International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

The Effect of Grouping Data by using Arithmetic Mean on the Significance of the Multiple Linear Regression Model

Adil Mousa Younis Waniss[1], Abdulmajied Ali Balkash[2]

[1]Associate Professor, Department of Management Information Systems, College of Administration and Economics, Qassim University, *Buraidah, P.O. Box 6633*, KSA
Email: *younisadil2002[at]gmail.com*

[2]Professor, Department of Management Information Systems, College of Administration and Economics, Qassim University, *Buraidah, P.O. Box 6633*, KSA,
Email: *d.balkash[at]gmail.com*

Abstract: ***This paper is completely focused on the impact of data aggregation on the multiple linear regression estimation particularly on the significance of the model and the multiple coefficient of determination , we generate random numbers from Excel and built multiple linear regression model with two independent variables ,we compared ungrouped multiple linear regression model with different five aggregated linear regression models, we find that, both F test and R-square are increasing with the increase of aggregation, except in case of grouping by less than five grouped where a Sharpe decreases happened on both of them. Aggregation gives misleading results for the coefficient of determination, but doesn't affect the significance of the different models. We find that grouping should not goes above twenty groups nor below five groups. The most important flinging is that; all ANOVA tables gives same p- values.***

Keywords: Aggregation, multiple linear regression. Coefficient of determination, F-test, ungrouped model

1. Introduction

The aggregation problem is a common problem in data analysis in nearly all fields of study, including social science and some fields of physics. In its broadest definition, the aggregation problem is the loss of information that occurs when aggregate, or large-scale, data is replaced by individual, or small-scale, data. Data aggregation is the process of collecting data to present it in summary form. This information is then used to conduct statistical analysis and can also help company executives make more informed decisions about marketing strategies, price settings, and structuring operations, among other things (Lukas Racickas 2023).

The use of random estimates in regression models has been gaining more attention in recent years (4, 5, 6). New ideas on combining different procedures for estimation, coding, forecasting and learning have recently been considered in statistics and several related fields, leading to a number of very interesting results (Y Yang Bernoulli, 2004 projecteuclid.org).

The purpose of this paper is to demonstrate the effect of aggregation on the significance of the multiple linear regression when the data is ungrouped against grouped data. We will show on the following paragraphs different types of grouped data (15,10,8, 6,5,4) class intervals, along with their ANOVA tables and coefficients of determination for each class interval and compare them with their cross ponding ungrouped measures with discussion of results. Throw out these paragraphs a detailed explanation of how aggregation is done by Excel and what types of functions we used.

2. Multiple Linear Regression for Un-grouped data

The multiple linear regression model is

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + \varepsilon \qquad (1)$$

Where y is the dependent variable ,are the independent x_i variables, $\beta_o, \beta_1, \beta_2, , \cdots, \beta_{k-1}$ are the regression coefficients of the model and ε is the residuals, to minimize the sum of the squares of the residuals we use the least square method to estimate the regression coefficients of the model. where, K -1 is the number of the independent variables, and K is the number of the regression coefficients (kor & altun,2020).

$S^2(b_0), S^2(b_1), S^2(b_2), \cdots, S^2(b_{k-1})$ are the variances of errors of the regression coefficients, n is sample size and

$$\sigma^2 = \frac{SSE}{n-k} \qquad (2)$$

Is the estimated variance of the estimated model?
The estimated model is:

$$\hat{y} = b_o + b_1 x_1 + b_2 x_2 + \cdots + b_{k-1} x_{k-1} \qquad (3)$$

The parameters of the model are estimated by the least square method as follows:

$$B = A^{-1} \cdot C \qquad (4)$$

The matrix is defined as follows:

$$A=\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}\cdot x_{2i} \\ \sum x_{2i} & \sum x_{1i}\cdot x_{2i} & \sum x_{2i}^2 \end{bmatrix},\ B=\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix},$$

$$C=\begin{bmatrix} \sum y_i \\ \sum x_{1i}\cdot y_i \\ \sum x_{2i}\cdot y_i \end{bmatrix} \quad (5)$$

The matrix variance is

$$\sigma^2=\frac{SSE}{n-k}\cdot A^{-1} \quad (6)$$

$$SST=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2,$$

$$SSR=B^T\times C-n\bar{y}^2, SST-SSR=\sum_{i=1}^{n}y_i^2-B^T\times C$$

$$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

2.1 The Multiple Coefficient of Determination

The multiple coefficient of determination is the measure which tell us how much of the dependent variable is explained by the independent variables, it lies between zero and one and it increase whenever we add a new independent variable to the model (Forst,2023).

$$R^2=\frac{SSR}{SST}=1-\frac{SSE}{SST} \quad (7)$$

And the correlation coefficient is:

$$r=\sqrt{R^2} \quad (8)$$

2.2 The Test of Significance of the Multiple Regression Model is:

Table 1: ANOVA for Multiple Linear Regression Model

SS	*df*	*MSS*	Test Statistic
SSR	$k-1$	$MSR=SSR/k-1$	$F=MSR/MSE$
SSE	$n-k$	$MSE=SSE/n-k$	
SST	$n-1$		

3. Multiple Linear Regression for Grouped Data

To calculate the above 7 equations for grouped data we aggregate the data according to the following steps:

3.1 Data is aggregated for different class intervals, 15, 10, 8, 6, 5 and 4 class intervals with upper and lower boundary and frequency for each class interval.

3.2 Each class interval frequencies will be represented by mean value instead of mid-point as usual.

3.3 For calculating the ANOVA table and the coefficient of determination, we use the same procedures as for ungrouped data but we take in account the frequencies.

3.4 The grouping of the data is don only for the independent variable y while the grouping of the independent variables x_1, x_2 is done automatically according to the rule of aggregation .

3.5 The parameters of the grouped model are estimated by the least square method for 15 class interval and 4 class interval with their means and frequencies as follows:

$$A=\begin{bmatrix} \sum_{i=1}^{15} f_i & \sum_{i=1}^{15}\bar{x}_{1i}\cdot f_i & \sum_{i=1}^{15}\bar{x}_{2i}\cdot f_i \\ \sum_{i=1}^{15}\bar{x}_{1i}\cdot f_i & \sum_{i=1}^{15}\bar{x}_{1i}^2\cdot f_i & \sum_{i=1}^{15}\bar{x}_{1i}\cdot\bar{x}_{2i}\cdot f_i \\ \sum_{i=1}^{15}\bar{x}_{2i}\cdot f_i & \sum_{i=1}^{15}\bar{x}_{1i}\cdot\bar{x}_{2i}\cdot f_i & \sum_{i=1}^{15}\bar{x}_{2i}^2\cdot f_i \end{bmatrix},$$

$$B=\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix},\quad C=\begin{bmatrix} \sum_{i=1}^{15}\bar{y}_i\cdot f_i \\ \sum_{i=1}^{15}\bar{x}_{1i}\bar{y}_i\cdot f_i \\ \sum_{i=1}^{15}\bar{x}_{2i}\cdot\bar{y}_i\cdot f_i \end{bmatrix} \quad (9)$$

For ANOVA table with 15 class intervals

$$SST=\sum_{i=1}^{15}\bar{y}_i^2\cdot f_i-\sum_{i=1}^{15}f_i\cdot\bar{\bar{y}}_i^2,$$

$$SSR=B^T\times C-\sum_{i=1}^{15}f_i\bar{\bar{y}}^2, SSE=SST-SSR \quad (10)$$

$$A=\begin{bmatrix} \sum_{i=1}^{4} f_i & \sum_{i=1}^{4}\bar{x}_{1i}\cdot f_i & \sum_{i=1}^{4}\bar{x}_{2i}\cdot f_i \\ \sum_{i=1}^{4}\bar{x}_{1i}\cdot f_i & \sum_{i=1}^{4}\bar{x}_{1i}^2\cdot f_i & \sum_{i=1}^{4}\bar{x}_{1i}\cdot\bar{x}_{2i}\cdot f_i \\ \sum_{i=1}^{4}\bar{x}_{2i}\cdot f_i & \sum_{i=1}^{4}\bar{x}_{1i}\cdot\bar{x}_{2i}\cdot f_i & \sum_{i=1}^{4}\bar{x}_{2i}^2\cdot f_i \end{bmatrix},$$

$$B=\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix},\quad C=\begin{bmatrix} \sum_{i=1}^{4}\bar{y}_i\cdot f_i \\ \sum_{i=1}^{4}\bar{x}_{1i}\bar{y}_i\cdot f_i \\ \sum_{i=1}^{4}\bar{x}_{2i}\cdot\bar{y}_i\cdot f_i \end{bmatrix} \quad (11)$$

For ANOVA table with 4 class intervals

$$SST=\sum_{i=1}^{4}\bar{y}_i^2\cdot f_i-\sum_{i=1}^{4}f_i\cdot\bar{\bar{y}}_i^2,$$

$$SSR=B^T\times C-\sum_{i=1}^{4}f_i\bar{\bar{y}}^2, SSE=SST-SSR \quad (12)$$

Similarly, we follow the same procedures for the class, 10, 8 and 6.

4. Application and Results

We use Excel (2010) to generate 130 random numbers and we defined them as is the demand for a certain commodity are the price of the commodity and monthly income rate respectively.

4.1. Calculation of Un-grouped data:

We calculate the multiple regression model for ungrouped data from Excel (2010), using both of "Linest" and "Regression Statistics" (Greg Harvey, Microsoft Excel 2010).

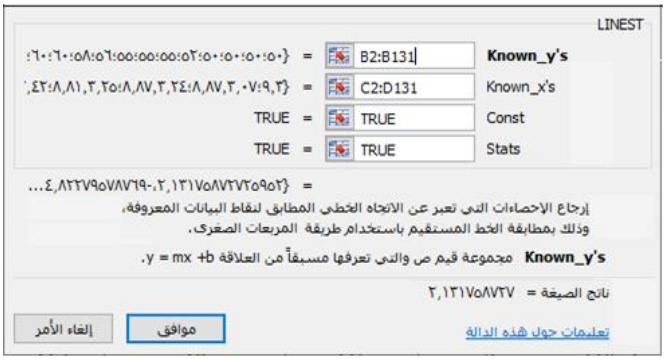


Table (2): Statistics for ungrouped Multiple Linear Regression Model

Multiple R	R Square	Adjusted R Square	Standard Error	n
0.990396	0.980883	0.980582	2.634596	130

Table (3): ANOVA for Multiple Linear Regression Ungrouped Model

S.O.V	*SS*	*df*	*MSS*	Test Statistic	Significance F
Regression	45231.01	2	22615.51	3258.203	7.4E-110
Residual	881.5193	127	6.941097		
Total	46112.53	129			

4.2. Calculation of Grouped data:

The calculation for 15 class intervals grouping is done by Excel as follows:

No	Upper-boundary	f	MEANY	MEANX1	MEANX2
1	55	8	52.125	8.82625	3.33
2	60	10	59.4	8.364	4.05
3	65	9	64.667	7.841	4.71
4	70	9	68.667	7.514	5.76
5	75	15	73.667	6.972	7.16
6	80	6	78.333	6.502	8.38
7	85	10	83.7	6.01	9.37
8	90	16	88.438	5.19375	10.94
9	95	14	94.143	4.249	12.29
10	100	10	99.4	3.486	13.54
11	105	4	104.75	2.985	14.40
12	110	5	107.6	2.786	14.64
13	115	6	113.5	2.573	15.18
14	120	6	118.5	2.182	15.82
15	125	2	125	2.005	16.80
		130	10949	731.99	1245.63

$$A=\begin{bmatrix} \sum_{i=1}^{15} f_i & \sum_{i=1}^{15} \bar{x}_{1i}\cdot f_i & \sum_{i=1}^{15} \bar{x}_{2i}\cdot f_i \\ \sum_{i=1}^{15} \bar{x}_{1i}\cdot f_i & \sum_{i=1}^{15} \bar{x}_{1i}^2\cdot f_i & \sum_{i=1}^{15} \bar{x}_{1i}\cdot \bar{x}_{2i}\cdot f_i \\ \sum_{i=1}^{15} \bar{x}_{2i}\cdot f_i & \sum_{i=1}^{15} \bar{x}_{1i}\cdot \bar{x}_{2i}\cdot f_i & \sum_{i=1}^{15} \bar{x}_{2i}^2\cdot f_i \end{bmatrix}$$

$$=\begin{bmatrix} 130 & 732 & 1246 \\ 732 & 4685 & 5931 \\ 1246 & 5931 & 14032 \end{bmatrix}$$

$$C=\begin{bmatrix} \sum_{i=1}^{15} \bar{y}_i\cdot f_i \\ \sum_{i=1}^{15} \bar{x}_{1i}\bar{y}_i\cdot f_i \\ \sum_{i=1}^{15} \bar{x}_{2i}\cdot \bar{y}_i\cdot f_i \end{bmatrix}=\begin{bmatrix} 10949 \\ 56615 \\ 114619 \end{bmatrix}$$

$$A^{-1}=\begin{bmatrix} 25.42 & -2.4 & -1.24 \\ -2.4 & 0.23 & 0.12 \\ -1.24 & 0.12 & 0.06 \end{bmatrix}$$

$$B=A^{-1}\times C=\begin{bmatrix} 25.42 & -2.4 & -1.24 \\ -2.4 & 0.23 & 0.12 \\ -1.24 & 0.12 & 0.06 \end{bmatrix}\times\begin{bmatrix} 10949 \\ 56615 \\ 114619 \end{bmatrix}$$

$$=\begin{bmatrix} 94.517 \\ -5.158 \\ 1.963 \end{bmatrix}=\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

The estimated grouped model with 15 class interval is:

$$y=94.517-5.158x_1+1.963x_2$$

And for ANOVA table, we have:

$$SST=\sum_{i=1}^{15}\bar{y}_i^2\cdot f_i-\sum_{i=1}^{15} f_i\cdot\bar{\bar{y}}_i^2,$$

$$SST=967965.96-922158.47=45807.49$$

$$SSR=B^T\times C-\sum_{i=1}^{15} f_i\bar{\bar{y}}^2,$$

$$SSR=\begin{bmatrix}94.517 & -5.158 & 1.963\end{bmatrix}\times\begin{bmatrix}10949 \\ 56615 \\ 114619\end{bmatrix}$$

$$SSR=-84.22308^2\times 130=45081.38956$$

$$SSE = SST - SSR = 45807.49 - 45081.39 = 726.098$$

$$\bar{\bar{y}}_i = 8422308$$

$$\sigma = \sqrt{\frac{SSE}{n-k} \cdot A^{-1}} =$$

$$\sqrt{\frac{726.09775}{130-3} \cdot \begin{bmatrix} 25.42 & -2.4 & -1.24 \\ -2.4 & 0.23 & 0.12 \\ -1.24 & 0.12 & 0.06 \end{bmatrix}}$$

$$= \begin{bmatrix} 12.05586 & & \\ & 1.1385 & \\ & & 0.590 \end{bmatrix}$$

$S(b_0) = 12.05586$, $S(b_1) = 1.1385$, $S(b_2) = 0.590$

and $R^2 = \frac{SSR}{SST} = \frac{45081.38956}{45807.49} = 0.984149$

and $R = \sqrt{0.984149} = 0.9920428$

For the test of significance for the grouped model, the statistical test *F* is:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1} = \frac{0.984149}{1-0.984149} \cdot \frac{127}{2} = 3942.538$$

$$F(\alpha; k-1; n-k) = F(0.05; 2; 127) = 3.068$$

Which is clearly highly significant, if we compare it with the results of significance of un-grouped model, we notice that, the significance of the grouped model has increased for the statistical test F, from 3258.20 to 3942.54, as well as for the coefficient of determination i.e.0.980883 for ungrouped with 0.984149 after grouping, similarly we repeat the same calculation for the different groups i.e. 10, 8, 6, 5 and 4 but, for the paper, we will presents the procedures for grouping with 4 class interval.

The calculation for 4 class intervals grouping is done by Excel as follows:

No	Upper-boundary	f	MEANY	MEANX1	MEANX2
1	69	30	59.7	8.255	4.18733
2	88	45	78.378	6.498	8.240
3	107	39	96.641	3.947	12.826
4	126	16	116.375	2.375	15.5625
		130	10949	731.99	1245.63

$$A = \begin{bmatrix} \sum_{i=1}^{4} f_i & \sum_{i=1}^{4} \bar{x}_{1i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{2i} \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{1i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{1i}^2 \cdot f_i & \sum_{i=1}^{4} \bar{x}_{1i} \cdot \bar{x}_{2i} \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{2i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{1i} \cdot \bar{x}_{2i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{2i}^2 \cdot f_i \end{bmatrix},$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}, \qquad C = \begin{bmatrix} \sum_{i=1}^{4} \bar{y}_i \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{1i} \bar{y}_i \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{2i} \cdot \bar{y}_i \cdot f_i \end{bmatrix}$$

$$A = \begin{bmatrix} \sum_{i=1}^{4} f_i & \sum_{i=1}^{4} \bar{x}_{1i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{2i} \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{1i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{1i}^2 \cdot f_i & \sum_{i=1}^{4} \bar{x}_{1i} \cdot \bar{x}_{2i} \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{2i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{1i} \cdot \bar{x}_{2i} \cdot f_i & \sum_{i=1}^{4} \bar{x}_{2i}^2 \cdot f_i \end{bmatrix}$$

$$= \begin{bmatrix} 130 & 731.99 & 1245.63 \\ 731.99 & 4642.231 & 6012.14 \\ 1245.63 & 6012.14 & 13872.02 \end{bmatrix} \Rightarrow$$

$$C = \begin{bmatrix} \sum_{i=1}^{4} y_i \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{1i} y_i \cdot f_i \\ \sum_{i=1}^{4} \bar{x}_{2i} \cdot y_i \cdot f_i \end{bmatrix} = \begin{bmatrix} 10949 \\ 57001.35 \\ 113879.99 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 43.14 & -4.07 & -2.11 \\ -7.07 & 0.38 & 0.19 \\ -2.11 & 0.19 & 0.10 \end{bmatrix}$$

$$B = A^{-1} \times C = \begin{bmatrix} 43.14 & -4.07 & -2.11 \\ -7.07 & 0.38 & 0.19 \\ -2.11 & 0.19 & 0.10 \end{bmatrix} \times$$

$$\begin{bmatrix} 10949 \\ 57001.35 \\ 113879.99 \end{bmatrix} = \begin{bmatrix} 82.08 \\ -3.99 \\ 2.57 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

The estimated grouped model with 4 class interval is:

$$y = 82.08 - 3.99_1 x_1 + 2.57 x_2$$

And for ANOVA table, we have

$$SST = \sum_{i=1}^{4} \bar{y}_i^2 \cdot f_i - \sum_{i=1}^{4} f_i \cdot \bar{\bar{y}}_i^2,$$

$$SST = 964291.4 - 922158.47 = 42132.9286$$

$$SSR = B^T \times C - \sum_{i=1}^{4} f_i \bar{\bar{y}}^2$$

$$SSR = \begin{bmatrix} 82.084 & -3.99 & 2.57 \end{bmatrix} \times \begin{bmatrix} 10949 \\ 57001.35 \\ 113879.99 \end{bmatrix} -$$

$$84.22308^2 \times 130 = 41579.84$$

$$SSE = SST - SSR = 42132.9286 - 41579.84 = 553.0866$$

$$\sigma = \sqrt{\frac{SSE}{n-k} \cdot A^{-1}} =$$

$$\sqrt{\frac{553.0866}{130-3} \cdot \begin{bmatrix} 43.14 & -4.07 & -2.11 \\ -7.07 & 0.38 & 0.19 \\ -2.11 & 0.19 & 0.10 \end{bmatrix}} =$$

$$\begin{bmatrix} 13.706 & & \\ & 1.29 & \\ & & 0.670 \end{bmatrix}$$

Hence:

$S(b_0) = 13.706$, $S(b_1) = 1.29$, $S(b_2) = 0.670$

And $R^2 = \frac{SSR}{SST} = \frac{41579.84}{42132.9286} = 0.98687282$

$R = \sqrt{0.98687282} = 0.9934147$

For the test of significance for the grouped model, the statistical test F is:

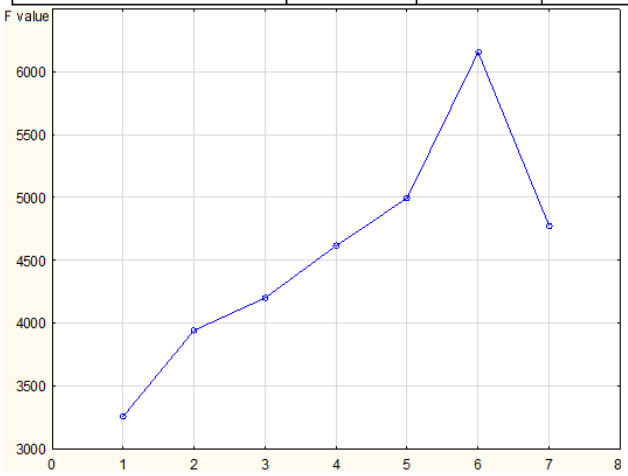
$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1} = \frac{0.98687282}{1-0.98687282} \cdot \frac{127}{2} = 4773.791$$

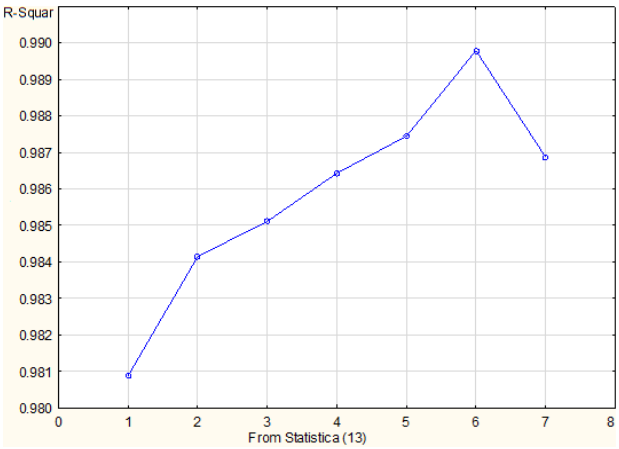
$$F(\alpha; k-1; n-k) = F(0.05; 2; 127) = 3.068$$

Which is clearly highly significant, if we compare it with the results of significance of un-grouped model, we notice that, the significance of the grouped model has increased for the statistical test F, from 3258.20 to 4773.7909, as well as for the coefficient of determination i.e.0.980883 for ungrouped with 0.98687282 after grouping. The followings tables and figures shows clearly the increments for both the values of grouped F statistics as well the coefficients of determination comparing with ungrouped once.

Table (6): F statistics and R-square for ungrouped and different grouped

	F- statistic	R Square	P-value
Ungrouped= 1	3258.20	0.980883	7.4E-110
15 class intervals grouped=2	3942.538	0.984149	5.1E-115
10 class intervals grouped=3	4201.65	0.985112	9.4E-117
8 class intervals grouped=4	4615.789	0.98643	2.6E-119
6 class intervals grouped=5	4994.400	0.987445	1.9E-121
5 class intervals grouped=6	6153.747	0.989786	3.8E-127
4 class intervals grouped=7	4773.7909	0.9868728	3.2E-120





4.3 Results Discussion

From table (6) and the above two Figures, we notice the followings:

1) Both F statistics and R-square are increasing as grouping is increasing from ungrouped throw 15 grouped up to 4 grouped.
2) In 4 grouped there is a sharp decreasing in both F statistics and R-square.
3) All the grouped as well as ungrouped linear regression models were significant.
4) All p-values indicates that significances were very height.
5) Groupings gives almost same results expect in case of height grouping i.e. 4 grouped.
6) Groupings has no effect on the F statistics i.e. the test of significance
7) Groupings has drastic effect on R-square values.
8) We Should use 15 to 20 class intervals for Grouping at most.
9) We Should use 5 class intervals for Grouping at least.
10) Aggregation or grouping gives misleading R-square results.

References

[1] *Lukas Racickas March 21,2023 aggregation: Definition, Benefits, and Examples*. *https://coresignal.com/blog/data- aggregation/*
[2] Aggregating regression procedures to improve performance Y Yang Bernoulli, 2004 project Euclid .org.
[3] [Adil M. Youniss, a PhD dissertation 2002].
[4] JUDITSKY, A. and NEMIROVSKI, A. (2000) Functional aggregation for nonparametric regression. Ann. Statist. 28 681–712. MR1792783
[5] YANG, Y. (2004). Aggregating regression procedures to improve performance. Bernoulli 10 25–47. MR2044592
[6] TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In Learning Theory and Kernel Machines. Lecture Notes in Artificial Intelligence 2777 303–313. Springer, Heidelberg.
[7] Greg Harvey, Microsoft Excel 2010 All-in-One for Dummies, Published by Wiley Publishing, Inc. 111 River Street Hoboken, NJ 07030-5774.
[8] Kor. K. & Altun, G., 2020. Is Support Vector Regression method suitable for predicting rate. Journal of Petroleum Science and Engineering,194.
[9] Frost, J., 2023. Statistics By Jim Making statistics intuitive. [Online] Available at: https://statisticsbyjim.com/regression/mean-squared-error-mse/.