The General Linear Model

http://marekrychlik.com/node/30

{The ¡strong¿general linear model ¡/strong¿(¡strong¿GLM¡/strong¿) captures the relationship between one or more response variables y, and several ¡strong¿design variables ¡/strong¿$x_1, x_2, \ldots, x_k$. One postulates a relationship between y and x_i which involves parameters β_j, $j = 1, 2, \ldots, k$:

{

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + e$$

where e represents the error. Typically, we assume that e has mean 0, i.e. that the error is unbiased.

{Variables x_j could be interval or ratio measurements (such as mass, time, temperature etc.), or can be ¡strong¿indicator variables¡/strong¿. An indicator variable has value of 0 or 1. A set of indicator variables may be assigned to indicate the membership of a unit to a particular treatment group. Thus, if there are t treatment groups, we would have t indicator variables. The i-th indicator variable of the set would have value of 1 for all units in group i and 0 otherwise.

{In the ¡a href="/node/28"¿meat packaging example¡/a¿ there are 4 treatment labels, and thus there are 4 indicator variables x_1-x_4. The GLM for this example would be (assuming there is no β_0):

{

$$y = \sum_{i=1}^{4} \beta_i x_i + e.$$

If we restrict the model to the i-th treatment group, we can write:

$$y_{ij} = \beta_i + \epsilon_{ij}$$

which is equivalent to the full model introduced before (formally setting $\mu_i = \beta_i$). Similarly, we can write the reduced model as a GLM:

{

$$y = \beta_0 + e.$$

{The parameters of the GLM are estimated using the least squares method.