APPENDIX 05 - Use of logistic regression in medical statistics

This appendix is taken from Statistical Models lecture notes, K. Rennolls, 1990, except the examples in A5.2 and A5.3.

A5.1 Introduction

The logistic model is

$$p_i=(\exp(eta_i))/(1+\exp(eta_i)), \text{ where } eta_i=b_0+b_1z_{i1}+...$$

and under the CANONICAL LINK we suppose that $y_i \sim B(p_i,n_i)$.

This is often written in the form

$$p_i=1/(1+\exp(-eta_i))=1/(1+\exp(-(b_0+b_1z_{i1}+...)))$$

If the variables z_i are risk variables, such as drugs, smoking, weight, cholestrol, body-fat, etc for some event then the logistic model relates the probability of having that event to those risk variables.

We can use any of the three methods:-

(a) Each individual has his own unique set of risk variables and outcome variable. Note that in GLIM, we must still define a vector that has reasonably large numbers.

(b) Each individual has has own unique set of risk variables and the outcome variable is 0 or 1. Note that in GLIM or GENSTAT we must still define an Nvariate N=1.

(c) The risk variables are classified. In GLIM or GENSTAT N<>1.

Often the probability of an event in any given interval is small. EG the probability of a sunburn in a 5 year study might be about 0.02 on average.

If we assume that the risk of an event is small for most of the population then $p_i=(\exp(eta_i))/(1+\exp(eta_i))$ is approximately $\exp(eta_i)$. That is, an EXPONENTIAL probability model.
Clearly this cannot be extrapolated too far! (it goes above 1).

A5.2 Relative Risk (RR)

The RR of a particular event with respect to a risk variable z is defined as the ratio $RR_z=pr(z+'1')/pr(z)$ where '1' represents an appropriate 'unit' of increase of the risk variable z. The relative risk RR_z will in general depend on the value of z at which it is calculated. As z tends to infinity, RR_z tends to 1.

HOWEVER if the probability model is exponential then RR_z is constant. So OK if risk level is small. Consider RR with respect to z_i under this assumption.

$$\text{RR of } z_i=p(z_i+'1')/p(z_i)$$
$$=\exp(b_0+b_1z_1+..+b_i(z_i+'1')+..) \;/\; \exp(b_0+b_1z_1+..+b_iz_i)$$
$$=\exp('1'b_i) = (\exp(b_i))^{'1'}$$

So the RR with respect to unit increase in x_i is just $\exp(b_i)$.

NB: If risk is high then RR=exp(b) <u>IS NOT VALID</u>

A5.2.1 Examples

(1) From New Scientist, This Week, Throw away the salt cellar, Page 13, Frank Lessen, Issue 1764, 13 April 1991

A reduction in salt intake by 3 grams for heart disease reduces the heart disease by 16%.

Use RR=exp(b). Relative risk = 0.84 = exp(-3b)
b=(-1/3) ln 0.84=0.0581. This is a rough estimate of the coefficient of salt intake in the logistic regression. If we increase the salt intake by 1 gram then we have a relative risk of 1.0598.

(2) From GLIM Program 7, the coefficient of weight loss is 0.08296 in the logistic regression. The relative risk is exp(0.08296)=1.0865.

A5.3 Confidence limits on relative risks

The 95% confidence interval of relative risk is given by:-
$\exp(b_i \pm 1.96 se(b_i))$ [1]

GLIM gives estimates of b_i and $se(b_i)$, but unfortunately it does not provide t.

A5.3.1 An example

From GLIM Program 7, the coefficient of weight loss in the logistic regression is 0.08296. What is the RR for a weight loss of 3 lbs?
RR=exp(0.08296*3)=1.283
Is it significant at 95% level? t=0.379. Critical value of t=2.776.
Not significant at 95% confidence level.
95% CI = exp(-0.345888*3) < RR < exp(0.511808*3)
0.354 < RR < 4.64

If the confidence interval includes 1, then the risk variable has no significant effect on the human.