SAULT COLLEGE

of Applied Arts and Technology

Sault Ste. Marie

COURSE OUTLINE

BIOMETRICS

BIO 308-4

revised June 1981 by H. Robbins

BIOMETRICS

BIO 308-4

COURSE OUTLINE

TEXT: "Introduction to Biostatistics" by Sokal and Rohlf

Unit #1 (8 hrs.) - Single classification ANOVA -- calculation of sources of variation, construction of ANOVA table, calculation of F-ratio, interpretation of results for Model I and Model II ANOVA, comparison with student's t-distribution, a priori and a posteriori tests.

Unit #2 (8 hrs.) - Two-way ANOVA, its calculation, construction of ANOVA table, interpretation of results, interaction between variables, significance testing, replication, assumptions of ANOVA, alternative non-parametric tests in lieu of ANOVA.

Unit #3 (10 hrs.) - Regression and correlation, straight-line regression and its calculation, Model I and Model II, replication, combination ANOVA - regression analysis and interpretation of results, tests of significance, calculation of correlation coefficient, confidence limits

Unit #4 (6 hrs.) - Chi-square test for goodness of fit, test of independence, use of computers in biology, preparation of data for computer analysis

Unit #5 (12 hrs.) - Population estimation methods, mark-recapture, sources of error, multiple mark-recapture methods, importance of sample size, survey removal method

(a) Prior to looking at the results, discuss what we might test for in the data i.e. - what would you expect to influence the total number of park users in any particular week over the past ten years? It is necessary to ask these pertinent questions prior to looking at the data (before doing the experiment) in order to use the more powerful tests. We will use these questions to tell us what to test for.

(b) Is this a model 1 or model II or mixed ANOVA?

(c) Does the superintendent have any justification for increasing or decreasing his staff in any week during the summer? i.e. are there differences in park usages as a result of the week of the summer?

(d) If the answer to the above (c) questions are affirmative, should he have more staff in June, July or August or will two of these months be the same?

(e) Can he justify increasing his staff over the whole summer based on the assumption that park attendance increases annually? Base this firstly on the last ten years then on the last five years.

Data for Question #2

	JUNE WEEK		JULY WEEK				AUGUST WEEK			
	3	4	1	2	3	4	1	2	3	4
1975	5.1	5.9	10.8	7.3	15.1	20.1	20.1	16.1	10.0	7.4
1974	4.9	5.1	3.1	7.7	15.1	25.0	19.0	15.0	9.9	8.4
1973	7.1	4.9	9.7	10.1	16.0	19.0	11.1	20.1	9.9	8.0
1972	4.4	4.4	7.6	7.9	17.0	12.1	17.6	15.5	9.9	7.1
1971	3.6	3.8	6.6	7.0	16.5	16.5	16.5	17.1	9.0	5.9
1970	2.5	2.5	6.8	7.0	18.0	12.8	17.1	12.9	8.9	6.1
1969	2.1	2.1	5.9	7.1	10.8	8.9	12.0	7.1	7.9	2.3
1968	1.0	1.9	5.1	6.0	10.1	15.6	15.7	8.9	8.9	1.1
1967	N.A.	0.8	4.4	4.5	8.8	7.1	8.9	4.1	N.A.	N.A.
1966	0.8	1.0	2.1	N.A.	3.4	4.4	4.1	2.9	N.A.	N.A.

N.A. -- Figures unavailable or park not open

3. Dissolved oxygen values were determined in 6 locations in a bay and in the open part of a lake. The results arranged according to location and depth follow (ppm):

	Open Lake				Protected Bay	
	Leeward Side		Windward Side		Leeward Side	Windward Side
Location	A	B	A	B	A	B
Surface	8.8	9.0	9.0	9.0	9.0	8.9
10 m	8.4	8.4	8.9	8.8	8.0	8.6
20 m	8.4	8.5	8.6	8.6	7.0	8.4
30 m	7.5	7.4	8.4	8.5	5.0	5.5
bottom	6.0	---	8.1	8.0	1.5	4.4

Note: Assume model 1, one-way ANOVA (ignore depths)

(a) Suggest two a priori questions for which you would like to know the answers (ignore depth).

(b) Ignoring the effects of depth, are there differences between locations?

(c) Answer your a priori questions and include in the above analysis.

(d) Ask one a posteriori question based on the results of your analysis and answer it.

(e) Treat the above data as a two-way analysis of variance and analyze.

4. Rainbow and brook trout were placed together in each of 10 raceways at a hatchery to determine which species grows more quickly in the cold hatchery water (average weight in grams after 1 year):

	Brook trout	Rainbow trout
1.	14.1	16.8
2.	20.8	20.5
3.	10.4	11.4
4.	10.7	14.1
5.	12.1	12.9
6.	12.4	12.1
7.	14.1	16.9
8.	9.9	9.8
9.	15.0	18.0
10.	12.9	14.1

Is there a difference between the two species? What kind of ANOVA is this?

5. An experimenter is evaluating the effects of various types of tags on the growth of three species of fish. The results follow (increase in weight in grams/week).

Largemouth Bass		Bluegill	Croaker
Atkins	22.5	12.8	6.9
Petersen	30.1	17.2	10.8
Spaghetti	28.9	15.8	9.1
Strap	25.5	14.1	8.1
Control (untagged)	40.1	23.2	12.5

Analyze to determine the effects of the tags on growth. (From: Evaluation of various tagging methods on several fishes and estuarine fishes in Louisiana).

6. In a study, designed to estimate the size of a spawning population of perch, it was suspected that the r/m ratio (fraction of fish recaptured) might have varied with time because of changes in emigration and immigration later in the run. Five different tag colours were also used. Analyze completely and state your conclusions. In this question, the date reflects changes in population size in the spawning area and tag colour reflects date of marking of the fish. For example: lilac tags were applied on May 1, 2 and 3.

Date caught	Lilac May 1,2,3	Green May 3,4	Yellow May 4,5	Grey May 5,6	White May 6-10
May 10	.034	.032	.036	.027	-
May 11	.022	.032	.028	.017	.018
May 12	.012	.015	.012	.010	.023
May 13	.006	.008	.010	.012	.016
May 14	.001	.005	.008	.000	.006
May 15	.001	.002	.002	.000	.003

(arc. sign transformation of r/m)

From: Estimates of a population of spawning perch and of the efficiency of trap and gill-net fishing.

7. In table 1, fish species diversity indexes are listed for littoral zone species in Lake Memphremagog. A is the value for permanent littoral zone species, B is the value index for the total catch including young-of-the-year which may leave at a later date. The values above the four locations are the levels of total phosphorus

(mg/m^3) obtained. The author wishes to determine whether the level of phosphorus influences the diversity index for fish. You should consider each year separately and then combine the two years in an overall analysis.

Table 1: Fish species diversity (H) values for littoral zone fish communities in four regions of Lake Memphremagog. A, permanent littoral zone species; B, total catch, young-of-the-year included.

	(28.5) Newport Bay		(20.0) South Basin		(12.1) Central Basin		(10.0) North Basin	
	A	B	A	B	A	B	A	B
1974								
June 21	2.03	2.12	2.34	2.24	.49	.38	.92	.32
July 30	1.49	1.91	2.33	1.00	1.27	1.54	1.23	1.35
Aug. 26	2.48	2.71	.55	1.36	.73	1.47	1.66	1.79
Oct. 6	2.73	1.89	1.92	1.97	.95	2.23	1.95	2.63
Mean	2.18	2.15	1.78	1.64	.86	1.40	1.44	1.54
1975								
May 22	1.53	1.52	2.13	2.13	1.46	1.44	1.49	.99
July 14	1.61	1.47	2.28	2.41	1.66	2.12	1.62	2.11
Aug. 30	2.44	2.37	2.40	2.97	2.21	2.28	2.17	.68
Oct. 22	2.47	1.54	.34	1.51			1.39	1.46
Mean	2.01	1.73	2.00	2.26	1.77	1.95	1.66	1.31
Grand Mean	2.09	1.94	1.89	1.95	1.32	1.68	1.55	1.42

From: Species diversity of littoral zone fishes along a phosphorus production gradient in Lake Memphremagog, Quebec-Vermont.

8. Perform Wilcoxin's signed-ranks, and the sign test on the data in question (4) above. Compare the results of the three methods of analysis and explain why you obtained different conclusions.

Correlation and Regression

9. State whether analysis of the following problems is most appropriate by correlation and/or regression and why.

   (a) It is suspected that an industrial user of water is altering the temperature before returning it to the river. The temperature is recorded at intervals downstream from the factory and is found to decrease with distance from the factory.

9. (b) Two species of fish migrate up a river at about the same time of year. A trap records the numbers moving upstream on a daily basis. The investigator wishes to know whether the two species respond in the same way to river conditions (temperature and discharge).

(c) A park superintendent in a remote region is responsible for two parks, one in which his office is located and one 50 miles down a nearby road. There are no phones. He wishes to devise a method of determining when the distant park is full by surveying his own park and the road to both parks.

(d) In a hatchery, the superintendent wants to determine whether heating the water will improve his cost/production ratio. That is, he wishes to determine under what conditions his production costs will be lowest.

10. In fisheries the usual method of determining the size of fish at different ages is by back calculation of scales. A relationship between scale radius and fish length is developed. Determine this relationship from the following data:

Scale radius (cm)	Fish length (cm)
.050	37.5
.040	30.4
.151	52.2
.044	31.5
.095	44.5
.008	10.5
.019	20.4
.034	27.2
.041	31.5
.024	22.4
.026	23.1
.075	38.9

11. A common method of population census of fish in streams is the removal method. The stream is shocked three or four times and each time the fish are removed and counted. The number of fish removed per unit of effort (one pass through with the shocker) is plotted against the cumulative number of fish captured. Determine the total population size in the section of stream by extrapolating the graph to the point where the catch/unit effort becomes negligible. Graph the results also.

Cum. catch	C.U.E.
0	450
450	40
490	6
496	1

12. In the summer stream water temperatures tend to be highest on clear sunny days and lowest on cool rainy days. An investigator suspects that stream temperature and turbidity may vary together. Determine if this is so.

Temperature	Turbidity (ppm)
11.0	.05
5.0	.15
2.9	.10
1.6	.30
1.2	.60

13. Deer are kept in an enclosure and fed prepared food of different protein and vitamin levels to determine if this influences the weight lost over the winter.

Vitamin & Protein Content (ppm)	Weight Loss (kg)
5.0	15.5
5.0	12.1
5.0	4.1
10.0	4.8
10.0	12.4
10.0	8.1
20.0	8.1

14. The amount of phosphorus in a freshwater lake is limited since much of it is often tied up in the vegetation and algae. Does the following data indicate an interrelationship between amount of algae and phosphorus in the water?

Algae (gm/m^3)	Phosphorus (ppm)
1.5	15
1.4	15
1.5	21
.5	28
1.9	10
.2	35
1.2	25
2.1	5

15. Grand River specific conductance values were recorded over a period of 10 years to determine trends in water quality. The results follow. Is there a significant trend in specific conductance over the ten year period?

Year	Specific Conductance ($\mu s/cm$)
1967	520
1968	520

Year	Specific Conductance (μs/cm)
1968	550
1968	530
1968	575
1968	490
1969	410
1969	510
1969	610
1969	600
1969	590
1970	410
1970	790
1972	750
1972	710
1972	610
1972	610
1972	550
1972	590
1973	610
1973	615
1973	605
1973	560
1973	480
1974	580
1974	650
1974	625
1974	620
1974	615
1974	480
1974	590
1975	800
1975	760
1975	625
1975	610
1975	590
1975	580
1975	560

From: Surface water quality in Canada: an overview.

16. In a study of the fecundity of brook trout, the researcher wishes to find the dependence of number of ova on body weight. The data follows:

No. of ova	Body weight (g)
2500	650
1300	350
1250	250
975	220
820	190
900	180

No. of ova	Body weight (g)
200	50
210	55
250	55
375	75
375	75
400	80
400	80
450	100
500	175
520	150
520	100
600	120
625	115
600	100

Determine the relationship. (From: fecundity of brook trout from a coastal stream in Prince Edward Island).

17. In a study of a Quebec population of brook trout, it was suspected that fecundity and egg diameter varied together and that perhaps both were dependent on length of fish. The data follows:

Fecundity	Egg diameter (mm)
109	3.52
193	4.19
271	4.22
356	4.32
426	4.16

Analyze and state your conclusions. (From: a population study of brook trout).

18. Observations of the wintering of waterfowl in the Toronto area have been made over many years. An example of the results for the mallard and black duck follows. Determine for each species whether there is a significant trend in numbers of birds over-wintering at Toronto.

Year	Mallard	Black Duck
1947	228	823
1948	295	1886
1949	403	1002
1950	612	952
1951	610	1075
1952	756	1238
1953	782	1218
1954	592	1289
1955	881	795
1956	781	847
1957	675	481

Year	Mallard	Black Duck
1958	604	595
1959	742	628
1960	781	1353
1961	1556	1304
1962	2428	1104
1963	1667	958
1964	1665	1508
1965	2295	1598
1966	2364	1934
1967	3373	1354
1968	2493	949
1969	2101	347
1970	2371	567
1971	2386	824
1972	2564	1044
1973	3219	1751
1974	4482	1104
1975	4374	989
1976	4025	735

From: population trends in waterfowl wintering in the Toronto region, 1929-1976

19. In the above question (18), do the numbers of the two species of ducks vary together?

20. In the annual waterfowl census at Toronto, the investigators wished to determine how close the numbers censused during the mid-winter census varied with the numbers observed during the Christmas census. The data follows for the Canada Goose only. What is your conclusion?

Year	Mid-winter census	Christmas census
1963	0	0
1964	65	1
1965	108	115
1966	140	222
1967	246	266
1968	275	487
1969	229	183
1970	354	636
1971	608	408
1972	680	1102
1973	1118	848
1974	894	747
1975	1788	1672
1976	1397	1909

From: Population trends in waterfowl wintering in the Toronto region, 1929-1976.

CORRELATION

SAMPLE QUESTION

21. In a laboratory experiment on the activity level of smallmouth and largemouth bass, it was suspected that selection of black sustrate over white, when an equal amount of each was available, was associated with changes in activity level of the fish. As the fish moved around more (were more active) they were less inclined to be selective of the black substrate because they were less closely associated with the substrate a) Determine if this is so for each species b) Determine if both species were equally affected.

It has been previously shown that both activity level and selection of substrate are dependent on environmental factors such as temperature, light intensity and fish age.

SMALLMOUTH BASS		LARGEMOUTH BASS	
Activity	Number over Black Substrate	Activity	Number over Black Substrate
.85	10.3	.88	12.5
.90	10.1	.70	13.0
.15	16.5	.05	16.1
.50	14.0	.01	16.0
.44	14.1	.00	17.5
.36	13.0	.21	16.0
.15	17.0	.35	12.8
.01	19.0	.41	12.9
.05	19.1	.60	12.4
.09	18.8	.30	14.1
.35	15.3	.77	12.0
.40	12.0	.65	14.5
.72	11.1	.45	12.1
.00	19.0	.25	12.7
.65	10.1	.90	10.1

22. In question #7, using the phosphorus levels given in brackets, determine the relationship between phosphorus level and diversity index for each year for permanent species and total population. Which relationships are significant?

23. An investigator is attempting to determine whether a new method of control for spruce budworm is working. In the region in general, the probability that a spruce bough is infested is known to be 0.85. Eight sample plots located in the test area were surveyed. The results follow:

Plot	No. Infected	No. not infected
1	17	17
2	44	19
3	40	30
4	24	20
5	31	10
6	15	12
7	26	20
8	40	38

Test whether this method of control is beneficial in killing spruce budworm.

24. Chi Square

Of 445 adult deer from northern Ontario, 45% weigh more than 50 kg, 35% more than 100 kg and 20% less than 100 kg. A corresponding study of 50 adult deer from Long Point in S. Ontario found 10% weighed more than 50 kg, 20% more than 100 kg and 70% less than 100 kg. Is the weight of deer independent of location?

25. It was suspected by an experimenter that tag colour might have an influence on the rate of tag loss since fish tended to bite off the tags. Largemouth bass were tagged with Floy tags of 3 colours as follows:

	Total Tagged	No. Lost Tags
Brown	38	14
Green	42	16
Orange	80	34

Is tag loss independent of tag colour? (From: Differential retention of five Floy tags on largemouth bass in hatchery ponds.)

26. Three species of forage fish were tested for vulnerability to predation by smallmouth bass. The data follows. Are the different species predated upon, at the same rate, by smallmouth bass?

	Total	Common Shiner	Hornyhead Chub	White Sucker
No. exposed	81	29	29	23
No. eaten	21	14	6	1

From: Vulnerability of three species of forage fish to predation by smallmouth bass in a hatchery trough.

27. **Petersen & Schnabel Population Estimates**

Table 1 lists results of a study to determine the estimates of the size of a spawning population of perch. Determine, using both the Petersen and the Schnabel methods, the estimates of population size for each day, beginning May 2, for each tag colour. Also calculate the 95% confidence limits. Accumulate the number marked for each colour up to and including the day previous to the date for which the estimation is made. For example: the first Schnabel estimate of population size would be based on a catch of 538, with m at 96 and r at 10. The first Petersen estimate would be for May 4 as the instantaneous recapture rate is required.

Table 1. *Daily catch, number marked (m), number recaptured (r) and fraction recaptured (r/m).*

Date	Catch	Catch per trap	Lilac			Green			Yellow			Grey			White			Fin-cut	
			m	r	r/m	m	r	r/m	m	r	r/m	m	r	r/m	m	r	r/m	m	r
1.5	98	4.9	96	1															
2.5	538	23.9	525	10															
3.5	361	13.6	59	26		271	5												
4.5	752	27.9		43	0.063	394	27		283	5									
5.5	580	28.9		49	0.072		33	0.050	214	28		252	1						
6.5	611	27.9		52	0.076		34	0.051		40	0.080	442	15		28	0			
7.5	340	15.3		21	0.031		24	0.036		10	0.020		31	0.045	253	0			
8.5	663	26.5		46	0.068		52	0.078		29	0.058		43	0.062	478	15			
9.5	553	25.5		34	0.050		34	0.051		18	0.036		27	0.039	409	31			
10.5	412	16.8		23	0.034		21	0.032		18	0.036		19	0.027	298	33			
[illegible]	320	15.1		15	0.022		21	0.032		14	0.028		12	0.017		27	0.018	225	6
[illegible]	269	14.2		8	0.012		10	0.015		6	0.012		7	0.010		34	0.023	193	11
13.5	199	10.5		4	0.006		5	0.008		5	0.010		8	0.012		23	0.016	143	11
14.5	87	4.6		1	0.001		3	0.005		4	0.008		0	0		9	0.006	60	8
15.5	59	3.1		1	0.001		1	0.002		1	0.002		0	0		5	0.003	44	7
Total	5,842		680	334	0.49	665	270	0.41	497	178	0.36	694	163	0.23	1466	177	0.12	665	43

Computational Formulae for Regression Analysis

(single Y for each value of X)

Sum of squares of X $= \varepsilon x^2 = \varepsilon X^2 - \frac{(\varepsilon X)^2}{n}$

Sum of squares of Y $= \varepsilon y^2 = \varepsilon Y^2 - \frac{(\varepsilon Y)^2}{n}$

Sum of products $= \varepsilon xy = \varepsilon XY - \frac{(\varepsilon X)(\varepsilon Y)}{n}$

Explained sum of squares $= \varepsilon \hat{y}^2 = \frac{(\varepsilon xy)^2}{\varepsilon x^2}$

Unexplained sum of squares $= \varepsilon d^2_{y.x} = \varepsilon y^2 - \varepsilon \hat{y}^2 = \varepsilon y^2 - \frac{(\varepsilon xy)^2}{\varepsilon x^2}$

Regression coefficient $= b_{y.x} = \frac{\varepsilon xy}{\varepsilon x^2}$

Y-intercept $= a = \bar{Y} - b_{y.x}\bar{X}$