The Meaning and Interpretation of Interaction

Garrett Fitzmaurice, ScD
From the Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA

INTRODUCTION

In many scientific endeavors, it is common to examine the effects of more than a single explanatory variable on the response or outcome variable of interest. For example, in a placebo-controlled clinical trial examining the effects of a new cholesterol-lowering drug, the investigators may also classify participants in the study according to their obesity status (obese versus non-obese). When two or more explanatory variables are considered simultaneously, one could ask whether their joint effect on the response variable is simply the "sum of the parts" (or individual effects). When their joint effect is discernibly larger or smaller than the "sum of the parts," we say that there is an "interaction" among the explanatory variables. Although the concept of interaction is relatively easy to explain,[1] scientific journals abound with articles in which there is apparent confusion about the precise meaning and interpretation of the term *interaction*. In this column, I discuss the definition and meaning of the term *interaction* as used in statistics and consider how to appropriately interpret an interaction.

THE DEFINITION OF INTERACTION

The term *interaction* has a very precise statistical meaning and refers to how the effect on the response of one explanatory variable depends on the level of one or more other explanatory variables. That is, interaction is said to arise when the effect of one explanatory variable depends on the particular level or value of another explanatory variable. To return to the example introduced earlier, if the effect of the cholesterol-lowering drug is greater among individuals who are classified as obese than among those who are classified as non-obese, we say that there is an interaction between treatment (new drug versus placebo) and obesity status. This is simply another way of saying that the effect of treatment is different at the two levels of the obesity factor and vice versa. Conversely, if the effect of the cholesterol-lowering drug is the same, regardless of the obesity status of the participants in the study, then we say that there is no interaction. This simple example is used throughout this column to illustrate some of the main aspects of interpreting interaction. Also, for ease of exposition, I restrict my discussion of interaction to the simple case where there are only two explanatory variables. When there are just two explanatory variables, their interaction is often referred to as a *two-way* or *two-factor* interaction on the response variable. However, it should be kept in mind that the concept of interaction generalizes in a natural way to more than two explanatory variables.

TESTING FOR INTERACTION

The general definition of interaction given above implies that, if there is no interaction among two explanatory variables, then the effect of one explanatory variable is constant or remains the same across all levels or values of the other. Next, I consider how to test for interaction among explanatory variables. For simplicity, the following discussion of testing for interaction focuses on linear regression of the response variable, Y, on two explanatory variables, say X_1 and X_2. However, the procedure for testing for interaction generalizes to more than two explanatory variables and to other types of statistical models (e.g., logistic regression and survival analysis).

Consider the simple example introduced earlier. For this example, suppose that the response variable, Y, is a measure of the amount of low-density lipoprotein cholesterol (LDL-C) in serum at the completion of the study. Let X_1 denote the treatment received, with $X_1 = 1$ if individuals are randomized to the cholesterol-lowering drug and $X_1 = 0$ if individuals are randomized to the placebo. Let $X_2 = 1$ if individuals are classified as obese and $X_2 = 0$ if individuals are classified as non-obese. One of the simplest models for describing the effects of X_1 and X_2 on the mean of Y is:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

This model assumes that there is no interaction between X_1 and X_2 on the mean of the response. To see why this is the case, consider the effect of X_1 at any two values of X_2, say $X_2 = 0$ and $X_2 = 1$. When $X_2 = 0$, the linear regression model given above can be written as

$$E(Y) = \beta_0 + \beta_1 X_1.$$

That is, a single unit change in X_1 results in a change of β_1 in the mean response. Similarly, when $X_2 = 1$, the linear regression model can be written as

$$E(Y) = (\beta_0 + \beta_2) + \beta_1 X_1.$$

In fact, for any value of X_2, the effect of X_1 on the mean response is the same. Thus, the effect of X_1 does not depend on the levels or values of X_2. Similarly, because there is a complete symmetry to the definition of interaction, it can also be shown that the effect of X_2 on the mean response is the same for all levels or values of X_1.

The standard procedure for testing for interaction is to add one or more terms to the model that represents interaction effects. When there are only two explanatory variables, X_1 and X_2, testing is achieved by creating a new variable that is the product of the two explanatory variables ($X_1 \cdot X_2$) and then adding this product to the model that already contains the original explanatory variables. The expanded or augmented model then becomes

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \cdot X_2.$$

Except on very rare occasions, the original explanatory variables should always be included in the model that contains their product. To see why this augmented model allows for interaction between X_1 and X_2, consider again the effect of X_1 at any two values of X_2, say $X_2 = 0$ and $X_2 = 1$. When $X_2 = 0$, the augmented model given above can be written as

$$E(Y) = \beta_0 + \beta_1 X_1.$$

That is, a single unit change in X_1 results in a change of β_1 in the mean response. However, when $X_2 = 1$, the augmented model can be written as

$$E(Y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1,$$

and a single unit change in X_1 now results in a change of $(\beta_1 + \beta_3)$ in the mean response. Thus, the effect of X_1 depends on the levels or values of X_2 and vice versa. A formal test

Correspondence to: Garrett Fitzmaurice, ScD, Department of Biostatistics, Harvard School of Public Health, Room 423, 665 Huntington Avenue, Boston, MA 02115, USA.

Nutrition 16:313–314, 2000 0899-9007/00/$20.00

PII S0899-9007(99)00293-2

for interaction is constructed by assessing the adequacy of the original model (without the product of the two explanatory variables). If there is no interaction, the original model should be adequate for the data at hand. This can be determined by testing the null hypothesis, H_0: $\beta_3 = 0$. A test of this hypothesis can be constructed, for example, by comparing the estimate of β_3 to its standard error.

TABLE I.

HYPOTHETICAL EXAMPLE OF A TREATMENT-BY-OBESITY STATUS INTERACTION THAT CAN BE REMOVED BY A LOGARITHMIC TRANSFORMATION

	Original scale		Logarithmic scale	
	Placebo	New drug	Placebo	New drug
Obese	200	150	5.298	5.010
Non-obese	140	105	4.942	4.654

SOME ASPECTS OF INTERPRETATION OF INTERACTION

Thus far, I have shown that the general definition of interaction given above implies that, if there is no interaction between two explanatory variables, then the effect of one explanatory variable is constant across all levels of the other. However, there is an important subtlety that needs to be recognized in this definition of interaction. In statistics, interaction is defined in terms of a specific model for the response variable (e.g., linear regression) and thus can depend on the measurement scale chosen for the response variable. For example, in the linear regression model given above, interaction is defined implicitly in terms of differences between the responses. That is, there is an interaction when the difference between the mean response for any two values of X_1 depends on the values of X_2 and vice versa. However, if the response variable is transformed, e.g., by taking logarithms, the definition of interaction on the transformed response would be somewhat different. The point being made here is that, by some suitable transformation of the response variable or the mean of the response variable, it may often be possible to remove interaction from the original measurement scale.

Consider the hypothetical summary statistics from a clinical trial of the new cholesterol-lowering drug displayed in Table I. On the original scale of measurement, there is an interaction between treatment and obesity status. That is, when compared with placebo, the effect of treatment with the new cholesterol-lowering drug is a reduction in LDL-C of 50 mg/dL for those who are classified as obese. Conversely, for those classified as non-obese, the effect of treatment with the new cholesterol-lowering drug is a reduction in LDL-C of only 35 mg/dL. However, on the logarithmic scale, there is no interaction. That is, on the logarithmically transformed scale, the effect of treatment with the new cholesterol-lowering drug is a reduction of 0.288 (or a reduction by a factor of 0.75, or $\exp[-0.288]$), regardless of obesity status. There is no interaction on the logarithmically transformed scale because the proportional reduction in LDL-C due to treatment with the new cholesterol-lowering drug is the same (0.75 or $\exp[-0.288]$) for those classified as obese and non-obese. On the original scale, treatment effects are expressed in terms of differences or absolute reductions; on the logarithmic scale, treatment effects are expressed in terms of proportional reductions. There is interaction on the original scale but no interaction on the logarithmic scale. Thus, in this example, the presence or absence of interaction depends entirely on the scale of measurement chosen. On the untransformed scale, there is interaction; however, on the logarithmic scale, there is not.

SUMMARY

In this column, I have considered the precise definition and meaning of the term *interaction* as used in statistics. From a practical point of view, it should be clear that, when there is no interaction, a simple and economical description of the effect of an explanatory variable can be given. When there is interaction, however, the description of the effect of one explanatory variable must also take into account the levels or values of other explanatory variables. In a future column, I will discuss how to present results when there is an interaction, and I will demonstrate how careful tabulation and plotting of the results is a crucial aspect of explaining the substantive meaning of an interaction. Finally, the statistical use of the term *interaction* is closely related to the medical use of the term *synergy*. In medicine, synergy[2] (derived from the Greek *sunergos*, meaning "working together") refers to the interaction of two (or more) drugs or agents whose combined effect at given concentrations is greater than the sum of their individual effects. Thus, synergy can simply be thought of as a special case or type of interaction.

REFERENCES

1. Cox DR. Interaction. Int Stat Rev 1984;52:1
2. Berenbaum MC. What is synergy? Pharmacol Rev 1989;41:93