E

TWC/23/18
ORIGINAL: English
DATE: June 15, 2005

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

Twenty-Third Session
Ottawa, June 13 to 16, 2005

USE OF GENERALIZED LINEAR MODELS IN DUS

LOGISTIC REGRESSION APPROACH

Document prepared by experts from Kenya

USE OF GENERALIZED LINEAR MODELSIN DUS

LOGISTIC REGRESSION APPROACH

NGENY JOHN MARK

KENYA PLANT HEALTH INSPECTORATE SERVICE

KENYA

LOGISTIC REGRESSION

- Used when response variables are qualitative.
- Still used when response variables are quantitative

BINARY RESPONSES

Denote the probability of an event occurring as p.

Probability of non-occurrence is 1-p.

Then odds ratio is the ratio is defined as:

$$O = \frac{p}{1 - P}$$

Assuming the data comes from a logit model, then

$$p = \frac{1}{1 + e^{-\beta x}}$$

BINARY (Cont.)

Then $O = \frac{p}{1-P} = \left(\frac{1}{1+e^{-\beta x}}\right) \Big/ \left(1 - \frac{1}{1+e^{-\beta x}}\right) = e^{\beta x}$

Thus $\log\left(\frac{p}{1-P}\right) = \beta x$

MORE THAN TWO RESPONSES

- ➔ Example of 3 response variables and 5 independent class variables
- ➔ Order the response variables
- ➔ Let p_i= event is response I, I=1,2,3 and $p_{1}+ p_{2} + p_{3} =1$
- ➔ Use of cumulative logit model
  - ➔ For j=1,2,3 let $F_{ij} = \sum_{m=1}^{j} p_{im}$ be the probability of jth response variable for the ith class variable.
  - ➔ Then $\log\left(\frac{F_{ij}}{1-F_{ij}}\right) = \underline{\beta}\,\underline{x}_i = \beta_{i1} + \beta_{i2} + \beta_{i3}$

EXAMPLE:

- ✓ DUS on Snap (French) bean.

EXAMPLE (Cont.)

- ✓ Response variable: Shape of curvature of pod with 3 states

1 CONCAVE

2 S-SHAPED

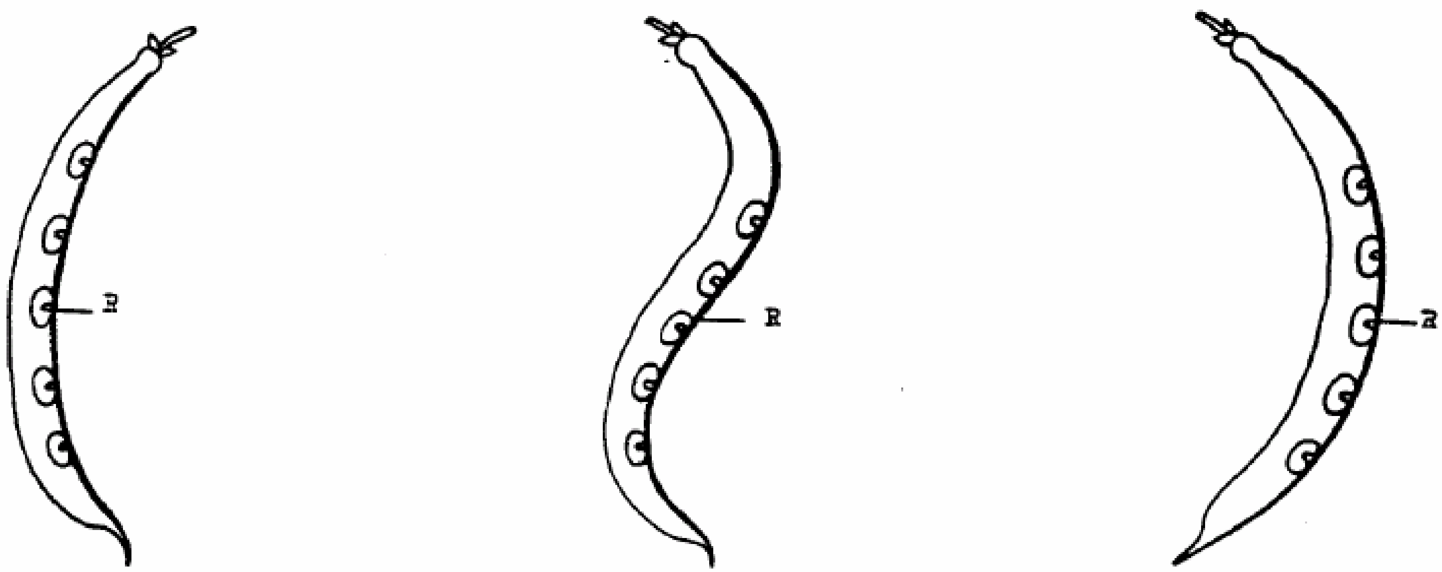
3 CONVEX

Ad/Add./Zu 28

Pod: shape of curvature

Gousse: forme de la courbure

Hülse: Art der Krümmung



1	2	3
concave concave konkav	s-shaped en S s-förmig	convex convexe konvex

R - dorsal suture
- suture dorsale
- Rückennaht

Data collected as frequencies

		2003			2004			COMBINED		
ENTRY	VCODE	1	2	3	1	2	3	1	2	3
1	R1	40	.	.	34	6	.	74	6	.
2	C1	21	8	11	23	11	6	44	19	17
4	R2	40	.	.	34	6	.	74	6	.
7	R3	.	40	.	32	8	.	32	48	.
8	C2	.	.	40	8	1	31	8	1	71

- R1,R2 and R3 are reference varieties
- C1 and C2 are candidate varieties

METHOD 1: GLM

The SAS

The GLM Procedure

VCODE	Estimate	Lower	Upper
C1	1.6625	0.8466	2.4784
C2	2.7875	1.9716	3.6034
R1	1.0750	0.2591	1.8909
R2	1.0750	0.2591	1.8909
R3	1.6000	0.7841	2.4159

METHOD 1 (Cont.)

Least Squares Means for Effect VCODE

t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: SCORE

i/j	1	2	3	4	5
1		-2.20572	1.151875	1.151875	0.12254
		0.0548	0.2791	0.2791	0.9052
2	2.205717		3.357592	3.357592	2.328257
	0.0548		0.0084	0.0084	0.0449
3	-1.15187	-3.35759		0	-1.02933
	0.2791	0.0084		1.0000	0.3302
4	-1.15187	-3.35759	0		-1.02933
	0.2791	0.0084	1.0000		0.3302
5	-0.12254	-2.32826	1.029335	1.029335	
	0.9052	0.0449	0.3302	0.3302	

METHOD 2: LOGISTIC

Analysis of Maximum Likelihood Estimates

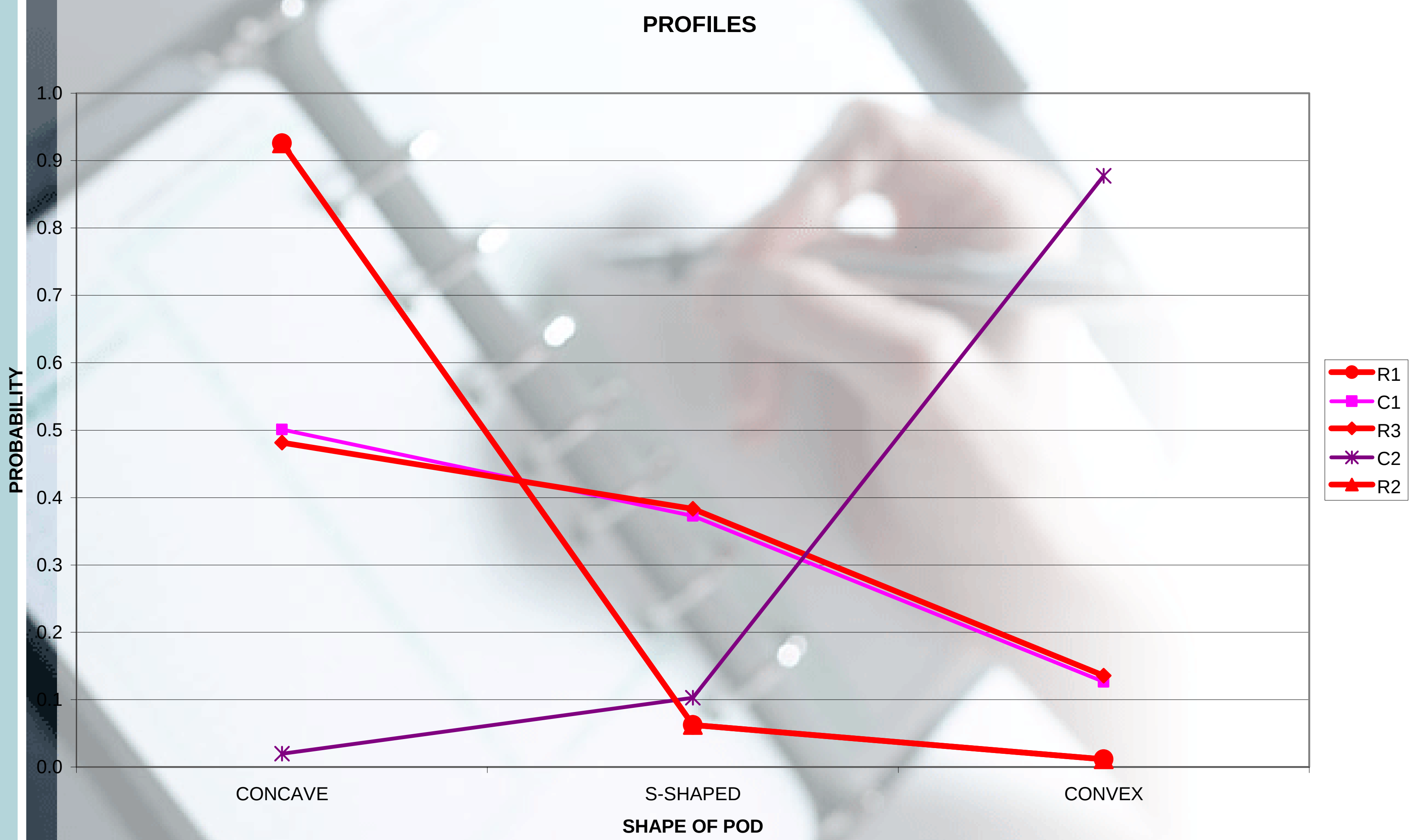
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	0.2164	0.1622	1.7803	0.1821
Intercept	2	1	2.1441	0.2116	102.6809	<.0001
VCODE	C1	1	-0.2129	0.2248	0.8963	0.3438
VCODE	C2	1	-4.1127	0.3362	149.6566	<.0001
VCODE	R1	1	2.3084	0.3677	39.4141	<.0001
VCODE	R2	1	2.3084	0.3677	39.4141	<.0001

METHOD 2 (Cont)

Contrast Rows Estimation and Testing Results

Contrast	Type	Row	Estimate	Standard Error	Alpha	Lower Limit	Upper Limit
C1	PARM	1	-0.2129	0.2248	0.05	-0.6535	0.2278
C1	EXP	1	0.8083	0.1817	0.05	0.5202	1.2558
C2	PARM	1	-4.1127	0.3362	0.05	-4.7716	-3.4538
C2	EXP	1	0.0164	0.00550	0.05	0.00847	0.0316
R1	PARM	1	2.3084	0.3677	0.05	1.5877	3.0291
R1	EXP	1	10.0582	3.6983	0.05	4.8926	20.6777
R2	PARM	1	2.3084	0.3677	0.05	1.5877	3.0291
R2	EXP	1	10.0582	3.6983	0.05	4.8926	20.6777
C1 Vs R1	PARM	1	-2.5212	0.4788	0.05	-3.4596	-1.5829
C1 Vs R2	PARM	1	-2.5212	0.4788	0.05	-3.4596	-1.5829
C2 Vs R1	PARM	1	-6.4211	0.5813	0.05	-7.5604	-5.2817
C2 Vs R2	PARM	1	-6.4211	0.5813	0.05	-7.5604	-5.2817

METHOD 2 (Cont)



SOME ISSUES TO PONDER

Interpretation of results

- ♦ No direct mapping of estimates to what was observed
- ♦ What do estimates mean?

Computational consideration

- Estimation of parameters by iterative process.
- If convergence is not met then what?

'

'

'