**Logistic Regression**

**1. Origins of Logistic Regression**

The logistic function was developed in the nineteenth century to describe population growth and the behaviour of autocatalytic chemical processes. Now, the growth rate of a quantity *W* (*t*) over time is given by:

*Ẇ* (*t*) = d*W* (*t*) / d*t*  (1.1)

Assuming that *Ẇ* (*t*) is proportional to *W* (*t*) we get:

*Ẇ* (*t*) = *β W* (*t*), *β* = *Ẇ* (*t*) / *W* (*t*) (1.2) where *β* refers to the constant rate of growth, thus leading to the exponential growth model

*W* (*t*) = *A* e*β t* (1.3)

where the initial value *W* (0) may replace *A*.

However, the exponential growth model is devoid of any upper limit. This problem was approached by the Belgian astronomer and statistician Alphonse Quetelet (1795-1874) and his student Pierre-Francois Verhulst (1804-1849) through the inclusion of an additional term, representing the increasing resistance to further growth, in equation (2) as follows:

*Ẇ* (*t*) = *β W* (*t*) − *φ* (*W* (*t*)) (1.4) Experimentation with varied forms of *φ* led to the following model when *φ* is a quadratic function:

*Ẇ* (*t*) = *β W* (*t*) (Ω − (*W* (*t*)) (1.5)

where Ω refers to the upper limit or saturation level of *W*.

In the above model, *Ẇ* (*t*) is proportional to both *W* (*t*) and (Ω − (*W* (*t*)). Expressing *W* (*t*) as a proportion {*P*(*t*)} of Ω, or *P*(*t*) = {*W* (*t*) / Ω}, results in the following differential equation:

*P*(*t*) = *β P*(*t*) {1− *P*(*t*)} (1.6)

the solution of which is given by

Equation (7) was named as the ***logistic*** function by Verhulst. In regression analysis, α and *β*  may be interpreted as the intercept and the regression coefficient (or the slope of the regression line), respectively.

The logistic function was rediscovered in 1920 by Raymond Pearl (1879-1940), the Director of the Department of Biometry and Vital Statistics at John Hopkins University, and his deputy Lowell J. Reed (1886-1966) while studying the United States’ population increase (Cramer, 2002).

**2. The Logic of Logistic Regression**

Discrete or qualitative rather than continuous or quantitative events are common in many social phenomena; for example, an event happens or does not happen, a person’s life can be changed in a variety of ways that involve a characteristic, event, or choice, or large social entities such as groups, organizations and nations can arise or disintegrate, become insolvent, confront, revolt, and so on. A dichotomous indicator or dummy variable is the most common way to represent discrete binary phenomena (Pampel, 2000).

**2.1 Linear Regression with Dichotomous Dependent Variable**

On the surface, a binary qualitative dependent variable seems to be appropriate for use in multiple linear regression. The dependent variable assumes only two values of zero and one. However, the estimated values for regression are in the form of mean proportions or probabilities conditional on the values of the independent variables. The regression coefficients may be interpreted as the increase or reduction in the estimated probability of possessing a characteristic or experiencing an event due to a unit change in the independent variables. However, such linear regression faces some problems.

**2.1.1 Conceptual problems**

* Probabilities and proportions, by definition, cannot exceed one (ceiling) or fall below zero (floor). Nonetheless, the linear regression line may stretch upward into positive infinity (or extend downward towards negative infinity) as the values of the independent variables increase (or decrease). A model may provide illogical and useless predicted values of the dependent variable greater than one and less than zero based on the slope of the regression line and the observed values of the independent variables.
* Typically, linear regression is based on the assumption of additivity, which states that the influence of one independent variable on the dependent variable remains constant regardless of the values of the other independent variables. Although models may include chosen product terms to accommodate non-additivity, a dummy dependent variable will almost certainly break the additivity assumption for all possible combinations of the independent variables. When the value of one independent variable reaches a level sufficient to push the probability of the dependent variable close to one (or close to zero), the impacts of other variables have little effect. Thus, the ceiling and floor impose an intrinsic, non-additive and interactive nature on the impact of all independent variables (Pampel, 2000).

**2.1.2 Statistical problems**

* Linear regression with dichotomous dependent variable violates the normality assumption (each value of the independent variables in the population is associated with a normal distribution of error terms around the predicted value of the dependent variable) and the homoscedasticity assumption (the dispersion of the error terms for each value of the independent variables is similar) (Pampel, 2000).

**2.2 Transformation of probabilities into logits**

The relationship between the probability of a dummy dependent variable and an independent variable is inherently non-linear. A constantly changing curve, such as the S-shaped curve, represents the relationship more smoothly and adequately than a straight line. Although other non-linear functions may depict the S-shaped curve, the logistic or logit transformation has gained popularity due to its desirable characteristics and relative simplicity.

Let *Yi* be the observations (i=1,2,3,...,n) of a dichotomous dependent variable (*Y*) representing some event and assuming the values of one (occurrence of the event) and zero (non-occurrence of the event) only. Given the probability of the occurrence of the event as *P* (*Yi* = 1) = *P*i , the probability of the non-occurrence of the event will be *P* (*Yi* = 0) = (1 − *Pi* ). The odds of the probability of the occurrence of the event relative to the probability of the non-occurrence of the event is then given by:

*O*i  = [ *P*i / (1 − *Pi* ) ] (2.1)

If *Pi* = 0, *Oi*  = 0 ; if 0 < *Pi*  < 1, 0 < *Oi*  < ∞ ; if *Pi*  = 1, *Oi* = ∞. So 0 ≤ *Oi*  ≤ ∞.

Odds are generally expressed implicitly as a ratio to one or as a single number. For instance, if the probability of an event equals 0.4, the odds are (0.4 / 0.6) or 0.667, indicating that the event occurs 0.667 times for each time it does not occur, or 667 occurrences per 1000 non-occurrences. Even though both probabilities and odds have a lower bound (floor) of zero, denoting the increasing likelihood of an event with increasingly large positive numbers, odds have no upper bound (ceiling), unlike probabilities. The transformation of probabilities into odds eliminates the probabilities’ ceiling value of one.

Now, the logit or logged odds is formed by taking the natural logarithm of the odds to eliminate the odds’ and hence probabilities’ floor value of zero, as:

*Li* = ln *Oi* = ln [ *Pi*/ (1 − *Pi* ) ] (2.2)

If *Oi* = 0 and *Li* = ∞; if 0 < *Oi* < 1, *Li* < 0; if *Oi* = 1, *Li* = 0; if 1 < *Oi*  ≤ ∞, 0 < *Li* ≤ ∞.   
So − ∞ ≤ *Li*  ≤ + ∞ (Pampel, 2000).

**2.3 Linearising the Non-linear**

The non-linear relationship between the independent variable (*X*) and the probability of the occurrence of the dichotomous dependent variable (*Y*) conditional on the observed values of *X* (*Xi*) expressed as *P* (*Xi*) [= *P* (*Yi* = 1 | *X* = *Xi* )],may be transformed into a linear regression function based on the logit of *P* (*Xi*) from equation (2.2) as:

*L* [ *P* (*Xi*) ] = ln [ *P* (*Xi*) / {1 − *P* (*Xi*) }] = *α* + *βXi*  (2.3)

where *α* and *β* denote the intercept and the regression coefficient (or the slope of the linear regression line), respectively.

Equating equations (1.7) and (2.3) and replacing *t* with *Xi,* we get:

(Pampel, 2000)

**3. Interpreting logistic regression coefficients**

The sign of *β* in equations (2.3) and (2.4) determines whether *P* (*X*)is increasing or decreasing as *X* increases. The rate of ascent or descent increases as the magnitude of *β*, that is | *β* |, increases; as *β* approaches zero, the curve flattens to a horizontal straight line. When *β* equals zero, *Y* is independent of *X*. For quantitative *X* with *β* greater than zero, the curve for *P* (*X*)has the shape of the cumulative distribution function of the logistic distribution, given by:

where *μ* and *τ* are the mean and standard deviation respectively of the logistic distribution.

Due to the symmetrical distribution of the logistic density, *P* (*X*)approaches one at the same rate that it approaches zero (Agresti, 2002). Multiple interpretations exist for *β* in terms of logged odds, odds and probabilities and the nature of the independent variable being continuous or dummy (Pampel, 2000).

**3.1 Logged Odds**

In the case of continuous independent variables, the logistic regression coefficients indicate the change in the projected logged odds of the occurrence of an event for one unit change in the independent variables. In contrast, an implicit comparison of the indicator group with the reference or excluded group is made by a one-unit change in the case of dummy independent variables. Browne (1997, p. 246), for example, uses logistic regression to forecast labor force participation for 922 female heads of home between the ages of 18 and 54 in 1989. For the continuous independent variable ‘Years employed’, the logistic regression coefficient of 0.13 indicates an increase in the logged odds of the dependent variable ‘Labor Force Participation’ by 0.13 for an additional year of employment. The author also compares the ‘Labor Force Participation’ with two dummy variables, namely ‘High school dropout’ and ‘High school graduate’ to that of the reference group consisting of women with some college education. These two dummy variables have correlations of -1.29 and -0.68, respectively, indicating that the logged odds of being in the labor force are 1.29 lower for high school dropouts than for those with some college and 0.68 lower for high school graduates than for those with some college education (Pampel, 2000).

**3.2 Odds**

Let us consider the following binary logistic regression model with two independent variables, *X*1 and *X*2 :

ln (*P* / 1 − *P*) = α + *β*1 *X*1 + *β*2 *X*2 (3.1)

where αis the intercept term, and *β*1 and *β*2 are the regression coefficients.

Exponentiating both sides we get:

(*P* / 1 − *P*) = (3.2)

The equation (3.2) determining the odds is multiplicative, where the predicted value of the dependent variable does not change when multiplied by a coefficient of one. Therefore, the effect of each independent variable on the odds can be measured by taking the antilog of the regression coefficients.

The percentage change in the odds for a one-unit change in the independent variable (with regression coefficient ‘*β‘*) which is given by:

% Δ = (e *β* – 1) \* 100 (3.3)

presents a more meaningful interpretation of the regression coefficients.

With reference to Browne's study mentioned above, the exponentiated coefficient for the:

* continuous independent variable ‘Years employed’, *e*0.13 or 1.14 indicates that a one-year increase in employment multiplies the odds of the dependent variable ‘Labor Force Participation’ by 1.14 or increases the odds by a factor of 1.14 or 14%;
* dummy variable ‘High school dropout’, *e*−1.29 or 0.28, indicates that a one-unit increase in the variable multiplies the odds of ‘Labor Force Participation’ by 0.28 or the odds are 0.28 times or 72% smaller than those with some college education (reference group);

dummy variable ‘High school graduate’, *e*−0.68 or 0.51, indicates that the odds of ‘Labor Force Participation’ are 0.51 times or 49% lower than those with some college education (Pampel, 2000).

**3.3 Probabilities**

Due to the non-linear and non-additive nature of the interactions between the independent variables and probabilities, they cannot be adequately described by a single coefficient. Instead, the influence on probability distributions must be determined for a given value or group of values, the choice of which is determined by the researcher’s concerns and the nature of the data (Pampel, 2000).

**3.3.1 Continuous Independent Variable**

Calculating the linear slope of the tangent of a non-linear curve at every single point is a simple approach to determine the effect of a continuous independent variable on probability. The slope of the tangent line is given by the partial derivative of the non-linear equation linking the independent variables to the probabilities. The partial derivative, which measures the change in the probability for an infinitesimally small change in *Xk* (*k* = 1,2)and defines the slope of the tangent line or the change in the tangent line due to a one-unit change in *Xk* at that value, is obtained from equation (3.1) as:

*( ∂ P* / ∂ *Xk* ) = *β k* \* *P* \* (1 − *P*) (3.4)

Considering Browne’s study mentioned above, the logistic regression coefficient for ‘Years employed’ equals 0.13; the mean of the dependent variable, the expected probability of ‘Labor Force Participation’ equals 0.83; and the probability of not participating equals 0.17. From equation (3.4), the partial derivative is given by (0.13 \* 0.83 \* 0.17) or 0.18, which indicates that an increase of one year of employment increases the probability of participation by 0.018 or almost 2% at the mean. The marginal effect reaches its maximum value of 0.032 when *P* = 0*.*5.

Long (1997) discusses several alternative methods to present a more complete summary of the range of effects of an independent variable on probabilities (Pampel, 2000). However, disagreements exist among analysts on the usefulness of even calculating a single partial derivative given the constraints of describing a non-linear and non-additive relationship with a single coefficient (DeMaris et al., 1990; DeMaris, 1993; Roncek, 1993).

**3.3.2 Dummy Independent Variable**

In the case of dummy independent variables, the tangent line for infinitesimally small changes makes little sense. However, it is possible to compute expected probabilities for each of the two groups and then measure the group differences in probabilities by subtracting the two probabilities. The steps involve the following calculations:

* Logit for the omitted group: *Lo* = ln (*Po* / (1 − *Po*)) (3.5)
* Logit for the dummy variable group: *Ld* = *L*o + *βd*  (3.6)

where *βd*  is the logistic regression coefficient

* Probability for the dummy variable group: *Pd*  = {1 / (1 + )} (3.7)
* The difference in probabilities: *Pd − Po*  (3.8)

With reference to Browne’s study mentioned earlier, the above calculations using the mean or expected probability of the dependent variable ‘Labor Force Participation’ (*Po* = 0.83), regression coefficient of ‘High school dropout’ (*bd* = −1.29), and women with some college education as the omitted group, are shown below:

* Logit for women with some college education: *Lo* = ln (0.83 / 0.17) = 1.586
* Logit for ‘High school dropout’: *Ld* = 1.586 −1.29 = 0.296;
* Probability for ‘High school dropout’: *Pd*  = {1 / (1 + 0.7438)}= 0.573;
* Difference in probabilities: *Pd − Po*  = 0.573 − 0.83 = 0.257

Thus, high school dropouts have a probability of participating which is 0.257 lower than those with some college education (Pampel, 2000).

**3.3.3 Predicted Probabilities for Continuous Independent Variables**

Predicted probabilities may be used as the partial derivative for continuous variables, just as they can be used for dummy variables. However, changes in projected probability reflect the real impact of a discrete change in the independent variable *X*, for example, one unit, rather than the influence on the tangent line suggested by an instantaneous or infinitesimally small change in *X* (Pampel, 2000). As a result, some favour using projected probabilities over the partial derivative when dealing with discrete changes (Kaufman, 1996; Long, 1997). The predicted probabilities are calculated using the same procedures for dummy variables, except that *X* is substituted for the omitted group and (*X* + 1) for the dummy variable group. The following sequential steps may be followed:

* Calculating Logged odds of *P* (i.e., the logit before the change in X)
* Adding *X*’s logistic regression coefficient to the starting logit and computing the probability for the new logit
* Subtracting the starting probability (at *X)* from the second probability (at *X* + 1) shows the effect of a one-unit change in *X* on the predicted probability at *P*

Considering the above-mentioned Browne’s study, we see *P* = 0.83, the coefficient for ‘Years employed’ = 0.13, the logit at *P* = ln (0.83 / 0.17) or 1.586. Adding the coefficient to this logit yields 1.716 (1.586 + 0.13). From equation (3.7),the probability for (Years employed + 1) is calculated as 0.848. The difference between 0.848 and 0.83 equals 0.018 which indicates that a one-year increase in ‘Years employed’ increases the probability of ‘Labor Force Participation’ by 0.018 at it’s mean (Pampel, 2000).

**4. Estimation and Model fit of Binary Logistic Regression Model**

**4.1** **Estimation of Binary Logistic Regression Model**

**4.1.1 Maximum Likelihood Estimation (MLE)**

Due to the binary nature of the dependent variable, the error term does neither have a normal distribution nor equal variances for the values of the independent variables. As a result, the estimating approach derived from the Ordinary Least Squares (OLS) criterion, which involves minimizing the sum of the squared deviations between the observed and predicted values of the dependent variables, fails to produce efficient estimates. Therefore, rather than using OLS, logistic regression depends on maximum likelihood algorithms to estimate the coefficients.

For logistic regression, the estimation of the regression coefficients begins with an expression for the likelihood of observing the pattern of occurrences (*Y* = 1) and non-occurrences (*Y* = 0) of an event or characteristic in a given sample. This expression, termed the likelihood function, depends on unknown logistic regression parameters. Maximum Likelihood Estimation (MLE) finds the model parameters that provide the maximum value for the likelihood function, thereby identifying the estimates for model parameters that are most likely to give rise to the pattern of observations in the sample data. The maximum likelihood function in logistic regression is given by:

(4.1) where *LF* refers to the likelihood function; *Yi* refers to the observed value of the binary dependent variable for case *i* ; *Pi,* whichrefers to the predicted probability for case *i* is given by *Pi* = { 1 / (1 + ) } where *Li* is the logged odds determined by the unknown regression coefficient *β* and the independent variables; Π refers to the multiplicative equivalent of the summation sign (Σ) meaning that *LF* multiplies the values for each case. Thus, the aim is to identify *β* values producing  *Li* and *Pi* values that maximize *LF.*

To avoid the problem of dealing with tiny numbers due to the multiplication of probabilities, the likelihood function can be converted into a logged likelihood function by taking the natural logarithm of both sides of equation (4.1) and simplifying as follows:

(4.2)

In practice, MLE aims to find those *β* values that have the greatest likelihood of maximizing the log-likelihood function (Pampel, 2000).

Since the likelihood equations are usually non-linear in *β*, general-purpose iterative methods, such as Newton-Raphson and Fisher Scoring methods, are used for estimating *β*.

**4.1.1.1 Newton-Raphson Method (NRM)**

The Newton-Raphson method is an iterative procedure for solving non-linear equations, such as those whose solution defines the maximum point of a function. Starting with a guess about the answer, it obtains a second estimate by approximating the function to be maximized by a second-degree polynomial in the region of the initial guess and then determining the location of the polynomial’s maximum value. Then it uses another second-degree polynomial to estimate the function in the vicinity of the second guess. The third guess is the position of the function’s maximum. The approach creates a succession of guesses in this fashion. When the function is appropriate and the original estimate is reasonable, they converge to the position of the maximum. To determine the value of at which the function *L***(β)** is to be maximised, let **u***′* **=** (∂*L***(β)**, (∂*L***(β)**,…); **H**, called the *Hessian matrix*, denote the matrix comprising the entries = (∂2*L*(**β**)∂) and is also known as the *observed information matrix*; **β** ( *t* ) , **u** ( *t* )  and **H (** *t* ) be **β, u** and **H** at step *t* in the iterative process (*t* = 0,1,2,…). The following relationship, assuming that **H (** *t* ) is non-singular, is obtained:

**β** ( *t* + 1 ) = **β** ( *t* ) − **u** ( *t* ) (4.3)

(Agresti, 2002, p. 143)

In the case of a logistic regression model with binary dependent variable *Y* and one independent variable *X*, as explained in section 2.2, **β, u** and **H** are given by:

, **u** = , **H** =

**4.1.1.2 Fisher Scoring Method (FSM)**

The Fisher Scoring Method (FSM), resembling theNewton-Raphson method (NRM), is an alternative iterative method for solving likelihood equations. The distinction between the two methods is concerned with the *Hessian matrix*. The NRM uses the *Hessian matrix* itself, also called the *observed information matrix*. In contrast, the FSM uses the *expected* v*alue* of this matrix, called the *expected information* *matrix*.

Let refer to the approximation *t* for the maximum likelihood estimate of the expected information matrix, that is, has elements −, evaluated at **β** ( *t* ) . The formula for FSM is given by:

**β** ( *t* + 1 ) = **β** ( *t* ) − **u** ( *t* ) (4.4)

(Agresti, 2002, p. 145-146)

**4.2** **Model fit of Binary Logistic Regression Model**

The various approaches to test the Goodness of Fit (GoF) of a binary logistic regression model are discussed below:

**(A) Information Criteria Test (ICT).** This is a single statistic for comparing models; better-fitted models have lower values of the same information criterion (Hilbe, 2015). The two widely used ICTs are:

Akaike Information Criterion (AIC) test defined by the statistic:

AIC = −2 ln *L* + 2*p* (4.5)

Bayesian Information Criterion (BIC) or Schwarz Criterion (SC) defined the test statistic:

BIC (or SC) = −2 ln *L* + *p* ln *n* (4.6)

where *L* is the likelihood function, *p* is the number of parameters, and n is the sample size.

**(B) Deviance and Pearson .** In this approach, the data represents the fit of the ideal model possible- the saturated model having a separate parameter for each observation. It is tested whether allparameters that are in the saturated model but not in the estimated model equal zero.

Deviance: D = 2 (4.7)

Pearson : (4.8)

= observed counts

= counts in the estimated model

D and follow chi-square distribution with degrees of freedom equal to the difference between the number of profiles (parameters estimated in the saturated model) and estimated parameters.

However, if the logistic regression model satisfies either of the following conditions:

* includes continuous variables
* includes many independent variables
* independent variables have a considerable number of categories

Deviance and Pearson do not follow distribution.

In that case, the **Hosmer-Lemeshow (HL)** test is recommended.

The HL statistic is calculated as follows:

* Estimating the individual probabilities
* Sorting within response and dividing the set into ten groups with a similar number of observations (*k* =1,2,3,…,10)
* Using the groups to calculate expected counts and compare with observed counts using Pearson statistic, which approximately follows distribution with *v*=10−2=8 degrees of freedom

**(C) Pseudo-R2.** Several other GoF statistics have been suggested as analogues of the R2 statistic often used in linear regression. In these formulae, L0 is the likelihood of regression with only an intercept, L1 is the likelihood of the model estimated, and n is the sample size.

* **McFadden R2 :**  (4.9)
* **Cox-Snell R2 :**  (4.10)

The drawback of this statistic is that its upper bound is not one, but .

* **Normalized Cox-Snell R2 :** can be normalized using the following transformation by Nagelkerke (1991): (4.11)

(Long & Freese, 2014; Pampel, 2000)

**5. Ordinal Logistic Regression vs. Multinomial Logistic Regression**

**Table 5.1: Ordinal Logistic Regression vs. Multinomial Logistic Regression**

| **Ordinal Logistic Regression** | **Multinomial Logistic Regression** |
| --- | --- |
| **Logit Model:**  Ordinal logistic regression is appropriate when the outcome variable is ordinal and has more than two ordered categories. Cumulative Logit Model or Proportional Odds Model is a particular type of model that considers the ordering of categories and assumes that the odds ratio is invariant to where the outcome categories are dichotomized.  Let the model include one ordinal outcome variable *Y* having 3 ordered categories (*Y* = 0, 1, 2) and onedichotomousindependent variable *X*1 (*X*1 = 0, 1). Then there are two ways to dichotomize the outcome: (*Y* ≥ 1 vs. *Y* < 1; *Y* ≥ 2 vs. *Y* < 2). With this categorization of *Y*, the logit model is given by two regression equations:  = + *β*1 *X*1  = + *β*1 *X*1  **Conditional Probabilities:**  ) =  ) =  **Odds Ratio (OR):**  OR1 = =  OR2 = | **Logit Model:**  When the outcome variable is nominal having more than two unordered categories, multinomial logistic regression is appro-priate. Let the model include one outcome variable *Y* having 3 unordered categories (*Y* = 0, 1, 2) and onedichotomousindependent variable *X*1 (*X*1 = 0, 1). Also, let the condi-tional probabilities be:  P (*Y* = 0 | *X*1) = *P*0 ; P (*Y* = 1 | *X*1) = *P*1 ;  P (*Y* = 2 | *X*1) = *P*2 .  Two regression equations give the logit model:  ln = + *β*11 *X*1  = + *β2*1 *X*1  **Conditional Probabilities:**  The probabilities are given by:  *P*0 =  *P1*=  *P2*=  **Odds Ratio (OR):**  OR1 = =  OR2 = |

*Source:* Author compilation from Kleinbaum and Klein (2010), Agresti (2002).

**6. Probit Analysis vs. Logistic Regression**

Probit analysis uses scores from the cumulative standard normal distribution instead of calculating logged odds from the logistic distribution. Logistic regression employs the logistic curve, while probit analysis uses the cumulative standard normal curve. Corresponding to the formula in logistic regression for the logged odds, *Li* = In( *Pi / (1− Pi* )*,* the formula for probit analysis identifies the *inverse* of the cumulative standard normal distribution. If the cumulative standard normal distribution is represented by Φ, then *P* = Φ (Z), and Z = Φ−1 ( *P* ), where Φ−1 refers to the inverse of the cumulative standard normal distribution. Although a simple formula cannot represent it, the inverse of the cumulative standard normal distribution transforms probabilities into linear Z scores representing the dependent variable in probit analysis. With probits as the dependent variable, the estimated coefficients show the change in *z* score units of the inverse of the cumulative standard normal distribution rather than the probability change. Even though independent variables have a non-linear relationship to the probabilities, z scores from the probit transformation are linear. Although the logit curve reaches the floor and ceiling somewhat quicker than the probit curve, the differences are insignificant. With simple formulae, one can convert probabilities into logged odds and vice versa in logistic regression. Probit analysis is more challenging because of the complicated formula for the standard normal curve. This means that programs must apply an arbitrary normalization to set the scale of logit and probit variables, as they have no inherent scaling properties. Using probit analysis, the standard deviation of the error is equal to 1, but using logit analysis, it is around 1.814. Probit and logit coefficients cannot be directly compared because of their different error variances. The logit coefficients will be about 1.8 times greater than the probit coefficients. It is possible to divide the logit coefficients by that factor to keep the units equivalent. However, logistic regression and probit coefficients will differ because of the modest discrepancies between the logistic and normal curves. Probit analysis does not have the same multiplicative odds coefficients as logistic regression, contributing to its increased popularity. Given the transformed units of the dependent variable, probit coefficients are interpretable in the same way as other regression coefficients. They illustrate the linear and additive change in the probit transformation’s z-score units (i.e., the inverse of the cumulative standard normal distribution) when the independent variables are changed by one unit. Perhaps even less evident than logged odds, the cumulative normal distribution’s standard units are of little interpretative significance (Pampel, 2000).

**References**

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). John Wiley & Sons.

Browne, I. (1997). Explaining the Black-white gap in labor force participation among women heading households. American Sociological Review, 62(2), 236-252. [https://doi.org/10.2307/2657302](https://doi.org/10.2307/2657302%20%20)

Cramer, J. (2003). The Origins of Logistic Regression. *Tinbergen Institute Discussion Paper*.

<https://doi.org/10.2139/ssrn.360300>

DeMaris, A. (1993). Odds versus probabilities in Logit equations: A reply to Roncek. *Social Forces, 71*(4), 1057-1065. <https://doi.org/10.2307/2580130>

DeMaris, A., Teachman, J., & Morgan, S. P. (1990). Interpreting logistic regression results: A critical commentary. *Journal of Marriage and the Family*, *52*(1), 271-277.

[https://doi.org/10.2307/352857](https://doi.org/10.2307/352857%20%20)

Hosmer, D. W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons.

Kaufman, R. L. (1996). Comparing effects in dichotomous logistic regression: A variety of standardized coefficients. *Social Science Quarterly,* 77, 90-109.

Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression: A Self-Learning Text* (3rd ed.). Springer Science & Business Media.

Long, J. S., & Long, J. S. (1997). Regression Models for Categorical and Limited Dependent Variables. SAGE.

Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using Stata* (3rd ed.). Stata Press.

Pampel, F. C. (2020). *Logistic Regression: A Primer*. SAGE Publications.

Roncek, D. W. (1993). When will they ever learn that first derivatives identify the effects of continuous independent variables or “Officer, you can’t give me a ticket, I wasn’t speeding for an entire hour”. *Social Forces*, *71*(4), 1067-1078. [https://doi.org/10.2307/2580131](https://doi.org/10.2307/2580131%20%20)