AMS 315
Data Analysis
Chapter Twelve Study Guide
Multiple Regression and the General Linear Model
Spring 2023

Context

The statement of the theory of multiple linear regression in matrix terms is given in Section 12.9. Understanding the results of this chapter is much easier if you invest in learning how to use these tools. I will present the theory in matrix form in this chapter's study guide. We assume that $Y = X\beta + \sigma Z$, where Y is an $n\times 1$vector of observations, X is an $n\times p$ matrix of known constants, β is a $p\times 1$ vector of unknown but constant parameters, $\sigma > 0$, and Z is an $n\times 1$vector of $NID(0,1)$ random variables. The sum of squares function is given by $SS(b) = (Y - Xb)^T (Y - Xb) = Y^T Y - 2b^T X^T Y + b^T X^T Xb.$ The OLS estimate $\hat{\beta}$ is any solution of the system of normal equations: $(X^T X)\hat{\beta} = X^T Y$. If $(X^T X)^{-1}$exists, then the OLS estimate is unique and $\hat{\beta} = (X^T X)^{-1} X^T Y$.

The procedures in this chapter provide tools to deal with the association of one continuous dependent variable and an arbitrary number of independent variables.

12.1. Introduction and Abstract of Research Study

This section contains a statement of the multiple linear regression model and key definitions. It is an important section with key definitions.

12.2 The General Linear Model

This section contains the statement of the multiple linear regression model without using matrix notation. You should work on Section 12.9.

12.3 Estimating Multiple Regression Coefficients

Ordinary Least Squares (OLS) estimates are the ones most used. If $(X^T X)^{-1}$exists, then the OLS estimate is unique, and $\hat{\beta} = (X^T X)^{-1} X^T Y$.

12.4 Inferences in Multiple Regression

Definition 12.2 is important. The overall F test statistic and the coefficient of determination (multiple correlation coefficient squared) $R^2_{y.x_1\cdots x_k}$ are important statistics. The supplemental material contains an example examination problem worked out in detail.

12.5 Testing a Subset of Regression Coefficients

This is a common procedure. The most common example is to test the contribution of variables sequentially.

12.6. Forecasting Using Multiple Regression

This was not covered in lectures and will not be in either this examination or the final.

12.7. Comparing the Slopes of Several Regression Lines

The Caspi et al. paper on the class Blackboard uses the approach of this section to test for a gene-environment interaction ($G \times E$). This paper had an enormous impact. A finding must be replicated for it to have accepted over the long haul. Unfortunately, this paper was not replicated as reported in the Risch et al. paper on the class Blackboard. Both papers are valuable as examples of the ongoing importance of the techniques that you are studying.

12.8. Logistic Regression

There will be no formal questions on this section. It describes well the analysis of data in which the dependent variable is an indicator variable. Studies with the dependent variable being an indicator variable (e.g., patient died, patient had a relapse) are extremely common in medical research. I will cover it in detail when we discuss Chapter 10 later in the course.

12.9 Some Multiple Regression Theory

This section is of fundamental importance. Mastering these techniques will increase your productivity enormously. In class, I added the definition of the variance-covariance matrix of a random vector, which is $vcv(Y) = E[(Y - E(Y))(Y - E(Y))^T]$. From this, we derived in class the result that $vcv(MY) = Mvcv(Y)M^T$.

12.10 Research Study: Evaluation of the Performance of an Electric Drill

I would prefer that you study the Caspi et al. and Risch et al. papers as examples of the issues that come up using the techniques of this chapter.

12.11 Summary and Key Formulas

This is a valuable summary.

Class Supplemental material

Let the correlation matrix of (Y, x_1, x_2) be

$$\begin{pmatrix} 1 & \rho(y,x_1)=\rho_{y1} & \rho(y,x_2)=\rho_{y2} \\ \rho(y,x_1)=\rho_{y1} & 1 & \rho(x_1,x_2)=\rho_{12} \\ \rho(y,x_2)=\rho_{y2} & \rho(x_1,x_2)=\rho_{12} & 1 \end{pmatrix}$$

The *partial correlation* between Y and x_2 controlling for x_1 is defined to be

$\rho_{y2.1}=\dfrac{\rho_{y2}-\rho_{y1}\rho_{12}}{\sqrt{(1-\rho_{y1}^2)(1-\rho_{12}^2)}}$. Analogous definitions hold for the Pearson product moment correlations.

> *Example Examination Problem*: A study collects the values of (Y, x_1, x_2) on 400 subjects. The total sum of squares for Y is 1000. The correlation between Y and x_1 is 0.67; the correlation between Y and x_2 is 0.50; and the correlation between x_1 and x_2 is 0.25.
>
> *a.* Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on x_1 and the sum of squares due to the regression on x_2 after including x_1.
>
> *b.* Test the null hypothesis that both $\beta_2 = 0$ and $\beta_1 = 0$; that is, the null hypothesis is that there is no association between Y and these two independent variables.
>
> *c.* Test the null hypothesis that the variable x_2 does not improve the fit of the model once x_1 has been included against the alternative that the variable does improve the fit of the model. Report whether the test is significant at the 0.10, 0.05, 0.01 levels of significance.

Solution: a. The sum of squares due to the regression on x_1 has 1 degree of freedom and is equal to $SS(x_1) = r_{y1}^2 SS(Total) = (0.67)^2 \times 1000 = 448.9$. The partial correlation coefficient of x_2 with y after controlling for x_1 is

$$\rho_{y2.1}=\frac{\rho_{y2}-\rho_{y1}\rho_{12}}{\sqrt{(1-\rho_{y1}^2)(1-\rho_{12}^2)}}=\frac{0.50-0.67\times0.25}{\sqrt{(1-0.67^2)(1-0.25^2)}}=\frac{0.3325}{\sqrt{0.5511\times0.9375}}=\frac{0.3325}{0.7188}=0.463$$

The sum of squares due to the regression on x_2 after including x_1 also has 1 degree of freedomis $SS(x_2 \mid x_1) = r_{y2.1}^2(1-r_{yx_1}^2)SS(Total) = (0.463)^2 \times 0.5511 \times 1000 = 118.1$. The sum of squares for error has 400-3=397 degrees of freedom and is

$$SS(Error) = SS(Total) - SS(x_1) - SS(x_2 \mid x_1) = 1000 - 448.9 - 118.1 = 1000 - 567.0 = 433$$

b. The sum of squares using both x_1 and x_2 has 2 degrees of freedom and is equal to $SS(x_1, x_2) = SS(x_1) + SS(x_2 \mid x_1) = 448.9 + 118.1 = 567.0$. The F-test for this hypothesis is $F=\frac{SS(x_1,x_2)/2}{SS(Error)/397}=\frac{567.0/2}{433.0/397}=\frac{283.5}{1.091}=259.9$. The critical value for the 0.01 level of significance is more than 4.61 (for two numerator and infinite denominator degrees of freedom) and 4.69 (for two numerator and 240 denominator degrees of freedom). I reject the null hypothesis that there is no association between *Y* and these two independent random variables.

c. The *F*-test for this hypothesis is $F=\frac{MS(x_2\mid x_1)}{MS(Error)}=\frac{118.1/1}{433.0/397}=\frac{118.1}{1.091}=108.3$ with 1 numerator and 397 denominator degrees of freedom. The critical value for the 0.01 level of significance is more than 6.63 (for one numerator and infinite denominator degrees of freedom) and 6.74 (for one numerator and 240 denominator degrees of freedom). I reject the null hypothesis that x_2 does not improve the fit of the model once x_1 has been included.

Complete Mediation and Complete Explanation Causal Models

In analyzing research data from engineering or physical sciences studies, the independent variables typically operate at the same time. Given this, the fact that a partial regression coefficient is an estimate of a partial derivative strongly indicates to the user that caution is warranted in the interpretation of a partial regression coefficient. In social science and epidemiological research, however, the independent variables may operate at different points of time. For example, x_1 may describe a variable measured when the participant was between ages 5 and 6, and x_2 may describe a variable measured when the participant was between the ages of 8 and 9. The time-ordering of the independent variables is a crucial consideration in the interpretation of partial regression coefficients.

For example, often one sees that ρ_{y2} appears significant (that is, x_2 has a significant *F* statistic in a multiple regression analysis or the r_{y2}, the Pearson product moment correlation, is significant) but that $\rho_{y2.1}$ does not appear significant. For example, in multiple regression analysis, the variable x_2 does not have a significant F-to-enter once x_1 is in the regression equation. There is a fundamental paper (Simon, 1954, available on JSTOR and on the Blackboard site) that you should download and read.

Simon points out that when one has a common cause model (or *explanation*), the independent variable x_1 precedes both x_2 and *y* with regard to operation impact. Then if x_1 "causes" x_2 and if x_1 "causes" *y*, then there will be a "spurious" correlation ρ_{y2} (this correlation will be non-zero even though x_2 has no causal relation to *y*) and $\rho_{y2.1}$ will be zero. For example, consider G. B. Shaw's correlation between the number of suicides in England in a given year and the number of churches of England in the same year.

In a causal chain model, the independent variable x_2 operates before and causes x_1 and x_1 operates before y and causes y. Simon also points out that, when the model is a causal chain (or *mediation*), one also observes that ρ_{y2} will be non-zero and $\rho_{y2.1}$ will be zero (even though x_2 causes y through the mediation of x_1). Both causal modeling situations have the same empirical facts. Deciding which interpretation is valid requires clarifying the sequence of operation of the variables. In practice, the relevant partial correlation may not be essentially 0. In this event, researchers speak of partial explanation and partial mediation.

Example Past Examination Questions

Common Information for Questions 1, 2, and 3

A research team sought to estimate the model $E(Y) = \beta_0 + \beta_1 x + \beta_2 w$. The variable Y was a measure of depression of a participant observed at age 25; the variable x was a measure of anxiety shown by the participant at age 18; and the variable w was a measure of the extent of traumatic events experienced by the participant before age 15. They observed values of y, x, and w on $n = 800$ subjects. They found that the standard deviation of Y, where the variance estimator used division by $n-1$, was 12.2. The correlation between Y and w was 0.31; the correlation between Y and x was 0.14; and the correlation between x and w was 0.41.

1. Compute the partial correlation coefficients $r_{Yx\bullet w}$ and $r_{Yw\bullet x}$.
   Answer: $r_{Yx\cdot w} = 0.0149$ and $r_{Yw\cdot x} = 0.2797$

2. Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on w and the sum of squares due to the regression on x after including w. Test the null hypothesis that $\beta_1 = 0$ against the alternative that the coefficient is not equal to zero. That is, test whether x adds significant additional explanation after using w. Report whether the test is significant at the 0.10, 0.05, and 0.01 levels of significance.

   Answer: The analysis of variance table is given by

   Analysis of Variance Table

Source	DF	SS	MS	F Statistic
Regression on w	1	11428.52	11428.52	
Regression on x|w	1	23.86	23.86	0.18
Error	797	107470.78	134.84	
Total	799	118923.16		

   The value of the test statistic is $F_{x|w} = 0.18$. Since $F_{0.10,1,797} = 2.71^+$, $F_{0.05,1,797} = 3.84^+$ and $F_{0.01,1,797} = 6.63^+$, we accept the null hypothesis that *x* does not add significant explanation after including *w* at the 0.10 level.

3. What interpretations can you make of these results in terms of causal models?

Answer: It is an explanation model.

End of application of common information

Common Information for Questions 4, 5, and 6

A research team sought to estimate the model $E(Y) = \beta_0 + \beta_1 x + \beta_2 w$. The variable Y was a measure of the extent of criminal behavior of a participant observed at age 30; the variable x was a measure of the rebelliousness shown by the participant at age 12; and the variable w was a measure of delinquency shown at age 18. They observed values of y, x, and w on $n = 1500$ subjects. They found that the standard deviation of Y, where the variance estimator used division by $n-1$, was 15.7. The correlation between Y and w is 0.62; the correlation between Y and x is 0.35; and the correlation between x and w is 0.58.

4. Compute the partial correlation coefficients $r_{Yx\bullet w}$ and $r_{Yw\bullet x}$.
   Answer: $r_{Yx\cdot w} = -0.015$ and $r_{Yw\cdot x} = 0.5465$

5. Compute the analysis of variance table for the multiple regression analysis of Y. Include the sum of squares due to the regression on w and the sum of squares due to the regression on x after including w. Test the null hypothesis that $\beta_1 = 0$ against the alternative that the coefficient is not equal to zero. That is, test whether x adds significant additional explanation after using w. Report whether the test is significant at the 0.10, 0.05, and 0.01 levels of significance.

   Answer: The analysis of variance table is given by:

   Analysis of Variance Table

Source	DF	SS	MS	F Statistic
Regression on w	1	142031.38	142031.38	
Regression on x|w	1	51.18	51.18	0.34
Error	1497	227405.95	151.91	
Total	1499	369488.51		

   The value of the test statistic is $F_{x|w} = 0.34$. Since the critical value for (1,1497) degrees of freedom is slightly over 2.71, we accept the null hypothesis that x does not add significant explanation after including w at the 0.10 level.

6. What, if any, interpretations can you make of these results in terms of causal models?

   Answer: It is a mediation model.

End of application of common information

7. The $n \times 1$ vector Y has a multivariate normal distribution. The expected value of Y is given by $E(Y) = X\beta$, where X is an $n \times p$ matrix of constants and β is a $p \times 1$ vector of unknown coefficients. The variance-covariance matrix V of Y is a symmetric, positive-definite and invertible $n \times n$ matrix of known constants. The matrix $(X^T V^{-1} X)^{-1}$ exists. Find the variance-covariance matrix of $W = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$.

   Answer: Note that $W = MY$, where $M = (X^T V^{-1} X)^{-1} X^T V^{-1}$, so that $\mathrm{VCV(W)} = (X^T V^{-1} X)^{-1}$