Visualizing Multiple Regression

Edward H. S. Ip
University of Southern California

Journal of Statistics Education Volume 9, Number 1 (2001)

Copyright © 2001 by Edward H. S. Ip, all rights reserved.
This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

Key Words: Average stepwise regression; Teaching statistics; Type I and Type II sums of squares; Venn diagram.

Abstract

Several examples are presented to demonstrate how Venn diagramming can be used to help students visualize multiple regression concepts such as the coefficient of determination, the multiple partial correlation, and the Type I and Type II sums of squares. In addition, it is suggested that Venn diagramming can aid in the interpretation of a measure of variable importance obtained by average stepwise selection. Finally, we report findings of an experiment that compared outcomes of two instructional methods for multiple regression, one using Venn diagrams and one not.

1. Introduction

One of the topics students encounter in statistics courses at both the undergraduate and the graduate level is multiple regression. This paper shows how the Venn diagram can be employed as a useful visual aid to help students understand important and fundamental concepts in multiple regression such as \( R^2 \), partial correlation, and Type I and II sums of squares. Introduced by Venn (1880), the Venn diagram has been popularized in texts on elementary logic and set theory (e.g., Suppes 1957). However, the use of Venn diagrams in the field of statistics has been quite limited. In a recent example, Shavelson and Webb (1990) used them in generalizability studies to make visually accessible the partitioning of total variance into components. Moreover, the Venn diagram has been used to illustrate correlation and regression (e.g., Pedhazur 1997; Hair, Anderson, and Tatham 1992, p. 47). While there are also good applications of Venn diagrams in a number of statistics texts (e.g., Agresti and Finlay 1997), just seeing them does not necessarily inform the lecturer about critical issues in creating them. The purpose of this article is to illustrate in a variety of ways that more extensive use of Venn diagrams can be made in the classroom. Their clearest application in these contexts requires examples with no more than three independent variables whose interrelationships explicitly avoid suppressor variable effects.

2. Venn Diagramming

A Venn diagram for regression displays the total sum of squares (TSS) as a rectangular box. Sums of squares (SS) of individual variables are depicted as ovals. Whenever numerical examples are demonstrated, shapes should be drawn to scale so that the effects of the variables can be interpreted accurately.

2.1 Coefficient of Determination \( R^2 \)

The coefficient of determination \( R^2 \) is the ratio of the sum of squares of regression (SSR), the total area covered by ovals, and TSS, the area of the rectangle. The case in which the variables are uncorrelated can be represented by separated ovals in the Venn diagram. For example, Figure 1a shows what happens when the variables \( x_1 \) and \( x_2 \) are uncorrelated. It is clear from the figure that \( R^2 = r_{x_1}^2 + r_{x_2}^2 \).
When the variables are correlated and contain redundant information, they can be represented by overlapping ovals. The overlapping part indicates the redundant information shared between the two related variables. A dataset is taken from the Student Edition of Minitab for Windows (McKenzie, Schaefer, and Farber 1995, p. T-21) to illustrate this situation. It consists of data on the annual salary (in thousands of dollars) of employees in a company. The predictor variables are gender and Nsuper, the number of staff under supervision by an individual. The sums of squares are SS(gender) = 337, SS(gender|Nsuper) = 212, SS(Nsuper) = 1494, and SS(gender, Nsuper) = 1706. These SS are graphically represented in Figure 1b. With the aid of the diagram, instructors can actually point to a piece that represents a particular SS. The ratio of "ground covered" by the ovals to the total area of the rectangle equals $R^2$. Since adding ovals (variables) always increases the "ground covered," the concept that "$R^2$ will always increase as a result of adding variables" can be easily appreciated by students with the aid of the diagram.

The validity of Figure 1b in illustrating the "overlap" of predictive information depends crucially on the fact that \( SS(\text{gender}) + SS(\text{Nsuper}) > SS(\text{gender, Nsuper}) \). Unfortunately, although the inequality \( SS(x_1) + SS(x_2) > SS(x_1, x_2) \) holds most of the time in practice, exceptions do occur, and when they do, the areas of overlap are not positive. This will be discussed in Section 3. In this section, it will be assumed that the overlapping areas are all positive.

### 2.2 Generalization of $R^2$

Various forms of generalization of $R^2$ can be found in the literature. One generalization is described in Pedhazur (1982). The generalized $R^2$ is a measure of the predictive power of a variable after partialing out another. The square of the partial correlation, as the measure is called, is defined as

\[
r^2_{y|x_1} = \frac{SS(x_2 | x_1)}{TSS - SS(x_1)} = \frac{SS(x_1, x_2) - SS(x_1)}{TSS - SS(x_1)}.
\]

A visual representation of $R^2$ in Figure 2 indicates the SSR contributed by $x_1$ and $x_2$ (shaded area in Figure 2a) when both variables are included in the regression model. Partialing out $x_1$ is equivalent to taking out the piece of SS that belongs to $x_1$ and treating the remaining area as the new TSS (Figure 2b). The residualized SS that is explained by $x_2$ can be represented by the shaded area, the ratio of which to the eclipsed TSS is the squared partial correlation $r^2_{x_1|x_2}$, sometimes referred to as the coefficient of partial determination in the regression context.
The notion of partial correlation can readily be extended to the multiple variable case with the aid of a Venn diagram. It takes little effort to complete the generalization of the multiple partial correlation to one that partials out more than one variable. Suppose there are four variables, $x_1, x_2, x_3, x_4$. Figure 3 shows a rectangle after partialing out both $x_1$ and $x_2$. The squared multiple partial correlation of $x_3$ and $x_4$ is based on the ratio of the area covered by $(x_3, x_4)$ and the eclipsed TSS. Generalizing the concept to more variables can be illustrated using Figure 3. For example, the multiple partial correlation of $(x_3, x_4)$ with two variables $(x_1, x_2)$ partialed out is given by

$$r_{y_{x_{1}, x_{2}}|x_{3}, x_{4}}^2 = \frac{SS(x_3, x_4 | x_1, x_2)}{TSS - SS(x_1, x_2)} = \frac{R_{y_{x_{1}, x_{2}}|x_{3}, x_{4}}^2 - R_{y_{x_{1}, x_{2}}}^2}{1 - R_{y_{x_{1}, x_{2}}}^2}.$$
There are several types of sums of squares used in the literature on linear models. The most commonly used SS reported in statistical packages are the Type I and Type II SS. A discussion of SS and related references can be found in the SAS/STAT User's Guide (SAS Institute Inc. 1990). The Type I SS is the SS of a predictor after adjusting for the effects of the preceding predictors in the model. For example, when there are three predictors, and their order in entering the equation is $x_1, x_2, x_3$, the Type I SS are SS($x_1$), SS($x_2 | x_1$), and SS($x_3 | x_2, x_1$). The Type I SS would not be the same if the variables entered the equation in a different order. The fact that Type I SS are model-order dependent is illustrated by the Venn diagram in Figure 4. The Type I SS of $x_2$ in (a) and (b) are, respectively, SS($x_2 | x_1$) and SS($x_2 | x_1, x_3$). The diagram helps instructors explain the arbitrariness of using the incremental SS such as the Type I SS, or the incremental $R^2$ in procedures such as forward selection designed to isolate the variable(s) of importance.

Figure 4. Type I SS for $x_2$ (Shaded Region) When the Order is (a) $x_1, x_2, x_3$; (b) $x_1, x_3, x_2$.

### 2.4 Type II SS

When the SS for each predictor is adjusted for all the other predictors in the regression equation, the resulting SS is called the Type II SS. In the three-predictor example, the Type II SSs are SS($x_1 | x_2, x_3$), SS($x_2 | x_1, x_3$), and SS($x_3 | x_2, x_1$). Each Type II SS represents the effect of the predictor when it is treated as the last predictor that enters the equation. See Figure 5 for an illustration.

![Type II SS Diagram](image-url)
Venn diagramming illustrates not only the Type II SS, but also the effect of multicollinearity. When multicollinearity exists between predictors, the effect of each predictor, as measured by its Type II SS, and thus when treated as the "last predictor in," may be insignificant even when the predictor is a significant one on its own. Chatterjee and Price (1977, p. 144) provide an example using achievement data that illustrates this. The response variable is a measure of achievement, and the three continuous predictors are indexes of family, peer group, and school. The first twenty data points in the example were used in a regression analysis, and the breakdown of the SS is shown in Table 1. The total SS equals 87.6, and $R^2 = 0.324$. The Venn diagram for this example appears in Figure 6. The "ground not covered" by any variable represents the SS for error (SSE) and is 59.2.

Table 1. SS of Partitions in the Venn Diagram in Figure 6

| Variable                  | SS  |
|---------------------------|-----|
| family only               | 0.8 |
| peer group only           | 8.3 |
| school only               | 0.4 |
| family and peer group only| 0.7 |
| family and school only    | 4.2 |
| school and peer group only| 3.3 |
| family, school, and peer group | 10.7 |
| Total SSR                 | 28.4 |

The $F$ statistic is given by $[\text{SS(family, peer group, school)/df(model)}]/[\text{SSE/df(error)}]$. This ratio is proportional to (area covered) / (area not covered) in the Venn diagram. For this example, $F = 2.55$ with $df = 3, 16$, and is significant at the $\alpha = 0.1$ level. However, none of the $t$-tests for the individual predictors are significant at the $\alpha = 0.1$ level. The $p$-values for family, peer group, and school are 0.648, 0.153, and 0.753, respectively. Note that a $t$ test for an individual variable -- family, for example -- is equivalent to an $F$ test ($df = 1, 16$) with its $F$ statistic being proportional to $\text{SS(family) / SSE}$, the area covered by family as the last predictor in, divided by the area not covered. The Venn diagram in Figure 6 illustrates that given the great deal of overlap among the variables (multicollinearity), even...
when the "ground covered" jointly by all three is substantial (leading to a significant overall F-test), the additional "ground covered" by each variable given the others may not be significant (leading to insignificant t-tests).

### 2.5 Average Stepwise Regression

Kruskal (1987) suggests an average stepwise approach for assessing the relative importance of a variable. When $k$ explanatory variables are present in a model, there are $k!$ possible orderings in which the variables can enter into regression. A variable's contribution to $R^2$ can be evaluated by averaging over all possible orderings. This approach avoids the pitfall of depending on the Type II SS or, equivalently, the incremental $R^2$, where the variable is entered last. The Venn diagram helps students visualize what really occurs when the incremental $R^2$s for all possible orderings are averaged. Figure 7 illustrates the situation.

![Venn Diagram](image)

Figure 7. Venn Diagram Showing SS in Average Stepwise Regression.

Consider the variable $x_1$. Denote the areas covered by only one variable ($x_1$ itself, labeled "1"), two overlapping variables (labeled "2"), three overlapping variables (labeled "3") by $A_0, A_1, A_2$, etc. When the incremental $R^2$ is calculated for all $k!$ possible orderings, the piece that does not overlap with any other variable, $A_0$, appears every time. The pieces that overlap with only one other variable appear $k!/2$ times because in half of the $k!$ orderings $x_1$ enters the regression model before the other overlapping variables. In general, the area that overlaps with $r$ other variables ($1 \leq r \leq k - 1$) appears in the $k!$ possible orderings $k!/(r + 1)$ times. Therefore, the average contribution in incremental SS of $x_1$ is given by

$$ A_0 + \frac{1}{2} A_1 + \frac{1}{3} A_2 + \cdots + \frac{1}{k} A_{k-1}. $$

Because $SS(x_1) = A_0 \cup A_1 \cup \cdots \cup A_{k-1}$, the average stepwise approach produces a value that is the sum of the contributions of various pieces from $r^2_{y|x_1}$, weighted down harmonically by the number of times it overlaps with other variables plus one. The Venn diagram helps students visualize the relationship. Students should have no difficulty comparing this value to the Type II SS, which is represented by the area covered by $x_1$ alone.

### 3. Limitations of Using Venn Diagrams to Illustrate Regression Concepts

A number of authors point out that the overall $R^2$ for a model may be greater than the sum of the partial $R^2$s for a subset of variables. For example, Hamilton (1987) provides a geometric argument for why sometimes $R^2 > r^2_{y|x_1} + r^2_{y|x_2}$. In addition, Kendall and Stuart (1973, p.
describe an extreme example in which \( r^2_{yx_1} = 0.00, r^2_{yx_2} = 0.18, R^2 = 1.00, \) and the correlation between \( x_1 \) and \( x_2 \) is -0.9. This dataset is presented in Table 2.

Table 2. Example of Suppressor Variable (Kendall and Stuart 1973)

| \( x_1 \) | \( x_2 \) | \( y \) |
|----------|----------|------|
| 2.23     | 9.66     | 12.37|
| 2.57     | 8.94     | 12.66|
| 3.87     | 4.40     | 12.00|
| 3.10     | 6.64     | 11.93|
| 3.39     | 4.91     | 11.06|
| 2.83     | 8.52     | 13.03|
| 3.02     | 8.04     | 13.13|
| 2.14     | 9.05     | 11.44|
| 3.04     | 7.71     | 12.86|
| 3.26     | 5.11     | 10.84|
| 3.39     | 5.05     | 11.20|
| 2.35     | 8.51     | 11.56|
| 2.76     | 6.59     | 10.83|
| 3.90     | 4.90     | 12.63|
| 3.16     | 6.96     | 12.46|

A variable that increases the importance of the others is called a suppressor variable (e.g., Pedhazur 1982, p. 104). When a suppressor variable is present, Venn diagramming may not be suitable. Specifically, in a case in which there are only two predictors, the inequality \( R^2 > r^2_{yx_1} + r^2_{yx_2} \) is equivalent to \( SS(x_1, x_2) > SS(x_1) + SS(x_2) \). On a Venn diagram, this implies that the overlapping area, indicated by \( SS(x_1, x_2) - SS(x_1 | x_2) - SS(x_2 | x_1) = SS(x_1) + SS(x_2) - SS(x_1, x_2) \) is negative.

When there are three variables, every non-overlapping and overlapping piece in a Venn diagram corresponds to a function of the SS of the multiple regression of subsets of variables \( \{x_1\}, \{x_2\}, \{x_3\}, \{x_1, x_2\}, \ldots, \{x_1, x_2, x_3\} \). Figure 8 shows the seven mutually exclusive pieces of SS for three variables.
Figure 8. Partition of Areas When There Are Three Variables.

The piece that is labeled "6" corresponds to \( SS(x_3 \mid x_1) - SS(x_3 \mid x_1, x_2) \), or equivalently,

\[
SS(x_1, x_3) - SS(x_1) - SS(x_1, x_2, x_3) + SS(x_1, x_2),
\]

(1)

and the piece that is labeled "3" (where all variables overlap) corresponds to

\[
SS(x_1) + SS(x_2) + SS(x_3) - SS(x_1, x_2) - SS(x_2, x_3) - SS(x_1, x_3) + SS(x_1, x_2, x_3).
\]

(2)

There is no guarantee that expressions such as (1) and (2) will always be positive. Although we can think of areas as being negative, this may lead to difficulty in interpretation. Furthermore, when there are four variables or more, it is not possible to show all the combinations of overlaps with ovals or any other convex figures. For these reasons, Venn diagramming to demonstrate numerical results, especially when there are more than two variables, may not be illuminating.

4. Effectiveness for Instructional Purposes

Despite its limitations, we believe that Venn diagramming is a valuable tool that can be used when concepts of multiple regression are introduced and described in the classroom. We performed an experiment to assess the efficacy of the Venn diagram approach in the instruction of multiple regression. We selected two large undergraduate statistics classes taught by the author and another professor in the spring semester of 1999 at the University of Southern California. The class size of each session was approximately equal to 270. Venn diagramming was used in the author’s class (the treatment session) but not in the other class (the comparison session). In the final exams of both instructors, a common question (included in the Appendix) concerning multicollinearity was included. To eliminate possible bias due to different emphases in lectures or familiarity with wording introduced by the author, the instructor from the comparison session wrote the actual problem after all lectures were completed. A teaching assistant, who was not informed about the purpose of the experiment, graded the same question from both sessions on a 4-point scale. Because each instructor wrote up his/her own exam, and the teaching assistant worked for only one instructor, it was not possible to conceal which instructor wrote which exam.

| Table 3 | Summary of Two-Sample \( t \)-test (Two-Sided) for Treatment and Comparison Groups |
|---------|----------------------------------------------------------------------------------|

\( p \)-value of the two-sided two-sample \( t \)-test was 0.014 with 197 degrees of freedom, and therefore the test was significant at the \( \alpha = 0.05 \) level. We examined the individual scores and found that a lot of students either obtained full credit or no credit at all. We were concerned that the difference might be due to a discrepancy in the absentee rates of the two sessions. However, further investigation revealed that the discrepancy in absentee rates was slight. To examine the possible instructor effect, we also compared student evaluations of both instructors -- even though we were certain that there were stylistic differences between them -- and the ratings (on a 5-point scale) for both instructors were close (both above 4.0). We acknowledge that there were confounding factors, the effects of which cannot be completely isolated. These possible effects include differences in student ability between the two classes and the bias unintentionally introduced from review and coaching sessions by the instructors.
The evidence regarding the efficacy of the Venn diagramming approach was statistically significant, but not extremely strong. We did note, however, that in the treatment session, some students used phrases such as "overlapping in predictive power" or even drew a Venn diagram to illustrate multicollinearity. It is possible that these students used the Venn diagram as a mnemonic to aid their recall for an explanation. Finally, it must be emphasized that the result of the experiment should not be seen as offering definitive evidence for the universal value of Venn diagramming. The instructional value inherent in its use may vary as a function of instructor, student, and institutional characteristics.

5. Conclusion

This article discusses how Venn diagramming can be used as a teaching aid in classroom instruction of topics such as $R^2$ and the Type I and Type II SS in multiple regression. The limitations of its use are also discussed. Clearly, students should be aware of these limitations. However, when the goal is to help students grasp concepts in multiple regression and to enable them to explain these concepts to others, Venn diagramming is an effective tool. This observation is substantiated by a small-scale study.

Acknowledgments

The author thanks Professor Catherine Sugar for her help with the experiment. He also thanks the referees and the Associate Editor for their constructive comments.

Appendix

The printout below shows a multiple regression of employee's salary on years of professional experience and job approval rating. The regression equation is Salary = 20 + 2 Years + 3 Rating.

| Predictor   | Coef  | Stdve | t-ratio | F    |
|-------------|-------|-------|---------|------|
| Constant    | 20    | 2.0   | 10.00   | .0000|
| Years       | 2     | 1.5   | 1.33    | .1000|
| Rating      | 3     | 3.0   | 1.00    | .1657|
| S=1.00      | R-sq=.414 | R-sq(adj.)=.345 |

Analysis of variance

| Source   | DF | SS  | MS  | F    | P   |
|----------|----|-----|-----|------|-----|
| Regression | 2  | 12.00 | 6.00 | 6.00 | 0.0107 |
| Error    | 17 | 17.00 | 1.00 |     |     |
| Total    | 19 | 29.00 |     |     |     |

a. A manager at the company says that the overall regression is useful for predicting salary. Say briefly what test you would use to determine this and use the printout to justify the conclusion.

b. The manager further notes that tests show neither years of experience nor job approval ratings appear significant. Explain this using values from the printout. Again, no calculations are required.

c. What do the results of part (b) say about the usefulness of experience and job approval rating as predictors of salary?

d. The manager is confused that the model is useful, but neither of the predictors is significant. Can you explain to her what might have caused this result?*

* Only part (d) was used in the experiment.
References

Agresti, A., and Finlay, B. (1997), *Statistical Methods for the Social Sciences* (3rd ed.), Upper Saddle River, NJ: Prentice Hall.

Chatterjee, S., and Price, B. (1977), *Regression Analysis By Example*, New York: Wiley.

Hair, J., Anderson, R., and Tatham, R. (1987), *Multivariate Data Analysis with Readings* (2nd ed.), NY: Macmillan.

Hamilton, D. (1987), "Sometimes $R^2 > r^2_{x_1} + r^2_{x_2}$. Correlated Variables Are Not Always Redundant," *The American Statistician*, 41, 129-132.

Kendall, M., and Stuart, A. (1973), *Advanced Theory of Statistics* (Vol. 2; 3rd ed.), NY: Hafner.

Kruskal, W. (1987), "Relative Importance by Averaging Over Orderings," *The American Statistician*, 41, 6-10.

McKenzie, J., Schaefer, R., and Farber, E. (1995), *The Student Edition of Minitab for Windows*, Reading, MA: Addison-Wesley.

Pedhazur, E. J. (1997), *Multiple Regression in Behavioral Research: Explanation and Prediction* (3rd ed.), Fort Worth, TX: Holt, Rinehart & Winston.

SAS Institute Inc. (1990), *SAS/STAT User's Guide* (Vol. 1), Version 6, Cary, NC: Author.

Shavelson, R. J., and Webb, N. M. (1990), *Generalizability Theory -- a Primer*, London: Sage Publications.

Suppes, P. (1957), *Introduction to Logic*, Princeton, NJ: Van Nostrand.

Venn, J. (1880), "On the Diagrammatic and Mechanical Representation of Propositions and Reasonings," *The London, Edinburgh, and Dublin Philosophy Magazine and Journal of Science*, 5, 1-18.

Edward H. S. Ip
Marshall School of Business
University of Southern California
Bridge Hall 401
Los Angeles, CA 90089-1421
eddie.ip@marshall.usc.edu

---

Volume 9 (2001) | Archive | Index | Data Archive | Information Service | Editorial Board | Guidelines for Authors | Guidelines for Data Contributors | Home Page | Contact JSE | ASA Publications

---