International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 3, Issue 2 (Mar-Apr 2015), PP. 01-06

ESTIMATING R^2 SHRINKAGE IN REGRESSION

Dr.Samira Muhamad Salh
University of sulamina kurdstan-iraq
Samira196902@yahoo.com

Abstract- The effectiveness of various analytical formulas for estimating R^2 Shrinkage in multiple regression analysis was investigated. Two categories of formulas were identified estimators of the squared population multiple correlation coefficient (ρ^2) and those of the squared population cross-validity coefficient (ρ_c^2). The authors compeered the effectiveness of the analytical formulas for determining R^2 shrinkage, with squared population multiple correlation coefficient and number of predictors after finding all combination among variables, maximum correlation was selected to computed all two categories of formulas. The results indicated that Among the 6 analytical formulas designed to estimate the population ρ^2 , the performance of the (Olkin & part formula-1 for six variable then followed by Burket formula & Lord formula-2 among the 9 analytical formulas were found to be most stable and satisfactory.

Keywords- (multiple regression, cross-validity, multiple correlation coefficient, Linear models shrinkage regression, R^2 shrinkage).

I. INTRODUCTION

Multiple regression is a widely used analytic technique for investigating the relationship between a dependent (or criterion) variable and a set of independent (or predictor) variables. However, researcher have long recognized that, in the process of optimizing the weighting of the independent variables for a sample, sampling chance or random error tends to be capitalized (Cohen & Cohen, 1983; Stevens, 1996; Wherry,1931). That optimizing process from which the multiple regression equation is derived causes the sample multiple correlation coefficient R to be systemically higher than the corresponding population parameter ρ .When the equation is applied to an independent sample other than the one from which the equation is obtained (i.e., cross-validation), the predictive power drops off. In multiple regressions, that phenomenon is referred to as statistical bias (Glass &Hopkins, 1996; Stevens, 1996).to determine the generalizability or the predictive power of a sample regression equation; researchers have developed different approaches to model validation (Darlington, 1968; Herzberg, 1969; Uhl & Eisenberg, 1970). There are two major approaches: empirical and analytical methods. Empirical methods estimate the average predictive power of a sample regression equation on other samples (cross-validation). Typical empirical methods for this purpose are data splitting, multicross-validation, jackknife, and bootstrap methods (Ayabe, 1985; Cummings, 1982; Kromrey& Hines, 1995; Krus & Fuller, 1982). Analytical methods adjust the statistical bias to yield the corrected sample R^2 .In the present article, we focus on the effectiveness of the various analytical formulas for correcting the upward bias of the sample R^2 in regression analysis Over the decades, a variety of correction formulas for sample R^2 shrinkage have been proposed (e.g., Browne, 1975; Darlington, 1968; Ezekiel, 1929; Lord, 1950; Nicholson, 1948; Stein, 1960; Wherry, 1931). However, there appears to be a lack of consensus in the literature on which method is most appropriate under what circumstances for estimating statistical bias in multiple regression. Some researchers have suggested that the Browne formula may be superior to others for estimating shrinkage in multiple regression (Kromrey & Hines, 1996), whereas others have suggested that both the Nicholson/Lord formula and the Okin and Pratt formula work equally well (Huberty & Mourad, 1980). Several factors may have contributed to the inconsistent findings. In the literature, considerable confusion exists over various analytical formulas. For example, in several studies the Ezekiel formula was mistakenly cited as the Wherry formula (Ayabe, 1985; Kennedy, 1988; Krus& Fuller, 1982; Schmitt, 1982; Stevens, 1996). Other authors failed to distinguish between ρ^2 (the squared population multiple correlation coefficient, or the population coefficient of determination) and ρ_c^2 (the squared population coefficient of cross-validation).Distinguishing between the two parameters is important, because an analytical method estimating sample R^2 shrinkage for one might not be accurate for the other.

Beyond those discrepancies, there are some problematic methodological issues for estimating statistical bias in multiple regressions. First, different types of shrinkage estimates have been used in different studies: Some authors have used only two or three analytical formulas (Ayabe, 1985; Kromrey& Hines, 1996; Krus & Fuller, 1982; Uhl& Eisenberg, 1970), whereas others have used more (Claudy, 1978; Cummings, 1982; Huberty & Mourad, 1980; Kennedy, others have used more (Claudy, 1978; Cummings, 1982; Huberty & Mourad, 1980; ennedy, 1988). Different conclusions may have been drawn because of the limited number of analytic formulas used in particular studies.

II. STATISTICAL BIAS

There are two major reasons for researchers to apply the multiple regression procedure: (a) to estimate the population multiple correlation coefficient from a sample and (b) to Predict the same criterion variable in new samples drawn from the same population (Claudy, 1978). Quantitative researchers have long recognized that when a multiple correlation coefficient is derived from a given sample, its value tends to be "deceptively" large and it is a "positively biased" estimates of the population multiple correlation coefficients (Carter, 1979; Larson, 1931; Wherry, 1931). Furthermore, when a sample multiple regression equation is applied to a new sample, it usually fits the new sample less well than it fit the sample from which the regression equation was derived (Larson, 1931; Stevens, 1996). If the regression equation from a sample can neither estimate the population parameter accurately nor predict well when applied to other samples, then the purposes of multiple regression are not fulfilled. Two types of "shrinkage R^2 ".

III. ESTIMATING POPULATION MULTIPLE CORRELATION COEFFICIENT (ρ)

One type of shrinkage occurs when one is estimating the squared population correlation coefficient ρ^2 from a sample R^2. Least squares criterion-which can be attributed to the work of Karl Gauss (1777-1855) more than 150 years ago-is the statistical principle widely used to model the linear relationship between the dependent variable and a set of independent variables. One of the basic assumptions of the multiple regression models is that the values of the independent Variables are known constants and are fixed by the researcher before the experiment. Only the Dependent variable is free to vary from sample to sample. That regression model is called the fixed linear regression model. However, in social and behavioral sciences, the values of independent variables are rarely fixed by the researchers and are also subject to random errors. Therefore, a second regression model for applications has been suggested, in which both dependent and independent variables are allowed to vary (Binder, 1959; Park & Dudycha, 1974).That model is called the random model (or correction model). Although the maximum likelihood estimates of the regression coefficients obtained from the random and fixed models are the same under normality assumptions, their distributions are very different. The random model is so complex that more research is needed before it can be accepted in place of the commonly used fixed linear regression model. Therefore, the fixed model is usually applied, even when the assumptions are not met completely (Claudy, 1978).Such applications of the fixed regression model with assumptions violated would cause "over fitting," because the random error introduced from the less-than-perfect sample data tends to be capitalized in the process. As a result, the sample multiple correlation coefficients obtained that way tends to overestimate the true population multiple correlation (Claudy, 1978; Cohen & Cohen, 1983; Cummings, 1982).

IV. ESTIMATING COEFFICIENT OF CROSS-VALIDATION (ρ_c)

The second type of shrinkage occurs when we use the regression weights derived from one sample to predict the criterion variable for a new sample drawn from the same population. When the regression weights derived from one sample are applied to a new sample, a multiple correlation coefficient called R_c is obtained. R_c is the validity estimate of the original sample regression equation in another sample, and it is an estimator of the population cross-validity coefficient ρ_c . The expected value of R_c [E(R_c)] over many samples would approach or equal ρ_c [E(R_c)= ρ_c] (Claudy,1978; Cummings, 1982; Herzberg, 1969). Because the population regression equation in the population usually functions better than the sample regression equation in the population, the value of ρ tends to be greater than ρ_c ($\rho_c < \rho$). Also, the sample multiple correlation coefficient is a positively biased estimator of the population multiple correlation coefficient [ρ < E(R)]. Thus, the relationship between the values of the two population parameters (ρ and ρ_c) and the two sample estimates (R *and* R_c) can be summarized as follows (Claudy, 1978; Cummings, 1982; Herzberg, 1969):

$$E(R_c) = \rho_c < \rho < E(R)$$

As is generally known, the sample multiple correlation coefficient *R* is used as the estimator for both ρ_c and ρ , but it is actually larger than either ρ_c or ρ . *R* is a positively biased estimator of ρ and an even more positively biased estimator of ρ_c (Cummings, 1982). Therefore, one must "shrink" or "correct" the estimator *R* to adjust for the positive bias in estimating either parameter in multiple regression analysis.

V. SHRINKAGE REGRESSION

Shrinkage regression refers to shrinkage methods of estimation or prediction in regression situations, useful when there is multicollinearity among there regresses. With a term borrowed from approximation theory, these methods are also called regularization methods. Such situations occur frequently in environmetric studies, when many chemical, biological or other explanatory variables are measured, as when Branco et al.[3]

Suppose we have *n* independent observations of $(x, y) = (x_1, x_2, ...x_p, y)$ from a standard multiple regression models:

$Y = \alpha + \beta' x + \varepsilon$ $\text{var}(\varepsilon) = \sigma^2$ (see Linear models).

The ordinary least squares (OLS) estimator can be written:

$$b_{OLS} = S_{xx}^{-1} S_{xy} \qquad (1)$$

Where S_{xx} is the sum of squares and products matrix of the centered *x* variables and S_{xy} is the vector of their sums of products with *y*. The OLS estimator is the best fitting and minimum variance linear unbiased estimator (best linear unbiased estimator, BLUE), with variance:

$$\text{var}(b_{OLS}) = \sigma^2 S_{xx}^{-1} \qquad (2)$$

That OLS yields the best fit does not say, however, that b_{OLS} is best in a wider sense, or even a good choice. We will discuss here alternatives to be used when the *X* variables are (near) multicollinear, that is when there are linear combinations among them that show little variation. The matrix S_{xx} is then near singular, so var(b_{OLS}) will have very large elements. Correspondingly, the components of b_{OLS} may show unrealistically large values. Under exact collinearity, b_{OLS} is not even uniquely defined. In these situations it can pay substantially to use shrinkage methods that trade bias for variance. Not only can they give more realistic estimates of, but they are motivated even stronger for construction of a predictor of *Y* from *X*.

VI. ANALYTICAL FORMULAS FOR ESTIMATING R^2 SHRINKAGE

Estimating R^2 Shrinkage and correcting for the statistical bias in sample multiple regression have been discussed extensively

in the literature (Browne, 1975; Cohen & Cohen, 1983; Huberty & Mourad, 1980; Krus & Fuller,1982; Larson,1931; Stevens,1996; Wherry,1931; Wishart,1930). Researchers have proposed various shrinkage formulas to estimate either ρ^2 (the squared population multiple correlation coefficient) or $\rho_c{}^2$ (the squared population coefficient of cross-validation). However, there has been some confusion about both the origins and purposes of these different formulas (Cummings, 1982; Huberty & Mourad, 1980;Kromrey& Hines, 1996; Newman, McNeil, Garver, & Seymour, 1979).In a review of the literature, we identified 15 such shrinkage formulas. We present and review those formulas based on the parameters they are estimating. In the following presentation of these analytical formulas, N is the sample size, ρ is the number of predictor variables, *R* is the sample multiple correlation coefficient, ρ is the population multiple correlation coefficient, ρ_c is the population cross-validity coefficient, and $\hat{R}$ is the corrected R obtained from an analytical formula.

VII. ESTIMATORS OF ρ^2

From our literature review, we identified the following (6) formulas for estimating the squared population multiple correlation coefficient ρ^2.

1) The Smith formula:

$$\hat{R}^2 = 1 - \frac{N}{1-N}\left(1-R^2\right) \qquad (3)$$

This formula was originally developed by Smith and was presented by Ezekiel in 1928 (Wherry, 1931).

2) The Wherry formula --I

$$\hat{R}^2 = 1 - \frac{N-1}{N-P-1}\left(1-R^2\right) \qquad (4)$$

This formula was actually proposed by Ezekiel as an estimator of ρ^2 (Ayabe,1985; Cohen & Cohen, 1983; Cummings, 1982; Huberty & Mourad, 1980; Kromrey & Hines, 1996; Newman et al., 1979). However, in the literature it has been cited with different names, listed here in decreasing order of frequency: the Wherryformula (Ayabe, 1985; Kennedy, 1988; Krus & Fuller, 1982; Schmitt,1982; Stevens, 1996), the Ezekiel formula (Huberty & Mourad, 1980; Kromrey & Hines, 1996), the Wherry/McNemer formula (Newman et al.,1979), and the Cohen/Cohen formula (Kennedy, 1988). The Wherry formula-2 was also cited in one study as an estimator for cross-validation (Kennedy, 1988). This formula is currently being implemented by popular statistical packages for computing the adjusted R^2 in multiple regression procedures (e.g.,SAS/STAT User's Guide,1990; SPSS User's Guide, 1996).

The Wherry formula -II

$$\hat{R}^2 = 1 - \frac{N-1}{N-P}\left(1-R^2\right) \qquad (5)$$

This formula was presented by Wherry (1931) but was cited in one study as the McNemer formula (Newman et al., 1979). In the literature, it is usually confused with the Wherry formula-1. A few studies have correctly cited it as the Wherry formula (Cummings, 1982; Huberty & Mourad, 1980; Kromrey &Hines, 1996; Uhl & Eisenberg, 1970).

3) The Olkin and Pratt formula

$$\hat{R}^2 = R^2 - \frac{P-1}{N-P-1}\left(1-R^2\right) - \frac{2(N-3)}{(N-P-1)(N-P+1)}\left(1-R^2\right)^2$$

or

$$\hat{R}^2 = 1 - \frac{(N-3)\left(1-R^2\right)}{(N-P-1)}\left[1 + \frac{2\left(\left(1-R^2\right)\right)}{N-P+1}\right] \qquad (6)$$

these three formulas are basically the same equation in different algebraic forms, and they are all approximations of Olkin and Pratt's (1958) unbiased estimate of the squared multiple correlation p^2. These formulas were cited as the Olkin and Pratt formula in several studies (Ayabe, 1985; Claudy, 1978; Huberty & Mourad, 1980; Krus& Fuller, 1982) and were cited as the Herzberg formula in one study (Cummings, 1982).

4) The Pratt formula.

Another approximation of the unbiased estimate was also presented by Pratt (personal communication to E. E. Cureton, October 20, 1964, citedClaudy, 1978); it was used in two studies (Claudy, 1978; Cummings, 1982):

$$\hat{R}^2 = 1 - \frac{(N-3)\left(1-R^2\right)}{(N-P-1)}\left[1 + \frac{2\left(\left(1-R^2\right)\right)}{N-P-2.3}\right]$$

The Claudy formula-

$$\hat{R}^2 = 1 - \frac{(N-4)\left(1-R^2\right)}{(N-P-1)}\left[1 + \frac{2\left(1-R^2\right)}{N-P+1}\right] \qquad (7)$$

Claudy (1978) suggested that this formula gives a better estimation of the population multiple correlation coefficient than both the Pratt and the Herzberg approximations of the Olkin and Pratt formula for estimating ρ^2.

VIII. ESTIMATOR OF $\rho_c{}^2$ OR ρ_c

From our literature review, we identified the following (9) formulas designed for estimating the population cross-validity coefficient $\rho_c{}^2$ or ρ_c.

1) the Lord formula-I

$$\hat{R}^2 = 1 - \frac{N+p+1}{N-P-1}\left(1-R^2\right) \qquad (8)$$

Researchers developed this formula to estimate the population cross-validity coefficient p^2 (Newman et al., 1979; Uhl& Eisenberg, 1970). It has been cited most as the Lord formula (Newman et al., 1979; Uhl& Eisenberg, 1970); however, in one study it was referred to as the Uhl and Eisenberg formula (Cummings, 1982).

The Lord formula-II

$$\hat{R}^2 = 1 - \frac{(N+p+1)(N-1)}{(N-P-1)N}\left(1-R^2\right) \qquad (9)$$

This formula was developed by Lord and Nicholson independently, and it has been cited as either the Lord formula (Kennedy, 1988; Newman et al., 1979) or the Nicholson formula (Schmitt, 1982). It was also cited as the Herzberg formula in one study (Cummings, 1982).

2) The Burket formula.

$$\hat{R}^2 = \frac{NR^2 - P}{R(N - P)} \quad (10)$$

This formula was first presented by Burket (1964) as a direct estimate of the population validity coefficient ρ_c rather than the squared population cross-validity coefficient $\rho_c{}^2$.

The formula was also called weight validity.

3)The Darlington or Stein formula.

$$\hat{R}^2 = 1 - \left(\frac{N-1}{N-P-1}\right)\left(\frac{N-2}{N-P-2}\right)\left(\frac{N+1}{N}\right)\left(1-R^2\right) \quad (11)$$

This formula was developed as an estimator of cross-validation coefficient $\rho_c{}^2$,and it has been referred to as either the Darlington formula or the Stein formula(Cummings, 1982; Darlington, 1968; Kennedy, 1988; Kromrey& Hines, 1996;Newman et al., 1979; Schmitt, 1982; Stein, 1960; Stevens, 1996).

4) The Browne formula.

$$\hat{R}^2 = \frac{(N-p-3)\rho^4 + \rho^2}{(N-2\rho-2)\rho^2 + \rho}\left(1-R^2\right) \quad (12)$$

In this formula, ρ_c is the squared population multiple correlation coefficients. Researchers have suggested that ρ^2 be estimated by either the (Wherry formula-I) or the Olkin and Pratt formula (Schmitt, 1982).The Claudy formula-I. Claudy (1978) proposed (3) different formulas for estimating the population ρ^2 or $\rho_c{}^2$. The Claudy formula-I takes the form:

$$R^2 = \left(2\rho - R\right)^2 \quad (13)$$

Researchers have also suggested that ρ be estimated by the Wherry formula-l(Cummings, 1982).

The Claudy formula-II. Claudy (1978) proposed another formula for estimating the population $\rho_c{}^2$:

$$\hat{R}^2 = 1 - \left(\frac{N-1}{N-P-1}\right)\left(\frac{N-2}{N-P-2}\right)\left(\frac{N-1}{N}\right)\left(1-R^2\right) \quad (14)$$

This formula was presented as “the Darlington formula" (Claudy, 1978), but the only difference between the original formula in Darlington's study and the formula in several other similar studies is the minus or plus sign in the second part.That difference may be attributable to either a misprint or an incorrect citation.

5) The Rozeboom formula-I. In the literature, there are two forms of the Rozeboom formula that were developed as estimators of cross-validity coefficient $\rho_c{}^2$.The Rozeboom formula-I takes the form (Rozeboom, 1978):

$$\hat{R}^2 = 1 - \frac{N+P}{(N-P)}\left(1-R^2\right) \quad (15)$$

Rozeboom formula-II. the Rozeboom formula-2 takes the form (Rozeboom,1981):

$$\hat{R}^2 = \rho^2\left[1 + \left(\frac{P}{N-P-2}\right)\left(\frac{1-\rho^2}{\rho^2}\right)\right]^{-1} \quad (16)$$

In this formula, ρ^2 is the squared population multiple correlation coefficient. Schmitt (1982) suggested that ρ^2 be estimated by either the Wherry formula-I or the Olkin and Pratt formula. after reviewing these various analytical formulas for correcting statistical bias, we found two possible reasons for the confusion in the literature about the different analytical formulas. First, there are many correction formulas and names associated with them. We reviewed 15 formulas in the present study. For some of those formulas, multiple names were used for the same formula or thesame name was used for different formulas. Second, some of the formulas were developed as estimators of ρ^2, and some of them were developed as estimators of $\rho_c{}^2$, but the distinction has not always been made clear.

Application part:

In this part, we through the application of the data is form a 1982 national academy of science publication report rating the "scholarly quality" of research programs in the humanity physical science and social science* .the data to be presented are the quality rating of 46 research doctoral programs in psychology, as well as six potential correlates of the quality rating .here is a description of the variables: QUALITY mean rating scholarly quality of program faculty NFACULTY number of faculty members in program as of December 1980 NGRADS number of program graduates from 1975 through 1980 PCTSUPP percentage of program graduates from 1975-1979 that received fellowships or training graduated support during their graduate education PCTGRANT percent of faculty members holding research grants from the alcohol ,drug abuse and mental health administration ,the national institute of health or national science foundation at any time during 1978-1980 NARTIC number of published articles attributed to program faculty member 1978-1980 PCTPUB percent of faculty with one or more published articles from 1978-1980 .{data published in Applied multivariate statistics for the social sciences, "JAMES STEVENS" 2002 page 634}

(y_i) and six explanatory variables (X_{ij}) which data descript as follows:

y_i = QUALITY

X_{i1} = NFACULTY

X_{i2} = NGRADS

X_{i3} = PCTSUPP

X_{i4} = PCTGRANT

X_{i5} = NARTIC

X_{i6} = PCTPUB

The results of methods with statistical analysis as follows:

Fist / we can find combination multiple regression model, to computed (R, R-Square, Adjusted R Square) between variables and results models summary shows in table (1):

Table (1) model summary for combination between variable

No.	Model	No. variable	R	R Square	Adjusted R Square	S.E of the Estimate
1	**Predictors: (Constant), x6, x2, x4, x3, x1,x5**	**6**	**.891**	**.793**	**.761**	**4.921**
2	**Predictors: (Constant), x5, x4, x3, x2, x1**	**5**	**.889**	**.790**	**.763**	**4.901**
3	Predictors: (Constant), x6, x2, x4, x3, x1	5	.867	.752	.721	5.319
4	Predictors: (Constant), x4, x1, x3, x2	4	.826	.682	.651	5.953
5	Predictors: (Constant), x5, x3, x2, x1	4	.844	.713	.685	5.654
6	Predictors: (Constant), x6, x2, x3, x1	4	.816	.665	.633	6.105
7	**Predictors: (Constant), x5, x4, x3, x2**	**4**	**.880**	**.774**	**.752**	**5.019**
8	Predictors: (Constant), x6, x2, x4, x3	4	.790	.625	.588	6.464
9	Predictors: (Constant), x6, x3, x4, x5	4	.872	.761	.737	5.162
10	Predictors: (Constant), x3, x2, x1	3	.723	.523	.489	7.198
11	Predictors: (Constant), x4, x1, x2	3	.787	.620	.593	6.426
12	Predictors: (Constant), x5, x2, x1	3	.778	.605	.577	6.552
13	Predictors: (Constant), x6, x2, x1	3	.779	.606	.578	6.541
14	Predictors: (Constant), x4, x2, x3	3	.735	.540	.508	7.067
15	Predictors: (Constant), x5, x3, x2	3	.841	.708	.687	5.633
16	Predictors: (Constant), x6, x2, x3	3	.742	.550	.518	6.994
17	**Predictors: (Constant), x5, x4, x3**	**3**	**.871**	**.758**	**.741**	**5.128**
18	Predictors: (Constant), x6, x4, x5	3	.835	.697	.675	5.737
19	Predictors: (Constant), x6, x3, x4	3	.767	.588	.558	6.692
20	Predictors: (Constant), x2, x1	2	.622	.387	.358	8.069
21	Predictors: (Constant), x3, x1	2	.721	.521	.498	7.135
22	Predictors: (Constant), x4, x1	2	.784	.615	.597	6.394
23	Predictors: (Constant), x5, x2	2	.768	.590	.571	6.594
24	Predictors: (Constant), x6, x2	2	.668	.446	.420	7.667
25	Predictors: (Constant), x6, x3	2	.711	.505	.482	7.245
26	**Predictors: (Constant), x5, x4**	**2**	**.831**	**.690**	**.676**	**5.735**
27	Predictors: (Constant), x6, x4	2	.688	.474	.449	7.474
28	Predictors: (Constant), x6, x5	2	.780	.608	.590	6.450
29	Predictors: (Constant), x6, x1	2	.778	.605	.587	6.476
30	Predictors: (Constant), x3, x2	2	.628	.394	.366	8.020
31	Predictors: (Constant), x4, x2	2	.647	.419	.392	7.854
32	Predictors: (Constant), x5, x3	2	.829	.687	.672	5.765
33	Predictors: (Constant), x4, x3	2	.707	.499	.476	7.289
(QUALTY) Dependent Variable: y						

Second /finding maximum correlation between variable to estimate (R^2 shrinkage) by applying (6) equation. Table (2) shows these results.

No.	Equation name	Number of variable				
		6 variable	5 variable	4 variable	3 variable	2 variable
1	Smith formula	0.76	0.76	0.75	0.74	0.67
2	Wherry formula 1	0.76	0.76	0.75	0.74	0.67
3	Wherry formula 2	0.77	0.76	0.75	0.74	0.68
4	Olkin & part formula 1	0.91	0.77	0.76	0.74	0.68
or	Olkin & part formula 2	0.76	0.77	0.76	0.75	0.68
or	Olkin & part formula 3	0.76	0.77	0.76	0.75	0.68
5	Partt formula	0.76	0.77	0.76	0.75	0.68
6	Claudy formula	0.77	0.77	0.76	0.75	0.65

Table (2) Eestimate R^2 Shrinkage

Third / finding and estimating (ρ_c^2) by applying (9) equation .Table (3) show results as follow:

Table (3) Estimate of ρ_c^2

No.	Equation name	Number of variable				
		6 variable	5 variable	4 variable	3 variable	2 variable
1	Lord formula 1	0.71	0.72	0.71	0.71	0.62
2	Lord formula 2	0.72	0.79	0.72	0.76	0.65
3	Burket formula	0.85	0.86	0.85	0.87	0.81
4	Darlington & Stein	0.72	0.72	0.67	0.71	0.65
5	Browne formula	0.53	0.52	0.28	0.52	0.19
6	Claudy formula 1	0.39	**0.54**	0.38	0.37	0.12
7	Claudy formula 2	0.72	**0.73**	0.72	0.72	0.66
8	Rezeboon formula 1	0.73	**0.73**	0.72	0.72	0.66
9	Rezeboon formula	0.72	0.53	0.52	0.72	0.64

IX. DISCUSSION

In above analysis we investigated the effectiveness of various analytical formulas designed to Estimate R^2 Shrinkage in multiple regression, analytical formulas were applied to the sample and the adjusted R^2 and R_c^2 were obtained and then compared with corresponding population parameters (**ρ^2** and ρ_c^2). first object was to compare the accuracy and usefulness of various analytical formulas for estimating the population (ρ^2). Among the 6 analytical formulas designed to estimate the population ρ^2 , the performance of the Olkin & part formula-1 for six variable then followed by Burket formula & Lord formula-2 among the 9 analytical formulas were found to be most stable and satisfactory.

REFERENCE

[1] Abdel H. El-Shaarawi and Walter W. Piegorsch. (2002)" Shrinkage regression" John Wiley & Sons, Ltd, Chichester.

[2] Carter, D.S.(1979). "Comparison of different shrinkage formulas in estimating population multiple correlation coefficients. educational and psychological measurement.39, 261-266.

[3] Chatterjee, S and Hadi,A.S.(1988)."Sensitivity Analysis in Regression" New York: john wiley.

[4] Frank E. Harrell Jr, 2001," Regression Modeling Strategies, Springer Science, New York.

[5] Glass, G. V, & Hopkins,K.D. (1996). "Statistical Methods in Education and psychology. Needham height, MA. Allyn & Bacon.

[6] James Stevens, 2002"Applied multivariate statistics for the social sciences "Mahwah, new jersey, London. 4th edition.

[7] Kennedy, E. (1988).Estimation of the squared cross-validity coefficient in the context of best subset regression. Applied psychological measurement. 12(3), 231-237.

[8] Kruse.D. J. & Fuller,E.A. (1982) "Multicross –validation in regression analysis" Educational and psychological measurement ,40,101-112.

[9] Vinod,H.D. and Ullah,A.(1981)." Recent Advances in Regression Methods "New York: Marcel Dekker.

[10] Marvin H.J.Gruber, (1998)"Improving efficiency by shrinkage", Marcel Dekker ,USA.

[11] Neter, J. Kutner. M. H. Nachtsheim. C. J. & Wasserman, W. (1996)" Applied linear regression models. Chicagb:Iewin.

[12] Tukey, J. W. (1975). "Applied Statistics "R.P. Gupta. Amsterdam-New York: North Holland Publishing Company.

[13] Ping Yin, Xitao Fan (2001)"estimating shrinkage in multiple regression a comparison of different analytical methods" journal of experimental education.