ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.: 4 - Sayı/No: 2 : 195-204 (2003)

ARAŞTIRMA MAKALESİ/**RESEARCH ARTICLE**

GOODNESS OF FIT MEASURES IN BINARY PROBIT REGRESSION MODELS

Özlem ÖZARICI[1], Berna YAZICI[2]

ABSTRACT

In this study the goodness of fit measures in Probit models for binary outcomes, have been examined. Pseudo-R2 measures and other measures are introduced and a comparison of those different measures is held. The application of each of the goodness of fit measure on "the users' satisfaction factors from a university's web site data" is also given.

Key Words: Goodness of fit, Discrete choice, Probit regression model, Binary choice model, Pseudo-R2, Binary probit.

İKİ DÜZEYLİ PROBİT REGRESYON MODELLERİNDE UYUM İYİLİĞİ ÖLÇÜTLERİ

ÖZ

Bu çalışmada iki düzeyli Probit modeli için kullanılan uyum iyiliği ölçütleri üzerinde durulmuştur. Bu amaçla Yapay-R2 ölçütleri ve diğer ölçütler tanıtılmış ve söz konusu ölçütlerin bir karşılaştırılması ele alınmıştır. Ayrıca "kullanıcıların bir üniversitenin web sayfasından olan memnuniyet faktörleri verisi" üzerine, herbir uyum iyiliği ölçütünün uygulamasına yer verilmiştir.

Anahtar Kelimeler: Uyum iyiliği, Kesikli tercih, Probit regresyon modeli, İki düzeyli tercih modeli, Yapay-R2, İki düzeyli probit.

1. INTRODUCTION

Goodness of fit measure is an indicator of adaptation of the data to the evaluated regression equation. After evaluating a Probit regression equation and obtaining the parameter estimates, the proof of how well the variation in dependent variable is explained by the independent variables is given with this measure. How better the dependent variable is explained by the independent variables, the goodness of fit measure will be greater.

Last two decades, likelihood based measures, log-likelihood based measures, measures based on the predicted probabilities, measures based on the variance decomposition of the predicted probabilities have been used as Pseudo-R2 measures and other measures like Yule's Q criterion, likelihood ratio test, the sum of weighted squared residuals have been used as the goodness of fit measures for binary outcomes.

The important question that the researcher must take into account is which goodness of fit measure among those should be chosen. Different criteria should be considered to choose the suitable goodness of fit measure.

2. BINARY PROBIT REGRESSION MODEL

When the researcher is interested in a qualitative dependent variable, one of the alternative regression models will be the binary Probit model. If the dependent variable is a binary variable, in this case 0 and 1 are designated to the levels of the variable (Ozarici, 2002).

[1] Osmangazi University, Department of Statistics, Eskisehir, TURKEY.
[2] Osmangazi University, Department of Statistics, Eskisehir, TURKEY.
Fax: (222) 3204910, **E-mail:** bbaloglu@anadolu.edu.tr

Received: 27 Şubat 2003; **Revised:** 19 Haziran 2003; **Accepted:** 21 Ağustos 2003.

A typical approach postulates an underlying continuous variable y_i^*:

$$y_i^* = x_i'\beta + \varepsilon_i \qquad i=1,...,n \tag{1}$$

where x_i': row vector of independent variable values for ith observation,

β: unknown parameter column vector that will be estimated,

ε_i: random error term vec tor typically assumed to be i.i.d. $(\varepsilon_i \sim N(0,1))$

y_i, binary dependent variable that takes its values based on unobserved y_i^* variable is given by Eq.(2):

$$y_i = \begin{cases} 1 \text{ if } y_i^* > 0 \\ 0 \text{ otherwise} \end{cases} \quad i = 1,...,n \text{ (n: number of observations)} \tag{2}$$

So the probabilities of appearing the levels of the dependent variable:

$$\begin{aligned} P(y_i = 1) &= P(y_i^* > 0) P(x_i'\beta + \varepsilon_i > 0) = P(\varepsilon_i > -x_i'\beta) = \Phi(x_i'\beta) \\ P(y_i = 0) &= 1\text{-}P(y_i^* > 0) = 1\text{-}\Phi(x_i'\beta) \end{aligned} \tag{3}$$

where $\Phi(x_i'\beta)$ is the value of cumulative standard Normal distribution and defined as follows (Griffith, Hill and Judge, 1993):

$$P_i = P(y_i = 1|\mathbf{x}_i) = \Phi(\mathbf{x}_i'\beta) = \int_{-\infty}^{\mathbf{x}_i'\beta} (2\pi)^{-1/2}.\exp(-\varepsilon^2/2).d\varepsilon \tag{4}$$

Using Eq.(3) and Eq. (4), a likelihood function for binary Probit model is written as follows:

$$L = L(\beta) = \prod_{i=1}^{n} [\Phi(\mathbf{x}_i'\beta)]^{y_i}.[1 - \Phi(\mathbf{x}_i'\beta)]^{1-y_i} \tag{5}$$

and a log-likelihood function is given by Eq. (6):

$$\log L = \sum_{i=1}^{n} y_i.\log\Phi(\mathbf{x}_i'\beta) + (1 - y_i)\log[1 - \Phi(\mathbf{x}_i'\beta)] \tag{6}$$

If Φ is standard normal, the model is called a binary Probit (Veall, M. R. and Zimmerman, K. F., 1996).

3. PSEUDO-R^2 MEASURES

The multiple determination coefficient values change between 0 and 1 for classical regression models. It is impossible for R^2 to get a value near 1, for the models for binary dependent variables. This is only possible if all the predicted probabilities are equal to 0 or 1. So, some regression analysts say that R^2 is not a useful statistic, but others commonly use that statistic to evaluate the model performance. From this point of view, a few Pseudo-R^2 is suggested. Researchers must choose a Pseudo-R^2 measure among the suggested ones. When it is necessary to decide which measure must be used, first of all, researchers must decide what for that measure is going to be used: measure of variation, hypothesis test, classifying the dependent variable in correct way.

If the researchers are interested in the explained variation (the most common use of R^2 in regression analysis) it is best to use one of the Pseudo-R^2 measures suggested by McKelvey and Zavoina (1975), Aldrich-Nelson (1984), or McFadden (1974). In the following sections, these measures and some others are going to be mentioned.

3.1 LIKELIHOOD BASED MEASURES

Various goodness of fit measures are proposed as determination coefficients by researchers. In this study the goodness of fit measures that are related with the binary Probit model are examined.

Pseudo-R^2 goodness of fit measures that are based on likelihood functions are given in the following sections.

3.1.1 Pseudo-R^2 Proposed by Maddala

Pseudo-R^2 proposed by Maddala is calculated using following formula:

$$R^2_{Maddala} = 1 - \left(\frac{L_0}{L_1}\right)^{\frac{2}{n}} \tag{7}$$

where L_0: likelihood value of the zero model (with a regression constant only)

L_1: likelihood value of the alternative model

n: sample size

Theoretical range: $0 \le R^2_{Maddala} \le 1 - (L_0)^{\frac{2}{n}}$.

On the other hand, this Pseudo-R^2 value is also equal to maximum likelihood Pseudo-R^2.

The maximum likelihood R^2 expresses the model fit as a transformation of likelihood ratio χ^2 in analogous way to that of R^2 in ordinary least (R^2_{OLS}) square regression.

$$R^2_{ML} = 1 - \exp(-\frac{G^2}{n}) \tag{8}$$

where G^2 is likelihood ratio and will be defined in Section 3.2. The statistic in Eq. (8) can be thought as a transformation of the F statistic.

3.1.2 Pseudo-R^2 Proposed by Cragg-Uhler

Because of the limitation on the maximum value for maximum likelihood R^2 Cragg and Uhler (1970) proposed a relative index that can reach one:

$$R^2_{C/U} = \frac{L_1^{\left(\frac{2}{n}\right)} - L_0^{\left(\frac{2}{n}\right)}}{1 - L_0^{\left(\frac{2}{n}\right)}} \tag{9}$$

Standardization of the Maddala Pseudo-R^2 by its own maximum theoretical range: $0 \le R^2_{C/U} \le 1$.

3.2 LOG-LIKELIHOOD BASED MEASURES

In this section Pseudo-R^2 goodness of fit measures that are based on log-likelihood function are taken into account.

3.2.1 Pseudo-R^2 Measure Proposed by McFadden

This measure uses the two log-likelihood values in Pseudo-R2 suggested by Aldrich-Nelson and defined as follows:

$$R^2_{mf} = 1 - \frac{\log L_1}{\log L_0} \tag{10}$$

where $\log L_1$ is the log-likelihood value of the model and $\log L_0$ is the log-likelihood value if the non-intercept coefficients are restricted to zero, under the condition that all coefficients in the regression model are different from zero (McFadden and Lerman, 1981). The limit values of the measure are -1 and 1.

If $\log L_1 = \log L_0$ then $R^2_{mf} = 0$,

If $\log L_1 = 0$ then $R^2_{mf} = 1$.

On the other hand using R^2_{mf}, the null hypothesis of regression coefficients' equality to 0 can be tested with χ^2 test statistic. Under the null hypothesis

H_0: $\beta = 0$

If then If $n \to \infty$ then $-2\log L_0 R^2_{mf} \to \chi^2_{K-1}$.

R^2_{mf} is a popular goodness of fit measure since it has advantageous properties like R^2_{OLS} does.

3.2.2 Pseudo-R^2 Measure Proposed by Aldrich-Nelson

In this measure the log-likelihood ratios are used depending on two situations. First situation forms the likelihood value for the null hypothesis. Likelihood value for the null hypothesis is usually denoted by L_0. Second situation forms the likelihood model for full (saturated) model and usually denoted by L_1.

For Pseudo-R^2 suggested by Aldrich-Nelson the χ^2 statistic is used with K degrees of freedom (K; the number of estimated independent variables without the constant value).

Pseudo-R^2 suggested by Aldrich-Nelson (1984) is defined as follows:

$$R^2_{A/N} = \frac{-2\log\left(\frac{L_0}{L_1}\right)}{n - 2\log\left(\frac{L_0}{L_1}\right)} \tag{11}$$

If the model is not contributive for explaining the variation in the dependent variable, the log-likelihood value logL1 in the transformation will be equal to logL0. In this situation the nominator of Eq. (11) will be zero and so Pseudo-R^2 of Aldrich-Nelson will be zero.

If the model completely explains the variation in the dependent variable $\log L_1$ will be equal to 0. In this situation the formula of Aldrich-Nelson is reduced the following formula:

$$R^2_{A/N} = \frac{-2\log L_0}{n - 2\log L_0} \tag{12}$$

where; $\log L_0 = n_0(\log(n_0/n) + n_1(\log(n_1/n))$, n_0: number of observations for $y_i=0$ and n_1: number of observations for $y_i=1$ $(n=n_0+n_1)$.

3.2.3 Pseudo-R^2 Measure Proposed by Veall and Zimmerman

Veal and Zimmerman (1995) suggested a measure based on log-likelihood:

$$R^2_{v/z} = \frac{2[\log L_1 - \log L_0]}{2[\log L_1 - \log L_0] + n} \cdot \frac{2\log L_0 - n}{2\log L_0} \quad (13)$$

The second term here is a modification to obtain the upper limit value for Aldrich-Nelson measure.

3.2.4 Pseudo-R2 Measure Proposed by Ben-Akiva and Lerman

Ben-Akiva and Lerman (1985) note that, this measure will always increase when new variables are added to the model whether or not these variables contribute usefully to explain the data. Therefore this measure does not adequately account for desired parsimony in the selected specification. For this reason, the Ben-Akiva and Lerman adjust McFadden's R^2 measure to penalize the addition of variables.

$$R^2_{B-A/L} = 1 - \frac{\log L_1 - K}{\log L_0} \quad (14)$$

where K is the number of independent variables in the model. Using either measure, the best model is the one with the largest R^2, corresponding to the model that explains the most variation in the data. Further, unlike the likelihood ratio presented above, these tests can be used to compare models that cannot be expressed as restricted subsets of each other.

3.2.5 Pseudo-R2 Measure Proposed by Estrella

Estrella (1998) suggests that the measure should be directly related to the valid test statistic for the significance of all slope coefficients and the derivative of the measure with respect to the test statistic should comply with corresponding derivatives in a linear regression.

Estrella's measure is written as follows:

$$R^2_{Estrella1} = 1 - \left(\frac{\log L_1}{\log L_0}\right)^{-\frac{2}{n}\log L_0} \quad (15)$$

Estrella suggests an alternative measure that is given in Eq. (15):

$$R^2_{Estrella2} = 1 - [(\log L - K)/\log L_0]^{-\frac{2}{n}\ln L_0} \quad (16)$$

where $\log L_0$ is computed with null parameter values, n is the number of observations used, K represents the number of estimated parameters.

3.3 MEASURES BASED ON THE PREDICTED PROBABILITIES

3.3.1 Pseudo-R2 Measure Proposed by Efron

Lave (1970) studied on the selection of the Probit model and gave a theoretical base for the measure suggested by Efron. So in the literature this measure is usually called as Efron's measure (Efron, 1978).

Taking $\hat{P}_i$ as the prediction of P_i, this measure is given as follows:

$$R^2_{ef} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{P}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{(y-\hat{P})'(y-\hat{P})}{y'Ny} \quad (17)$$

where $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ (sample mean of y) and ,
$N = I_n - n^{-1}ss'$, $s = (1,1,\ldots,1)'$

Measure is calculated directly from the determination coefficient in linear regression by separating the total variation for binary outcomes, in two components as explained and (unexplained variation = SSE = $(y-\hat{P})'(y-\hat{P})$) Efron derived this decomposition for a model with grouped observations and group constants as the only explanatory variables. However this decomposition is not confirmed in the model that includes individual data and individual predicted probabilities. In this case the following equation can be assumed:

$$y'Ny = (y-\hat{P})'(y-\hat{P}) + (\hat{P}-\bar{y})'(\hat{P}-\bar{y}) + 2\hat{P}'(y-\hat{P}) - 2\bar{y}'(\bar{y}-\hat{P}) \quad (18)$$

So the claims on lower limit of R^2_{ef} 's being equal to 0 can be occurred.

3.3.2 Pseudo-R2 Measure Proposed by Achen

Achen suggests that, the measures defined above include some errors and those statistics cannot be used to decide whether one or a subgroup of regression coefficients is equal to 0 or not.

Aldrich-Nelson suggests that this is not a failure as mentioned by Achen, because the χ^2 statistic is similar to the F statistic in regression analysis.

Achen's measure has also advantage of having the explained variance for binary outcome (denoted as y_i by Achen).

The variance for binary outcomes ($var(y_i)$) is defined as P_iQ_i, where Pi the probability of y_i=1 and Q_i=(1-P_i). Under standard normal distribution characteristics of Probit coefficients, Pseudo-R^2 suggested by Achen is defined as follows (Hagle and Mitchell, 1992):

$$R^2 \frac{\frac{1}{n}\sum_{i=1}^{N}\frac{(\hat{P}_i-\bar{y})^2}{\hat{P}_i\hat{Q}_i}}{1+\frac{1}{n}\sum_{i=1}^{N}\frac{(\hat{P}_i-\bar{y})^2}{\hat{P}_i\hat{Q}_i}} \quad (19)$$

where $\hat{P}_i$ is estimate of P_i.

3.3.3 Average Probability of Correct Prediction

It is possible to get an idea about how accurate the data are predicted for binary outcomes. Depending on this idea the following measure is used:

$$R^2_{cp} = 1 - n^{-1}(\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) \quad (20)$$

where the predicted value

$\hat{y}_i = 1, \text{if } P(y_i = 1) > 0.5$

$\hat{y}_i = 0, \text{if } P(y_i = 1) < 0.5$

Two different problems can occur for this measure:

1- It doesn't matter the predicted probability is equal to 1 or 0, greater than 0.5 or less than 0.5, the value of R^2_{cp} does not change.

2- If the ratio of 1's are greater than 0.5 then the lack of fit can not be known. If the predicted value is 1 for each observation, the number of correctly predicted observation is greater. But the model misclassifies each observation with y_i=0. The way of handling this last problem is to notify either correct ratio of 1's or correct ratio of 0's. Another way is to calculate the mean probability of correct prediction. This calculation is as follows (Windmeijer, 1995):

$$R^{2*}_{cp} = n^{-1}\sum_i\left\{y_i\hat{P}_i + (1-y_i)(1-\hat{P}_i)\right\} \quad (21)$$

where y_i is the value of the dependent variable in case i, $\hat{P}_i$ are the predicted probabilities for y=1 in case i.

3.4 MEASURES BASED ON THE VARIANCE DECOMPOSITION OF THE PREDICTED PROBABILITIES

Pseudo-R^2 Measure Proposed by McKelvey-Zavoina

This measure is one of the most common ones in the applications. The measure takes values between 0 and 1, and it is interpreted like R^2_{OLS} for Linear Probability model.

The variance of binary Probit model's estimated coefficients and this way, explained variance is calculated. This quantity is denoted by var $(\hat{y}_i)$.

$$R^2 = \frac{var(\hat{y}_i)}{1+var(\hat{y}_i)} \quad (22)$$

Using the var $(\hat{y}_i)$ values, the Pseudo-R^2 suggested by McKelvey and Zavoina (1975) is obtained by Eq. (22).

4. OTHER MEASURES

The other measures that are commonly used by the researchers are given in this section.

4.1 Yule's Q Criterion

This measure is the ratio of the odds differences to the odds totals of research results (Agresti, 2002).

Predicted y for Probit model

		0	1	Total
Observed	0	a	b	a+b
y	1	c	d	c+d
	Total	a+c	b+d	a+b+c+d

Yule's Q criterion is calculated as follows:

$$Q = \frac{\frac{a}{b}-\frac{c}{d}}{\frac{a}{b}+\frac{c}{d}} = \frac{ad-bc}{ad+bc} \quad (23)$$

where,

a: the value of prediction is equal to 0, when the observed value of dependent variable is equal to 0,

b: the value of prediction is equal to 1, when the observed value of dependent variable is equal to 0, c: the value of prediction is equal to 0, when the observed value of dependent variable is equal to 1,

d: the value of prediction is equal to 1, when the observed value of dependent variable is equal to 1.

Q criterion is a measure that shows the appropriateness of observed probabilities and the predicted probabilities, takes values between 0 and 1, and denotes a strong relationship if it has close values to 1. This is required for the researchers.

4.2 Likelihood Ratio

Likelihood ratio, on the contrary with the Pseudo-R^2, does not give any indicator about the general content of the model to explain the variability in the dependent variable. The formula is as follows:

$$G^2 = -2\log\left(\frac{L_0}{L_1}\right) = (-2\log L_0) - (-2\log L_1) = -2(\log L_0 - \log L_1) \quad (24)$$

Taking L_0 and L_1 likelihood values, likelihood ratio has χ^2 distribution with the number of independent variables, degrees of freedom (Aldrich and Nelson, 1984). On the contrary with the Pseudo-R^2, likelihood ratio gives statistically significance measure of the improvement for the parameters that are added to the zero models.

Likelihood ratio also can be used to test various models or subgroup of variables for the same data. To test the hypothesis about regression coefficients equality to 0, likelihood ratio G^2 can be compared to χ^2 value. In this case the hypotheses are mentioned as follows:

$H_o = B_I = 0, \quad B_2 = 0,...,B_k = 0$

$H_1 = B_k \neq 0$, for at least one $k = 1,...,K$

If $G^2 > \chi^2$ (K-1,α) the null hypothesis is rejected. The test of contribution of the variables to model is possible with G^2 differences. This difference values are compared with the relevant χ^2 table value and so the decision is made about the importance of the contribution of the variables that are tested.

4.3 Sum of Weighted Squared Residuals

It is assumed that the sum of weighted squared residuals has asymptotically χ^2 distribution. However, McCullach (1986) and Windmeijer (1990) have shown that the sum of weighted squared residuals has Normally distributed and this measure is calculated as follows (Windmeijer, 1995):

$$T_n = \sum_{i=1}^{n} \frac{(y_i - \hat{P}_i)^2}{\hat{P}_i(1-\hat{P}_i)} \quad (25)$$

where $\hat{P}_i$ is the predicted value of P_i.

4.4 Correctly Classification Percent (CCP)

This measure shows the percent of correctly classification of the actual values due to the predicted values and calculated as follows:

Observed Value	Predicted Value			
		0	1	Total
	0	a	b	a+b
	1	c	d	c+d
	Total	a+c	b+d	N

$$CCP = \frac{axd}{bxc} \quad (26)$$

This measure is quite common in use.

4.5 Squared Sample Correlation Coefficient

The squared sample correlation coefficient of predicted and true probabilities can be calculated and interpreted as well:

$$r^2(y,\hat{P}) = \frac{[Cov(y,\hat{P})]^2}{var(y)\,var(\hat{P})} \quad (27)$$

If the squared correlation of predicted and true probabilities is one, this indicates the correct model is chosen (Veall and Zimmermann, 1996).

An application and a comparison for each goodness of fit measure that is mentioned in previous sections for binary Probit models are going to be given in the next section.

5. APPLICATION

Great advances in science and technology have rapidly changed the structure of the society as well as the method of business and education in the last decade. Especially internet has given more power in this variation. Corporations and people have to be aware of the influence of this variation. Recent changes in the internet technology have also brought new approaches to education. Especially universities use their internet services to give the students' examination and

Table 1. The Definition of the Variables.

Variable	Type of Variable	Levels of Variable	Variable Code
Satisfaction from the web site	Dependent	Yes-no	S
Capability of introducing the university	Independent	1, ... ,4	IC
Sufficiency of content	Independent	1, ... ,4	SC
Deficiency of the web site	Independent	0, ... , 14	DW
Process speed of the web site	Independent	1, ... ,4	PS
To be informed about the facilities of the web site	Independent	1, ... ,4	BI
The use of student information system	Independent	0, ... ,10	UI
Facility of the main page	Independent	1, ... ,4	FM

inspection results, course programs and etc. This compels students to use internet and web sites of the university they attend. So the users' pleasure from university's web site gains importance.

In this part of the study, the users' satisfaction factors from a university's web site are examined. To do this, the users are asked to answer the questionnaire on that web site. 90 users who joined the questionnaire constructed the sample for the study. The variables are given in Table 1.

Using 90 questionnaires and the forward stepwise regression, the results on Table 2 have been obtained. Results are obtained using LIMDEP statistical software.

Starting model includes IC with the constant. Variables are added to the model due to the increase that they provide on the log-likelihood. For all models constant, IC, SC and DW parameters are significant and other parameters are insignificant. However, since the study is related with how the Pseudo-R^2 values support that result, the insignificant variables are not removed. The Pseudo-R^2 values for 7 different models are given in Table 3.

Table 2. Results of Probit Regression Analysis, Coefficients, Standard Errors, t Statistics and P-values.

Model	Coefficient	Standard Error	t statistic	P-value
Constant	-2.7472	0.5052	-5.438	0.000
IC	1.0376	0.2189	4.738	0.000
Constant	-3.9249	0.7049	-5.568	0.000
IC	0.7099	0.2469	2.875	0.003
SC	0.9073	0.2672	3.395	0.001
Constant	-2.9118	0.8028	-3.627	0.000
IC	0.7659	0.2732	2.804	0.005
SC	0.9528	0.3073	3.101	0.002
DW	-0.1856	0.0594	-3.124	0.002
Constant	-3.4576	0.9379	-3.686	0.000
IC	0.7036	0.2889	2.435	0.014
SC	1.0443	0.3243	3.220	0.001
DW	-0.2117	0.0670	-3.159	0.001
PS	0.2649	0.1948	1.360	0.173
Constant	-3.6531	0.9958	-3.669	0.000
IC	0.7013	0.2863	2.450	0.014
SC	0.9873	0.3317	2.976	0.002
DW	-0.2130	0.0666	-3.200	0.001
PS	0.2169	0.2027	1.071	0.284
BI	0.1658	0.2413	.687	0.491
Constant	-3.3354	0.9883	-3.375	0.000
IC	0.6337	0.2981	2.125	0.033
SC	1.0185	0.3433	2.967	0.003
DW	-0.2014	0.0695	-2.898	0.003
PS	0.2176	0.2144	1.015	0.310
BI	0.2103	0.2511	0.838	0.402
UI	-0.1014	0.0748	-1.355	0.175
Constant	-3.5544	1.0539	-3.373	0.000
IC	0.5751	0.3159	1.820	0.068
SC	0.9819	0.3517	2.792	0.005
DW	-0.2045	0.0708	-2.891	0.003
PS	0.2322	0.2189	1.060	0.289
BI	0.2294	0.2537	0.904	0.365
UI	-0.0951	0.0757	-1.256	0.209
FM	0.1758	0.2582	0.681	0.495

Table 3. Pseudo-R^2 Results, Web Site Satisfaction Data.

Model	R^2_{ML}	$R^2_{C/U}$	$R^2_{McFadden}$	$R^2_{A/N}$	$R^2_{V/Z}$	$R^2_{B\ A/L}$	$R^2_{Estrella1}$
IC	0.2620	0.3746	0.2527	0.2330	0.4269	0.7127	0.2955
IC SC	0.3607	0.5155	0.3721	0.3091	0.5662	0.7599	0.4285
IC SC DW	0.4408	0.6301	0.4834	0.3676	0.6733	0.8036	0.5480
IC SC DW PS	0.4528	0.6473	0.5015	0.3761	0.6889	0.8083	0.5669
IC SC DW PS BI	0.4557	0.6514	0.5058	0.3782	0.6927	0.8103	0.5715
IC SC DW PS BI UI	0.4670	0.6677	0.5234	0.3862	0.7075	0.8188	0.5898
IC SC DW PS BI UI FM	0.4698	0.6716	0.5277	0.3882	0.7110	0.8210	0.5942

Model	$R^2_{Estrella2}$	R^2_{Efron}	R^2_{Achen}	R^{2*}_{cp}	$R^2_{M/Z}$	$r^2(y,\hat{P})$	R^2_{OLS}
IC	0.3372	0.3277	0.4102	0.145	0.4004	0.3684	0.2852
IC SC	0.4887	0.4111	0.7538	0.169	0.5599	0.3861	0.3882
IC SC DW	0.6246	0.5231	0.9820	0.193	0.7005	0.5194	0.4704
IC SC DW PS	0.6615	0.5251	0.9887	0.195	0.7251	0.4412	0.4782
IC SC DW PS BI	0.6843	0.5316	0.9908	0.196	0.7266	0.4073	0.4815
IC SC DW PS BI UI	0.7197	0.5545	0.9919	0.199	0.7380	0.4891	0.4991
IC SC DW PS BI UI FM	0.7416	0.5633	0.9941	0.201	0.7418	0.5323	0.5021

All measures in Table 3 show the same pattern. The variables added to the starting model are IC, SC, DW, PS, BI, UI and FM respectively. The Pseudo-R^2 values increase sharply when IC, SC and DW are added to the model. Then, the variables added to the model cause slight increase. These results support the findings about significant variables in Table 2.

R^2_{ML}, R^2_{OLS}, $R^2_{Estrella1}$, $R^2_{McFadden}$, R^2_{Efron} Pseudo-R^2 values give the similar results with each other.

Similarly, $R^2_{V/Z}$, $R^2_{M/Z}$, $R^2_{C/U}$ Pseudo-R^2 give the similar results with each other.

Other goodness of fit measures that are used for Probit Regression Analysis, Likelihood Ratio, Yule's Q criterion, correctly classification percent, sum of weighted squared residuals are calculated using LIMDEP statistical software and given in Table 4.

Examining the Yule's Q criterion, it is seen that, the values of the criterion sharply increase until the 3rd model that is significant. Then while the insignificant variables are added to the model, the criterion takes the smaller values. Similarly, χ^2 values are increase sharply and show slight increase after the 3rd model. Examining the correctly classification percent, it is seen that the 3rd model and the model that includes the all variables take the same correctly classification percent. That value is 89%, which is high.

It is an expected result that likelihood ratio value increases when the number of variables increases. It can be seen that each likelihood ratio result is significant. But, since it is known that similar last four variables are not included by the model, similar values are expected for likelihood ratio. The results supported that consideration.

The sum of weighted squared residuals must be as small as possible. It is seen that, when another variable is added to the model, until the fourth model, T_n values decrease sharply. After the fourth model, T_n values decrease slightly. T_n values show a different pattern when the last variable is added to the model. The value in question increases slightly. That shows T_n measure is not consistent for the data set used in this study.

Table 4. Other Goodness of Fit Measures

Model	LR	Yule's Q	T_n	CCP
IC	27.348	0.910	104.574	0.844
IC SC	40.270	0.913	74.151	0.844
IC SC DW	52.310	0.972	69.951	0.888
IC SC DW PS	54.260	0.942	59.329	0.866
IC SC DW PS BI	54.740	0.927	59.210	0.855
IC SC DW PS BI UI	56.640	0.950	58.560	0.877
IC SC DW PS BI UI FM	57.099	0.960	60.642	0.888

6. CONCLUSIONS

For the data set that is used in this study, since the results are not consistent for T_n and $r^2(y, \hat{P})$, in the choice of the measures, those two can be neglected. On the other hand, since T_n and likelihood ratio don't have limit values, they don't give the researcher a certain result and they are difficult to interpret in this way. However, since they give a general idea about the goodness of fit of the model, they recommended using with Pseudo-R^2 measures.

The choice among the Pseudo-R^2 measures are simply depends on easiness of their calculation and facility of use. There is no general agreement on how to assess the fit corresponding to practical significance. It is not so important to interpret one of those Pseudo-R^2, what important is not to use only one Pseudo-R^2 and use a few of them in one analysis.

For the choice of a Pseudo-R^2 Windmeijer (1995) uses two criteria. First is closeness to R^2_{OLS}, and second is closeness to $r^2(y, \hat{P})$. They advise to compare the Pseudo-R^2 values with those two criteria and choose the closest Pseudo-R^2. In this study R^2_{ML} is calculated instead of R^2_{OLS}. It is recommended that the researchers may also compare the Pseudo-R^2 values with R^2_{ML} and choose the closest one.

The studies of Veal-Zimmerman (1995) showed that R^{2*}_{cp} does not work better than R^2_{cp}. However those measures give a clue about model's reflecting the data correctly. In the application part of this study, against the problems that may occur with R^2_{cp}, the mean probability of correct prediction (R^{2*}_{cp}) is calculated and interpreted.

Among the Pseudo-R^2 measures, R^{2*}_{cp} has very small values for all models and R^2_{Achen} on the contrary has very big values (in other words, due to this measure, variables explain the model almost exactly), give an idea about the measures that they underestimate and overestimate respectively.

Besides $R^2_{M/Z}$, the most common Pseudo-R^2 in the literature, since they have similar values with that measure, $R^2_{V/Z}$, $R^2_{C/U}$ and $R^2_{Estrella}$ are also recommended.

Assessing the model fit researcher must use at least 2 or 3 Pseudo-R^2 as the best fit measure for binary Probit models.

REFERENCES

Agresti, A. (2002). *Categorical Data Analysis,* Second Edition, Wiley Series and Sons Inc.

Aldrich, J.H. and Nelson, F.D. (1984). *Linear Probability and Probit Models, Quantitative Applications in the Social Sciences,* Sage Publications, Newburry Park.

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand,* The MIT Press, Cambridge, MA.

Cragg, S.G. and Uhler, R. (1970). The Demand for Automobiles. *Canadian Journal of Economics* 3, 386-406.

Efron, B. (1978). Regression and ANOVA With Zero-one Data Measures of Residual Variation. *JASA* 73, 113-121.

Estrella, A. (1998). A New Measure of Fit for Equations with Dichotomous Dependent Variables. *Journal of Business and Economic Statistics* 16(2), 198-205.

Griffiths, W.E., Hill, R.C. and Judge, G.G. (1993). *Learning and Practicing Econometrics,* John Wiley and Sons Inc., p.855.

Hagle, T. M. and Mithchell, G.E. (1992). Goodness of Fit Measures for Probit Models. *Econometric Reviews* 14(1), 101-116.

McFadden, M. and Lerman, S.R. (1981). *Econometric Models of Probabilistic Choice, Structural Analysis of Discrete Data with Econometric Applications,* The MIT Press, Cambridge, MA.

McKelvey, R.D. and Zavoina, W. (1975). A Statistical Model for the Analysis of. Ordinal Level Dependent Variables *Journal of Mathematical Sociology* 4, 103-120.

Veall, M.R. and Zimmerman, K.F. (1995). Comments on Goodness of Fit Measures in Binary Choice Models *Econometric Reviews* 14(1), 117-120.

Veall, M.R. and Zimmerman, K.F. (1996). Pseudo-R^2 Measures for Some Common Limited Dependent Variable Models *Journal of Economic Surveys* 10(3), 241-259.

Ozarici, O. (2002). *Bivariate Probit Model with Full Observability and Heteroscedasticity and an Application,* Unpublished PhD Thesis, Osmangazi University, Eskisehir, Turkey.

Windmeijer, F.A. G. (1995). Goodness of Fit Measures in Binary Choice Models. *Econometric Reviews* 14(1), 101-116.

Berna Yazıcı, was born in Eskişehir, in 1972. She received her Bachelor degree from Department of Statistics in Anadolu University in 1993. She was appointed as a research assistant at the Department of Statistics in Anadolu University. She received her masters degree in 1996 and PhD degree in 2001 from the same University. She has been working as an assistant professor in Department of Statistics, Anadolu University since 2001.

Özlem Özarıcı-Alpu, was born in Eskişehir, in 1973. She received her Bachelor degree from Department of Statistics in Anadolu University in 1993. She was appointed as a research assistant at the Department of Statistics in Osmangazi University. She received her masters degree in 1996 and PhD degree in 2002 from the same University. She has been working as a research assistant in Department of Statistics, Osmangazi University.