SAULT COLLEGE
of Applied Arts and Technology
Sault Ste. Marie

COURSE OUTLINE

BIOMETRICS

FOR 301-4

revised June, 1981 by H. Robbins

TEXT:

Sokal, R. R., and F. J. Rohlf. 1973. "Introduction to Biostatistics" San Francisco, Freeman 368 p.

REFERENCE TEXTS:

Alder, H. L. and E. B. Roessler, 1972. Introduction to Probability and Statistics", San Francisco, Freeman 373 p.

Finney, D. J., 1966. "Experimental Design and Its Statistical Basis", Chicago, Univ. Chicago Press 169 p.

Giles, R. H. (Editor) 1971. "Wildlife Management Techniques", Washington, The Wildlife Society 633 p.

Ricker, W. E. 1968. "Methods for Assessment of Fish Production in Fresh Water". IBP Handbook No. 3, Blackwell, Oxford 313 p.

Snedecor, G. W., and W. G. Cochran, 1967. "Statistical Methods", 6th Edition. Ames, Iowa State University Press 593 p.

Sokal, R. R. and F. J. Rohlf, 1969. "Biometry, the Principles and Practice of Statistics in Biological Research", San Francisco, Freeman 776 p.

Steel, G. D. and I. H. Torrie, 1960. "Principles and Procedures of Statistics", Toronto, McGraw-Hill 481 p.

BIOMETRICS

FOR 301-4

COURSE OUTLINE

TEXT: "Introduction to Biostatistics" by Sokal & Rohlf

Unit #1 (4 hours)*	- Course introduction - The computer & how it works - The APL language, use of symbols - The calling of programs or workspaces - An example of a workspace and its use to solve a statistical problem.
Unit #2 (5 hours)	- Data in Biology & Descriptive Statistics - Samples & populations, variables, accuracy & precision, derived variables, frequency distributions, the handling of data - Arithmetic mean, median, mode, range, standard deviation, sample statistics, coding, coefficient of variation.
Unit #3 (6 hours)	- Probability, binomial and poisson distribution - Probability, random sampling, hypothesis testing, binomial and poisson distribution
Unit #4 (4 hours)	- The normal distribution - Continuous variables, derivation, properties and applications of the normal distribution, departures from normality.
Unit #5 (8 hours)	- Sampling, estimation & hypothesis testing - Distribution of means, confidence limits, student's t-distribution, samples and confidence limits, chi-square, confidence limits for variances, hypothesis testing.
Unit #6 (4 hours)	- Introduction to analysis of variance including the theory of its use, use of the F-distribution to compare variances partition of sums of squares and degrees of freedom in one-way ANOVA, Model I and Model II ANOVA.

* Hours do not include time for problem solving during classes.

Practical problems are selected from various natural resource disciplines.

STUDENT EVALUATION

Term Tests:

Term tests will be written at the end of each of units 2, 4, 5, and 6. In addition to homework questions assigned for the students' own practice, some of these questions will be assigned for computation on the APL terminal. These questions will be handed in when complete and will make up approximately 25% of the course mark.

Equipment - An electronic calculator having the simpler mathematical functions, including square roots, is mandatory for classroom and test purposes. A statistical calculator is not permitted under any circumstances.

Students will receive grades based on their course average and consistency of performance. Each student must, however, complete all of the course requirements. The pass mark in each unit is 50%.

List of Formulae - Biometrics

	Ungrouped data	Frequency Distribution
Arithmetic mean $(\bar{Y})$-(μ)	$\frac{\Sigma Y}{n}$	$\frac{\Sigma(f\ Y)}{\Sigma f}$
Mean deviation (MD)	$\frac{\Sigma \lvert Y - \bar{Y} \rvert}{n}$	$\frac{\Sigma(\lvert Y - \bar{Y} \rvert f)}{\Sigma f}$
Standard deviation (s)*(σ)	$\sqrt{\frac{\Sigma(Y - \bar{Y})^2}{n - 1}}$	$\sqrt{\frac{\Sigma(Y - \bar{Y})^2 f}{n - 1}}$
Variance $(s\ )(\sigma^2)$	$\frac{\Sigma(Y - \bar{Y})^2}{n - 1}$	$\frac{\Sigma[(Y - \bar{Y})^2 f]}{n - 1}$
Sum of squares (SS)	$\Sigma(Y - \bar{Y})^2$	$\Sigma[(Y - \bar{Y})^2 f]$
Coefficient of variation (CV)	$\frac{s}{\bar{Y}}$ or $\frac{\sigma}{\mu}$ X100	

For computational purposes:

$$SS = \Sigma(Y - \bar{Y})^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$$

$$SS = \Sigma[(Y - \bar{Y})^2 f] = \Sigma f Y^2 - \frac{(\Sigma f Y)^2}{n}$$

N. B.

The latter terms in each of these formulae is called the correction factor (CT)

$$\text{Median for grouped data} = C \times \left(\frac{\frac{n}{2} - f_c}{f_m} \right) + L$$

C = class width
L = mean of lower boundary of median class and upper boundary of next lower class
Y = individual values
Y = in frequency table the class <u>mark</u>
n = total number of observations
f = frequency of values in each class
Σ = "the sum of"
f_c = cumulative frequency up to and including the class immediately below median class
f_m = frequency of the median class

*Note: your text uses y in these formulas $y = Y - \bar{Y}$

also $n = \Sigma f$ in a frequency distribution

Biometrics Problems

First Term

Probability & Binomial Distribution

1. In an experiment in the development of the splake, a fish geneticist is attempting to determine if genes originating from both the brook trout and lake trout are equally influential in determining the characteristics of the offspring. He selects ten characteristics which are different in each of the two species and cross breeds adults. He then examined 2000 of the offspring when they had reached a suitable size and noted the number (out of ten) of characteristics typical of each species in each fish.

(a) Assuming that the parents of each species exert an equal effect on the characteristics, how many offspring would he expect to obtain having half brook trout and half lake trout characteristics?

(b) How many would have all ten characteristics of lake trout?

(c) How many would be <u>more</u> like brook trout than lake trout?

(d) If lake trout genes (characteristics) tend to show up twice as often as brook trout genes, how many offspring will be <u>more</u> like <u>brook trout</u>?

2. The population of deer in an area is high but the investigator wishes to know the probability of encountering one. He knows there are on the average 5 deer in each 10 square miles and that in this population they are always found in pairs. Assuming that the pairs are completely independent of each other, what is the probability that each of three one square mile plots selected at random will have a deer in it?

3. A zoo wishes to obtain fish from a lake for its aquarium. It wishes to be 99% sure that it will have at least 2 males and at least 3 females for later breeding purposes. It is known that the ratio of males to females in the population is 45:55 but the sex cannot be determined at the time of sampling. How many fish will the collectors require?

Normal & Student's t-Distribution

4. We are sampling, periodically, young bass from a nest site in a river. Other nest sites are nearby in which spawning has taken place at various times during the previous three weeks. After hatching, bass emerge from the nest bottom and school in a loose group over the nest site. <u>After a few more days they disperse along the shoreline with individuals from various nest sites mixing with each other</u>. Since the fish in each nest hatch at different times, their lengths will not be the same on the average.

Population A is a list of the lengths of fish from nest number 42, a location in which spawning occurred quite late in relation to the surrounding nests. Subsequently these fish are smaller than the others. It is quite certain that all the lengths in this population are lengths of fish spawned in nest 42. The distribution of these lengths, since they are taken from fish from the same population, should be normally distributed.

Population B is a list of the lengths of fish from the same nest site but taken a few days later.

A. Using the graphic method, test whether each population is normally distributed. In population B, what appears to have occurred? Explain your answer fully.

B. In population B, how many fish lengths would we expect to fall between 1.10 and 1.19 mm. Compare this with the observed number and explain why there is a difference.

5. Estimate the number of fish required in (3) above using the normal or student's t.

6. An investigator sexing waterfowl tabulates his totals and finds the following:

Mallard - 75 male and 60 female
Teal - 300 male and 320 female
Wooduck - 30 male and 32 female

Do the above data each follow a 50:50 sex ratio?

7. The lengths of a fish species are found to be normally distributed around a mean of 29.5 cm with a standard deviation of 4.1 cm. In a certain locality, an investigator measures 500 fish. If this population is the same size as the species in general, how many lengths would fall between 32.5 and 35.6 cm?

8. In the previous question (7), the investigator finds an average length of 32.5 cm with standard deviation of 3.1 cm in 105 fish. Would you say this group of fish had the same average length as the general population?

9. A wildlife investigator is attempting to determine whether a group of geese he has captured belong to the giant Canada or lesser Canada group. It is known that the average weight of giant Canada is 18 pounds with a standard deviation of 4.0 lbs. His group of 10 geese weighed 12 lbs. with a standard deviation of 5 lbs. Are they giant Canadas?

POPULATION A (in mm)

.87
.89
.81
.87
.90
.89
.75
.86
.85
.88
.83
.85
.79
.85
.81
.84
.81
.81
.80
.81
.86
.79
.87
.84

$\bar{Y}$ = 0.86
S = .039
N = 24

POPULATION B (in mm)

1.14
1.25
1.29
1.30
1.39
1.30
1.02
1.40
1.20
1.11
1.45
1.10
1.08
1.32
1.21
1.02
1.01
1.19
1.13
1.05
1.05
1.31
1.04
1.10
1.05
1.23
1.10
1.31
1.20
1.11
1.05
1.04
1.02

$\bar{Y}$ = 1.20
S = .131
N = 33

10. Pellet counts and browse surveys are done by counting the number of items recorded in randomly sampled plots. The data from the tally sheets follows:

Plot Number	Number of Items Recorded
1	144
2	55
3	0
4	412
5	75
6	70
7	89
8	100
9	50
10	31
11	10

(a) Determine the 95% upper and lower confidence limits for the mean of the sample.

(b) How would you increase the reliability of your estimates?

(c) Calculate the 95% confidence limits for the CV of the sample means.

11. In question (9) above, in a sample of 25 geese, how many would we expect to weigh between 12 and 16 lbs if they belonged to the giant Canada population? Assume the same standard deviation.

12. The ratio of male to female deer in a population is 1:1. Out of one hundred deer found in a wintering area in early spring, 58 are male. Is there reason to believe that males can survive the winter better than females (5% level)?

13. The average height of trees from a site type is 44.5 feet with a standard deviation of 10.5 feet. A sample of 25 trees is measured and found to average 49.5 feet. Set the 99% confidence limits and state whether the sample was biased.

14. In question (13) above, state the null hypothesis. If in fact the average height of trees in the area from which the sample was taken was 47.5 feet, show using diagrams the <u>approximate</u> probability of committing a type II error.

Chi-Square Distribution

15. A sample of 25 female animals was found to have a weight standard deviation of 250.5 lbs. Males are known to have a standard deviation of 200.9 lbs. Were the variances in the weights of the two sexes identical?

16. Answer question (15) above using the f-distribution.

17. In the splake experiment, the brook and lake trout were crossed and 144 F_1 hybrids were obtained. These were inbred and 100 F_2 hybrids were obtained. The size of these offspring were determined after two years in the hatchery (cm).

	n	Y	s
F_1	144	34.5	1.15
F_2	100	31.5	4.20

Is there a significantly greater amount of variability in femur lengths among the F_1 than among the F_2 hybrid splake?

18. Specific conductance $\mu S/cm$) (micro statohms/cm) was measured in the S. Saskatchewan River prior to 1967 as part of a survey of Canada's water quality levels. A dam was completed on the river in 1967 and again conductance values were taken for a further 6 year period. The values before and after follow.

Was there a difference as a result of the dam in the variability of conductance readings?

Before		After	
625	380	410	395
625	350	410	390
600	375	415	385
600	350	410	385
580	350	410	385
500	325	410	385
500	330	410	385
480	325	410	385
480	325	415	350
470	315	410	375
450	300	450	375
450	290	440	375
450	280	435	380
440	280	435	380
440	275	435	380
425	270	435	385
425	275	430	280
400	260	420	210
400	265	420	
395	265	420	
390		420	
380		390	
380		395	

From: Surface water quality in Canada an overview.

19. Suppose the standard deviation of stature in males is 2.48 in. One hundred male students chosen at random in a large university are measured, and their average height is found to be 68.52 in. Determine the 99% confidence limits for the mean height of the men in this university.

20. Two different samples from the same population are used to find confidence limits for the mean of the population. The first sample gives confidence limits of 11 and 21, the second 13 and 25. If the same level of significance is used, what is the ratio of the size of the second sample?

21. The average monthly income of families in a certain city is to be estimated from a random sample selected from a directory. If the average income is to be estimated within $100.00 with a probability of 90% (that is, if we desire to be 90% confident that the sample mean and the mean income of all families do not differ by more than $100.00), find the least size of the sample required for such an estimate. Assume it is known that the standard deviation of the incomes of the families in the city is $200.00.

22. It is known from long experience that 5% of certain articles produced are defective and have to be discarded. A new man who has produced 600 of these articles has made 42 defective articles. Does this cast doubt on the man's ability to perform the job?

23. At a university, student records kept over a period of many years indicate that 64% pass the entrance examination in mathematics. During the fall of 1959, of 400 freshmen taking the test, 70% passed. Was this a significant improvement in the students' aptitude in mathematics?

24. Of 64 offspring of a certain cross between guinea pigs, 8 are black and the others are not black. According to the genetic model, these numbers (black and not black) should be in the ratio 3:13. Are the observed data consistent with the model at the 5% level of significance?

25. The height of adults in a certain town has a mean of 65.42 inches with a standard deviation of 2.32 inches. A sample of 144 adults living in the slum district is found to have a mean height of 64.82 inches. Does this indicate that the residents of the slums are significantly retarded in growth on the basis of the 1% level of significance?

26. The daily wages in a particular industry are normally distributed with a mean of $13.20 and a standard deviation of $2.50. If a company in this industry employing 40 workers pays these workers on the average $12.20, can this company be accused of paying inferior wages at the 1% level of significance?