Statistical Invariance for the Logit and Probit Models

Christos P. Kitsos[1,*], Thomas L. Toulias[1], Effie Papageorgiou[2]

[1] Department of Informatics, Technological Educational Institute of Athens, Greece

[2] Department of Medical Laboratories, Technological Educational Institute of Athens, Greece

*Correspondence to xkitsos@teiath.gr

Abstract. *The target of this paper is to present an Affine Geometry point of view for the statistical invariance, applicable to the logit model. A number of properties are introduced for the logit model which, for certain* p*-values, are equivalent to the probit model.*

1 Introduction

There are cases where the multiple logit model needs a transformation, in order for the experiment design problem to be faced in practice, see [9] among others. In those cases, it has been proved in [18] that an affine transformation is actually needed when the canonical form of a logit model is investigated. Moreover, an affine transformation can act through experimentation and can preserve statistical invariance for a number of transformations.

In this paper we extend the affine transformation to the multiple logit model. The well-known relation of the (simple linear) logit with the probit model is employed to introduce invariance through an affine transformation. This group of transformations is defined for the logit model and can act to the probit model due to the equivalence of the logit and probit model when $p \in [0.2,\ 0.8]$. In Section 2, we discuss the distance and other useful geometrical ideas, which are the main line of thought in a number of statistical problems.

The affine transformations point of view is briefly discussed, under the logit model, in Section 3. In Section 4, the probit and logit model are applied in a biometrical problem concerning breast cancer.

2 Geometrical Insight

In principle, there are statistical approaches, both in design theory and estimation, with a strong mathematical insight. There also cases where, despite the fact that the developed procedure "works well", there is lack of theoretical insight. This paper is based on these cases: It tries to offer a strong mathematical background for the proposed method.

Through the usual Euclidean distance, i.e. $d(x,y) = \Sigma_{i=1}^{n}(x_i - y_i)^2$, $x, y \in \mathrm{R}^n$, there are a number of other distance measures, such that

$$d_1(x,y) = e^{d(x,y)} - 1, \quad x, y \in \mathrm{R}^n, \quad \text{or}$$

$$d_2(x,y) = \min\{|x_i - y_i|,\ \ i = 1,\dots,n\}, \quad x = (x_i), y = (y_i) \in \mathrm{R}^n.$$

Example 1. The idea of the Euclidean distance is acting in statistical theory, not only in estimation; see the pioneering paper by Blyth, [5]. Recall the Latin Square of order $k \in \mathrm{N}$: By definition, it is a $k \in \mathrm{N}$ array in which each of the k symbols (usually Latin letters) occurs once in each row and once in each column. It can be proved that for $m \geq 2$, the $m \times m$ array is defined by

$$L(i, j) = i + j, \quad i, j \in \mathrm{Z}_m,$$

with $\mathbb{Z}_m$ being the set of integers modulo m, is a Latin Square.

For a $k \times k$ Latin Square, recall the main part of an ANOVA Table when the Analysis of Variance (or Regression Analysis) is adopted, see [8] among others:

Source of Variation	*df*	SS
Rows	$k-1$	$\left\Vert \bar{y}_{\bullet j \bullet} - \bar{y}_{\bullet\bullet\bullet} \right\Vert^2$
Columns	$k-1$	$\left\Vert \bar{y}_{\bullet\bullet l} - \bar{y}_{\bullet\bullet\bullet} \right\Vert^2$
Treatment	$k-1$	$\left\Vert \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet} \right\Vert^2$
Error	$\left\Vert \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet} \right\Vert^2$	by subtraction
Total	$k^2 - 1$	$\left\Vert y - \bar{y}_{\bullet\bullet\bullet} \right\Vert^2$

df: degrees of freedom, SS*: sum of squares*

Table 1.1: ANOVA (main part).

In fact, the sum of squares offers a very illustrating example of the extension of the Pythagorean Theorem.

Topology is a more “primitive” concept than the one of distance. The topology introduced by $d(x, y)$ on R^n is the same with that of $d_1(x, y)$, $d_2(x, y)$, i.e. the topology, in general, is not dependent on the form of the defined distance on R^n.

Example 2. Recall that, in binary problems as in [22], the relative risk is actually working as a $d_1(x, y)$ distance.

Now, the idea is to think of a set of points, say S, locally resembling R^n. This is behind when we say that the set S is a manifold:

(i) each point of which, say *P*, can have an open neighborhood,

(ii) which has a continuous 1-1 and onto map, say *f*, of an open set of R^n (for some *n*).

We emphasize that the map *f* needs to be 1-1; it is not required for the lengths or the angles to be preserved, but it only states that $f(U) \subseteq \mathrm{R}^n$ for the manifold *M*. The pair (U, f) is known as a chart. So, we only express a practical need: to present the set S "like" R^n, in terms of a co-ordinate chart mapping. For any *P* of S we define (U, f) and $f(P)$ is considered with coordinates $\left(x_1(P), x_2(P), \ldots, x_n(P)\right)$, see [19].

Given two charts (U, f), (V, g) and (V, g) then: $g \circ f : \mathrm{R}^n \to \mathrm{R}^n$ with

$$(x_1, x_2, \ldots, x_n) \mapsto (y_1, y_2, \ldots, y_n), \quad y_i = y_i(x_1, \ldots, x_n), \quad i = 1, \ldots, n.$$

Example 3. An affine space L in $\mathbb{R}^2$ is defined by the mean value of two elements x_1, x_2. Any two points such that $x_1 + x_2 = 2\overline{x}$ or $x_2 = 2\overline{x} - x_1$ define the line L_1. If we add a constant h to both values we are moved to line L_2 as

$$\tfrac{1}{2}[(\mathrm{x}_1 + h) + (x_2 + h)] = \overline{x} + h\,.$$

Notice that the parallel line to orbit L_1 passing through the origin O is a vector space, while $L_1, L_2, \ldots$ are not see Fig. 2.1. For the orbit L_2 it holds $L_2 = L_1 + \text{const.} \times L_1$.

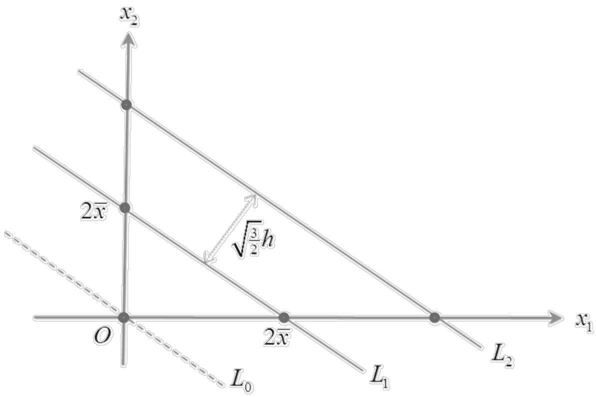

Figure 2.1: L_0 is a vector space, L_1 represents $\frac{1}{2}(x_1 + x_2) = \overline{x}$, L_2 represents $L_1 + \sqrt{3/2}h$.

An orbit is defined from the so called "group of transformations", say G discussed in Section 3, and the affine character described above is discussed by Fraser in [7], while for the Logit model see Kitsos (2006, 2011).

3 Invariant Statistical Applications

Statistical invariance has attracted a lot of interest in bibliography; see the pioneering book by Lehman [20]. Among the statistical applications with no mathematical justification, there is a very important one. Haines et al. in [9] consider the logistic dose response model for the proportion of "successes" p:

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \quad x_1, x_2 \geq 0.$$

This application is one of the few considering an extension of the simple logit model. To deal with the difficulties of the chosen model, see [9], they simply replaced $\beta_i d_i$, $i = 1,2$ with z_i, $i = 1,2$ and they came across known results, evaluating the Fisher's information,

$$\mathrm{I}(\theta) = \mathbf{v}\mathbf{v}^{\mathrm{T}}, \text{ with } \mathbf{v} = \left[\frac{\exp(\theta\mathbf{z}^{\mathrm{T}})}{1+\exp(\theta\mathbf{z}^{\mathrm{T}})}\right]^{1/2}\mathbf{z}, \tag{1}$$

and $\boldsymbol{\theta} = (\beta_0, 1, 1)$, $\mathbf{z} = (1, z_1, z_2)$. The relation (1) is in line with the theoretical assumption: for the logit and probit model, there is always a vector *w* such that $\mathrm{I}(\theta) = \mathbf{w}\mathbf{w}^{\mathrm{T}}$, see [16]. The transformation of the vector $\boldsymbol{\beta}$ to vector $\boldsymbol{\theta}$ and $\mathbf{x} = (1, x_1, x_2)$ to $\mathbf{z} = (1, z_1, z_2)$ has been the idea to work with. Here are the main steps. For the logit transformation of the proportion of successes $p(x)$, considered as a function of the input vector x, of the form

$$y = \log\left\{\frac{p(x)}{1-p(x)}\right\} = \beta_1 + \beta_1 x_2 + \ldots + \beta_p x_p + \sigma e, \tag{2}$$

$e \sim \mathcal{N}(0,1)$, the group (under matrix multiplication) of the affine transformations *G* is formed as, Kitsos (2011),

$$G = \left\{ g = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ \beta_1 & \beta_2 & \ldots & \beta_p & \sigma \end{pmatrix} = \begin{pmatrix} \mathbf{I}_p & \mathbf{O}^{\mathrm{T}} \\ \boldsymbol{\beta} & \sigma \end{pmatrix}, \; g \in \mathrm{R}^{p+1} \right\}, \tag{3}$$

where we denote

$$\mathbf{I}_p = \mathrm{diag}(1,\ldots,1) \in \mathrm{R}^{p\times p} \text{ and } \mathbf{O} = (0,\ldots,0) \in \mathrm{R}^p, \quad \sigma \in \mathrm{R}^+.$$

It can be proved that $\exists g^{-1}:\ gg^{-1} = g^{-1}g = \mathrm{id.}$, with

$$g^{-1} = \begin{pmatrix} \mathbb{I}_p & \mathbb{O}^{\mathrm{T}} \\ \sigma^{-1}\boldsymbol{\beta} & \sigma \end{pmatrix}. \tag{4}$$

We need a transformation of model (2) to move on to a set where affine transformations are possible. Therefore (2) can be written technically, and this is the key idea, as

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \\ e \end{pmatrix} = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ \beta_1 & \beta_2 & \ldots & \beta_p & \sigma \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_p' \\ e' \end{pmatrix}, \text{ or} \tag{5}$$

$$\Psi = gE, \tag{6}$$

with $g \in G$ and the definition of Ψ and E being obvious. Therefore, the orbit $G\psi$ is

$$G\psi = \left\{ \sum_{i=1}^{p} \beta_i x_i + \sigma e, \quad \sigma \in \mathrm{R}^+ \right\}.$$

From the affine geometry it is known that for “another model” ω (practically another hyperplane), it holds

$$G\psi \cap G\omega = \varnothing, \text{ or } G\psi = G\omega. \tag{7}$$

Hyperplanes $G\psi$ and $G\omega$ are either identical or disjoint (we are referring to a completely different experimental situation). Notice that, as neither $\beta_i \geq 0$, $i = 1, 2, \ldots, p$, nor $\sum_{i=1}^{p} \beta_i = 1$, the set $G\psi$ cannot be considered as an affine simplex, but expands a p-dimensional hyperplane.

Under this theoretical framework, the following can be proved, see [16].

Theorem 3.1. *Consider any positive definite matrix* $\mathbf{M}$ *with* $\det \mathbf{M} \neq 0$ *and a vector* $\mathbf{c}$ *with appropriate dimensions such that* $\mathbf{N} := \mathbf{c}^{+}\mathbf{M}^{-1}\mathbf{c}$ *is valid. Then* $\mathbf{N}$ *remains invariant if* $\mathbf{c}$ *and* $\mathbf{M}$ *transformed under g.*

Corollary 3.1. *If* $\mathbf{M} = \mathbf{M}(\theta, \xi)$ *is the average per observation information matrix, with* ξ *being a design measure, then* $\mathbf{N} = \mathbf{N}(\theta, \xi)$ *is the transformed form of* $\mathbf{M} = \mathbf{M}(\theta, \xi)$ *Moreover, if* $p = 2$*, the co-optimal design remains invariant under g.*

The above Theorem 3.1 provides evidence that the D-optimal design criteria remain invariant under g transformation. Still we need to prove that $g \in G$ can be transformed, so that the parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ can be $\boldsymbol{\theta} = (\beta_1, 1, 1, \ldots, 1)$.

Theorem 3.2. *There is a transformation* $g \in G$ *as in (2.3), with element* $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ *with parameter vector equal to* $\boldsymbol{\theta} = (\beta_1, 1, 1, \ldots, 1)$.

Proof. See [16].

This result is a fundamental one, see [9], where the transformation was used arbitrarily with no further explanation. Moreover, this transformation to $\boldsymbol{\theta} = (\beta_1, 1, 1, \ldots, 1) \in \mathrm{R}^{p}$ provides evidence that there exists always the so-called "canonical form" of the logit model, see [6, 14, 16]. Now, as far as the relation between the logit and probit is concerned, the following discussion clarifies their equivalence. Therefore, the affine transformation in Theorems 3.1 and 3.2 holds also for the probit case. Indeed:

For the "linear predictor" $\eta_i = \eta(x_i) = \beta_0 + \beta_1 x_i$, $i = 1, 2, \ldots, n$, expressed through the Logistic distribution (and approximated with the corresponding Normal distribution), it holds

$$\begin{aligned} p_i = L(x_i) &= \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \\ \approx F(x_i) &= \tfrac{1}{\sqrt{2\pi \operatorname{var}}} \int_{-\infty}^{x_i} e^{-\frac{1}{2\operatorname{var}}(u-\mu)^2} du = \tfrac{1}{2} + \tfrac{1}{2} \operatorname{erf}\left(\frac{x_i - \mu}{\sqrt{2 \operatorname{var}}} \right), \end{aligned}$$

where L is the cumulative distribution function (c.d.f.) of the Logistic distribution $\mathcal{L}(\mu, \sigma)$ with mean $\mu = -\beta_0 / \beta_1$, scale parameter $\sigma = \beta_1^{-1}$ or variance $\tau = \frac{1}{3}(\pi\sigma)^2 = \frac{1}{3}\pi^2 \beta_1^{-2}$, F is the c.d.f. of the Normal distribution $\mathcal{N}(\mu, \tau)$, i.e. with same mean and variance as the Logistic distribution, and $\operatorname{erf}(.)$ being the known error function. Therefore, the relation between the logit and probit model is given by

$$\begin{aligned} x_i = L^{-1}(p_i) &= \frac{1}{\beta_1} \left(\log \frac{p_i}{1 - p_i} - \beta_0 \right) \\ \approx F^{-1}(p_i) &= \mu + \sqrt{2 \operatorname{var}} \operatorname{erf}^{-1}(2p_i - 1). \end{aligned}$$

Equivalently, using the Probit and Logit functions, we get

$$
\begin{aligned}
x_i &= \mathrm{Logit}\left(\tfrac{p_i-\mu}{\sigma}\right) = \frac{1}{\beta_1}\left[\mathrm{Logit}(p_i)-\beta_0\right] \\
&\approx F^{-1}(p_i) = \mathrm{Probit}\left(\tfrac{p_i-\mu}{\sqrt{\mathrm{var}}}\right) = \mu+\sqrt{\tau}\,\mathrm{Probit}(p_i).
\end{aligned} \tag{8}
$$

Figure 3.1 explains the number, as well as the kind of transformations involved in binary response data sets, to be approached by the canonical form. The flow chart of the procedure is the following:

(1) Data $\mapsto$ (2) Probability Model $\mapsto$ (3) Logit $\mapsto$
$\mapsto$ (4) Regression Estimates $\Leftrightarrow$ (5) Affine Model $\mapsto$
$\mapsto$ (6) Estimates (through the Affine Transformation) $\mapsto$
$\mapsto$ (7) Canonical Form of Logit $\Leftrightarrow$
$\Leftrightarrow$ Probit when $p \in [0.2,\ 0.8]$ $\mapsto$ Return to $x,\ \beta$ $\mapsto$ (2).

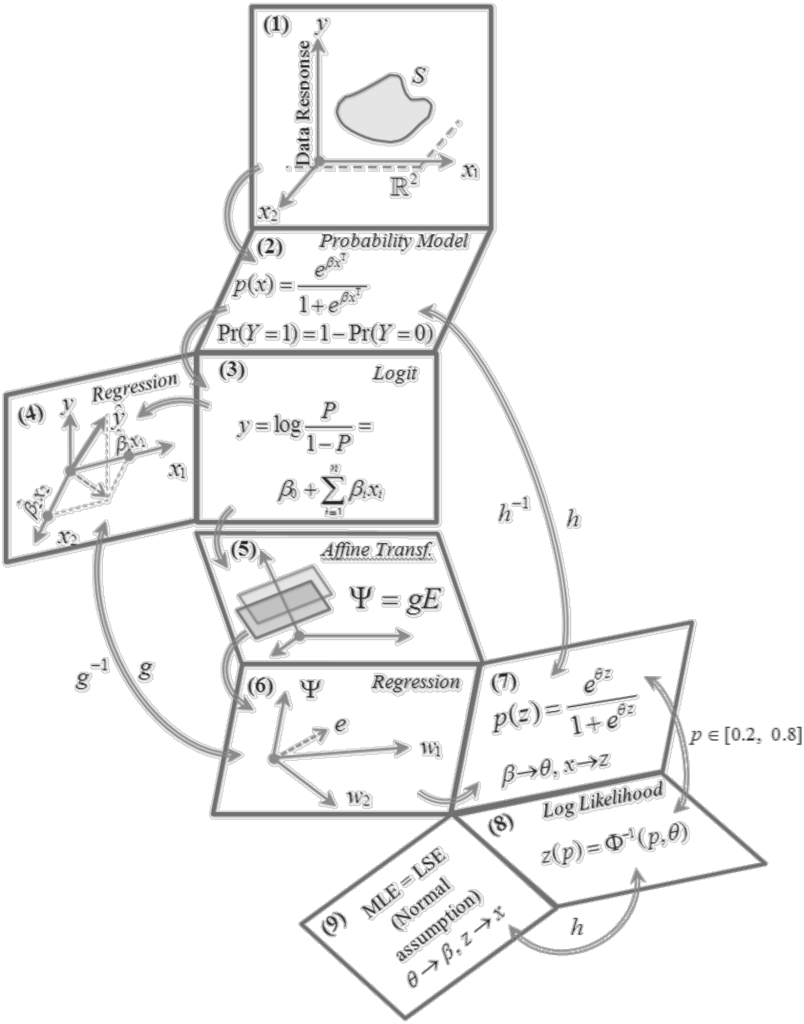

Figure 2.1: The transformation to Canonical Form of the Logit model.

4 Breast Cancer Application

Consider the data set and analysis discussed in [15] for the breast cancer problem. Table 4.1 and 4.2 summarizes, for the specific data set, the analysis of the one variable logit model, considering the binary response y, and the corresponding input variables, as appeared in the following Tables.

Variables	**β**	se(**β**)	exp(**β**)
Time (β_1)	-0.9925	0.4805	0.3707
Constant (β_0)	-0.1542	0.2782	

Table 4.1: The contribution of “Time”.

Variables	**β**	se(β)	exp(β)
Start (β_1)	-1.4016	0.4204	4.0615
Constant (β_0)	-0.7885	0.2412	

Table 4.2: The contribution of “Start”.

Applying (3) for the case of Table 4.1, the affine transformation G acts on vector $E = (1, x', e')^{\mathrm{T}}$ obtaining vector $\Psi = (1, x, e)^{\mathrm{T}}$, i.e.

$$\begin{pmatrix} 1 \\ x \\ e \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -0.1542 & -0.9925 & \sigma \end{pmatrix} \begin{pmatrix} 1 \\ x' \\ e' \end{pmatrix}, \text{ i.e.}$$

$$y = -0.1542 - 0.9925x' + \sigma e' .$$

Recall that the Logit and Probit functions are defined as the quantile functions Λ^{-1} and Φ^{-1} of the standardized Logistic $\mathcal{L}(0,1)$ and Normal $\mathcal{N}(0,1)$ distribution, i.e. $\mathrm{Logit}(p) := \Lambda^{-1}(p) =$

$\log\{p/(1-p)\}$, and $\text{Probit}(p) := \Phi^{-1}(p) = \sqrt{2}\,\text{erf}^{-1}(2p-1)$, $p \in [0,\ 1]$.

In Table 4.3 the logit values $x_k = L^{-1}(p_k)$ for $p_k = k/10 \in [0,\ 1]$, $k = 1, 2, \ldots, 9$ are presented where $L^{-1} = L^{-1}(p)$ is the quantile function of the Logistic distribution $\mathcal{L}(\mu, \sigma) = \mathcal{L}(-0.1554, 1.0076)$ for the random variable (r.v.) X describing "Time", i.e. with mean $\mu_X = \mu = -\beta_0/\beta_1 = 0.1554$ and scale parameter $\sigma = -\beta_1^{-1} = 1.0076$. We work with the estimates β_0 and β_1 from Table 4.1. For the variance σ_X^2, recall that $\sigma_X^2 = \frac{1}{3}(\pi\sigma)^2 = 3.3398$.

Table 4.3 also provides the probit values $z_k = F^{-1}(p_k)$, with $p_k \in [0,1]$ as above. We consider $F^{-1} = F^{-1}(p)$ as the quantile function of the usual Normal distribution $\mathcal{N}(-0.1554, \sigma^{*2})$ of the same mean variance as the Logistic one. Thus, we let r.v. $Z \sim \mathcal{N}(-0.1554, \sigma^{*2})$ with scale parameter σ^{*2} such that the variances of X and Z being equal, i.e. $\sigma_Z^2 = \sigma_X^2 = 3.3398$. Therefore, using the $\text{Probit}(.)$ notation, it holds

$$x_k = \frac{1}{\beta_1}\left[\text{Logit}(p_k) - \beta_0\right]$$

$$\approx z_k = \text{Probit}\left(\tfrac{p_k - \mu}{\sigma_\gamma}\right) = \mu + \sigma^{*2}\text{Probit}(p_k), \quad k = 1, 2, \ldots, 9.$$

p_i	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
x_k	-2.369	-1.552	-1.009	-0.564	-0.155	0.253	0.698	1.241	2.058
z_k	-2.497	-1.693	-1.114	-0.618	-0.155	0.308	0.803	2.187	2.187

Table 4.3: Logit values $x_k = L^{-1}(p_k)$, and probit values $z_k = F^{-1}(p_k)$ for the contribution of "Time".

Recall that the mean absolute relative error, or MARE (%), between logit and probit values, is defined as

$$\mathrm{MARE} = \tfrac{100}{9}\sum_{k=1}^{9}\left|\frac{x_k - z_k}{x_k}\right|, \quad k = 1, 2, \ldots, 9\,,$$

For the r.v. “Time” it has been evaluated MARE = 9.85% which is fairly good, while for the r.v. “Start” the corresponding MARE = 27.2%.

Table 4.4 present the logit/probit estimations for the r.v. describing “Start” where the corresponding estimates are in Table 4.2.

p_i	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
x_k	-2.130	-1.552	-1.167	-0.852	-0.563	-0.273	0.042	0.426	1.005
z_k	-2.221	-1.652	-1.241	-0.890	-0.563	-0.235	0.116	0.527	1.096

Table 4.3: Logit values $x_k = L^{-1}(p_k)$, and probit values $z_k = F^{-1}(p_k)$ for the contribution of “Start”.

Finally, the fact that the probit model provides a “good” fitting for the logit model of “Time” and “Start” is also supported by the higher R^2 coefficient, given by,

$$R^2 = 1 - \frac{\sum_{k=1}^{9}(x_k - z_k)^2}{\sum_{k=1}^{9}(x_k - \overline{x}_k)^2}\,,$$

which measures the goodness-of-fit between the logit x_k and the probit z_k values. For random variables “Time” and “Start” we obtain both $R^2_{\text{Time}} = R^2_{\text{Start}} = 0.9935$, i.e. very close to 1.

In this paper we discussed the invariance of a group of transformations for the logit and probit model. The empirical estimations validate the theoretical background. The above discussion may be extended, through a generalized form of the common probit function, which is part of our future work.

Acknowledgments

We would like to thank the referees for their useful comments, which improve the final form of this paper.

Bibliography

1. Aranda-Ordaz RD (1981): On two families of transformations to additivity for binary response data. *Biometrika* 68: 357-364.
2. Bechger TM, Verhelst ND, Verstrales HHFM (2001): Identifiability of nonlinear logistic test models. *Psyciometrika* 68(3): 357-372.
3. Berkson J (1955): Maximum likelihood and minimum χ^2 estimates of the logistic function. *Journal of American Stat. Association* 50: 130-162.
4. Bliss CI (1935): The calculation of the dosage mortality curve. *Annals of Applied Biology* 22(1): 134-167.
5. Blyth CR (1970): On the inference and decision models of Statistics. *Ann. Math. Stat*. 41: 1034-1058.
6. Ford I, Torsney B, Wu CFJ (1992): The use of canonical form in the construction of locally optimal designs for non-linear problems. *JRSS*, 54(2): 569-583.
7. Fraser DAS (1968): The Structure of Inference. *John Wiley*.
8. Graybill FA (1976): Theory and Applications of the Linear Model. *Duxbury*.
9. Haines ML, Kabera G, Ndlovu P, O'Brien ET (2007): D-optimal designs for logistic regression in two variables. In: Diaz JM, Torsney B (Hrsg.): Advances in Model-Oriented Design and Analysis (MODA8); *Physica-Verlag*, pp. 91-98.

10. Hosmer DW, Lemeshow S (1989): Applied Logistic Regression. *John Wiley, NY*.
11. Jaynes ET (1957): Information Theory and Statistical Mechanics. *Phys. Rev*. 106: 620–630.
12. Kitsos CP (1999): Optimal designs for estimating the percentiles of the risk in multistage models in carcinogenesis. *Biometrical Journal* 41(1): 33-43.
13. Kitsos CP (2005): Optimal designs for bioassays in carcinogenesis. In: Edler L, Kitsos CP (Hrsg.): Quantitative Methods for Cancer and Human Health Risk Assessment; John Wiley, England, pp. 267-279.
14. Kitsos CP (2006): On the logit methods for Ca problems. In: Statistical Methods for Biomedical and Technical Systems, Limassol, Cyprus; pp. 335-340.
15. Kitsos CP (2010): Estimating the relative risk for breast cancer. *Biometrical Letters* 47(2): 133-146.
16. Kitsos CP (2011): Invariant canonical form for the multiple logistic regression. *Mathematics in Engineering, Science and Aerospace* (MESA) 38(1): 267-275.
17. Kitsos CP (2014): Statistical Invariance due to affine transformations. To appear.
18. Kitsos CP, Toulias TL (2014): A min-max entropy algorithm for relative risk problems. To appear.
19. Kobayashi S, Nomizu K (1963): Fundations of Differential Geometry, Volume I. *John Wiley*.
20. Lehmann EP, Romano JP (2005): Testing Statistical Hypotheses (3rd Edition). *Springer*.
21. Muller CH, Kitsos CP (2004): Optimal design criteria based on tolerance regions. In: Di Bucchianno A, Lauter H, Wynn H (Hrsg.): Advances in Model-Oriented Design and Analysis (MODA7); *Physica-Verlag*, pp. 107-115.
22. Rigas AG, Kotti VK (2005): Logistic regrassion methods and their implementation. In: Edler L, Kitsos CP (Hrsg.): Quantitative Methods for Cancer and Human Health Risk Assessment; John Wiley, England, pp. 355-369.