

A FUNÇÃO LOGÍSTICA COM UMA OU MAIS VARIÁVEIS
INDEPENDENTES PARA DADOS DE RESPOSTA QUÂNTICA

Elizabete Rodrigues de Oliveira

Orientadora: SONIA VIEIRA

Dissertação apresentada ao Instituto
de Matemática, Estatística e Ciência
da Computação da Universidade Esta-
dual de Campinas, como requisito par-
cial para a obtenção do título de
Mestre em Estatística

CAMPINAS
Estado de São Paulo - Brasil
1979

UNICAMP
BIBLIOTECA CENTRAL

À meus pais

AGRADECIMENTOS

Agradeço à Profa. Sonia Vieira todo apoio dado durante a elaboração deste trabalho.

Ao meu marido, Mário Alcides de Oliveira Scafí, agradeço a revisão do texto original.

Ao Prof. Rodolfo Hoffmann, agradeço a elaboração do programa para computadores utilizado neste trabalho.

Agradeço ao Prof. Frederico M. Wiendl pelo fornecimento dos dados apresentados como exemplo na seção 4.3.

A Sra. Elisa S. Peron agradeço a cuidadosa datilografia deste trabalho.

ÍNDICE

	Página
RESUMO	vi
SUMMARY	vii
1. INTRODUÇÃO	1
1.1. Dados de Resposta Quântica	1
1.2. Análise Estatística de Dados de Resposta Quântica	7
1.3. Proposição	9
2. A FUNÇÃO LOGÍSTICA COM UMA VARIÁVEL INDEPENDENTE, PARA DADOS DE RESPOSTA QUÂNTICA	10
2.1. Características da função	10
2.2. Estimativas dos parâmetros	12
2.2.1. O método de Berkson	12
2.2.2. O método dos mínimos quadrados ponderados não- lineares	16
2.2.3. O método da máxima verossimilhança	20
2.2.4. Considerações sobre as estimativas preliminares utilizadas no método dos mínimos quadrados pon- derados não-lineares e no método da máxima ve- rossimilhança	22
2.3. Variâncias e Covariâncias das Estimativas dos Parâme- tros	23
2.3.1. Usando as estimativas dos parâmetros obtidas pelo método de Berkson	23
2.3.2. Usando as estimativas dos parâmetros obtidas pelo método da máxima verossimilhança	24
2.4. Variâncias das Estimativas dos Lógites e Variâncias das Estimativas das Proporções	26

2.4.1. Variâncias das estimativas dos lógites	27
2.4.2. Variâncias das estimativas das proporções	27
3. A FUNÇÃO LOGÍSTICA COM DUAS OU MAIS VARIÁVEIS INDEPENDENTES PARA DADOS DE RESPOSTA QUÂNTICA	30
3.1. Introdução	30
3.2. Estimativas dos parâmetros	31
3.2.1. Uma extensão do método de Berkson	32
3.2.2. Método dos mínimos quadrados ponderados não-lineares	34
3.2.3. O método da máxima verossimilhança	38
3.3. Variâncias e covariâncias das estimativas dos parâmetros	40
3.3.1. Usando as estimativas dos parâmetros obtidas pela extensão do método de Berkson	40
3.3.2. Usando as estimativas dos parâmetros obtidas pelo método da máxima verossimilhança	41
3.4. Variâncias das Estimativas dos Lógites e Variâncias das Estimativas das Proporções	42
3.4.1. Variâncias das estimativas dos lógites	42
3.4.2. Variâncias das estimativas das proporções	43
4. RESULTADOS	45
4.1. Introdução	45
4.2. O ajuste da função logística com uma variável independente	46
4.3. O ajuste da função logística com duas variáveis independentes	50
5. PROGRAMA PARA COMPUTADORES	56
BIBLIOGRAFIA	69

RESUMO

A função logística, definida por

$$P_i = \left[1 + e^{-\underline{x}_i \underline{\beta}} \right]^{-1}$$

onde P_i representa a probabilidade no i -ésimo grupo, $\underline{\beta}$ é um vetor-coluna de $k+1$ parâmetros desconhecidos e $\underline{x}_i$ é um vetor-linha de $k+1$ constantes conhecidas, com primeiro elemento igual a 1, pode ser ajustada aos dados de resposta quântica.

Neste trabalho são apresentados os estimadores dos parâmetros da função logística usando o método de Berkson, o método dos mínimos quadrados ponderados não-lineares e o método da máxima verossimilhança. Também são apresentados os estimadores das variâncias e das covariâncias das estimativas dos parâmetros e estimadores das variâncias das proporções estimadas, tanto quando se usa o método de Berkson como quando se usa o método de máxima verossimilhança.

São fornecidos alguns exemplos e são discutidos os resultados do ajuste da função logística para dados de resposta quântica, existentes na literatura.

Consta, em anexo, um programa para computador, em FORTRAN IV, com instruções para uso, que ajusta a função logística, quando a variável dependente é uma proporção, e existem k variáveis independentes.

SUMMARY

The logistic function, defined by

$$P_i = \left[1 + e^{-\underline{x}_i \underline{\beta}} \right]^{-1}$$

where P_i is the probability of success in the i^{th} group, $\underline{\beta}$ is a column of $k+1$ unknown parameters and $\underline{x}_i$ is a row of $k+1$ known constants, with the first element equal to 1, can be fitted to quantal response data.

In this paper, estimators for parameters of the logistic function, by using Berkson's method, non-linear weighted least squares method and maximum likelihood method are presented. Estimators for variances and covariances of estimatives of parameters and estimators for variances of estimated proportions, by using both Berkson's method and maximum likelihood method are also presented.

Some examples are provided and results obtained by fitting the logistic functions to quantal response data from other papers are discussed.

A computer-program, in FORTRAN-IV, with instructions for use, which can fit a logistic function with k independent variables when the dependent variable is a proportion, is included herein.

1 - INTRODUÇÃO

1.1 - Dados de Resposta Quântica

Os dados de resposta quântica, no conceito clássico apresentado por REIERSOL (1961), são obtidos quando se consideram s grupos de n_i , onde $i=1,2,\dots,s$, ensaios binomiais independentes, se relaciona ao i -ésimo grupo uma variável não-aleatória X_i e se observa, nesse grupo, a frequência relativa de sucessos $p_i = m_i/n_i$. Então m_i representa o número de sucessos observados no i -ésimo grupo. A variável aleatória M_i tem distribuição binomial com parâmetros n_i e P_i .

Nos exemplos clássicos de resposta quântica, $X_i, i=1,2,\dots,s$, são quantidades, de uma droga ou de um estímulo, ministradas a s grupos de indivíduos e se observa, em cada grupo, a proporção de indivíduos afetados.

FINNEY (1971) apresenta um exemplo bastante típico de dados de resposta quântica. Esse autor analisou a ação inseticida da rotenona para Macrosiphoniella sanborni (pulgão em crisântemo). No experimento, conduzido por MARTIN (1942), foram utilizadas cinco concentrações

x_i da rotenona. A cada concentração foram submetidos n_i insetos e se observou o número m_i de insetos mortos. Os dados estão na tabela 1.1.

Tabela 1.1 - Ação inseticida da rotenona para Macrosiphoniella sanborni

Concentração (mg/l) (x_i)	Número de insetos sub metidos (n_i)	Número de insetos mor tos (m_i)	proporção de insetos mortos (p_i)
2,6	50	6	0,120
3,8	48	16	0,333
5,1	46	24	0,522
7,7	49	42	0,857
10,2	50	44	0,880

Entretanto, existem problemas totalmente diversos que também se enquadram no conceito clássico de dados de resposta quântica. É o caso do exemplo citado por AITCHISON e BENNET (1970), onde x_i se refere à classe de renda pessoal e se observa, na i -ésima classe, a proporção de indivíduos que possuem automóvel.

Outro exemplo, aparentemente diverso mas que também se enquadra no conceito clássico de dados de resposta quântica, foi encontrado em astronomia. Existe uma pequena área na lua para a qual, além de fotografias comuns, existe uma excelente fotografia tirada fora da atmosfera terrestre. São dados os números (m_i) de crateras detectadas na fotografia comum e os números (n_i) de crateras contadas na fotografia tirada fora da atmosfera, para cada diâmetro (x_i). Tais dados são apresentados na

tabela 1.2. As razões entre os números de crateras detectadas pelas diferentes fotografias, isto é, $p_i = m_i/n_i$, podem ser vistas como dados de resposta quântica.

Tabela 1.2 - Número de crateras não vulcânicas de uma região da lua.

Diâmetro (em mm) (x_i)	Número detectado na fotografia excelente(n_i)	Número detectado na fotografia comum(m_i)	Proporção (p_i)
1,0	10	2	0,200
1,5	17	11	0,647
2,0	26	21	0,808
2,5	14	10	0,714
3,0	16	15	0,937
4,0	8	8	1,000
5,0	2	2	1,000

Fonte: - SCOTT, E. - Dados não publicados, apresentados em uma apostila mimeografada do Departamento de Estatística da Universidade da California, Berkeley.

Também cabe perfeitamente no conceito clássico de dados de resposta quântica o experimento feito por BOCK e JONES (1968). Esses autores estudaram a discriminação do paladar para solução com diferentes concentrações de cloreto de sódio. Para obter dados relativos à discriminação do paladar, os autores fixaram, como padrão, uma solução com concentração de 1% (1g/100ml) e estabeleceram seis soluções-testes com

concentrações x_i ; $i=1,2,\dots,6$. Foram então apresentados pares de soluções, constituídos pela solução padrão e uma solução-teste, para indivíduos de ambos os sexos que se apresentaram, voluntariamente, como juízes. Para melhor assegurar a independência do julgamento, os pares de soluções foram apresentados aos juízes em ordem aleatória. Ainda, cada juiz fez, ao longo do experimento que durou duas semanas, várias sessões e em cada sessão testou apenas um par de soluções. Em cada sessão, o juiz recebia duas vasilhas, sem qualquer marca. Uma das vasilhas continha a solução-padrão e a outra continha uma solução-teste. Cada juiz era então instruído para tomar todo o conteúdo da vasilha à esquerda, lavar a boca com água destilada, descansar 3 segundos e então tomar todo o conteúdo da vasilha à direita; depois esse juiz informava qual das duas vasilhas continha a água mais salgada. Os dados coletados estão apresentados na tabela 1.3.

Tabela 1.3 - Dados sobre discriminação do paladar para soluções com diferentes concentrações de cloreto de sódio

Concentração da solução teste(mg/100ml) (x_i)	Número de juízes (n_i)	Número de juízes que optaram pela solução teste como mais sal- gada (m_i)	Proporção (p_i)
0,7	49	2	0,041
0,8	48	3	0,062
0,9	48	13	0,271
1,1	50	31	0,620
1,2	47	38	0,809
1,3	49	48	0,970

Entretanto, também podem ser obtidos dados de resposta quântica em situações diversas das discutidas até aqui. Na conceituação de OLIVER (1964) tem-se dados de resposta quântica quando se observa a difusão de uma característica na população, ao longo do tempo. Nesta conceituação está implícita a idéia de que, quando a variável independente, que é o tempo, tende para infinito, a proporção de indivíduos da população que apresentam a característica tende para um.

FINNEY (1971) apresenta o seguinte exemplo, que se enquadra nesta conceituação: um grupo de adolescentes, do sexo feminino, com idades variando entre 9 e 17 anos foi distribuído em 25 faixas etárias; para a i -ésima faixa foi obtido o ponto médio de classe (x_i) e se anotou, das n_i adolescentes, o número m_i das que já haviam tido a menarca. Neste exemplo, existe a implicação de que, à medida que aumenta a idade, aumenta a proporção de adolescentes que já tiveram a menarca. Os dados utilizados pelo autor, que foram obtidos por MILICER e SZCZTKA (1966) constam na tabela 1.4.

Tabela 1.4 - Estudo da menarca em 3918 adolescentes de Varsóvia

Ponto médio da faixa etária (x_i)	Número de adolescentes		Proporção (p_i)
	total (n_i)	com menarca (m_i)	
9,21	376	0	0,000
10,21	200	0	0,000
10,58	93	0	0,000
10,83	120	2	0,017
11,08	90	2	0,022
11,33	88	5	0,057
11,58	105	10	0,095
11,83	111	17	0,153
12,08	100	16	0,160
12,33	93	29	0,312
12,58	100	39	0,390
12,83	108	51	0,472
13,08	99	47	0,475
13,33	106	67	0,632
13,58	105	81	0,771
13,83	117	88	0,752
14,08	98	79	0,806
14,33	97	90	0,928
14,58	120	113	0,942
14,83	102	95	0,931
15,08	122	117	0,959
15,33	111	107	0,964
15,58	94	92	0,979
15,83	114	112	0,982
17,58	1049	1049	1,000

1.2 - Análise Estatística de Dados de Resposta Quântica

A análise estatística de dados de resposta quântica exige a formulação de modelos que descrevam a relação funcional entre a probabilidade de sucesso P_i e a variável não-aleatória x_i .

FINNEY (1971) apresentou um estudo exaustivo de um modelo baseado na função de distribuição normal. Então, a probabilidade de sucesso P_i , em função da variável não-aleatória x_i , pode ser descrita usando a função de distribuição normal reduzida, isto é,

$$P_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y_i} e^{-\frac{y^2}{2}} dy \quad (1.2.1)$$

A abscissa Y_i da distribuição normal reduzida, correspondente à probabilidade P_i , é denominada próbite. A relação funcional entre o próbite e a variável não-aleatória x_i é dada pela função linear

$$Y_i = -\frac{u}{\sigma} + \frac{1}{\sigma} x_i \quad (1.2.2)$$

que permite estimar u e σ .

Alguns autores definem próbite como Y_i+5 , onde 5 é um valor arbitrário usado apenas para evitar o sinal negativo. Esta definição tem apenas interesse histórico, porque foi feita por BLISS (1935), autor que primeiro descreveu, na literatura, dados de resposta quântica.

É importante observar que o ajuste da função normal reduzida, como é definida em (1.2.1) implica em admitir que a capacidade de

respostas dos indivíduos tem distribuição normal.

BERKSON (1944) propôs que se ajustasse, aos dados de resposta quântica, a função logística com assíntota 1, definida por

$$P_i = \frac{1}{1+e^{-(\alpha+\beta x_i)}} \quad (1.2.3)$$

O ajuste dessa função aos dados de resposta quântica fica bastante fácil se for feita a transformação:

$$L_i = \log \left(\frac{P_i}{1-P_i} \right) = \alpha + \beta x_i \quad (1.2.4)$$

BERKSON (1944) denominou lógite a variável transformada L_i , definida em (1.2.4), por analogia com a palavra próbite.

Diversos autores, conforme mostra a extensa revisão da literatura apresentada por FINNEY (1978), observaram que as análises de próbites e lógites, feitas sobre os mesmos dados experimentais, conduzem a resultados perfeitamente concordantes. JOHNSON e KOTZ (1970) observaram que tais resultados são necessariamente semelhantes, devido à semelhança das distribuições normal e logística.

Entretanto, é importante observar que os lógites são de aplicação direta, o que não acontece com os próbites, que exigem o uso de uma tabela de distribuição normal reduzida. Ainda, conforme argumentou BERKSON (1944), os lógites têm a vantagem de não exigir a pressuposição de que a capacidade de respostas dos indivíduos tem distribuição normal, feita quando se usa o modelo (1.2.1).

VIEIRA (1975) e VIEIRA e HOFFMANN (1979) propuseram que se ajustasse, aos dados de resposta quântica, a função de Gompertz, com assíntota igual a 1, definida por:

$$P_i = e^{-\alpha\beta^{x_i}} \quad (1.2.5)$$

O ajuste dessa função é mais fácil se for definida a transformada

$$G_i = -\log(-\log P_i) = A+Bx_i$$

onde $A = -\log \alpha$ e $B = -\log \beta$.

É interessante observar que as distribuições normal e logística são simétricas, mas a distribuição de Gompertz é positivamente as simétrica. Então, conforme discutem os autores, deve-se esperar que os resultados do ajustamento da função de Gompertz, a dados de resposta quântica, sejam diferentes dos resultados obtidos com análises de próbites e lôgites. É claro que a validade do método deve ser discutida principalmente em termos dos resultados obtidos com dados experimentais.

1.3 - Proposição

Este trabalho apresenta um estudo do ajuste da função logística para dados de resposta quântica. São considerados tanto os casos de uma variável independente como os casos de duas ou mais variáveis independentes. São dados exemplos, com base em dados da literatura e são fornecidos programas para computador.

2 - A FUNÇÃO LOGÍSTICA COM UMA VARIÁVEL INDEPENDENTE, PARA DADOS DE RESPOSTA QUÂNTICA

2.1 - Características da função

A função logística, definida por

$$Z = \frac{\gamma}{1+e^{-(\alpha+\beta x)}} \quad (2.1.1)$$

onde γ , α e β são parâmetros, $\gamma > 0$ e $\beta > 0$, foi indicada para o estudo descritivo do crescimento de populações humanas por VERHULST (1845), que a denominou de "curva logística". Muitos anos mais tarde, PEARL e REED (1920), sem conhecerem a contribuição de VERHULST (1845), obtiveram a mesma função.

A função logística (2.1.1) é monotonicamente crescente e fica entre duas assíntotas: $Z=0$ e $Z=\gamma$. O parâmetro γ , que é a distância entre as duas assíntotas, é denominado "nível de saturação". A função logística (2.1.1) tem ponto de inflexão de coordenadas $x_i = -\alpha/\beta$ e

$Z = \gamma/2$ e é radialmente simétrica em torno desse ponto. O aspecto gráfico da função, mostrada na figura 2.1, é de um S alongado, entre as retas $Z = 0$ e $Z = \gamma$. Esse aspecto, característico desta e de outras funções, define uma família de curvas denominadas assintótico-sigmóides, muito utilizadas no estudo descritivo de dados de crescimento.

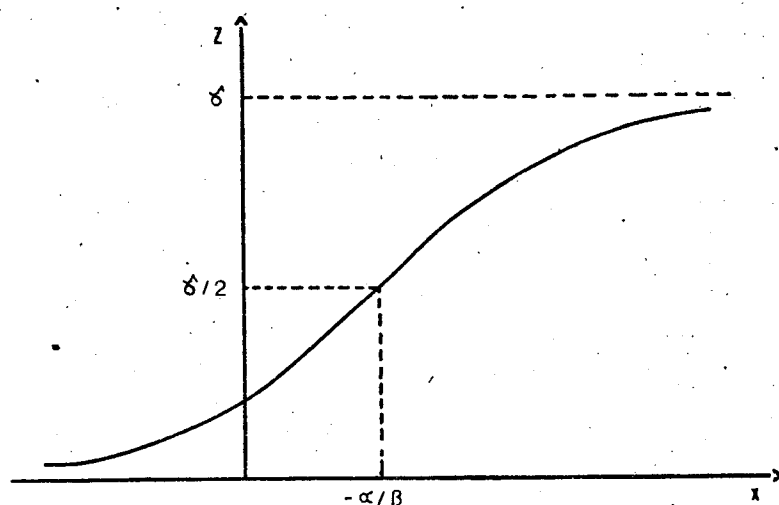


Figura 2.1 - Gráfico da função logística

Quando o nível de saturação γ é conhecido, ao invés de trabalhar com a variável Z , pode-se definir uma nova variável $P = Z/\gamma$, que representa a proporção do máximo valor possível de Z . Então, a função (2.1.1) fica definida como segue:

$$P_i = \left[1 + e^{-(\alpha + \beta x_i)} \right]^{-1} \quad (2.1.3)$$

A função logística (2.1.3) tem assíntotas $P = 0$ e $P = 1$.

Em diversos trabalhos BERKSON (1944, 1951, 1953, 1955) propôs que se ajustasse a função logística (2.1.3) aos dados de resposta quantitativa.

A função logística (2.1.3) é escrita por COX (1970) como segue:

$$p_i = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \quad (2.1.4)$$

2.2 - Estimativas dos parâmetros

O ajustamento da função logística para dados de resposta quântica pode ser feito através do método de BERKSON (1944), que propõe uma transformação para linearizar a função, ou pelos métodos clássicos de mínimos quadrados ou de máxima verossimilhança, também estudados pelo autor.

2.2.1 - O método de Berkson

Dado o modelo matemático da função logística (2.1.3), BERKSON (1944) obteve:

$$L_i = \log \frac{p_i}{1-p_i} = \alpha + \beta x_i \quad (2.2.1.1)$$

O autor denominou *lôgite* a transformada obtida em (2.2.1.1). Propõe então que, ao invés de se ajustar a função logística, partindo das frequências relativas p_i , se ajustasse uma regressão linear ponderada dos *lôgites* $\ell_i = \log(p_i/(1-p_i))$ contra x_i . Os fatores de ponderação são os inversos das variâncias de ℓ_i . Para obter as variâncias de ℓ_i , usa-se uma

aproximação através da série de Taylor^{1/}, desenvolvida em torno da média dos p_i ; isto é, em torno de $E(p_i) = P_i$. São usados apenas os dois primeiros termos da série.

Uma dada função $f(x)$, pode ser expandida através da série de Taylor em torno de um ponto x_0 . Considerando apenas os dois primeiros termos do desenvolvimento, vem:

$$f(x) = f(x_0) + f'(x_0)(x-x_0)$$

No ponto em que $x_0 = E(x)$, tem-se:

$$f(x) - f(E(x)) = f'(E(x)) (x-E(x))$$

Elevando ambos os membros ao quadrado e aplicando esperança matemática^{2/}, obtém-se:

$$V(f(x)) = [f'(E(x))]^2 V(x) \quad (2.2.1.2)$$

Lembrando que

$$\frac{\partial \ell_i}{\partial p_i} = \frac{1}{p_i(1-p_i)} \quad (2.2.1.3)$$

e

$$V(p_i) = \frac{P_i(1-P_i)}{n_i}, \quad (2.2.1.4)$$

substituindo $f(x)$ por ℓ_i em (2.2.1.2), vem:

^{1/} Ver, por exemplo, SILVA LEME (1965) pp.100-1.

^{2/} Ver, por exemplo, SILVA LEME (1965), pp. 100-1 ou SNEDECOR e COCHRAN (1967) pp.324-5.

$$V(\ell_i) = \frac{1}{n_i p_i (1-p_i)} \quad (2.2.1.5)$$

Essa é a variância assintótica de ℓ_i .

Como P_i são parâmetros desconhecidos, BERKSON (1944) propôs um estimador para as variâncias de ℓ_i , que se obtém substituindo os P_i , de (2.2.1.5), pelas frequências relativas p_i .

Para obter as estimativas dos parâmetros α e β da função logística, pelo método proposto por Berkson, usa-se o método dos mínimos quadrados ponderados não-lineares. Então, os quadrados dos desvios dos lógites observados, em relação aos lógites estimados pela regressão, devem ser ponderados por $n_i p_i (1-p_i)$. A soma de quadrados de desvios ponderados tem, assintoticamente, distribuição de χ^2 , razão porque BERKSON(1955) a denominou de " χ^2 de lógite" e usou a indicação que segue:

$$\chi^2_{(\text{lógite})} = \sum_i n_i p_i (1-p_i) (\ell_i - \hat{L}_i)^2 \quad (2.2.1.6)$$

onde $\hat{L}_i$ é a estimativa de L_i .

Minimizando (2.2.1.6), obtêm-se o sistema de equações normais:

$$\left. \begin{aligned} \hat{\alpha} \sum_i n_i p_i (1-p_i) + \hat{\beta} \sum_i n_i p_i (1-p_i) x_i &= \sum_i n_i p_i (1-p_i) \ell_i \\ \hat{\alpha} \sum_i n_i p_i (1-p_i) x_i + \hat{\beta} \sum_i n_i p_i (1-p_i) x_i^2 &= \sum_i n_i p_i (1-p_i) x_i \ell_i \end{aligned} \right\} \quad (2.2.1.7)$$

Fazendo $w_i = n_i p_i (1-p_i)$, de (2.2.1.7) se obtém:

$$\hat{\alpha} = \frac{(\sum_i x_i^2 w_i)(\sum_i \ell_i w_i) - (\sum_i x_i w_i)(\sum_i \ell_i x_i w_i)}{(\sum_i w_i) \left[(\sum_i x_i^2 w_i) - \frac{(\sum_i x_i w_i)^2}{\sum_i w_i} \right]} \quad (2.2.1.8)$$

$$\hat{\beta} = \frac{(\sum_i \ell_i x_i w_i) - \frac{(\sum_i \ell_i w_i)(\sum_i x_i w_i)}{\sum_i w_i}}{(\sum_i x_i^2 w_i) - \frac{(\sum_i x_i w_i)^2}{\sum_i w_i}} \quad (2.2.1.9)$$

É fácil verificar que (2.2.1.8) se reduz a $\hat{\alpha} = \bar{\ell} - \hat{\beta} \bar{x}$, onde $\bar{\ell}$ e $\bar{x}$ são as médias ponderadas dos ℓ_i e dos x_i , respectivamente.

Baseado na facilidade de resolução do sistema de equações (2.2.1.7), BERKSON (1944, 1951, 1953, 1955) recomendou enfaticamente esse método.

Entretanto, se $p_i=0$ ou $p_i=1$, não se definem os lógites observados $\ell_i = \log(p_i/(1-p_i))$. BERKSON (1955) propôs então o uso do que chamou de "regra do 2n"; isto é, propôs que se substituissem os valores $p_i=0$ por $p_i=1/2n_i$ e $p_i=1$ por $p_i=1-1/2n_i$.

SILVERSTONE (1957) verificou que as estimativas obtidas pelo método de Berkson são consistentes, no sentido de convergência em probabilidade e REIERSOL (1961) afirmou que a modificação introduzida nos dados pela "regra do 2n" não prejudica as propriedades assintóticas dos estimadores. Entretanto, na opinião de SILVERSTONE (1957), usando a

"regra do 2n", os estimadores obtidos não têm a propriedade de suficiência, no sentido de que nem toda a informação, necessária para a estimação dos parâmetros, está contida na amostra.

2.2.2 - O Método dos Mínimos Quadrados Ponderados Não-lineares

Dada a função logística, definida em (2.1.3), o método dos mínimos quadrados ponderados não-lineares, também apresentada por BERKSON (1944), consiste em obter, para estimativas^{3/} dos parâmetros α e β , os valores que minimizam a soma de quadrados dos desvios de cada p_i observado em relação à sua estimativa $\hat{P}_i$, ponderados pelo inverso da variância de p_i .

Dado que P_i são parâmetros desconhecidos, BERKSON (1944) usou, nos fatores de ponderação, as estimativas $\hat{P}_i$ de P_i . Então a soma de quadrados dos desvios ponderados ficou como segue:

$$S.Q.Desvios = \sum_i (p_i - \hat{P}_i)^2 \cdot \frac{n_i}{\hat{P}_i(1 - \hat{P}_i)} \quad (2.2.2.1)$$

onde

$$\hat{P}_i = \frac{1}{1 + e^{-(\hat{\alpha} + \hat{\beta}x_i)}} \quad (2.2.2.2)$$

Segue-se, de (2.2.2.2), que

3/ Neste trabalho, as estimativas dos parâmetros α e β serão sempre indicadas por $\hat{\alpha}$ e $\hat{\beta}$. Entretanto, conforme mostra o texto, os estimadores dependem do método utilizado, o que significa que cada método produz estimativas diferentes.

$$\frac{\partial \hat{P}_i}{\partial \hat{\alpha}} = \hat{P}_i(1-\hat{P}_i) \quad (2.2.2.3)$$

e

$$\frac{\partial \hat{P}_i}{\partial \hat{\beta}} = x_i \hat{P}_i(1-\hat{P}_i) \quad (2.2.2.4)$$

Para aplicar o método dos mínimos quadrados ponderados não-lineares, devem ser obtidas as derivadas parciais de (2.2.2.1) em relação a $\hat{\alpha}$ e $\hat{\beta}$. Igualando essas derivadas parciais a zero, obtêm-se o sistema de equações normais:

$$\left. \begin{aligned} \sum_i \frac{n_i(p_i - \hat{P}_i)(p_i - 2p_i \hat{P}_i + \hat{P}_i^2)}{\hat{P}_i(1-\hat{P}_i)} &= 0 \\ \sum_i \frac{n_i x_i (p_i - \hat{P}_i)(p_i - 2p_i \hat{P}_i + \hat{P}_i^2)}{\hat{P}_i(1-\hat{P}_i)} &= 0 \end{aligned} \right\} \quad (2.2.2.5)$$

O sistema não-linear de equações (2.2.2.5) não tem solução explícita. Esse sistema pode ser resolvido pelo método de Newton-Raphson^{4/}, que consiste em calcular correções sucessivas a partir de uma solução preliminar, até que se obtenha uma correção que possa ser considerada desprezível.

Então, para uma dada função f , o método aproxima $f(\underline{x})$, onde $\underline{x}$ é um vetor n -dimensional, por:

^{4/} Ver, por exemplo, HOFFMANN e VIEIRA (1977), pp. 286-9.

$$f(\underline{x}) = f(\underline{x}^0) + \left[\frac{\partial f}{\partial x_1} \right]_{\underline{x}=\underline{x}^0} \Delta x_1 + \left[\frac{\partial f}{\partial x_2} \right]_{\underline{x}=\underline{x}^0} \Delta x_2 + \dots + \left[\frac{\partial f}{\partial x_n} \right]_{\underline{x}=\underline{x}^0} \Delta x_n \quad (2.2.2.6)$$

onde $\underline{x}^0$ é um vetor de dimensão n , tomado como solução preliminar;

$x_1, x_2, \dots, x_n$ são, respectivamente, a primeira, a segunda, a n -ésima coordenadas do vetor $\underline{x}$; $\left[\frac{\partial f}{\partial x_i} \right]_{\underline{x}=\underline{x}^0}$, onde $i=1,2,\dots,n$, são os valores das de-

rivadas de f em relação a $x_1, x_2, \dots, x_n$ para $\underline{x}=\underline{x}^0$ e $\Delta x_i = x_i - x_i^0$, onde x_i^0 , $i=1,2,\dots,n$, é a i -ésima coordenada do vetor $\underline{x}^0$.

Então, lembrando as derivadas parciais dadas em (2.2.2.3) e (2.2.2.4) e fazendo

$$F(\hat{\alpha}, \hat{\beta}) = \sum_i \frac{n_i(p_i - \hat{p}_i)(p_i - 2\hat{p}_i \hat{p}_i + \hat{p}_i)}{\hat{p}_i(1 - \hat{p}_i)}$$

e

$$G(\hat{\alpha}, \hat{\beta}) = \sum_i \frac{n_i x_i (p_i - \hat{p}_i)(p_i - 2\hat{p}_i \hat{p}_i + \hat{p}_i)}{\hat{p}_i(1 - \hat{p}_i)}$$

o sistema de equações (2.2.2.5) pode ser resolvido pelo método de Newton-Raphson, que consiste em obter, por processo iterativo, a solução do seguinte sistema de equações lineares:

$$\left. \begin{aligned} \left[\frac{\partial F}{\partial \hat{\alpha}} \right]_{\substack{\hat{\alpha}=\alpha^0 \\ \hat{\beta}=\beta^0}} \Delta \hat{\alpha} + \left[\frac{\partial F}{\partial \hat{\beta}} \right]_{\substack{\hat{\alpha}=\alpha^0 \\ \hat{\beta}=\beta^0}} \Delta \hat{\beta} &= - F(\alpha^0, \beta^0) \\ \left[\frac{\partial G}{\partial \hat{\alpha}} \right]_{\substack{\hat{\alpha}=\alpha^0 \\ \hat{\beta}=\beta^0}} \Delta \hat{\alpha} + \left[\frac{\partial G}{\partial \hat{\beta}} \right]_{\substack{\hat{\alpha}=\alpha^0 \\ \hat{\beta}=\beta^0}} \Delta \hat{\beta} &= - G(\alpha^0, \beta^0) \end{aligned} \right\} \quad (2.2.2.7)$$

onde

$$\frac{\partial F}{\partial \hat{\alpha}} = \sum_i n_i r_i,$$

$$\frac{\partial F}{\partial \hat{\beta}} = \sum_i n_i x_i r_i,$$

$$\frac{\partial G}{\partial \hat{\alpha}} = \sum_i n_i x_i r_i,$$

$$\frac{\partial G}{\partial \hat{\beta}} = \sum_i n_i x_i^2 r_i,$$

$$\Delta \hat{\alpha} = \hat{\alpha} - \alpha^0,$$

$$\Delta \hat{\beta} = \hat{\beta} - \beta^0,$$

e

$$r_i = \frac{2p_i \hat{p}_i (\hat{p}_i - p_i \hat{p}_i + p_i) - \hat{p}_i^2 - p_i^2}{\hat{p}_i (1 - \hat{p}_i)}.$$

Substituindo os valores das derivadas parciais, o sistema de equações (2.2.2.7), fica como segue:

$$\sum_i n_i r_i^0 \Delta \hat{\alpha} + \sum_i n_i x_i r_i^0 \Delta \hat{\beta} = - F(\alpha^0, \beta^0) \quad (2.2.2.8)$$

$$\sum_i n_i x_i r_i^0 \Delta \hat{\alpha} + \sum_i n_i x_i^2 r_i^0 \Delta \hat{\beta} = - G(\alpha^0, \beta^0)$$

onde

$$r_i^0 = \left[r_i \right]_{\substack{\hat{\alpha} = \alpha^0 \\ \hat{\beta} = \beta^0}}$$

2.2.3 - O método da máxima verossimilhança

Este método consiste em obter, para estimativas dos parâmetros α e β da função logística definida em (2.1.3), os valores $\hat{\alpha}$ e $\hat{\beta}$ que maximizam a função de verossimilhança.

Para obter a função de verossimilhança $\mathcal{L}$, basta lembrar que, a variável não-aleatória x_i estão relacionados n_i ensaios binomiais independentes, nos quais se observam $m_i = n_i p_i$ sucessos. Como $M_i \sim B(n_i, P_i)$, então

$$\mathcal{L} = \prod_i \binom{n_i}{m_i} P_i^{m_i} (1-P_i)^{n_i-m_i} \quad (2.2.3.1)$$

Como $\mathcal{L}$ é uma função monotonicamente crescente, o ponto de máximo da função coincide com o ponto de máximo do logaritmo neperiano da função. Então, aplicando logaritmo neperiano, vem:

$$\Lambda = \log \mathcal{L} = \sum_i \left[\log \binom{n_i}{m_i} + m_i \log P_i + (n_i - m_i) \log (1-P_i) \right] \quad (2.2.3.2)$$

Fazendo as derivadas parciais de Λ em relação a $\hat{\alpha}$ e a $\hat{\beta}$ obtem-se, lembrando as derivadas parciais dadas em (2.2.2.3) e (2.2.2.4), o sistema de equações:

$$\left. \begin{aligned} \sum_i n_i (p_i - \hat{P}_i) &= 0 \\ \sum_i n_i x_i (p_i - \hat{P}_i) &= 0 \end{aligned} \right\} \quad (2.2.3.3)$$

O sistema de equações (2.2.3.3) não tem solução explícita, mas pode ser resolvido pelo método de Newton-Raphson.

Então, lembrando as derivadas parciais dadas em (2.2.2.3) e (2.2.2.4), fazendo

$$H_1(\hat{\alpha}, \hat{\beta}) = \sum_i n_i (p_i - \hat{p}_i)$$

e

$$H_2(\hat{\alpha}, \hat{\beta}) = \sum_i n_i x_i (p_i - \hat{p}_i)$$

o método de Newton-Raphson, aplicado ao sistema de equações (2.2.3.3), consiste em resolver, por processo iterativo um sistema de equações lineares similar ao (2.2.2.8), onde:

$$\frac{\partial H_1}{\partial \hat{\alpha}} = - \sum_i n_i \hat{p}_i (1 - \hat{p}_i)$$

$$\frac{\partial H_1}{\partial \hat{\beta}} = - \sum_i n_i x_i \hat{p}_i (1 - \hat{p}_i)$$

$$\frac{\partial H_2}{\partial \hat{\alpha}} = - \sum_i n_i x_i \hat{p}_i (1 - \hat{p}_i)$$

$$\frac{\partial H_2}{\partial \hat{\beta}} = - \sum_i n_i x_i^2 \hat{p}_i (1 - \hat{p}_i)$$

Portanto, deve ser resolvido o seguinte sistema de equações lineares:

$$\left. \begin{aligned} \sum n_i t_i^0 \Delta \hat{\alpha} + \sum n_i x_i t_i^0 \Delta \hat{\beta} &= H_1(\alpha^0, \beta^0) \\ \sum n_i x_i t_i^0 \Delta \hat{\alpha} + \sum n_i x_i^2 t_i^0 \Delta \hat{\beta} &= H_2(\alpha^0, \beta^0) \end{aligned} \right\} \quad (2.2.3.4)$$

onde

$$t_i^0 = \left[\hat{p}_i (1 - \hat{p}_i) \right]_{\substack{\hat{\alpha} = \alpha^0 \\ \hat{\beta} = \beta^0}}$$

e

α^0 e β^0 são soluções preliminares, obtidas pelo método de Berkson ou pelo método de mínimos quadrados ponderados não-lineares.

Lembrando que as dificuldades para a solução do sistema de equações (2.2.3.3) praticamente desapareceram para os usuários de computador, SILVERSTONE (1957) recomendou que as estimativas de α e β fossem obtidas, sempre que possível, através do método da máxima verossimilhança que, aliás, é o único indicado por JOHNSON e KOTZ (1970). É interessante lembrar aqui que o método de máxima verossimilhança produz estimativas consistentes e assintoticamente eficientes.

2.2.4 - Considerações sobre as estimativas preliminares utilizadas no método dos mínimos quadrados ponderados não-lineares e no método da máxima verossimilhança

Tanto o método dos mínimos quadrados ponderados não-lineares como o método da máxima verossimilhança conduzem a sistemas de equações que só podem ser resolvidos por processos iterativos.

O número de iterações que se fazem necessárias para chegar a correções que possam ser consideradas desprezíveis depende, em grande parte, dos valores das estimativas preliminares disponíveis. É importante frisar que o método de Berkson conduz a estimativas preliminares bastante boas, o que assegura convergência bastante rápida do processo, toda vez que essa convergência é possível, tanto quando se usa o método dos mínimos quadrados ponderados não-lineares como quando se usa o método da máxima verossimilhança.

2.3 - Variâncias e Covariâncias das Estimativas dos Parâmetros

2.3.1 - Usando as estimativas dos parâmetros obtidas pelo método de Berkson

O método de Berkson consiste em ajustar uma regressão linear ponderada do logite contra x_i . Então, as variâncias e covariâncias das estimativas $\hat{\alpha}$ e $\hat{\beta}$ dos parâmetros α e β de (2.2.1), dadas em (2.2.1.7) e (2.2.1.8), são obtidas pelo método usual em modelos lineares^{5/}. Então

$$\hat{V}(\hat{\alpha}) = \frac{\frac{\sum_i x_i^2 w_i}{\sum_i w_i}}{(\sum_i x_i^2 w_i) - \frac{(\sum_i x_i w_i)^2}{\sum_i w_i}} s^2 \quad (2.3.1.1)$$

^{5/} Ver, por exemplo, DRAPER e SMITH (1966) pp. 44-55.

$$\hat{V}(\hat{\beta}) = \frac{1}{\left(\sum_i x_i^2 w_i \right) - \frac{\left(\sum_i x_i w_i \right)^2}{\sum_i w_i}} s^2 \quad (2.3.1.2)$$

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = \frac{-\frac{\sum_i x_i w_i}{\sum_i w_i}}{\left(\sum_i x_i^2 w_i \right) - \frac{\left(\sum_i x_i w_i \right)^2}{\sum_i w_i}} s^2 \quad (2.3.1.3)$$

2.3.2 - Usando as estimativas dos parâmetros obtidas pelo método da máxima verossimilhança.

As variâncias e covariâncias das estimativas dos parâmetros, obtidas pelo método da máxima verossimilhança, são dadas assintoticamente pelo limite inferior de Cramer-Rao, ou seja, são dadas pelo inverso da matriz de informação^{6/}.

Então, dada a função de verossimilhança, definida em (2.2.3.1), a matriz de informação W é:

$$W = \begin{bmatrix} \sum_i n_i P_i (1-P_i) & \sum_i n_i x_i P_i (1-P_i) \\ \sum_i n_i x_i P_i (1-P_i) & \sum_i n_i x_i^2 P_i (1-P_i) \end{bmatrix} \quad (2.3.2.1)$$

^{6/} Ver, por exemplo, THEIL (1971) pp. 387-9.

A inversa dessa matriz, W^{-1} , é a matriz de variâncias e co variâncias assintóticas das estimativas de máxima verossimilhança dos parâmetros α e β . Então:

$$W^{-1} = \frac{1}{|W|} \cdot \begin{bmatrix} \sum_i n_i x_i^2 P_i (1-P_i) & - \sum_i n_i x_i P_i (1-P_i) \\ - \sum_i n_i x_i P_i (1-P_i) & \sum_i P_i (1-P_i) \end{bmatrix} \quad (2.3.2.2)$$

onde

$$|W| = \sum_i n_i P_i (1-P_i) \sum_i n_i x_i^2 P_i (1-P_i) - \left[\sum_i n_i x_i P_i (1-P_i) \right]^2 \quad (2.3.2.3)$$

Como os P_i são parâmetros desconhecidos, pode-se obter uma estimativa da matriz W^{-1} , usando $\hat{P}_i$ como estimativa de P_i . Então

$$\hat{W}^{-1} = \frac{1}{|\hat{W}|} \cdot \begin{bmatrix} \sum_i n_i x_i^2 \hat{P}_i (1-\hat{P}_i) & - \sum_i n_i x_i \hat{P}_i (1-\hat{P}_i) \\ - \sum_i n_i x_i \hat{P}_i (1-\hat{P}_i) & \sum_i \hat{P}_i (1-\hat{P}_i) \end{bmatrix} \quad (2.3.2.4)$$

Para verificar se foi obtido um bom ajustamento, calcula-se a soma dos quadrados dos desvios das proporções, ponderados pelo inverso das estimativas das variâncias de $\hat{P}_i$, isto é, calcula-se:

$$SQD = \sum_{i=1}^s (p_i - \hat{P}_i)^2 \frac{n_i}{\hat{P}_i (1-\hat{P}_i)} \quad (2.3.2.5)$$

Nos casos em que o modelo logístico adotado é verdadeiro, a soma de quadrados de desvios ponderados tem distribuição de χ^2 com $(s-2)$ graus de liberdade.

Lembrando que

$$E \left\{ \chi^2_{(s-2)} \right\} = s-2,$$

nos casos em que se obtêm uma soma de quadrados de desvios ponderados muito elevada, isto é, muito superior ao valor dos respectivos graus de liberdade, as estimativas das variâncias e covariâncias assintóticas das estimativas dos parâmetros devem ser corrigidas, isto é, devem ser multiplicadas por um "fator de heterogeneidade"^{7/} h , onde

$$h = \frac{SQD}{s-2} \quad (2.3.2.6)$$

Então, nesses casos, a matriz $\hat{M}$ das estimativas das variâncias e covariâncias assintóticas das estimativas dos parâmetros é dada pelo produto da matriz definida em (2.3.2.4) por h , isto é

$$\hat{M} = \hat{W}^{-1} h \quad (2.3.2.7)$$

2.4 - Variâncias das Estimativas dos Lógitos e Variâncias das Estimativas das Proporções

É possível obter as variâncias das estimativas dos lógitos, isto é, de $\hat{L}_i$; $i=1,2,\dots,s$ e as variâncias das estimativas das proporções, isto é, de $\hat{P}_i$; $i=1,2,\dots,s$, tanto quando se estimam os parâmetros pelo método de Berkson como quando se estimam os parâmetros pelo método da máxima verossimilhança.

^{7/} Ver FINNEY (1971) pp. 70-2.

2.4.1 - Variâncias das estimativas dos lógitos

Como $\hat{L}_i = \hat{\alpha} + \hat{\beta}x_i$, segue-se que:

$$V(\hat{L}_i) = V(\hat{\alpha}) + x_i^2 V(\hat{\beta}) + 2x_i \text{cov}(\hat{\alpha}, \hat{\beta})$$

Lembrando as estimativas das variâncias e da covariância de $\hat{\alpha}$ e $\hat{\beta}$, dadas em (2.3.1.1), (2.3.1.2) e (2.3.1.3), vem:

$$\hat{V}(\hat{L}_i) = \frac{\frac{\sum_i x_i^2 w_i}{\sum_i w_i} + x_i^2 - 2x_i \frac{\sum_i x_i w_i}{\sum_i w_i}}{\left(\sum_i x_i^2 w_i \right) - \frac{\left(\sum_i x_i w_i \right)^2}{\sum_i w_i}} s^2 \quad (2.4.1.1)$$

É fácil ver que (2.4.1.1) pode ser reescrita como segue:

$$\hat{V}(\hat{L}_i) = \left[\frac{1}{\sum_j w_j} + \frac{(x_i - \bar{x})^2}{\sum_j (x_j - \bar{x})^2 w_j} \right] s^2 \quad (2.4.1.2)$$

onde

$$\bar{x} = \frac{\sum_i x_i w_i}{\sum_i w_i}$$

2.4.2 - Variâncias das estimativas das proporções

As variâncias de $\hat{P}_i$, $i=1,2,\dots,s$, obtidos pelo método de Berkson e pelo método da máxima verossimilhança, distinguem-se entre si somente no que se refere às variâncias e à covariância de $\hat{\alpha}$ e $\hat{\beta}$.

Para obter as variâncias de $\hat{P}_i$; $i=1,2,\dots,s$, utiliza-se uma aproximação por série de Taylor, como segue:

$$\hat{P}_i = \hat{P}_i(\alpha^0, \beta^0) + \left[\frac{\partial \hat{P}_i}{\partial \alpha} \right]_{\substack{\hat{\alpha}=\alpha_0^0 \\ \hat{\beta}=\beta_0^0}} (\hat{\alpha}-\alpha^0) + \left[\frac{\partial \hat{P}_i}{\partial \beta} \right]_{\substack{\hat{\alpha}=\alpha_0^0 \\ \hat{\beta}=\beta_0^0}} (\hat{\beta}-\beta^0)$$

Fazendo $\alpha^0 = \alpha$ e $\beta^0 = \beta$, vem:

$$\hat{P}_i - P_i = P_i(1-P_i)(\hat{\alpha}-\alpha) + x_i P_i(1-P_i)(\hat{\beta}-\beta)$$

Elevando ambos os membros ao quadrado e aplicando esperança, obtém-se:

$$V(\hat{P}_i) = P_i^2(1-P_i)^2 V(\hat{\alpha}) + x_i^2 P_i^2(1-P_i)^2 V(\hat{\beta}) + 2x_i P_i^2(1-P_i)^2 \text{Cov}(\hat{\alpha}, \hat{\beta})$$

Como os P_i ; $i=1,2,\dots,s$, são parâmetros desconhecidos, substitui-se P_i por $\hat{P}_i$ e se obtém a estimativa da variância de $\hat{P}_i$, como segue:

$$\widehat{V}(\hat{P}_i) = \hat{P}_i^2(1-\hat{P}_i)^2 \left[\widehat{V}(\hat{\alpha}) + x_i^2 \widehat{V}(\hat{\beta}) + 2x_i \widehat{\text{Cov}}(\hat{\alpha}, \hat{\beta}) \right] \quad (2.4.1.3)$$

2.4.2.1 - Variâncias das estimativas das proporções quando os parâmetros são estimados pelo método de Berkson

Lembrando as estimativas das variâncias e da covariância de $\hat{\alpha}$ e $\hat{\beta}$, dadas em (2.3.1.1), (2.3.1.2) e (2.3.1.3) vem, de (2.4.1.3):

$$\widehat{V}(\widehat{P}_i) = \widehat{P}_i^2(1-\widehat{P}_i)^2 \left[\frac{\frac{\sum x_i^2 w_i}{i} + x_i^2 - 2x_i \frac{\sum x_i w_i}{i}}{\frac{(\sum x_i w_i)^2}{i} - \frac{\sum w_i}{i}} \right] s^2$$

Reescrevendo, vem:

$$\widehat{V}(\widehat{P}_i) = \widehat{P}_i^2(1-\widehat{P}_i)^2 \left[\frac{1}{\sum_j w_j} + \frac{(x_i - \bar{x})^2}{\sum_j (x_j - \bar{x})^2 w_j} \right] s^2 \quad (2.4.2.1)$$

2.4.2.2 - Variâncias das estimativas das proporções quando os parâmetros são estimados pelo método da máxima verossimilhança

Lembrando as estimativas das variâncias e da covariância assintóticas de $\hat{\alpha}$ e $\hat{\beta}$, dadas em (2.3.2.4), tem-se:

$$\widehat{V}(\widehat{P}_i) = v_i \left(\sum_i n_i x_i^2 \widehat{P}_i(1-\widehat{P}_i) + x_i^2 \sum_i n_i \widehat{P}_i(1-\widehat{P}_i) - 2x_i \sum_i n_i x_i \widehat{P}_i(1-\widehat{P}_i) \right) \quad (2.4.2.1)$$

onde

$$v_i = \frac{\widehat{P}_i^2(1-\widehat{P}_i)^2}{|\widehat{W}|}$$

Nos casos em que é necessário usar o "fator de heterogeneidade", as estimativas das variâncias e da covariância assintóticas de $\hat{\alpha}$ e $\hat{\beta}$ são dadas em (2.3.2.7). Nesses casos, as variâncias de $\widehat{P}_i$ são dadas por (2.4.2.1), mas onde

$$v_i = \frac{\widehat{P}_i^2(1-\widehat{P}_i)^2 h}{|\widehat{W}|}$$

3 - A FUNÇÃO LOGÍSTICA COM DUAS OU MAIS VARIÁVEIS INDEPENDENTES PARA DADOS DE RESPOSTA QUÂNTICA

3.1 - Introdução

A função logística para dados de resposta quântica, definida em (2.1.3) pode ser generalizada como segue:

$$P_i = \left[1 + e^{-x_i \beta} \right]^{-1} \quad (3.1.1)$$

onde P_i representa a probabilidade de sucesso no i -ésimo grupo, x_i é o vetor-linha de dimensão $k+1$ de variáveis independentes para a i -ésima observação, com primeiro elemento igual a 1 e β é o vetor-coluna de dimensão $k+1$ dos parâmetros desconhecidos.

Como exemplo de aplicação da função logística com duas ou mais variáveis independentes para dados de resposta quântica, pode-se lembrar que, em farmacologia, pode haver interesse em estudar a proporção de camundongos que apresentam convulsão em função de doses crescentes de duas ou mais drogas potencialmente convulsionantes.

Podem ainda ser lembrados exemplos em psicologia, tais como: o estudo da proporção de respostas, obtidas de grupos de sujeitos submetidos a um certo estímulo, em função do tempo de treino desses sujeitos e do número de reforçadores; o estudo da proporção de respostas de grupos de sujeitos, em função do tempo decorrido entre a resposta e o reforçador e em função da idade dos sujeitos.

Todos estes exemplos se referem a estudos onde às variáveis independentes são fixas. Entretanto, a função logística para dados de resposta quântica também pode ser ajustada nos casos em que uma ou mais variáveis independentes podem ser vistas como variáveis aleatórias. Uma revisão bibliográfica sobre o assunto pode ser vista em BISHOP, FIENBERG e HOLLAND (1971). Como exemplo, pode-se lembrar que, em sociologia rural, pode haver interesse em estudar a proporção de proprietários rurais que adotam certa técnica agrícola (como por exemplo, o uso de curvas de nível) em função do grau de escolaridade do proprietário e da área da propriedade. As variáveis independentes podem ser vistas, neste caso, como variáveis aleatórias.

3.2 - Estimativas dos parâmetros

O ajuste da função logística com duas ou mais variáveis independentes aos dados de resposta quântica pode ser feito ou através de uma extensão do método sugerido por BERKSON (1944), ou através do método dos mínimos quadrados ponderados não-lineares, ou através do método de máxima verossimilhança.

3.2.1 - Uma extensão do método de Berkson

Usando a sugestão feita por BERKSON (1944), facilmente se obtém:

$$L_i = \log \frac{p_i}{1-p_i} = \underline{x}_i \underline{\beta} \quad (3.2.1.1)$$

onde $\underline{x}_i$ e $\underline{\beta}$ são os mesmos vetores de (3.1.1)

Os fatores de ponderação dos quadrados dos desvios são $n_i p_i (1-p_i)$.

É necessário definir

$$\underline{\ell} = \begin{bmatrix} \ell_1 \\ \ell_2 \\ \vdots \\ \ell_s \end{bmatrix} \quad (3.2.1.2)$$

onde

$\ell_i = \log(p_i/(1-p_i))$ e lembrar que, conforme consta na seção (2.2.1),

$$\frac{\partial \ell_i}{\partial p_i} = \frac{1}{p_i(1-p_i)}$$

e

$$V(p_i) = \frac{p_i(1-p_i)}{n_i}$$

Então, adotando procedimento similar àquele adotado para a obtenção de (2.2.1.5), vem:

$$V(\underline{x}) = \begin{bmatrix} \frac{1}{n_1 P_1 (1-P_1)} & 0 & \dots & 0 \\ 0 & \frac{1}{n_2 P_2 (1-P_2)} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{n_s P_s (1-P_s)} \end{bmatrix}$$

Substituindo os parâmetros desconhecidos P_i pelas frequências relativas p_i , conforme a sugestão feita por BERKSON (1944), vem:

$$\widehat{V}(\underline{x}) = \begin{bmatrix} \frac{1}{n_1 p_1 (1-p_1)} & 0 & \dots & 0 \\ 0 & \frac{1}{n_2 p_2 (1-p_2)} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{n_s p_s (1-p_s)} \end{bmatrix} \quad (3.2.1.3)$$

Pode-se então fazer o ajuste da regressão linear ponderada da variável transformada em função de $\underline{x}_i$.

Minimizando a função:

$$(\underline{\ell} - \underline{\hat{\ell}})' [\widehat{V}(\underline{\ell})]^{-1} (\underline{\ell} - \underline{\hat{\ell}}) \quad (3.2.1.4)$$

onde $\hat{\ell}_i = \underline{x}_i' \underline{\hat{\beta}}$ e

$$\underline{\hat{\ell}} = \begin{bmatrix} \hat{\ell}_1 \\ \hat{\ell}_2 \\ \vdots \\ \hat{\ell}_s \end{bmatrix}$$

obtem-se o sistema de equações normais

$$\underline{X}' [\widehat{V}(\underline{\ell})]^{-1} \underline{X} \underline{\hat{\beta}} = \underline{X}' [\widehat{V}(\underline{\ell})]^{-1} \underline{\ell} \quad (3.2.1.5)$$

onde $\underline{X}$ é a matriz cujas linhas são os vetores $\underline{x}_i$ de (3.1.1).

Então, as estimativas de $\underline{\beta}$, são dadas por^{8/}:

$$\underline{\hat{\beta}} = (\underline{X}' [\widehat{V}(\underline{\ell})]^{-1} \underline{X})^{-1} \underline{X}' [\widehat{V}(\underline{\ell})]^{-1} \underline{\ell} \quad (3.2.1.6)$$

3.2.2 - Método dos mínimos quadrados ponderados não-lineares

O método dos mínimos quadrados ponderados não-lineares consiste em obter, como estimativa do vetor $\underline{\beta}$, o vetor $\underline{\hat{\beta}}$ que minimiza a soma dos quadrados dos desvios $(p_i - \hat{P}_i)$, ponderados pelos inversos das variâncias de p_i , dadas por $P_i(1-P_i)/n_i$.

^{8/} Ver, por exemplo, HOFFMANN e VIEIRA (1977), pp. 204-6 ou SEARLE (1971) pp. 86-7.

Entretanto, dessa maneira se obtêm um sistema de equações que não tem solução, porque os valores P_i , usados nos fatores de ponderação, são desconhecidos.

Para que o sistema tenha solução, utiliza-se nos fatores de ponderação, as estimativas $\hat{P}_i$ de P_i , isto é, minimiza-se a função:

$$\sum_i (p_i - \hat{P}_i)^2 \cdot \frac{n_i}{\hat{P}_i(1 - \hat{P}_i)} \quad (3.2.2.1)$$

onde

$$\hat{P}_i = \frac{1}{1 + e^{-x_i \hat{\beta}}} \quad (3.2.2.2)$$

Como

$$\frac{\partial \hat{P}_i}{\partial \hat{\beta}'} = \hat{P}_i(1 - \hat{P}_i) x_i, \quad (3.2.2.3)$$

fazendo a derivada de (3.2.2.1) em relação ao vetor $\hat{\beta}'$, obtêm-se o seguinte sistema de equações:

$$\sum_i \left\{ \frac{(p_i - \hat{P}_i)n_i(p_i - 2\hat{P}_i)}{\hat{P}_i(1 - \hat{P}_i)} \right\} x_i = \underline{0} \quad (3.2.2.4)$$

onde $\underline{0}$ é um vetor linha de dimensão $k+1$, cujos elementos são todos iguais a zero.

A solução para o sistema de equações (3.2.2.4) é encontrada através de um processo iterativo, pelo método de Newton-Raphson de

maneira análoga a utilizada na seção (2.2.2).

Seja

$$F_j = \sum_i \left\{ \frac{(p_i - \hat{p}_i) n_i (p_i - 2\hat{p}_i \hat{p}_i + \hat{p}_i)}{\hat{p}_i (1 - \hat{p}_i)} \right\} x_{ij} \quad (3.2.2.5)$$

aproximando-se F_j , pela série de Taylor, obtém-se

$$F_j(\underline{\beta}) = F_j(\underline{\beta}^0) + \sum_{h=0}^k \left[\frac{\partial F_j}{\partial \hat{\beta}_h} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_h; \quad h=0,1,\dots,k \quad (3.2.2.6)$$

que é um sistema linear de $(k+1)$ equações, análogo ao sistema (2.2.2.8) e que pode ser reescrito como segue

$$\left. \begin{aligned}
 & \left[\frac{\partial F_0}{\partial \hat{\beta}_0} \right]_{\hat{\beta}=\beta^0} \Delta\beta_0 + \left[\frac{\partial F_0}{\partial \hat{\beta}_1} \right]_{\hat{\beta}=\beta^0} \Delta\beta_1 + \dots + \left[\frac{\partial F_0}{\partial \hat{\beta}_k} \right]_{\hat{\beta}=\beta^0} \Delta\beta_k = -F_0(\beta^0) \\
 & \left[\frac{\partial F_1}{\partial \hat{\beta}_0} \right]_{\hat{\beta}=\beta^0} \Delta\beta_0 + \left[\frac{\partial F_1}{\partial \hat{\beta}_1} \right]_{\hat{\beta}=\beta^0} \Delta\beta_1 + \dots + \left[\frac{\partial F_1}{\partial \hat{\beta}_k} \right]_{\hat{\beta}=\beta^0} \Delta\beta_k = -F_1(\beta^0) \\
 & \vdots \\
 & \left[\frac{\partial F_k}{\partial \hat{\beta}_0} \right]_{\hat{\beta}=\beta^0} \Delta\beta_0 + \left[\frac{\partial F_k}{\partial \hat{\beta}_1} \right]_{\hat{\beta}=\beta^0} \Delta\beta_1 + \dots + \left[\frac{\partial F_k}{\partial \hat{\beta}_k} \right]_{\hat{\beta}=\beta^0} \Delta\beta_k = -F_k(\beta^0)
 \end{aligned} \right\} (3.2.2.7)$$

onde

$$\frac{\partial F_j}{\partial \beta_h} = \sum_i n_i x_{ih} x_{ij} \left\{ \frac{2p_i \hat{p}_i (\hat{p}_i - p_i \hat{p}_i + p_i) - \hat{p}_i^2 - p_i^2}{\hat{p}_i (1 - \hat{p}_i)} \right\}$$

$$\Delta\beta_h = \hat{\beta}_h - \beta_h^0 ; h=0,1,\dots,k$$

e β^0 é um vetor de estimativas preliminares dos parâmetros, usado para iniciar o processo iterativo.

Tomando como estimativa inicial o vetor $\hat{\beta}$ obtido pelo método de Berkson, inicia-se o processo iterativo, de maneira similar à já descrita na seção 2.2.2.

3.2.3 - O método da máxima verossimilhança

Definida a função de verossimilhança como na seção 2.2.3, isto é

$$\mathcal{L} = \prod_i \binom{n_i}{m_i} p_i^{m_i} (1-p_i)^{n_i-m_i}, \quad (3.2.3.1)$$

aplicando logaritmo neperiano, vem:

$$\Lambda = \log \mathcal{L} = \sum_i \left[\log \binom{n_i}{m_i} + m_i \log p_i + (n_i - m_i) \log(1-p_i) \right] \quad (3.2.3.2)$$

Fazendo a derivada de Λ em relação ao vetor $\underline{\beta}'$ e lembrando a derivada definida em (3.2.2.3), facilmente se obtém:

$$\frac{\partial \Lambda}{\partial \underline{\beta}'} = \sum_i m_i (1 - \hat{p}_i) \underline{x}_i - \sum_i (n_i - m_i) \hat{p}_i \underline{x}_i \quad (3.2.3.3)$$

Como $m_i = p_i n_i$, pode-se escrever:

$$\frac{\partial \Lambda}{\partial \underline{\beta}'} = \sum_i n_i p_i (1 - \hat{p}_i) \underline{x}_i - \sum_i n_i (1 - p_i) \hat{p}_i \underline{x}_i,$$

donde se obtém, rearranjando e igualando a zero, o seguinte sistema de equações:

$$\sum_i n_i (p_i - \hat{p}_i) \underline{x}_i = \underline{0} \quad (3.2.3.4)$$

onde $\underline{0}$ é o mesmo vetor definido em (3.2.2.4)

A solução para o sistema (3.2.3.4) é obtida através de um processo iterativo. Seja

$$G_j = \sum_i n_i (p_i - \hat{p}_i) x_{ij}; \quad j=0,1,\dots,k \quad (3.2.3.5)$$

aproximando-se G_j , pela série de Taylor, obtêm-se:

$$G_j(\underline{\beta}) = G_j(\underline{\beta}^0) + \sum_{h=0}^k \left[\frac{\partial G_j}{\partial \hat{\beta}_h} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_h; \quad j=0,1,\dots,k \quad (3.2.3.6)$$

que é um sistema linear de $(k+1)$ equações, análogo ao sistema (2.2.2.8) e pode ser reescrito como:

$$\left. \begin{aligned} \left[\frac{\partial G_0}{\partial \hat{\beta}_0} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_0 + \left[\frac{\partial G_0}{\partial \hat{\beta}_1} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_1 + \dots + \left[\frac{\partial G_0}{\partial \hat{\beta}_k} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_k &= -G_0(\underline{\beta}^0) \\ \left[\frac{\partial G_1}{\partial \hat{\beta}_0} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_0 + \left[\frac{\partial G_1}{\partial \hat{\beta}_1} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_1 + \dots + \left[\frac{\partial G_1}{\partial \hat{\beta}_k} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_k &= -G_1(\underline{\beta}^0) \\ \left[\frac{\partial G_k}{\partial \hat{\beta}_0} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_0 + \left[\frac{\partial G_k}{\partial \hat{\beta}_1} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_1 + \dots + \left[\frac{\partial G_k}{\partial \hat{\beta}_k} \right]_{\underline{\beta}=\underline{\beta}^0} \Delta \beta_k &= -G_k(\underline{\beta}^0) \end{aligned} \right\} \quad (3.2.3.7)$$

onde

$$\frac{\partial G_j}{\partial \hat{\beta}_h} = - \sum_i n_i x_{ih} x_{ij} \hat{p}_i (1 - \hat{p}_i)$$

$$\Delta \beta_h = \hat{\beta}_h - \beta_h^0; \quad h = 0,1,\dots,k$$

e $\underline{\beta}^0$ é uma estimativa preliminar de $\underline{\beta}$, usado para iniciar o processo iterativo.

Tomando como estimativa inicial o vetor $\underline{\hat{\beta}}$, obtido pelo método de Berkson ou pelo método dos mínimos quadrados ponderados não-lineares, inicia-se o processo iterativo, procedendo-se de maneira similar à já descrita na seção 2.2.2.

3.3 - Variâncias e covariâncias das estimativas dos parâmetros

3.3.1 - Usando as estimativas dos parâmetros obtidas pela extensão do método de Berkson

O método de Berkson consiste em ajustar uma regressão linear ponderada dos lógites contra x_i . Então, as variâncias e covariâncias das estimativas dos elementos de $\underline{\beta}$ definido em (3.1.1), dadas em (3.2.1.6), são obtidas pelo método usual em modelos lineares^{9/}, ou seja:

$$V(\underline{\hat{\beta}}) = E(\underline{\hat{\beta}} - E(\underline{\hat{\beta}}))(\underline{\hat{\beta}} - E(\underline{\hat{\beta}}))'$$

Lembrando as estimativas dos elementos de $\underline{\beta}$, dadas em (3.2.1.6) e lembrando as estimativas das variâncias de $\underline{\ell}$, isto é, $\hat{V}(\underline{\ell})$, dadas em (3.2.1.3), vem:

$$\hat{V}(\underline{\hat{\beta}}) = \left[X' (\hat{V}(\underline{\ell}))^{-1} X \right]^{-1} s^2 \quad (3.3.1.1)$$

^{9/} Ver, por exemplo, SEARLE (1971) pp. 90 ou HOFFMANN e VIEIRA (1977) pp. 206-7.

3.3.2 - Usando as estimativas dos parâmetros obtidas pelo método da máxima verossimilhança

As variâncias e covariâncias das estimativas dos parâmetros obtidas pelo método da máxima verossimilhança são dadas, assintoticamente, pelo limite inferior de Cramer-Rao, ou seja, são dadas pelo inverso da matriz de informação^{10/}.

Então dada a função de verossimilhança, definida em (3.2.3.1), a matriz de informação é:

$$W = \sum_i n_i P_i (1-P_i) x_i' x_i \quad (3.3.2.1)$$

É fácil verificar que os elementos da matriz de informação são dados por

$$w_{hj} = \sum_i n_i P_i (1-P_i) x_{ij} x_{ih} \quad (3.3.2.2)$$

A inversa dessa matriz, W^{-1} é a matriz de variâncias e covariâncias assintóticas das estimativas de máxima verossimilhança dos elementos de β .

Como os P_i são parâmetros desconhecidos, pode-se obter uma estimativa da matriz W^{-1} , usando $\hat{P}_i$ como estimativa de P_i . Então

$$\hat{W}^{-1} = \left[\sum_i n_i \hat{P}_i (1-\hat{P}_i) x_i' x_i \right]^{-1} \quad (3.3.2.3)$$

^{10/} Ver, por exemplo, THEIL (1971) pp.387-9.

Como já foi mencionado na seção 2.3.2, nos casos em que a soma de quadrados de desvios ponderados, definida em (2.3.2.5), $\bar{e}$ é muito elevada, ou seja, muito superior ao valor dos respectivos graus de liberdade, as estimativas das variâncias e covariâncias assintóticas das estimativas dos parâmetros devem ser corrigidas, isto é, multiplicadas pelo "fator de heterogeneidade" h , definido em (2.3.2.6).

Nesses casos, a matriz $\hat{M}$, das estimativas das variâncias e covariâncias assintóticas das estimativas de máxima verossimilhança dos elementos de $\underline{\beta}$, é dada pelo produto da matriz, definida em (3.3.2.3), pelo "fator de heterogeneidade" h , isto é:

$$\hat{M} = \hat{W}^{-1} h \quad (3.3.2.4)$$

3.4 - Variâncias das Estimativas dos Lógites e Variâncias das Estimativas das Proporções

Podem ser obtidas as variâncias das estimativas dos Lógites isto é, de $\hat{L}_i$; $i=1,2,\dots,s$ e as variâncias das estimativas das proporções, isto é, de $\hat{P}_i$; $i=1,2,\dots,s$, tanto quando se estimam os parâmetros pelo método de Berkson como quando se estimam os parâmetros pelo método da máxima verossimilhança.

3.4.1 - Variâncias das estimativas dos lógites

Como $\hat{L}_i = \underline{x}_i \hat{\underline{\beta}}$, facilmente se obtém:

$$V(\hat{L}_i) = \underline{x}_i V(\hat{\underline{\beta}}) \underline{x}_i'$$

Lembrando a estimativa da matriz de variâncias e covariâncias dada em (3.3.1.1), facilmente se obtém:

$$\widehat{V}(\underline{\hat{L}}_i) = \underline{x}_i \left[\underline{x}' (\widehat{V}(\underline{\hat{L}}))^{-1} \underline{x} \right]^{-1} \underline{x}_i' s^2 \quad (3.4.1.1)$$

3.4.2 - Variâncias das estimativas das proporções

As variâncias das proporções $\hat{P}_i$; $i=1,2,\dots,s$, obtidas pelo método de Berkson, distinguem-se daquelas obtidas pelo método da máxima verossimilhança apenas no que se refere às variâncias de $\underline{\hat{\beta}}$.

Para obter as variâncias de $\hat{P}_i$; $i=1,2,\dots,s$, usa-se uma aproximação através da série de Taylor, como segue:

$$\hat{P}_i = \hat{P}_i(\underline{\hat{\beta}}^0) + \left[\frac{\partial \hat{P}_i}{\partial \hat{\beta}_0} \right]_{\underline{\hat{\beta}}=\underline{\hat{\beta}}^0} (\hat{\beta}_0 - \beta_0) + \left[\frac{\partial \hat{P}_i}{\partial \hat{\beta}_1} \right]_{\underline{\hat{\beta}}=\underline{\hat{\beta}}^0} (\hat{\beta}_1 - \beta_1) + \dots + \left[\frac{\partial \hat{P}_i}{\partial \hat{\beta}_k} \right]_{\underline{\hat{\beta}}=\underline{\hat{\beta}}^0} (\hat{\beta}_k - \beta_k)$$

Fazendo $\underline{\hat{\beta}}^0 = \underline{\hat{\beta}}$, obtém-se:

$$\hat{P}_i - P_i = \sum_j \underline{x}_{ij} P_i (1 - P_i) (\hat{\beta}_j - \beta_j)$$

que pode ser escrito como segue:

$$\hat{P}_i - P_i = P_i (1 - P_i) \underline{x}_i (\underline{\hat{\beta}} - \underline{\beta})$$

Elevando ao quadrado ambos os membros, vem:

$$(\hat{P}_i - P_i)^2 = P_i^2 (1 - P_i)^2 \underline{x}_i (\underline{\hat{\beta}} - \underline{\beta}) (\underline{\hat{\beta}} - \underline{\beta})' \underline{x}_i'$$

Aplicando-se esperança matemática, obtém-se:

$$V(\hat{P}_i) = P_i^2(1-P_i)^2 \underline{x}_i' V(\underline{\beta}) \underline{x}_i$$

Como os P_i ; $i=1,2,\dots,s$, são parâmetros desconhecidos, são substituídos por $\hat{P}_i$, obtendo-se a estimativa da variância de $\hat{P}_i$, como segue:

$$\hat{V}(\hat{P}_i) = \hat{P}_i^2(1-\hat{P}_i)^2 \underline{x}_i' \hat{V}(\underline{\beta}) \underline{x}_i \quad (3.4.2.1)$$

onde $\hat{V}(\underline{\beta})$ é a estimativa da matriz de variâncias e covariâncias do vetor $\underline{\beta}$.

3.4.2.1 - Variâncias das estimativas das proporções quando os parâmetros são estimados pelo método de Berkson

Lembrando a matriz de variâncias e covariâncias dada em (3.3.1.1), obtêm-se, de (3.4.2.1)

$$\hat{V}(\hat{P}_i) = \hat{P}_i^2(1-\hat{P}_i)^2 \underline{x}_i' \left[X' (\hat{V}(\underline{\beta}))^{-1} X \right]^{-1} \underline{x}_i s^2 \quad (3.4.2.1.1)$$

3.4.2.2 - Variâncias das proporções quando os parâmetros são estimados pelo método da máxima verossimilhança

Lembrando a matriz de variâncias e covariâncias dada em (3.3.2.3) e usando $\hat{P}_i$ como estimativa de P_i , tem-se, de (3.4.2.1):

$$\hat{V}(\hat{P}_i) = \hat{P}_i^2(1-\hat{P}_i)^2 \underline{x}_i' \left[\sum n_i \hat{P}_i (1-\hat{P}_i) \underline{x}_i' \underline{x}_i \right]^{-1} \underline{x}_i \quad (3.4.2.2.1)$$

Nos casos em que é necessário usar o "fator de heterogeneidade", tem-se, de (3.4.2.1) e lembrando (3.3.2.4):

$$\hat{V}(\hat{P}_i) = \hat{P}_i^2(1-\hat{P}_i)^2 h \underline{x}_i' \left[\sum n_i \hat{P}_i (1-\hat{P}_i) \underline{x}_i' \underline{x}_i \right]^{-1} \underline{x}_i \quad (3.4.2.2.2)$$

4 - RESULTADOS

4.1 - Introdução

Os dados de resposta quântica têm grande importância em biologia, pois aparecem no estudo da eficiência de drogas em geral, quer usadas como medicamentos, quer usadas como inseticidas, bactericidas, etc..

Entretanto, a biologia não dispõe de teorias que determinem a escolha da forma matemática da função que deva ser ajustada aos dados de resposta quântica. Por outro lado, a literatura mostra que tanto a logística como a acumulada da curva normal dão bons ajustes. Ainda, diversos trabalhos mostram que os resultados de ambos os ajustes, feitos sobre os mesmos dados, são bastante concordantes, mas a logística é de aplicação mais direta.

Os capítulos 2 e 3, do presente trabalho, mostram que a análise estatística de dados de resposta quântica, feita através da logística, é relativamente complexa. Isto explica o uso, bastante generalizado, de processos meramente gráficos para a descrição do fenômeno

biológico, estudado através de dados de resposta quântica. Também existem processos simplificados, como os de Reed-Muench e de Spearman-Kärber, descritos por FINNEY (1971), que facilitam os cálculos em detrimento do rigor estatístico.

Do ponto de vista do estatístico, a forma mais adequada de estudar dados de resposta quântica é através da análise de regressão, que permite não só a descrição do fenômeno biológico como também a inferência estatística.

Entretanto, os dados de resposta quântica, originalmente definidos no campo da biologia por BLISS (1937), também aparecem em outras áreas do conhecimento. Isto justifica sua maior divulgação, tanto no que se refere aos aspectos metodológicos como no que se refere aos aspectos de cálculo, relativamente complexos, que devem ser feitos.

Neste capítulo são apresentados os resultados do ajuste da função logística para dados de resposta quântica, encontrados na literatura.

4.2 - O ajuste da função logística com uma variável independente

A função logística foi ajustada aos dados referentes à ação inseticida da rotenona sobre Macrosiphoniella sanborni (pulgão em crisântemo), que constam na tabela 1.1.

Na tabela 4.2.1, são apresentados as estimativas dos parâmetros, as estimativas das variâncias e das covariâncias das estimativas

dos parâmetros, as somas de quadrados dos resíduos e os "quadrados médios", com 3 graus de liberdade, obtidos através dos três métodos descritos. É conveniente assinalar que no método de Berkson, o "quadrado médio" (s^2) do resíduo é obtido através da soma de quadrados ponderados dos lógitos. Entretanto, no método dos mínimos quadrados ponderados não-lineares e no método da máxima verossimilhança, os "quadrados médios" (s^2) são obtidos através da soma de quadrados ponderados das proporções. Para simples comparação, também foram obtidas as somas de quadrados ponderados das proporções, a partir dos lógitos, no método de Berkson.

Tabela 4.2.1 - Estimativas dos parâmetros, estimativas das variâncias e das covariâncias das estimativas dos parâmetros, somas de quadrados dos resíduos e "quadrados médios" (s^2) com 3 graus de liberdade, obtidos através dos três métodos descritos

Estatísticas	Método		
	Berkson	M.Q.P.	M.V.
$\hat{\alpha}$	-4,83641	-4,83727	-4,88691
$\hat{\beta}$	7,05742	7,05970	7,14618
S.Q.Res.lóg.	1,39407	-	-
S.Q.Res.prop.	1,40681	1,40678	1,42179
s^2	0,46469	0,46893	0,47393
$\hat{V}(\hat{\alpha})$	0,40550	-	0,41336
$\hat{V}(\hat{\beta})$	0,77208	-	0,79701
Côv($\hat{\alpha}$, $\hat{\beta}$)	-0,54141	-	-0,55593

Os resultados apresentados na tabela 4.2.1 mostram que as somas de quadrados dos resíduos têm valores inferiores e bastante próximos aos respectivos graus de liberdade. Portanto, não deve ser feita a multiplicação desse valor pelo "fator de heterogeneidade" h . Aliás, FINNEY (1971), trabalhando com os mesmos dados experimentais, mas usando a acumulada da curva normal (próbite), não utilizou o fator h .

É conveniente observar que os resultados apresentados na tabela 4.2.1 mostram que as estimativas dos parâmetros, obtidas através dos três métodos, são bastante próximas. Também deve ser notado que as somas de quadrados dos resíduos, cujos valores são bastante próximos da respectiva esperança assintótica, indicam que a função logística se ajustou bastante bem aos dados. Também deve ser observado que a menor soma de quadrados dos resíduos corresponde ao método dos mínimos quadrados ponderados não-lineares, mas que a soma de quadrados ponderados das proporções, obtida a partir dos lógites, no método de Berkson, tem valor muito próximo à soma de quadrados obtida pelo método de mínimos quadrados ponderados não-lineares. Outros indicadores da qualidade do ajuste são as grandezas dos "quadrados médios", bastante pequenas em ambos os métodos e os valores das estimativas das variâncias das estimativas dos parâmetros, pequenos em relação aos valores das estimativas dos parâmetros.

Parece razoável afirmar que um bom ajuste é condição suficiente para obter a concordância de resultados, qualquer que seja o método, dos três descritos, utilizado. Com base nesta afirmativa e nos resultados apresentados na tabela 4.2.1, pode-se recomendar, nos casos em

que o ajuste é bom, o uso do método de Berkson apenas, desde que não se disponha de um computador.

Na tabela 4.2.2 são apresentadas as proporções observadas, as proporções estimadas através dos três métodos e as estimativas dos desvios-padrões das proporções, estimadas pelo método de Berkson e pelo método da máxima verossimilhança.

Tabela 4.2.2 - Proporções observadas e proporções estimadas, através dos métodos de Berkson, mínimos quadrados ponderados não-lineares e da máxima verossimilhança e estimativas dos desvios-padrões das proporções estimadas pelos métodos de Berkson e da máxima verossimilhança, para os dados da tabela 1.1.

Ensaio	P	Método de Berkson		M.M.Q.P.		M.M.V.	
		$\hat{P}$	$s(\hat{P})$	$\hat{P}$	$\hat{P}$	$s(\hat{P})$	
1	0,12000	0,12923	0,03359	0,12770	0,12924	0,03327	
2	0,33333	0,32199	0,04212	0,32220	0,32209	0,04180	
3	0,52174	0,53922	0,03996	0,54228	0,53941	0,03976	
4	0,85714	0,80533	0,03586	0,80971	0,80551	0,03582	
5	0,88000	0,90735	0,02642	0,91058	0,90747	0,02610	

É fácil ver, observando os resultados apresentados na tabela 4.2.2, que as estimativas das proporções, obtidas pelos três métodos,

são bastante concordantes entre si e bastante próximas das proporções observadas. Ainda, as estimativas dos desvios-padrões das estimativas das proporções são relativamente pequenas. Isto confirma a observação feita anteriormente, de que o ajuste da função logística, para esses dados, é bastante bom.

4.3 - O ajuste da função logística com duas variáveis independentes

WIENDL, SGRILLO e WALDER (1979) estudaram a influência da idade pupal na radiosensibilidade de Ceratitis capitata (Wied.) às irradiações gama.

De acordo com esses autores, a Ceratitis capitata (Wied.), conhecida como mosca-do-Mediterrâneo, é uma das mais importantes pragas das plantas cultivadas, prejudicando mais de 50 plantas frutíferas.

Os métodos químicos de controle mostram-se, de uma maneira geral, eficientes no início, mas se tornam obsoletos com o passar dos anos, além de contribuir para a degradação do meio ambiente.

Então tem sido estudada, em todo o mundo, a viabilidade da aplicação da Técnica do Inseto Estéril (TIE), que visa controlar eficientemente a praga, sem causar a poluição do meio ambiente.

Dentre os vários problemas a serem estudados para a aplicação eficiente da TIE, consta o item da dose esterilizante. Esta deve ser a menor possível, a fim de não prejudicar nem os hábitos nem a biologia da espécie.

Com a finalidade de estudar dados relativos à influência da idade na radiosensibilidade, os autores conduziram um experimento utilizando pupas de Ceratitis capitata (Wiend.), provenientes de uma criação do setor de Entomologia do CENA-ESALQ-USP.

Nesse experimento, foram utilizadas as seguintes doses de radiação gama: 4,1; 8,2; 12,7; 16,8; 20,4; 25,0; 29,1 krad, em idades variando de 3 a 8 dias. Para cada dose e para cada idade foram utilizadas 95 pupas.

Os dados obtidos se referem ao número de insetos que emergiram das pupas submetidas à diferentes quantidades de irradiação em diferentes idades e estão na tabela 4.3.1.

Na tabela 4.3.2 são apresentados as estimativas dos parâmetros, as estimativas das variâncias e das covariâncias das estimativas dos parâmetros, as somas de quadrados dos resíduos e os "quadrados médios" (s^2), com 40 graus de liberdade, obtidos através dos três métodos descritos. É conveniente frisar que, no método de Berkson, o "quadrado médio" do resíduo é obtido através da soma de quadrados ponderados dos lógites, enquanto que no método dos mínimos quadrados ponderados não-lineares e no método da máxima verossimilhança os "quadrados médios" são obtidos através da soma de quadrados ponderados das proporções. Entretanto, para comparação, também foi obtida a soma de quadrados ponderados das proporções, partindo dos lógites, no método de Berkson.

Tabela 4.3.1 - Número exposto e número de insetos que emergiram das pupas submetidas a diferentes quantidades de irradiação em diferentes idades

Ensaio	Nº exposto	Nº afetado	Prop.afetado	Dose	Idade
1	95	0	0,00000	4,1	3
2	95	83	0,87368	4,1	4
3	95	87	0,91578	4,1	5
4	95	92	0,96842	4,1	6
5	95	89	0,93684	4,1	7
6	95	93	0,97894	4,1	8
7	95	1	0,01052	8,2	3
8	95	19	0,20000	8,2	4
9	95	83	0,87368	8,2	5
10	95	90	0,94736	8,2	6
11	95	88	0,92631	8,2	7
12	95	88	0,92631	8,2	8
13	95	0	0,00000	12,7	3
14	95	5	0,05263	12,7	4
15	95	63	0,66315	12,7	5
16	95	81	0,85263	12,7	6
17	95	93	0,97894	12,7	7
18	95	85	0,89473	12,7	8
19	95	0	0,00000	16,8	3
20	95	0	0,00000	16,8	4
21	95	36	0,37894	16,8	5
22	95	72	0,75789	16,8	6
23	95	93	0,97894	16,8	7
24	95	95	1,00000	16,8	8
25	95	0	0,00000	20,4	3
26	95	0	0,00000	20,4	4
27	95	4	0,04210	20,4	5
28	95	44	0,46315	20,4	6
29	95	83	0,87368	20,4	7
30	95	89	0,93684	20,4	8
31	95	0	0,00000	25,0	3
32	95	0	0,00000	25,0	4
33	95	0	0,00000	25,0	5
34	95	13	0,13684	25,0	6
35	95	63	0,66315	25,0	7
36	95	92	0,96842	25,0	8
37	95	0	0,00000	29,1	3
38	95	0	0,00000	29,1	4
39	95	0	0,00000	29,1	5
40	95	2	0,02105	29,1	6
41	95	26	0,27368	29,1	7
42	95	79	0,83157	29,1	8

Tabela 4.3.2 - Estimativas dos parâmetros, estimativas das variâncias e covariâncias das estimativas dos parâmetros, somas de quadrados dos resíduos e "quadrados médios"(s^2), com 40 graus de liberdade, obtidos através dos três métodos descritos, para os dados da tabela 4.3.1

Estatística	Método		
	Berkson	M.Q.P.	M.V.
$\hat{\beta}_0$	-3,83914	-4,34912	-6,22751
$\hat{\beta}_1$	-0,18085	-0,17999	-0,26662
$\hat{\beta}_2$	1,23696	1,30101	1,90326
S.Q.Res.lógites	286,53997	-	-
S.Q.Res.prop.	407,55809	395,67609	1183,34131
s^2	7,34718	9,89190	30,34208
$\hat{V}(\hat{\beta}_0)$	0,60686	-	1,87622
$\hat{V}(\hat{\beta}_1)$	0,00054	-	0,00334
$\hat{V}(\hat{\beta}_2)$	0,02042	-	0,11998
$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$	0,00021	-	0,03126
$\text{Cov}(\hat{\beta}_0, \hat{\beta}_2)$	-0,09721	-	-0,41284
$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$	-0,00154	-	-0,01551

Os resultados apresentados na tabela 4.3.2 mostram que as somas de quadrados dos resíduos têm valores bastante superiores aos respectivos graus de liberdade. Portanto, é razoável multiplicar esse valor pelo "fator de heterogeneidade" h , conforme recomenda FINNEY (1971).

Então, as estimativas das variâncias e das covariâncias das estimativas dos parâmetros estão "corrigidas".

É conveniente observar que as estimativas dos parâmetros, obtidas através dos três métodos e apresentadas na tabela 4.3.2, são diferentes entre si. Neste caso, é razoável utilizar, como estimativas dos parâmetros, os valores obtidos através do método de máxima verossimilhança. Este método tem melhores propriedades, conforme discute SILVERSTONE (1957). Ainda, é o único método recomendado por JOHNSON e KOTZ (1970). Também deve ser notado que o método de mínimos quadrados ponderados não-lineares conduz a uma soma de quadrados dos resíduos relativamente menor do que aquela obtida pelo método de Berkson. Portanto, parece razoável afirmar que, para estes dados, os métodos iterativos conduzem a melhores resultados.

Por outro lado, os "quadrados médios" são relativamente grandes, indicando que o ajuste da função logística, para estes dados, é apenas razoável.

Tabela 4.3.3 - Proporções observadas e proporções estimadas, através dos métodos de Berkson, mínimos quadrados ponderados não-lineares e da máxima verossimilhança e estimativas dos desvios-padrões das proporções estimadas pelos métodos de Berkson e da máxima verossimilhança, para os dados da tabela 4.3.1

Ensaio	P	Método de Berkson		M.M.Q.P.		M.M.V.	
		$\hat{P}$	$s(\hat{P})$	$\hat{P}$	$\hat{P}$	$s(\hat{P})$	
1	0,00000	0,29531	0,08839	0,23432	0,16647	0,08884	
2	0,87368	0,59079	0,08247	0,52920	0,57257	0,13336	
3	0,91579	0,83260	0,04254	0,80502	0,89986	0,05867	
4	0,96842	0,94486	0,01734	0,93814	0,98368	0,01428	
5	0,93684	0,98334	0,00674	0,98236	0,99753	0,00291	
6	0,97895	0,99511	0,00252	0,99514	0,99963	0,00055	
7	0,01053	0,16642	0,05749	0,12763	0,06274	0,03959	
8	0,20000	0,40751	0,07444	0,34954	0,30987	0,09871	
9	0,87368	0,70323	0,05080	0,66373	0,75073	0,08633	
10	0,94737	0,89087	0,02465	0,87879	0,95283	0,03023	
11	0,92632	0,96566	0,01100	0,96381	0,99267	0,00702	
12	0,92632	0,98978	0,00449	0,98988	0,99890	0,00141	
13	0,00000	0,08128	0,03196	0,06111	0,01977	0,01540	
14	0,05263	0,23360	0,05450	0,19294	0,11915	0,05250	
15	0,66315	0,51222	0,05183	0,46754	0,47570	0,08288	
16	0,85263	0,78344	0,03139	0,76333	0,85888	0,05562	
17	0,97895	0,92573	0,01770	0,92216	0,97609	0,01732	
18	0,89474	0,97724	0,00830	0,97753	0,99636	0,00386	
19	0,00000	0,04044	0,01790	0,03018	0,00671	0,00636	
20	0,00000	0,12680	0,03666	0,10257	0,04337	0,02630	
21	0,37895	0,33346	0,04839	0,29567	0,23319	0,06568	
22	0,75789	0,63283	0,03747	0,60659	0,67103	0,07251	
23	0,97895	0,85586	0,02630	0,84993	0,93189	0,03595	
24	1,00000	0,95339	0,01445	0,95413	0,98922	0,00938	
25	0,00000	0,02151	0,01060	0,01602	0,00258	0,00287	
26	0,00000	0,07040	0,02445	0,05641	0,01707	0,01329	
27	0,04211	0,20691	0,04225	0,18005	0,10431	0,04662	
28	0,46316	0,47336	0,04586	0,44646	0,43857	0,08178	
29	0,87368	0,75589	0,03792	0,74763	0,83974	0,06190	
30	0,93684	0,91430	0,02370	0,91583	0,97234	0,01000	
31	0,00000	0,00948	0,00536	0,00706	0,00076	0,00102	
32	0,00000	0,03191	0,01376	0,02546	0,00507	0,00515	
33	0,00000	0,10197	0,03050	0,08755	0,03303	0,02305	
34	0,13684	0,28119	0,05126	0,26058	0,18642	0,07421	
35	0,66316	0,57405	0,05918	0,56416	0,60584	0,10625	
36	0,96842	0,82279	0,04462	0,82622	0,91158	0,05075	
37	0,00000	0,00454	0,00290	0,00339	0,00025	0,00040	
38	0,00000	0,01546	0,00791	0,01233	0,00170	0,00210	
39	0,00000	0,05132	0,02008	0,04386	0,01132	0,01051	
40	0,02105	0,15709	0,04409	0,14419	0,07132	0,04558	
41	0,27368	0,39100	0,07243	0,38227	0,34000	0,12393	
42	0,83158	0,68866	0,07271	0,69446	0,77556	0,10679	

5 - PROGRAMA PARA COMPUTADORES

Neste capítulo é apresentada a listagem do programa para computadores que possibilita a obtenção das estimativas dos parâmetros da função logística pelos três métodos descritos; isto é, o método de Berkson, o método dos Mínimos Quadrados Ponderados Não-lineares e o método da Máxima Verossimilhança. O programa também fornece outros resultados, como as variâncias e covariâncias das estimativas dos parâmetros, as proporções estimadas, as variâncias das proporções estimadas, além de outras estatísticas.

A linguagem utilizada é o FORTRAN-IV e o programa pode ser utilizado em computadores IBM-1130 ou em computadores de maior porte. Utiliza-se a subrotina MINV do SSP (Scientific Subroutine Package) para a inversão de matrizes, cuja listagem também é apresentada.

O programa está originalmente dimensionado para um máximo de 4 variáveis independentes e 100 ensaios. Nos casos em que se deseja um ajuste com mais de 4 variáveis independentes e/ou mais de 100 ensaios, basta alterar as instruções DIMENSION.

O programa exige 3 leituras:

i) vetor XNOME: contém a denominação do problema. Admite um máximo de 80 caracteres, incluindo os espaços em branco. É o primeiro cartão a ser lido.

ii) valores de N, K, KLG e KVE, sendo cada um deles como segue:

N - número de ensaios, em formato I3

K - número de variáveis independentes, em formato I3

KLK - seleciona a transformação a ser realizada nas variáveis independentes. Pode assumir os valores:

0 - não realiza nenhuma transformação

1 - toma o logaritmo decimal das variáveis independentes

2 - toma o logaritmo neperiano das variáveis independentes

3 - toma a raiz quadrada das variáveis independentes

O valor de KLG deve ser perfurado em formato I1, na décima coluna.

KVE - calcula (KVE=1) ou não (KVE=0) os desvios-padrões dos lógites e das probabilidades estimadas. Deve ser perfurado em formato I1, na décima terceira coluna.

Os valores de N, K, KLG e KVE devem ser colocados no segundo cartão de dados a ser lido. Assim por exemplo, se $N=50$; $K=4$; $KLK=3$ e $KVE=1$, o cartão deve ser perfurado como segue:

— 5 0! — 4! — 3! — 1!

Finalmente a última leitura compreende o número de indivíduos expostos, o número de indivíduos afetados e os correspondentes valores das variáveis independentes. Existirão então N cartões, cada um deles com K+2 valores, em formato F8.0, sendo o primeiro valor (YN) referente ao número de indivíduos expostos, o segundo valor (YM), referente ao número de indivíduos afetados e os restantes K valores referentes as variáveis independentes.

```

DIMENSION YN(100),YM(100),X(4,100),B(5),V(5),C(5),DB(5)
DIMENSION Z(5,5),A(25),LL(5),NN(5)
DIMENSION XNOME(21)

```

```

LOGITO COM MAIS DE UMA VARIABEL INDEPENDENTE

```

```

PROGRAMADOR= R. HOFFMANN

```

```

101 FORMAT(20A4)

```

```

102 FORMAT(4X,20A4)

```

```

100 CONTINUE

```

```

PRIMEIRO CARTAO - NOME CARACTERIZANDO OS DADOS (COL. 1 A 80)

```

```

READ(2,101)(XNOME(I),I=1,20)

```

```

34 FORMAT('1')

```

```

35 FORMAT(1H )

```

```

READ(2,10)N,K,KLG,KVE

```

```

10 FORMAT(2I3,3X,I1,2X,I1)

```

```

N=NUMERO TOTAL DE ENSAIOS E K=NUMERO DE VARIAVEIS INDEPENDENTES

```

```

SE KLG=0 NAO TRANSFORMA AS VARIAVEIS INDEPENDENTES

```

```

SE KLG=1 CALCULA LOG DECIMAL DAS VARIAVEIS INDEPENDENTES

```

```

SE KLG=2 CALCULA LOG NEPERIANO DAS VARIAVEIS INDEP.

```

```

SE KLG=3 CALCULA RAIZ QUADRADA DAS VARIAVEIS INDEPENDENTES

```

```

SE KVE=1 CALCULA DESVIOS PADRES DOS LOGITOS E DAS PROBABILIDADES

```

```

ESTIMADAS - SE KVE=0, NAO CALCULA.

```

```

IF(N)31,31,30

```

```

30 CONTINUE

```

```

K=K+1

```

```

DO 40 J=1,N

```

```

40 READ(2,11)YN(J),YM(J),(X(I,J),I=2,K)

```

```

11 FORMAT(10F8,0)

```

```

YN(J)=NUMERO DE INDIVIDUOS EXPOSTOS AO TRAT. NO J-ESIMO ENSAIO

```

```

YM(J)=NUMERO DE INDIVIDUOS AFETADOS NO J-ESIMO ENSAIO

```

```

X(I,J)=VARIAVEIS INDEP. QUE CARACTERIZAM O TRATAMENTO

```

```

WRITE(3,34)

```

```

WRITE(3,102)(XNOME(I),I=1,20)

```

```

WRITE(3,35)

```

```

WRITE(3,333)KLG

```

```

333 FORMAT(' TRANSFORMACAO DAS VARIAVEIS INDEP., KLG=',I1)

```

```

WRITE(3,35)

```

```

WRITE(3,43)

```

```

DO 41 J=1,N

```

```

PO=YM(J)/YN(J)

```

```

41 WRITE(3,42)J,YN(J),YM(J),PO,(X(I,J),I=2,K)

```

```

42 FORMAT(1X,I4,2X,2F8,2,F10,5,5F17,7)

```

```

43 FORMAT(' ENSAIO',4X,'YN',6X,'YM',5X,'PROP.',10X,'X(1)',13X,'X(2)')

```

```

IF(KLG)31,301,302

```

```

302 GO TO (310,320,330),KLG

```

```

310 DO 311 I=2,K

```

```

      DO 311 J=1,N
311  X(I,J)=0.43429448*ALOG(X(I,J))
      GO TO 301
320  DO 321 I=2,K
      DO 321 J=1,N
321  X(I,J)=ALOG(X(I,J))
      GO TO 301
330  DO 331 I=2,K
      DO 331 J=1,N
331  X(I,J)=SQRT(X(I,J))
301  CONTINUE

```

```

C
C   ACRESCIMO DE UMA COLUNA DE 1'S NA MATRIZ X
C

```

```

      DO 140 J=1,N
140  X(1,J)=1.

```

```

C
C   METODO DE BERKSON - MINIMOS QUADRADOS PONDERADOS DOS LOGITOS
C

```

```

      DO 150 I=1,K
      C(I)=0.
      DO 150 L=1,K
150  Z(I,L)=0.
      SY=0.
      SQY=0.
      SF=0.
      SQRG=0.
      DO 151 J=1,N
      P=YM(J)/YN(J)
      IF(P)950,221,222
221  P=1./(2.*YN(J))
      GO TO 223
222  IF(P-1.)223,224,950
224  P=1.-1./(2.*YN(J))
223  CONTINUE
      Y=ALOG(P/(1.-P))
      F=YN(J)*P*(1.-P)
      SF=SF+F
      SY=SY+Y*F
      SQY=SQY+Y*Y*F
      DO 151 I=1,K
      C(I)=C(I)+X(I,J)*Y*F
      DO 151 L=1,K
151  Z(I,L)=Z(I,L)+X(I,J)*X(L,J)*F
      M=1
      DO 2 I=1,K
      DO 2 L=1,K
      A(M)=Z(I,L)
      2  M=M+1
      CALL MINV(A,K,D,LL,NN)
      M=1
      DO 802 I=1,K
      DO 802 L=1,K
      Z(I,L)=A(M)
802  M=M+1
      DO 4 I=1,K
      B(I)=0.
      DO 3 L=1,K
      3  B(I)=B(I)+Z(I,L)*C(L)

```

```

4  SQRG=SQRG+B(I)*C(I)
   SQRES=SQY-SQRG
   SQY=SQY-SY*SY/SF
   SQRG=SQRG-SY*SY/SF
   R2=SQRG/SQY
   GL=N-K
   QMRES=SQRES/GL
   DO 5 I=1,K
5  V(I)=SQRT(Z(I,I)*QMRES)
   WRITE(3,34)
   WRITE(3,102)(XNOME(I),I=1,20)
   WRITE(3,103)
103 FORMAT(4X,'DADOS DE RESPOSTA QUANTICA-LOGITO- P=1/(1+EXP(-XB))')
   WRITE(3,35)
   WRITE(3,71)N
71  FORMAT(2X,'NUMERO DE ENSAIOS=',I3)
   WRITE(3,35)
   WRITE(3,80)
80  FORMAT(' AJUSTE PELO MET.DE BERKSON-MIN.QUADR.LOGITOS PONDERADOS')
   WRITE(3,35)
   WRITE(3,72)
72  FORMAT(' ESTIMATIVAS DOS PARAMETROS E DESVIOS PADROES')
   WRITE(3,20)(B(I),I=1,K)
   WRITE(3,20)(V(I),I=1,K)
20  FORMAT(2X,5F20.10)
   WRITE(3,35)
   WRITE(3,73)SQY
73  FORMAT(' SOMA DE QUADRADOS TOTAL DOS LOGITOS=',F20.10)
   WRITE(3,74)SQRES
74  FORMAT(' SOMA DE QUADR. RESIDUAL DOS LOGITOS=',F20.10)
   WRITE(3,76)QMRES
76  FORMAT(' QUADRADO MEDIO RESIDUAL DOS LOGITOS=',F20.10)
   WRITE(3,75)R2
75  FORMAT(' COEFICIENTE DE DETERMINACAO=',8X,F20.10)
   WRITE(3,35)
   WRITE(3,16)
   QQY=0.
   QQ=0.
   DO 6 J=1,N
   PO=YM(J)/YN(J)
   P=YM(J)/YN(J)
   IF(P)950,231,232
231 P=1./(2.*YN(J))
   GO TO 233
232 IF(P-1.)233,234,950
234 P=1.-1./(2.*YN(J))
233 CONTINUE
   Y=ALOG(P/(1.-P))
   YE=0.
   DO 235 I=1,K
235 YE=YE+R(I)*X(I,J)
   PE=1./(1.+EXP(-YE))
   QQY=QQY+YN(J)*P*(1.-P)*(Y-YE)**2
   QQ=QQ+(YN(J)/(PE*(1.-PE)))*(PO-PE)**2
   WRITE(3,15)J,X(2,J),YN(J),YM(J),PO,PE,Y,YE
6  CONTINUE
15  FORMAT(2X,I3,7F16.7)
16  FORMAT(' ENSAIO',5X,'X(1)',8X,'YN=NO.EXPOSTO',3X,'YM=NO.AFETADO',4
1X,'PROP.AFETADA',3X,'PROP.ESTIMADA',5X,'LOGITO OBS.',5X,'LOGITO ES

```

```

2T.')
```

WRITE(3,35)
WRITE(3,18)

18 FORMAT(5X,'S.Q.POND.RES.DAS PROPORCOES')
WRITE(3,20)QQ
WRITE(3,35)
WRITE(3,918)

918 FORMAT(5X,'S.Q.POND.RES.DOS LOGITOS')
WRITE(3,20)QQY
WRITE(3,35)
WRITE(3,800)

800 FORMAT(' ESTIMATIVAS DAS VAR. E COVAR. DAS EST. DOS PARAMETROS')
DO 821 I=1,K
821 WRITE(3,20)(Z(I,L),L=1,K)
WRITE(3,35)
QQY=QQY/GL
WRITE(3,903)QQY

903 FORMAT(' (S.Q.RES.POND.LGTS)/(GRAUS DE LIB.)=',F20.10)
WRITE(3,35)
WRITE(3,812)

812 FORMAT(' EST.DAS VAR.E COVAR.DAS EST.DOS PARAMETROS COM CORRECAO P
1ELO QUI=QUADR.LGTS/(GR.DE LIB.)')
DO 831 I=1,K
DO 832 L=1,K
832 Z(I,L)=Z(I,L)*QQY
831 WRITE(3,20)(Z(I,L),L=1,K)

C
C
C
C

VARIANCIAS DAS EST. DOS LOGITOS E DAS PROBABILIDADES ESTIMADAS
NO METODO DE BERKSON

IF(KVE)31,901,902

902 WRITE(3,35)
WRITE(3,913)

913 FORMAT(' ENSAIO',3X,'PROP.ESTIMADA',7X,'S(P EST.) S(P EST.)CORR.'
1,11X,'LOGITO EST. S(LGT EST.) S(LGT EST.)CORR.')

904 FORMAT(2X,I3,2X,3F16.7,6X,3F16.7)
DO 905 J=1,N
PO=YM(J)/YN(J)
P=YM(J)/YN(J)
IF(P)950,931,932

931 P=1./(2.*YN(J))
GO TO 933

932 IF(P-1.)933,934,950
934 P=1.-1./(2.*YN(J))
933 CONTINUE
Y=ALOG(P/(1.-P))
VYC=0.
YF=0.
DO 906 I=1,K
YE=YE+B(I)*X(I,J)
DO 906 L=1,K

906 VYC=VYC+X(I,J)*X(L,J)*Z(I,L)
PE=1./(1.+EXP(-YE))
VPC=VYC*(PE*(1.-PE))*2
VY=SQRT(VYC/QQY)
VP=SQRT(VPC/QQY)
VPC=SQRT(VPC)
VYC=SQRT(VYC)

905 WRITE(3,904)J,PE,VP,VPC,YE,VY,VYC

```

901 CONTINUE
36 FORMAT(/)
WRITE(3,36)

C
C
C
METODO DOS MIN. QUADR. NAO LINEARES, PROCESSO ITERATIVO

WRITE(3,81)
81 FORMAT(' METODO DOS MINIMOS QUADR.PONDERADOS NAO LINEARES-PROCESS
10 ITERATIVO')
55 CONTINUE
DO 152 I=1,K
C(I)=0.
DO 152 L=1,K
152 Z(I,L)=0.
DO 153 J=1,N
PO=YM(J)/YN(J)
YE=0.
DO 154 I=1,K
154 YE=YE+B(I)*X(I,J)
PE=1./(1.+EXP(-YE))
DO 153 I=1,K
C(I)=C(I)+YN(J)*X(I,J)*(PE-PO)*(2.*PO*PE-PO-PE)/(PE*(1.-PE))
DO 153 L=1,K
HELP=PO*PO+PE*PE+2.*PO*PE*(PO*PE-PO-PE)
153 Z(I,L)=Z(I,L)+YN(J)*X(I,J)*X(L,J)*HELP/(PE*(1.-PE))
M=1
DO 7 I=1,K
DO 7 L=1,K
A(M)=Z(I,L)
7 M=M+1
CALL MINV(A,K,D,LL,NN)
APR=0.
M=1
DO 8 I=1,K
DB(I)=0.
DO 9 L=1,K
DB(I)=DB(I)+A(M)*C(L)
9 M=M+1
8 APR=APR+ABS(DB(I)/V(I))
APR=APR-.1E-05
WRITE(3,20)(DB(I),I=1,K)
CALL DATSW(1,ID)
GO TO (51,52),ID
52 IF(APR)51,51,53
53 CONTINUE
DO 54 I=1,K
54 B(I)=B(I)+DB(I)
GO TO 55
51 CONTINUE
WRITE(3,35)
WRITE(3,34)
WRITE(3,20)(DB(I),I=1,K)
WRITE(3,35)
WRITE(3,102)(XNOME(I),I=1,20)
WRITE(3,103)
WRITE(3,81)
WRITE(3,35)
WRITE(3,71)N
WRITE(3,35)

```

```

WRITE(3,82)
82 FORMAT(' ESTIMATIVAS DOS PARAMETROS')
WRITE(3,20)(B(I),I=1,K)
WRITE(3,35)
WRITE(3,16)
QQ=0.
DO 21 J=1,N
PO=YM(J)/YN(J)
P=YM(J)/YN(J)
IF(P)950,241,242
241 P=1./(2.*YN(J))
GO TO 243
242 IF(P=1.)243,244,950
244 P=1.-1./(2.*YN(J))
243 CONTINUE
Y=ALOG(P/(1.-P))
YE=0.
DO 245 I=1,K
245 YE=YE+B(I)*X(I,J)
PE=1./(1.+EXP(-YE))
QQ=QQ+(YN(J)/(PE*(1.-PE)))*(PO-PE)**2
21 WRITE(3,15)J,X(2,J),YN(J),YM(J),PO,PE,Y,YE
WRITE(3,35)
WRITE(3,18)
WRITE(3,20)QQ
WRITE(3,36)

```

C
C
C

METODO DA MAXIMA VEROSSIMILHANCA,PROCESSO ITERATIVO

```

WRITE(3,91)
91 FORMAT(' METODO DA MAX.VEROSSIMILHANCA,PROCESSO ITERATIVO')
555 CONTINUE
DO 652 I=1,K
C(I)=0.
DO 652 L=1,K
652 Z(I,L)=0.
DO 653 J=1,N
PO=YM(J)/YN(J)
YE=0.
DO 654 I=1,K
654 YE=YE+B(I)*X(I,J)
PE=1./(1.+EXP(-YE))
DO 653 I=1,K
C(I)=C(I)+X(I,J)*YN(J)*(PO-PE)
DO 653 L=1,K
653 Z(I,L)=Z(I,L)+X(I,J)*X(L,J)*YN(J)*PE*(1.-PE)
M=1
DO 507 I=1,K
DO 507 L=1,K
A(M)=Z(I,L)
507 M=M+1
CALL MINV(A,K,D,LL,NN)
APR=0.
M=1
DO 807 I=1,K
DO 807 L=1,K
Z(I,L)=A(M)
807 M=M+1
DO 508 I=1,K

```

```

      DB(I)=0
      DO 509 L=1,K
509  DB(I)=DB(I)+Z(I,L)*C(L)
508  APR=APR+ABS(DB(I)/V(I))
      APR=APR-.1E-05
      WRITE(3,20)(DB(I),I=1,K)
      CALL DATSW(2,ID)
      GO TO (551,552),ID
552  IF(APR)551,551,553
553  CONTINUE
      DO 554 I=1,K
554  R(I)=B(I)+DB(I)
      GO TO 555
551  CONTINUE
      WRITE(3,35)
      WRITE(3,34)
      WRITE(3,20)(DB(I),I=1,K)
      WRITE(3,35)
      WRITE(3,102)(XNOME(I),I=1,20)
      WRITE(3,103)
      WRITE(3,91)
      WRITE(3,35)
      WRITE(3,71)N
      WRITE(3,35)
      WRITE(3,82)
      WRITE(3,20)(B(I),I=1,K)
      WRITE(3,35)
      WRITE(3,16)
      QQ=0.
      DO 521 J=1,N
      PO=YM(J)/YN(J)
      P=YM(J)/YN(J)
      IF(P)950,741,742
741  P=1./(2.*YN(J))
      GO TO 743
742  IF(P-1.)743,744,950
744  P=1.-1./(2.*YN(J))
743  CONTINUE
      Y=ALOG(P/(1.-P))
      YE=0.
      DO 745 I=1,K
745  YE=YE+B(I)*X(I,J)
      PE=1./(1.+EXP(-YE))
      QQ=QQ+(YN(J)/(PE*(1.-PE)))*(PO-PE)**2
521  WRITE(3,15)J,X(2,J),YN(J),YM(J),PO,PE,Y,YE
      WRITE(3,35)
      WRITE(3,18)
      WRITE(3,20)QQ
      WRITE(3,35)
      WRITE(3,800)
      DO 822 I=1,K
822  WRITE(3,20)(Z(I,L),L=1,K)
      WRITE(3,35)
      QQ=QQ/GL
      WRITE(3,823)QQ
823  FORMAT(' (S.Q.RES.POND.PROP.)/(GRAUS DE LIB.)=',F20.10)
      WRITE(3,35)
      WRITE(3,826)
826  FORMAT(' EST. DAS VAR. E COVAR. DAS EST. DOS PARAMETROS COM CORREC

```

```

1AO PELO QUI-QUADR.PROP./ (GR. DE LIB.))
DO 824 I=1,K
DO 825 L=1,K
825 Z(I,L)=Z(I,L)*QQ
824 WRITE(3,20)(Z(I,L),L=1,K)
C
C   VARIANCIAS DAS EST. DOS LOGITOS E DAS PROBABILIDADES ESTIMADAS
C   PELO METODO DE MAX. VEROSSIMILHANCA
C
IF(KVE)31,911,912
912 WRITE(3,35)
WRITE(3,913)
DO 915 J=1,N
VYC=0.
YE=0.
DO 916 I=1,K
YE=YE+B(I)*X(I,J)
DO 916 L=1,K
916 VYC=VYC+X(I,J)*X(L,J)*Z(I,L)
PE=1./(1.+EXP(-YE))
VPC=VYC*(PE*(1.-PE))**2
VP=SQRT(VPC/QQ)
VY=SQRT(VYC/QQ)
VPC=SQRT(VPC)
VYC=SQRT(VYC)
915 WRITE(3,904)J,PE,VP,VPC,YE,VY,VYC
911 CONTINUE
WRITE(3,36)
GO TO 100
950 PAUSE 7777
C   PAUSE 7777=HA PROPORCAO OBS. NEGATIVA OU MAIOR DO QUE UM
GO TO 100
31 CALL EXIT
END

```

```

SUBROUTINE MINV(A,N,D,L,M)
DIMENSION A(1),L(1),M(1)
D=1.0
NK=-N
DO 80 K=1,N
NK=NK+N
L(K)=K
M(K)=K
KK=NK+K
BIGA=A(KK)
DO 20 J=K,N
IZ=N*(J-1)
DO 20 I=K,N
IJ=IZ+I
10 IF(ABS(BIGA)-ABS(A(IJ)))15,20,20
15 BIGA=A(IJ)
L(K)=I
M(K)=J
20 CONTINUE
J=L(K)
IF(J-K) 35,35,25
25 KI=K-N
DO 30 I=1,N
KI=KI+N
HOLD=-A(KI)
JI=KI-K+J
A(KI)=A(JI)
30 A(JI)=HOLD
35 I=M(K)
IF(I-K) 45,45,38
38 JP=N*(I-1)
DO 40 J=1,N
JK=NK+J
JI=JP+J
HOLD=-A(JK)
A(JK)=A(JI)
40 A(JI)=HOLD
45 IF(ABS(BIGA)-1.E-20)46,46,48
46 D=0.0
RETURN
48 DO 55 I=1,N
IF(I-K) 50,55,50
50 IK=NK+I
A(IK)=A(IK)/(-BIGA)
55 CONTINUE
DO 65 I=1,N
IK=NK+I
HOLD=A(IK)

```

```

      IJ=I-N
      DO 65 J=1,N
      IJ=IJ+N
      IF(I-K) 60,65,60
60    IF(J-K) 62,65,62
62    KJ=IJ-I+K
      A(IJ)=HOLD*A(KJ)+A(IJ)
65    CONTINUE
      KJ=K-N
      DO 75 J=1,N
      KJ=KJ+N
      IF(J-K) 70,75,70
70    A(KJ)=A(KJ)/BIGA
75    CONTINUE
      D=D*BIGA
      A(KK)=1.0/BIGA
80    CONTINUE
      K=N
100   K=(K-1)
      IF(K) 150,150,105
105   I=L(K)
      IF(I-K) 120,120,108
108   JQ=N*(K-1)
      JR=N*(I-1)
      DO 110 J=1,N
      JK=JQ+J
      HOLD=A(JK)
      JI=JR+J
      A(JK)=-A(JI)
110   A(JI)=HOLD
120   J=M(K)
      IF(J-K) 100,100,125
125   KI=K-N
      DO 130 I=1,N
      KI=KI+N
      HOLD=A(KI)
      JI=KI-K+J
      A(KI)=-A(JI)
130   A(JI)=HOLD
      GO TO 100
150   RETURN
      END

```

BIBLIOGRAFIA

- AITCHISON, J. e BENNET, J.A. (1970). Polychotomus Quantal Response by Maximum Indicat. Biometrika, London, 57:253-62.
- BERKSON, J. (1944). Application of the Logistic Function to Bio-assay. J.Am.Statist.Ass., Boston, 39:357-65.
- BERKSON, J. (1951). Why I Prefer Logits to Probits. Biometrics, Washington, 7:327-39.
- BERKSON, J. (1953). A Statistically Precise and Relatively Simple Method of Quantal Response, based on the Logistic Function. J.Am.Statist.Ass., Boston, 48:565-99.
- BERKSON, J. (1955). Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function. J.Am.Statist.Ass., Boston, 50:130-63.
- BISHOP, Y.M.M.; FIENBERG, S.E. e HOLLAND, P.W. (1975). Discrete Multivariate Analysis. Cambridge. The Mit Press.
- BLISS, C.I. (1935). The Calculation of the Dosage Mortality Curve. Ann. Appl.Biol., Cambridge, 22:134-67.

- BOCK, R.D. e JONES, L.V. (1968). The Measurement and Prediction of Judgment and Choise. San Francisco, Holden-Day, Inc.
- CRAMER, E.M. (1964). Some Comparisons of Methods of Fitting the Dosage Response Curve for Small Samples. J.Am.Statist.Ass., Boston, 59:779-93.
- COX, D.R. (1970). The Analysis of Binary Data. London, Methuen & Co.Ltd.
- DRAPER, N. e SMITH, H. (1966). Applied Regression Analysis. New York, John Wiley.
- FINNEY, D.J. (1971). Probit Analysis. 3a. ed., Cambridge. Cambridge University Press.
- FINNEY, D.J. (1978). Statistical Method in Biological Assay. 3a. ed. London, Griffin.
- FISK, P.R. (1961). The Graduation of Income Distributions. Econometrica, 29:171-85.
- GRIZZLE, J.E. (1961). A New Method of Testing Hypotheses and Estimating Parameters for the Logistic Model. Biometrics, Washington, 17:372-85.
- HITCHCOCK, S.E. (1966). Tests of Hypotheses about the Parameters of the Logistic Function. Biometrika, London, 53:535-44.
- HODGES, J.L. (1958). Fitting the Logistic by Maximum Likelihood. Biometrics, Washington, 14:453-61.
- HOFFMANN, R. e VIEIRA, S. (1977). Análise de Regressão: Uma Introdução à Econometria. São Paulo, Hucitec-Edusp.
- HSI, B.P. (1969). The Multiple Sample Up-and-down Method in Bioassay. J.Am.Statist.Ass., Boston, 64:147-62.

- JOHNSON, N.L. e KOTZ, S. (1970). Continuous Univariate Distributions- 2. Boston, Houghton Mifflin.
- LITTLE, R.E. (1968). A Note on Estimation for Quantal Response Data. Biometrika, London, 55:578-9.
- MARTIN, L. (1951). Probits Normaux Classiques et Probits Relatifs à la Répartition Triangulaire. Bulletin de l'Institut Agronomique et des Stations de Recherches de Gembloux, 19:364-86. Apud: FINNEY, D. J. (1971). Probit Analysis. 3a. ed. Cambridge, Cambridge University Press.
- MILICER, H. e SZCZOTKA, F. (1966). Age at Menarche in Warsaw Girls in 1965. Human Biology, 38:199-203. Apud: FINNEY, D.J. (1971). Probit Analysis. 3a. ed. Cambridge, Cambridge University Press.
- OLIVER, F.R. (1964). Methods of Estimating the Logistic Growth Function. Appl.Statist. London, 13:57-66.
- PEARL, R. e REED, L.J. (1920). On the Rate of Growth of the Population of the United States since 1790 and its Mathematical Representation. Proc.natn.Acad.Sci. U.S.A., Washington, 6:275-88.
- REIERSOL, O. (1961). Linear and Non-linear Multiple Comparisons in Logit Analysis. Biometrika, London, 48:359-65.
- SEARLE, S.R. (1971). Linear Models. New York, John Wiley.
- SILVA LEME, R.A. da (1965). Curso de Estatística: Elementos. 2a. ed. Rio de Janeiro. Ao Livro Técnico.
- SILVERSTONE, H. (1957). Estimating the Logistic Curve. J.Am.Statist.Ass., Boston, 52:567-77.

- SNEDECOR, G.W. e COCHRAN, W.G. (1967). Statistical Methods. 6a. ed. Ames. The Iowa State University Press.
- SOWDEN, R.R. (1972). On the First-Order Bias of Parameter Estimates in a Quantal Response Model under Alternative Estimation Procedures. Biometrika, London, 59:573-9.
- THEIL, H. (1971). Principles of Econometrics. New York. John Wiley.
- VERHULST, P.E. (1845). Recherches Mathématiques sur la Loi d'Accroissement de la Population. Académie de Bruxelles, Bruxelles, 18:1-38.
- VIEIRA, S. (1975). Estudo de Funções Assintótico-Sigmóides. FOP-UNICAMP, Piracicaba (Tese de Livre-docência).
- VIEIRA, S. e HOFFMANN, R. (1979). O Uso da Função de Gompertz para Dados de Resposta Quântica. Anais da Reunião Internacional de Biometria. Setembro, 1975. Piracicaba, 453-9.
- WIENDL, F.M.; SGRILLO, R.B. e WALDER, J.M.M. (1979). Influência da Idade Pupal na Radiosensibilidade às Radiações Gama de Ceratitis capitata (Wied.). Energ.Nucl.Agric., Piracicaba, 1:15-19.
- WILKS, S.S. (1962). Mathematical Statistics. New York, John Wiley.