Línuleg tölfræðilíkön (09.15.46)

Vikublað 5

1: Let C_a be the number of fish which are age determined to age a. Assume that the errors in age determination is such that a fish can move up or down one age group. Thus, the correct number is K_a, but the number of misreadings are either up from a to $a+1$ being U_a and down from a to $a-1$ misreadings are V_a and therefore the observed number is $C_a = K_a + U_{a-1} - U_a + V_{a+1} - V_a$. Assume that U_a and V_a are random variables with the same fixed variance, independent of age, a.

Find the correlation matrix of the vector $(C_1, \ldots, C_A)'$.

How would you set up a regression analysis in order to describe C_a as a second degree polynomial in age?

2: Consider the catch data in the following table (millions of fish per year and age group). Analyse this using some anova or ancova and investigate correlations of the residuals. Attempt to estimate (some of) the correlations and modify the model accordingly.

	1982	1983	1984	1985	1986	1987
3	3.29	3.55	6.75	6.46	20.64	11.00
4	20.81	10.91	31553	24.55	20.33	62.13
5	24.46	24.31	19.42	35.39	26.64	27.19
6	28.35	18.94	15.33	18.27	30.84	15.13
7	14.01	17.38	8.08	8.71	11.41	15.70
8	7.67	8.38	7.34	4.20	4.44	4.16
9	11.52	2.05	2.68	2.26	1.77	1.46

One possible model is

$$ln\,(C_{ay}) = \mu + \alpha_a + \beta_y + \gamma_{a-y}.$$

(note that the last factor is indexed by $a-y$ as an index).

(a) What is the meaning of the factors?

(b) How are they confounded?

(c) Can the model be used for estimating the variances and correlations in spite of the confounding?

3: Consider the linear model $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{jik},\ 1 \le k \le K_{ij},\ 1 \le i \le I\ 1 \le j \le J$, where ϵ_{ijk} are i.i.d. with zero mean.

First assume we want to use the constraints $\sum \alpha_i = 0$ and $\sum \beta_j = 0$, and that $K_{ij} = K$ is fixed.

(a) Solve the LS estimation problem using lagrange multipliers.

(b) Set up the appropriate matrices leading to the revised normal equations for the constraint and solve the LS estimation problem using matrix algebra.

4: What would the appropriate constraint be in (3) in the case of unequal sample sizes?

5: When the **x**-variables are highly correlated it may become important to reduce this collinearity to obtain more stable estimates. Solve this issue in the case of quadratic regression with $x_i = i, \quad i = 1, \ldots, 10$.

6: Derive the general solution for the estimated regression coefficients in multiple linear regression (with intercept) when the **x**-variables are completely uncorrelated. What happens when the scaling is further implemented so that each variable (including y) has zero mean and the standard deviation is 1?

7: Use the following data in a one-way analysis of variance. Use appropriate multiple comparisons to evaluate pairwise differences as well as to evaluate whether some quadratic effect exists.

		Replication(k)					
		1	2	3	n	mean	stdev
Factor	1	3.10	3.12	2.45	3	2.89	0.381
level	2	0.90	1.76	1.06	3	1.24	0.457
(i)	3	3.16	1.49	1.67	3	2.11	0.917

Do the exercise first by hand and repeat in R.

8: Derive a formula for $E[MSA]$ in the equal-replicate two-way layout.

9: Prove the orthogonality (and thus independence) of the residuals which form SSA and SSB in the single-replicate two-way layout.

10: The following table contains, for each year (t) measurements of temperature (T), shrimp abundance (U), shrimp juvenile abundance (I), shrimp cathch (Y), capelin biomass(B), growth of 4-year old cod (G), spawning stock biomass of cod (S) and biomass of juvenile cod (J).

Interpret these data.

t	T	U	I	Y	B	G	S	J
79	0.5	75.7	2313	1.1	3177	809	447	872
80	5.7	79.8	4747	3.1	2210	777	602	880
81	2.7	77.6	3217	2.1	1442	398	389	704
82	2.7	76.4	1909	1.7	1128	595	266	623
83	1.2	85.0	4368	6.1	2182	725	214	584
84	3.5	86.0	2418	12.2	3579	997	219	605
85	5.0	93.0	3930	12.2	3688	851	268	577
86	3.5	89.0	4943	17.1	3987	873	268	768
87	4.4	77.5	4309	24.6	3727	725	253	921
88	1.7	65.8	4089	20.7	2990	620	193	818
89	3.3	72.0	4994	18.1	2677	785	269	595
90	3.2	81.6	8180	19.4	2146	570	344	408
91	3.6	87.1	8406	26.1	2454	771	232	508
92	4.3	83.5	6376	27.4	3050	570	244	357
93	4.3	94.0	7192	30.1	3185	1004	224	358
94	4.7	104.6	9611	42.1	3119	675	276	292
95	0.3	87.6	9742	49.2	3700	857	380	189

11: A comparison was undertaken on the yield of four varieties of oats, A, B, C and D, measures in kg wet weight per square meter.

	A	B	C	D
n_i	6	6	3	3
$\sum_j x_{ij}$	26.8	30.4	12.6	16.3
$\sum_j x_{ij}^2$	122.14	155.44	53.90	89.45

a) Is there a significant difference between the varieties?

b) A and C are summer varieties but B and D are winter varieties. Is there a difference between the seasonal varieties?

c) Do all-pairwise comparisons between the varieties. What method should be used and why?

12: Phospate from three farms, measured as the percentage of dry weight yielded the following measurements:

Farm	n_i	$\bar{x}_i$	s_i
1	2	0.26	0.021
2	8	0.29	0.024
3	6	0.32	0.026

a) Is the difference between farms significant?

b) Do multiple comparisons between the values at the farms.

13: The following table gives the weight, sex and length of 25 herring, obtained in a sample in 1992.

Use weight and length on log-scale. Do appropriate analyses to investigate whether length and weight are significantly related, whether mean weight is variable by sex and whether the length-weight relationship is variable among the sexes, first the intercept, then the slope and finally either one.

w	s	l
166	1	28
165	0	28
59	1	21
271	1	33
178	0	29
67	0	22
312	1	36
257	0	33
38	1	18
91	1	23
210	1	31
256	1	33
163	1	28
128	0	26
122	0	26
223	1	31
287	1	33
61	0	22
102	1	25
60	0	21
53	0	20
66	0	22
149	0	27
165	0	28
74	1	22

Assorted exercises from the book: 9.1, 9.11, 11.15, 16.7, 21.12