	21.5	14.5	9.8
strain 2 means	18.5	15.5	13.5

2 way w replication (Factoral) ANOVA

4. A study was performed to compare the PCB concentrations in the milk of mothers from southern Quebec with that in mothers from the north. Fifty women in each category were studied.

North	16.5	21.2	15.8	...	19.5
South	4.6	5.9	6.2	...	3.3

unpaired t
1-way ANOVA

5. A study was performed to determine whether a particular clone of **Daphnia pulex** changes its distribution in the water column in the presence of predator substance.

Depth	**0**	**10**	**20**	**30**	**40**	**50**
# Daphnia without	2	10	16	19	5	3
#Daphnia with	1	6	13	14	14	15

χ^2 contingency

6. A study was performed to determine whether chipmunks will carry larger loads when food is located at a farther distance from their burrows. Food was placed at intervals of 8, 16, and 24 meters from the burrows of a chipmunk and the number of seeds taken at each distance was recorded. The experiment was repeated with 15 different chipmunks.

	8m	**16m**	**24m**
chipmunk 1	2	5	6
chipmunk 2	3	5	9
chipmunk 3	3	7	10
•			
chipmunk 15	2	4	7

2 way w.o. replication ANOVA
random block

57

7. A study was performed to look at species diversity (by the Shannon index) as a function of pH. One hundred lakes were sampled and in each the pH of and the diversity of zooplankton were measured.

pH	5.6	6.2	4.9	...	7.1
Diversity	1.79	1.92	1.32	...	2.11

correlation

8. A study was performed to determine the effect of exposure to mercury on glutathion levels in trout. Six artificial streams were used. Mercury was added at 1ppb to three and three were left as a control. Ten fish were studied in each condition. *Given below are the means (in arbitrary units) from each stream. Assume the original data is available for analysis.*

Control	**1 ppb Hg**
15	24
19	27
14	22

nested ANOVA

IV. For the following state the test statistic, the name of the sampling distribution, and the null and alternate hypotheses in words and symbols. 16pts

1. A particular species is known to have a mean number of eggs per clutch of 48.6 with a standard deviation of 2.5. A population exposed to metals form a smelter for 15 years has a mean of 34.5 with a standard deviation of 2.3 based on a sample size of 200. Is clutch size reduced in the population from the polluted site?

2 $z = \dfrac{\bar{X} - \mu}{\sigma_{\bar{X}}}$

1 normal

Ho: $\mu \geq 48.6$ clutch size is not reduce

$\mu < 48.6$ clutch size is reduced

2. The number of deaths from typhoid fever in a particular population is known to be 4.6. One year after an immunization program the number of deaths is 3. Has the immunization program been successful?

Poisson distributionHo:

$$P(x) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

Ho: mu > or = 4.6 immunization is not successful
H1: mu< 4.6 immunization is successful

3. In a study of the feeding performance of Heliconius caterpillars the following data were obtained for growth of the last instar on 5 different food species. Test for homoscedasticity.

Plant species	**1**	**2**	**3**	**4**	**5**
mean	3.2	4.6	5.3	3.8	6.2
s	0.3	0.4	0.25	0.2	0.35
n	20	20	20	20	20

$$\left(F_{max} = \frac{S_1^2}{S_5^2} \right)$$ we don't do this one anymore

or Levene's test

4. The probability of a child developing bronchitis in the first year of life is known to be 6%. In one area 3 children less than one years old out of 15 families developed bronchitis. Is this unusually high?

$$P(x) = \frac{n!}{(n-x)!\,x!} p^{x} q^{n-x}$$

V. Give the relationship between the following pairs of distributions 6 pts

1. Normal and t

as $n \to \infty$ $t \to$ normal

$t_{[\infty]} = z$

2. χ^2 and F

as $n \to \infty$ χ^2 approaches νF

$\chi^2 = \nu F$

3. Binomial and Poisson

as $n \to \infty$ & $p \to 0$ binomial approaches Poisson

VI. Which non-parametric test substitutes for each of the following parametric tests? 6pts

1. Single classification Anova

KrusKal Wallis

2. Paired t

Wilcoxon
or
sign if not symmetrical around median

3. Pearson (parametric) correlation

Spearmann

4. Unpaired t

Mann Whitney U

6c

VII. Which test statistic (or measure) is best for the following circumstances? Name and symbol (if possible). Explain very briefly. 10 pts

1. To indicate the proportion of variation of one variable associated with another in a study of heart rate vs temperature in the lab

2 coefficient of determination , r^2
or ✓✓

2. To indicate whether the incidence of a rare genetic defect is random, contagious, or uniform

2 coefficent of dispersion , $\frac{s^2}{\bar{X}}$

3. To indicate central tendency for index of activity measured on a scale of 1-4

2 median
ranked data

4. To indicate the error associated with a particular mean

1½ confidence limits
not dependent on sample size

2 — $s_{\bar{X}}$ is dependent on sample

61

VIII. Which transformation is usually most successful under the following circumstances? 4.5pts

1. Mean	16	20	31
Variance	1.2	2.0	4.5

log maybe

2. Binomial data

arc sine $\sqrt{p}$

3. Weight vs. length

log

BIOLOGY 322 FINAL EXAMINATION 2001 CLOSED BOOK

I. Answer the following. 27pts

1. Give an example of a biased statistic. Explain briefly.

s^2 with $\underline{n}$ as denominator underestimates the population variance as the numerator which is based on the range of the sample underestimates the population range

2. Which is generally more powerful- a random block Anova or a single- classification Anova? Why?

a random block ANOVA because it removes the variation due to a major confounding variable ∴

$F = \frac{MS_{tr}}{MS_{error}}$

1-way		random
SS groups		SS treat
		+ SS blocks
+ SS error ← larger		+ SS error ← smaller ∴ MS error usually smaller
SS Total		SS Total

3. Why are 95% confidence limits the most commonly calculated? (What is the problem with 90%? With 99%?).

compromise between accuracy + precision

90% not accurate (low %)

99% not precise (broad limits)

4. Explain why the null hypothesis in ANOVA implies that the parametric value of F=1 .

Ho: $\mu_1 = \mu_2 = \mu_3$

If all means are equal, we can use property 3 of sampling dist. of mean: $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

$F = \frac{MS\,gr}{MS\,error}$ ∴ $n\sigma^2_{\bar{x}} = \sigma^2$

$n\sigma^2_{\bar{x}}$ " MS groups, σ^2 " MS error

∴ $\frac{n s^2_{\bar{x}} \text{ est}}{s^2} = 1$

6

5. Under what circumstances can one have statistical significance without biological significance?

With large sample sizes one can find statistically significant correlations, differences etc., but the variable tested may explain only a small proportion of the dependent variable. For example one might have a very low r^2 but $p < 0.05$

6. Which of the following is the most powerful? The least? Explain briefly. Multiple t, Dunnetts, or Tukey tests?

most Multiple t - $\alpha > 0.05$

Dunnests 0.05
fewer comparisons than Tukey

least Tukey

7. Compare and contrast correlation and regression. (Model I)

reg | corr
line - dependent & independent | degree association - interdep. variables
manipulative (Model I) | always mensurative
can infer causation (Model I) | cannot infer causation

8. In a χ^2 test of the association between cadmium contamination and a particular congenital defect you obtain a p of 0.029.

a. Do you reject or fail to reject?

reject H_0 $p < 0.05$

b. Verbalize your conclusions.

There is significant association between cadmium & congenital defects

9. List the properties of the **Sampling distribution** of the mean.

(1) usually normal

(2) $\mu_{\bar{X}} = \mu$

(3) $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

II. Answer the following. 27pts

1. One what scale is each of the following measured? Which measure of central tendency is most appropriate? 4pts

a. dominance rank on a scale of 1-5

ordinal - median

b. temperature in Celsius

interval - mean - probably not highly skewed so mean has lowest sampling error & includes all values

2. Give the name of the sampling distribution and write the null and alternate hypotheses for all sets of tests for the following using symbols where appropriate. 10pts

a. 2- way Anova with replication

F dist

Ho: $\mu_1 = \mu_2 = \mu_3$
H_1: not all μs =

Ho: $\mu_A = \mu_B = \mu_C$
H_1: not all means =

Ho: no interaction between factors
H_1: interaction..

b. Nested Anova

F dist

Ho: $\mu_1 = \mu_2 = \mu_3$
H_1: not all μs =

Ho: no difference among subgroups within same group
H_1: difference among subgroups..

c. Correlation where the research hypothesis is that the relationship is positive

r or t dist

Ho: $\rho \leq 0$

H_1: $\rho > 0$

65

d. Goodness of fit to the Poisson χ^2

Ho: the data fit the Poisson

H₁: the data do not fit the Poisson

3. What are the **differences** in properties of these pairs of distributions? 4pts

a. binomial and Poisson

Poisson is a special case of binomial $\left\} \begin{array}{l} n \to \infty \\ p \to 0 \\ q \to 1 \end{array}\right.$

is always + skew, but binomial can be +, 0, -

b. χ^2 and F

χ^2 special case of F $\quad x(n-1)$

χ^2 $\quad$ F

$\mu = n-1$ $\quad$ $\mu \to 1$

$\sigma^2 \uparrow$ with $\uparrow n$ $\quad$ $\sigma^2 \downarrow$ with $\uparrow n$

only one df $\quad$ 2 df-numerator and denominator

4. Which test **statistic** (or measure) is best for the following circumstances? Name and symbol (if possible). Explain very briefly. 9pts

1. To indicate the proportion of variation in PCB levels associated with variation in body fat

r^2 – coefficient of determination – we want proportion; parametric valid

2. To measure the degree of aggregation (clumping)in copepods

$\frac{s^2}{\bar{X}}$ ← statistic $\quad$ coefficient of dispersion $\quad$ < 1 for regular; = 1 for random Poisson; > 1 for clumped

3. To indicate population dispersion for number of cells/ l (values>1000)

s $\quad$ standard deviation – we want to estimate how variable the population is (not how accurate a mean is; then we use $s_{\bar{x}}$ or CL) s easier to interpret than s^2

III. Consider the following experiment. A study was performed to determine whether addition of top predatory fish can restore lakes polluted with nutrients. Trout were added to one polluted lake and another similar lake was left as a control. One of the measures of effect was total phosphorous in µg/l. Phosphorous was measured in each lake on three dates. On each date 10 sites were chosen at random. Below is data for the year following trout addition. Given are the means and standard deviations of the 10 measurements. 6pts

	Lake with trout			Control Lake		
	June 6	**July 5**	**Aug. 21**	**June 6**	**July 5**	**Aug.21**
Mean	100.2	88.5	53.3	91.8	67.6	31.2
s	12.9	9.7	4.5	10.2	7.3	1.2
s^2	166.41	94.09	20.25	104.04	53.29	1.4

1. There is at least one glaring error in the data collection. What is it and how would you correct it?

Pseudoreplication: use > 1 lake in each treatment

2. Ignoring this consideration, which statistical test would you use?

2-way with replication (dates are same)

3. Would you need to transform this data? If so, which transformation is likely to be the most effective? Why?

Variances appear unequal and correlated with $\bar{X}^2$ or worse!!

$100.2 \xleftarrow{2X} 53.3$ $\quad 91.8 \xleftarrow{3X} 31.2$

$166.41 \xleftarrow{8X} 20.25$ $\quad 104.04 \xleftarrow{100X} 1.4$

use log

IV. For the following state why the data is not suitable for parametric analysis and what you can do to correct it. Include both non-parametric tests and transformations if appropriate. 8pts

1. a correlation between liver weight/body weight and body weight

not mathematically independent
use liver wt. vs. rest of body wt

2. a paired t where the dependent variable is # algal cells per field (values ranging from 0-9)

not normal (Poisson)
+ skew
s^2 correlated with $\bar{X}$ → heteroscedasticity
transform

sign test sign test more likely than Wilcoxon because probably very skewed

6/

V. For the following state the type of test (name) that would best be used to analyze the data. Where necessary assume a parametric test applies or that the data will be transformed. Where means are given assume the original data is available for analysis. The list of names follows. 16pts

Unpaired t	Single classification Anova	Regression (I)
Paired t	Random block Anova (2way w.o. replication)	Correlation
	Factoral Anova (2way with replication)	χ^2 Goodness of fit to the (specify)
	Nested (hierarchical) Anova	χ^2 Contingency

1. Mosquitoes require a blood meal to activate egg production. In a study of the activation mechanism, investigators looked at the effect of the repressor AaGATAr on production of the yolk precursor protein vitellogenin. Does the data below suggest that the repressor reduces production of vitellogenin? Data is amount of gene transcription.

	With repressor	Without repressor
Mean	95.2	39.8
S.D.	16.8	15.2

unpaired t (single ANOVA)

2. In a study of factors increasing susceptibility to extinction, investigators wished to determine whether two measures of rarity, a narrow geographical distribution and low population density, are themselves associated. Members of the order Carnivora were classified with regard to these two factors. Given are the number of species in each category.

	Widespread distribution	Narrow distribution
High population density	33	5
Low population density	15	16

χ^2 Contingency

3. A laboratory study was performed to determine if the metabolic efficiency of egg production in a particular strain of Drosophila decreases as temperature increases. The metabolic efficiency was measured at eight different controlled temperatures. Given are the means of 10 observations made at each temperature.

Temperature	16	18	20	22	24	26	28	30
Met. Efficiency	22.5	20.8	19.8	19.2	17.4	15.0	13.7	11.2

regression

4. A study was performed to compare the PCB concentrations in the milk of mothers from southern Quebec with that in mothers from the north. Four areas in southern Quebec and four

in Northern Quebec were selected at random and 10 women were measured per area. Given below are the means for each area.

North (mean)	16.5	21.2	15.8	19.5
South (mean)	4.6	5.9	6.2	3.3

Nested ANOVA
(areas in N + areas in S)

5. A study was performed to determine whether the male titmouse performs different activities at different heights. Ten birds were used and five observations were made on each activity for each bird. Given are the mean heights for the five observations.

	Singing	**Feeding**	**Preening**	**Resting**
bird 1	186	44	120	70
bird 2	172	40	124	63
:				
bird 10	191	55	144	90

2- way – (if don't use original data random block)
otherwise it is really repeated measures & I made a booboo
You would say 2 way w. replication

6. A study was performed on the nest site preferences in broad-tailed humming birds. The sites were determined for 54 females.

Aspen	**Spruce**	**Fir**	**Willow**
6	24	21	3

χ^2 Goodness of fit to uniform

7. A study was performed to determine whether the tat protein released from HIV-1 infected cells is associated with Kaposi's sarcoma. Of 78 patients with sarcoma 10 showed antibodies for tat; 21 of the 219 patients without sarcoma had tat antibodies.

χ^2 Contingency

8. A study was performed to determine the effect of exposure to mercury on glutathion levels in trout. Mercury and glutathione levels were measured on ninety randomly chosen fish.

Mercury	121	166	34	...	92
glutathione	66	85	13	...	47

Correlation

VI. For the following state the test statistic, the name of the sampling distribution, and the null and alternate hypotheses in words and symbols. 16pts

1. A particular species is known to have a mean weight of 48.6 with a variance of 6.45. A population exposed to metals from a smelter for 15 years has a mean of 34.5 with a variance of 5.49 based on a sample size of 200. Is the weight reduced in the population from the polluted site?

Normal dist

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

$H_0: \mu \geq \mu_0$ or $\mu \geq 48.6$

$H_1: \mu < \mu_0$ or $\mu < 48.6$

2. What test would you use on #1 data to determine whether weights are less **variable** in the polluted area?

χ^2 dist

$$\chi^2 = \frac{\nu s^2}{\sigma^2}$$

$H_0: \sigma^2 \geq 6.45$

$H_1: \sigma^2 < 6.45$

70

counted, low mean

3. The mean number of breeding individuals of a particular species was known to be 4.7 per hectare. After installation of hydroelectric installation, the following season only 1 breeding pair was observed in the acre around the installation. Does this represent a significant decrease?

Poisson

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$H_0: \mu \geq 4.7$

$H_1: \mu < 4.7$

4. Fifteen percent of the population is left-handed. In a survey of accidents at a particular factory, of the 23 individuals involved in minor accidents in two years, 5 were left handed. Is this unusually high?

Binomial

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

$H_0: p \leq 0.15$

$H_1: p > 0.15$

mpm .71

BIOLOGY 322 FINAL: EXAMINATION 2002 total pts 120

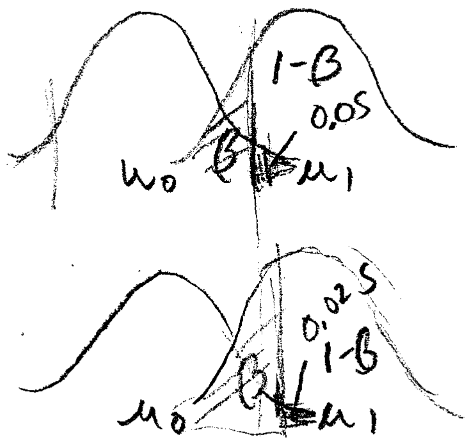
I. Answer the following. 74pts

1. Why are random samples preferred to regular ones? 2pts

Statistics assume them

2. Which is generally more powerful a one or a two-tailed t-test? Illustrate your answer with a sketch of power. 6pts

one



3. For a correlation what is the best indication of statistical significance? Biological significance? 4pts

stat: $p < 0.05$ - reject Ho

biol: assuming reject Ho, r^2 coefficient of determination

4. What is the test statistic formula for the intuitive method of a single classification Anova? What is the parametric value of this statistic if you reject Ho? 5pts

$$F = \frac{n\, s^2_{\bar{x}}}{s_p^2}$$

where s_p^2 = pooled s^2

if reject Ho $F_0 > 1$

72

5. Why is it inappropriate to use multiple t tests to compare 4 means? Use the binomial distribution to calculate the maximum chance of any type one error. 5pts

α over whole set too large

one comparison P(Type I) = 0.05 P(not type I) = 0.95

P(not type I) for 6 ~~4XX~~ comparisons = $(0.95)^{\cancel{4X}\,6}$

P(any type I) for 4 comparisons = $1-(0.95)^{\cancel{4X}\,6}$

6. Why is it inappropriate to use model I regression on a mensurative experiment? Include the meaning of a mensurative experiment in your explanation. 3pts

Model I regression takes a $\perp$ to the line & assumes X is fixed or controlled by an experimentor. In a mensurative expt. the investigator does not control X.

7. In a 2x2 χ^2 test of the association between cadmium contamination and a particular congenital defect you obtain a χ^2 of 4.98. 8pts

a. Write the null hypothesis in words

Ho: congenital defect independent of Cd exposure

H1: " " dependent on " "

b. What is your p value?

df (r-1)(c-1) = (2-1)(2-1) = 1, 0.05 < p < 0.0

c. Do you reject or fail to reject?

reject Ho

d. Verbalize your conclusions.

Cd exposure significantly associated with ~~a affects the~~ the incidence of the defect.

8. How does one partition SS_{total} in a 2-way Anova with replication? 3 pts

$SS_{TOT} = SS_A + SS_B + SS_{A \times B} + SS_{error}$

9. What is the expected MS for the group factor in a nested Anova? 2pt

$\left(F = \frac{MS_{groups}}{MS_{subgroups}} \right)$ expected MS $= \sigma^2 + \sigma^2_{sub} + \sigma^2_{gr}$

10. One what scale is each of the following measured? Which measure of central tendency is most appropriate? 4pts

a. temperatures ranging from 0-10 C

interval mean

b. food choice: oligochaetes, isopods, amphipods, other

nominal mode

11. Give the name of the sampling distribution and write the null and alternate hypotheses for all sets of tests for the following using symbols where appropriate. 12pts

a. Nested Anova

F

$H_0: \mu_1 = \mu_2 = \mu_3$ Group

H_1: not all μs $=$

Ho: no difference in subgroups within same group Subgroup

H1: is " " " " " "

-74

b. To test whether two groups are equally variable

F - $H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

c. To compare ~~means~~ values of two <u>independent</u> groups of <u>ranked</u> data

Mann Whitney U

H_0: the groups are the same

H_1: the groups are not the same

d. Factoral Anova

$H_0: \mu_1 = \mu_2 = \mu_3$ for factor A & B

H_1: not all μ's =

interaction

H_0: no interaction between factors A & B

H_1: is " " " "

12. What is the relation between these pairs of distributions? (Hint: think of the properties). 4pts

a. binomial and Poisson

Poisson is binomial with $n \to \infty$, $p \to 0$, $q \to 1$

b. χ^2 and F

χ^2 is F with infinite df in denominator

x df

$\chi^2_\nu = \nu F_{[\nu, \infty]}$

6.4 7.5

13. Which **statistic** (or measure) is best for the following? Name and symbol/ formula. 8pts

a. To indicate the sampling error associated with PCB concentration in body fat

$s_{\bar{x}}$ standard error

b. To indicate population dispersion for body size in domestic cats and in cougars when you wish to compare their variability

cv coefficient of variability

c. To indicate whether the spatial distribution of milkweed plants is random

coefficient of dispersion $= \frac{s^2}{\bar{x}}$

but Morisitas is better

d. To indicate the degree of association between vitamin E intake and lifespan

r_{sp} spearman's rank correlation coefficient

14. In a test of the association between obesity and the response to pain, researches obtained a value of r of -.31 based on 10 individuals. Is there a significant negative relationship? 8pts

a. Write the null hypothesis in symbols.

$H_0: \rho \geq 0$

$H_1: \rho < 0$

b. Give the critical value.

$df = n - 2 = 10 - 2 = 8$ crit $= -0.549$

c. Do you reject or fail to reject?

$0.25 < p < 0.10$ or $|r_c| < |r_{crit}|$

fail to reject

d. Verbalize your conclusions.

no evidence of significant neg. association

76

$$F = \frac{MS\ treatment}{MS\ lakes}$$ $$F = \frac{MS_{[illegible]}}{MS\ error}$$

II. Consider the following experiment. A study was performed to determine whether addition of top predatory fish can restore lakes polluted with nutrients. Trout were added to 5 polluted lakes and 5 lakes were left as a control. One of the measures of effect number of *Anabaena* colonies per 0.1l. 12 pts

Sorry this is poorly worded

	Lakes with trout	Control lakes
	15	10
	8	14
	11	12
	9	12
	7	20
mean	10	13.6
variance	10	14.8

try cummulative mean

1. There are at least two errors in the data collection. What are they? Explain.

small sample size - no clear-cut answer
no mention of confounding variables being controlled
were the control + exp't'l assigned at random?

2. Would you need to transform this data? If so, which transformation is likely to be the most effective? Why?

counted data mean ≥ 10 - probably don't transform

3. Give the test statistic formula for the test(s) that you should use to analyze the data and give the null hypotheses using words and symbols.

if one-tail

$H_0: \mu_T \geq \mu_C$ # Anabaena not reduced
$H_1: \mu_T < \mu_C$ # Anabaena reduced

$$t = \frac{\bar{X}_A - \bar{X}_B}{s_{\bar{X}_A - \bar{X}_B}}$$

or do nested ANOVA if you have original data ~~lakes within~~

$H_0: \mu_T = \mu_C$
$H_1: \mu_T \neq \mu_C$

H_0: no variability among lakes within treatment
H_1: variability among lakes within treatment

6

77

III. For the following state why the data is not suitable for parametric analysis and what you can do to correct it. Include both non-parametric tests and transformations if appropriate. 8pts

1. A random block design on multiplicative data

violates normality
homoscedasticity
additivity

log transformation

Friedmann Chi ^{2}

2. A regression of population size on time

not linear

log transform of N

Spearmann rank correlation (we don't do nonparametric regression)

IV. For the following state the type of test (name) that would best be used to analyze the data. Where necessary assume a parametric test applies or that the data will be transformed. Where means are given assume the original data is available for analysis. The list of names follows. 16pts

Unpaired t	Single classification Anova		Regression (I)
Paired t	Random block Anova (2way w.o. replication)		Correlation
	Factoral Anova (2way with replication)	χ^2 Goodness of fit to the (specify)	
	Nested (hierarchical) Anova		χ^2 Contingency

1. In a study of factors increasing susceptibility to extinction, investigators wished to determine whether two measures of rarity, a narrow geographical range and low population density, are themselves associated. The geographical range in km and the mean population density in #/km^2 were measured on each of 20 species of the order Carnivora.

correlation

2. A study was performed on the nest site in broad-tailed humming birds. The investigators wished to determine whether nest site preference (aspen, fir, spruce or willow) varied with experience. The sites were determined for 19 experienced and 23 inexperienced females.

χ^2 contingency

ex inex

a
f
s
w

78

3. A study was performed to determine the effect of photoperiod and genotype on the latent period of infection of barley mildew isolate AB3. Fifty leaves of each of four genotypes were chosen and allocated at random to one of 5 different photoperiod treatments (10 leaves per photoperiod.) The response noted was the number of days until the appearance of visible symptoms.

Phot period 1 2 3 4 5

2 way - with replication : factorial ANOVA genotype A 10=n B C D

4. A study on the effects of crowding on hamsters measured serum corticosterone on artifical colonies of different densities. 7 colonies were established with densities of 1,2,3,4,5,6,7 juveniles per m^2. After 3 weeks, the mean corticosterone levels for each colony were measured.

regression (you controlled density)

5. In a study of the effects of copper on rainbow trout, the investigators wished to determine whether exposure to copper affects the number of mucus cells. Twenty fish were chosen and randomly assigned to either the copper treatment or the control. Four areas on the gill were chosen at random for each fish. The number of cells per square micron was counted in each of the four areas.

Cu 10 fish control 10 fish 4 measures per fish

Nested or hierarchical ANOVA

6. In a classical study of natural selection, investigators wished to determine whether mutations occur randomly (with regard to selection). A strain of bacteria unable to survive without glucose was inoculated onto 100 agar plates without glucose and the number of mutants per plate was counted.

χ^2 Goodness of fit to the Poisson

7. A study was performed to compare the effects of 4 hormone treatments on seed production in Sitka spruce. Ten trees were chosen at random and on each tree 4 similar branches were selected and assigned at random to one of the 4 treatments. Seed production on each branch was determined.

Treat 1 2 3 4
tree 1 n=1
:
10

Random Block ANOVA

8. In a study of the barnacle *Balanus balanoides* the investigators examined whether slope the tendency to settle on a substrate. Three conditions were used: horizontal, 30°, and 75°. The number settling on each was determined.

χ^2 Goodness of fit to uniform

V. For the following give the test statistic formula, the name of the sampling distribution, and the null and alternate hypotheses using symbols where appropriate. 12pts

1. A particular species is known to have a mean population density of 6.3 animals per km^2. A population exposed to metals from a smelter for 15 years has a density of three animals in a kilometer. Is the population size reduced in the population from the polluted site?

H_0: $\mu \geq 6.3$
H_1: $\mu < 6.3$

Poisson

$$P(X) = \frac{e^{-\lambda} \lambda^{X}}{X!}$$

$6.3 = \lambda$

$X = 1, 2, 3$

2. A study was performed to determine whether earthworms exposed to pollution behave differently from unexposed worms. Twenty exposed and twenty unexposed were allowed to choose between three soil types and the number choosing each was recorded.

	ex	unex
Soil 1	#	
2		
3		

χ^2 contingency

Ho: the choice of soil types is independent of exposure

H_1: Soil choice depends on exposure

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

3. A particular species of bird is known to have a mean weight of 98.2 g with a standard deviation of 5.6. After installation of hydroelectric installation, the following season the mean weight was found to be 95.2 with a standard deviation of 6.4. Does this represent a significant decrease?

Normal

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}$$

Ho $\mu \geq 98.2$

H_1: $\mu < 98.2$

4. The water snake has two forms, a banded and an unbanded form. The percentage of the banded form is known to be 62%. On one island only 4 of 18 snakes were banded. Are there fewer banded on the island?

Binomial

$$P(X) = \frac{n!}{(n-X)!X!} p^X q^{(n-X)}$$

7/80

Final exam (part of 2005)

IV. Test the following 3 hypotheses. 22 pts

1. An investigator wished to test the significance of the equation relating altitude and stream temperature. The equation is Y = 16.133 – 0.0041X with a standard error of the regression coefficient of 0.0032 based on 20 measurements. 9pts

a. State the hypotheses using words and symbols

$H_0: \beta = 0$ There is a relationship between altitude & stream T°, i.e. stream T° varies with altitude

$H_1: \beta \neq 0$ There is no relationship between altitude & stream T°, i.e. stream T° does not vary with altitude

b. State the level of significance

$\alpha = 0.05$ (2-tailed)

c. Find the test statistic and degrees of freedom.

$$t = \frac{b-0}{S_b} = \left|\frac{0.004}{0.0032}\right| = 1.25 \qquad df = n-2 = 20-2 = 18$$

t_{calc} (1.25) < t_{crit} (2.101) ∴ FTR H_0

d. Find p.

P_{calc} (0.50–0.20) > P_{crit} (0.05) ∴ FTR H_0

e. State the statistical conclusion

fail to reject H_0

f. Verbalize

We have no evidence that stream temperature does not vary with altitude.

5

2. A study was performed to determine whether two endangered species *P. palustris* and *T. europaeus* are found together. 120 1x1m quadrats were sampled. The number of quadrats with *T. europaeus* only, *P. palustris* only, neither and both were recorded. A χ^2 test yields a value of 6.21. 6pts

	0	+
T.e.		
T.p.		

a. State the hypotheses using words

H_0 : T. europaeus & P. palustris occurrence are independent of each other

H_1 : T. europaeus & P. palustris are contingent.

7/8

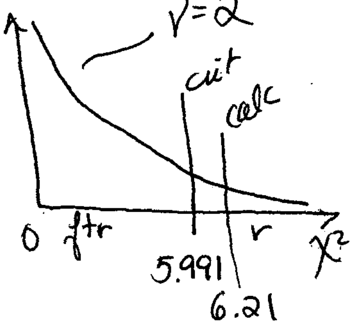
b. Find the degrees of freedom $df = (\#\text{columns}-1)(\#\text{rows}-1)$

$= (2-1)(2-1) = 1$

c. Find the critical value and sketch the reject and fail to reject regions.

$\chi^2_{crit\ 0.05,\ df=1} \ (\cancel{5.991})\ 3.841 < \chi^2 calc\ (6.21) \ \therefore$ R Ho

$P_{crit}\ (0.05) > p\ calc\ (0.05 - 0.025)\ \therefore$ R Ho



d. State the statistical conclusion

reject Ho

e. Verbalize

We have evidence that T. europaeus & P. palustris are contingent. (they are found together).

3. The normal count of white blood cells using a particular assay is known to have an average of 6.2. A particular individual has only 1. Is this value unusually low? 7pts

Counted data with low mean → use Poisson

a. State the hypotheses using symbols

Ho: $\mu \geq 6.2$

H_1: $\mu < 6.2$

b. State the level of significance

$\alpha = 0.05$ (1-tailed)

c. Calculate p.

$p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ $\quad p(0) = \dfrac{e^{-6.2}\, 6.2^0}{0!} = 0.002$ probability of having 1 or less = 0.015

$p(1) = \dfrac{e^{-6.2}\, 6.2^1}{1!} = \underline{0.013}$

$\Sigma\ 0.015$

d. State the statistical conclusion

$P_{calc}\ (0.015) < p\ crit\ (0.05)$

$\therefore$ reject Ho

e. Verbalize

We have evidence that the white blood cell count of this individual is ~~particularly~~ unusually low!