

What is the definition of instrumental variables and what are some examples of their use?

Authored by
stats writer

April 18, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the definition of instrumental variables and what are some examples of their use?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136534>

Instrumental variables (IVs) are statistical techniques used to address endogeneity, which is a common problem in observational studies where the relationship between two variables is distorted due to the presence of unobserved factors. IVs help to identify the causal relationship between two variables by using a third variable (the instrumental variable) that is correlated with the independent variable, but not directly with the dependent variable. This allows researchers to isolate the effect of the independent variable on the dependent variable without the influence of confounding factors.

Some examples of instrumental variables include natural experiments, such as random assignment in a clinical trial, and proxy variables that are correlated with the independent variable but do not directly affect the dependent variable, such as distance to a hospital in a study on healthcare utilization. Other common uses of IVs include studies on the effects of education on income, the impact of policies on economic outcomes, and the relationship between environmental factors and health outcomes. Overall, instrumental variables are an important tool in statistical analysis for studying causal relationships in complex systems.

Instrumental Variables: Definition & Examples

Often in statistics we're interested in estimating the effect that one variable has on another. For example, perhaps we want to know:

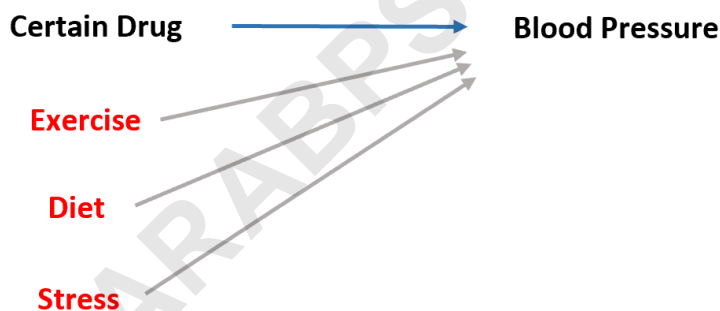
How does amount of time spent studying affect exam scores? How does a certain drug affect blood pressure? How does stress affect heart rate?

In each scenario, we want to understand whether or not some predictor variable affects a response variable. However, often there will be other variables that affect the relationship between the two variables.

For example, suppose we use a certain drug as our predictor variable and blood pressure as our response variable. We are only interested in the effect that the drug has on blood pressure:



However, other variables like time spent exercising, overall diet, and stress levels also affect blood pressure:



Thus, if we run a simple linear regression using the drug as our predictor variable and blood pressure as our response variable, we can't be sure that the regression coefficients will accurately capture the effect

that the drug has on blood pressure because outside factors (exercise, diet, stress, etc.) could also be playing a role.

One potential way to get around this problem is to use an instrumental variable.

What is an Instrumental Variable?

An instrumental variable is a third variable introduced into regression analysis that is correlated with the predictor variable, but uncorrelated with the response variable. By using this variable, it becomes possible to estimate the true causal effect that some predictor variable has on a response variable.

For example, suppose we want to estimate the effect that a certain drug has on blood pressure:

Certain Drug  Blood Pressure

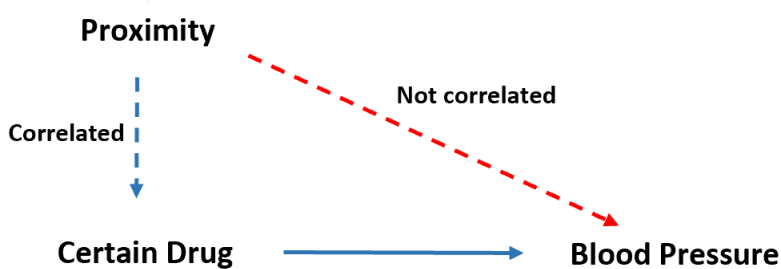
An example of an instrumental variable that we may use in this regression analysis is an individual's proximity

to a pharmacy.

This variable "proximity" would likely be highly correlated with whether or not the individual takes the certain drug because an individual wouldn't be able to obtain it in the first place if they don't live near a pharmacy.



However, the variable "proximity" is not expected to have any correlation with blood pressure. The only association it would have with blood pressure is through the predictor variable.



The way that we actually use an instrumental variable is through instrumental variables regression, sometimes called two-stage least squares regression.

Instrumental Variables Regression

Instrumental variables regression (or two-stage least squares regression) uses the following approach to estimate the effect that a predictor variable has on a response variable:

Stage 1: Fit a regression model using the instrumental variable as the predictor variable.

In our specific example, we would first fit the following regression model:

Certain drug = $B_0 + B_1(\text{proximity})$

We would then be left with predicted values for certain drug (\hat{cd}), which we'll call \hat{cd} .

Stage 2: Fit a second regression model using the predicted values for \hat{cd} .

Next, we'll fit the following regression model:

Blood pressure = $B_0 + B_1(\text{cdhat})$

If the regression coefficient for cdhat turns out to be statistically significant, then we can say that there is a causal effect of the drug on blood pressure.

The reason we can say this is because we solely used "proximity" to come up with cdhat and since we know that proximity should not be correlated with blood pressure, any significant correlation in the second stage regression can be attributed to the certain drug.

Cautions on Using Instrumental Variables

An instrumental variable should only be used if it meets the following criteria:

It is highly correlated with the predictor variable. It is not correlated with the response variable. It is not correlated with the other variables that are left out of the model (e.g. proximity is not correlated with exercise, diet, or stress).

If an instrumental variable does not meet this criteria, then it should not be used in the regression model because it will likely produce unreliable and biased

results.

Bonus: A Video Explanation of Instrumental Variables

The following video by Ashley Hodgson provides an excellent visual explanation of instrumental variables:

ARABPSYCHOLOGY.COM