

Research Methods II, Spring Term 2003  
Multiple Regression: Spatial interpretation

In this lecture we will be discussing multiple regression; this is an extension of the simple linear regression you learnt last year. To begin with we will review the concepts of simple regression, and then extend them to multiple regression.

Simple linear regression

Simple regression involves one continuous dependent variable (also called the criterion variable), and an independent variable (also called the predictor variable) that can be (but need not be) continuous as well. For example, we might want to see if we can predict how happy people are (the dependent or criterion variable) based on how much money they earn (the predictor variable). The first thing to do is to plot a (continuous) measure of happiness against income. If the relationship looks linear (i.e. it is not systematically curved), then we can fit a straight line to summarize the relationship. The equation of the straight line is:

$$\text{predicted happiness} = a + b \cdot \text{income} \quad \dots (1)$$

How much happiness do we predict someone has if they had zero income?

If income = 0, then  $b \cdot \text{income} = 0$ , so from equation (1)

predicted happiness = a

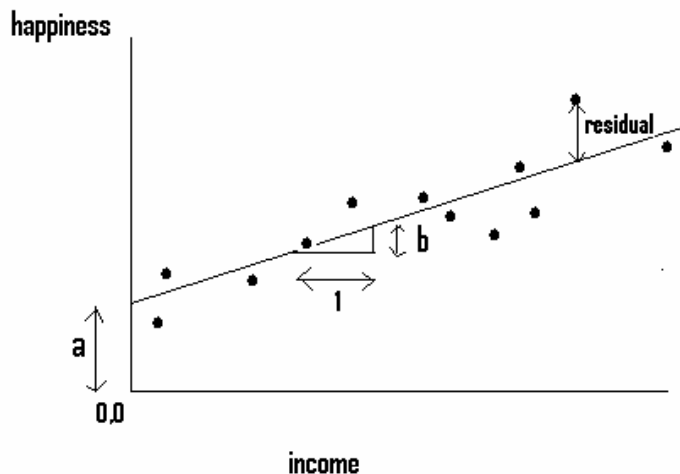
“a” is called the intercept; it’s how happy someone is with zero income.

If income increases by 1 unit, income = 1, so  $b \cdot \text{income} = b$ , so from equation (1)

predicted happiness = a + b

Predicted happiness has increased by b units (from a to a+b) when income increased by 1 unit (from 0 to 1). “b” is the slope of the straight line; it is how much happiness increases by a unit increase in income.

Any given person has an actual income and an actual happiness. Based on the income and our equation, we can calculate the predicted happiness for that person. If our data points lie on a perfect straight line, then the predicted happiness would equal the actual happiness. In fact, the actual happiness will be a little bit more or a little bit less than the predicted value. The difference between the actual and predicted values of happiness is called the residual for that subject. The residual is the error in prediction. In simple regression, we choose values of a and b so that the sum of the squared residuals as is small as possible.



### Multiple regression

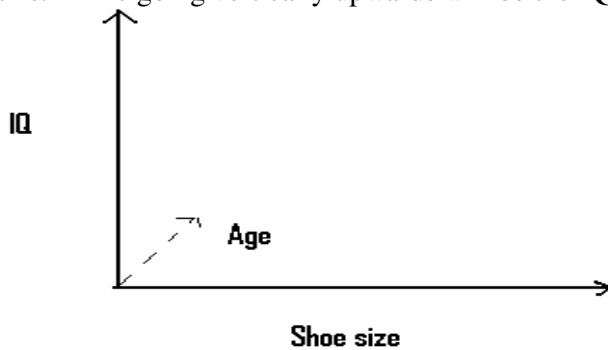
Hands up who thinks IQ is correlated with shoe size.

Well, it is.

As children get older their feet get bigger; their ability to perform IQ tests also gets better. So if you took some children between say 5 and 15 and plotted performance on an IQ test against shoe size you would get a significant slope. It's not that big feet cause large IQs or vice versa; the relationship between IQ and shoe size arises because of the relationship of both IQ and shoe size with age.

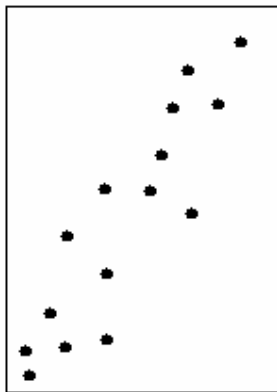
To fully understand the relationships between these variables we need to plot them in a three dimensional space.

Consider the table in front of you. The edge of the table going from left to right will be the shoe size axis. The edge of the table going out away from you will be the age axis. A line going vertically upwards will be the IQ axis.

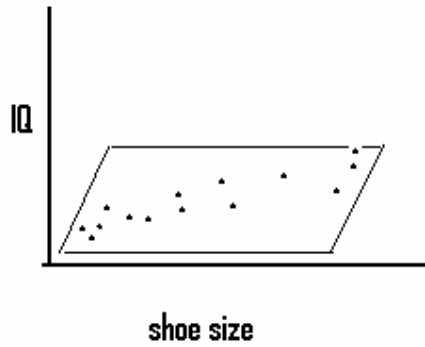


Consider now just shoe size and age. For every child, there will be a particular combination of shoe size and age, and that will correspond to a particular point on the table top. That child has a certain IQ, which is a certain height above the table top. So our data points occupy the three dimensional space above the table top.

Now we consider different configurations of these points. Take a sheet of paper and scatter points on it from one corner to the other diagonally opposite:



Now put this piece of paper over your table top thus:



The long edge of the paper is flat against the table edge that corresponds to shoe size, and the short edge of the paper is rising at an angle above the table edge that corresponds to age. Now imagine the points on the paper are just hanging in space just as they are positioned. In practice they wouldn't really be all in the plane of the paper; each point would also have some height a little bit above or a little bit below the plane, but we need not worry about that now.

Now imagine a bright light shining evenly from above so that each point casts a shadow on the table top. Look at these shadows. What relationship do they show between age and shoe size? This is what a plot would look like if all you had measured were age and shoe size. You would see a positive correlation between the two.

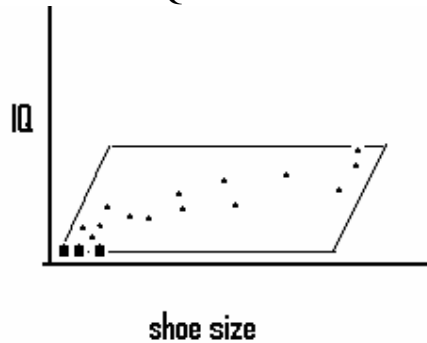
Now imagine the light shining directly towards you, projecting shadows on a screen standing vertically in front of you in the shoe size-IQ plane. The scatter plot you see is the first order relationship between shoe size and IQ ignoring age; you would see a positive correlation between shoe size and age.

Now imagine the light shining from the right projecting on a screen standing in a plane defined by the age axis and the IQ axis. The projection of the points on this screen would show a positive correlation between age and IQ.

So all pair-wise correlations between our variables are positive.

Now what relationship is demonstrated between shoe size and IQ *when age is held constant*?

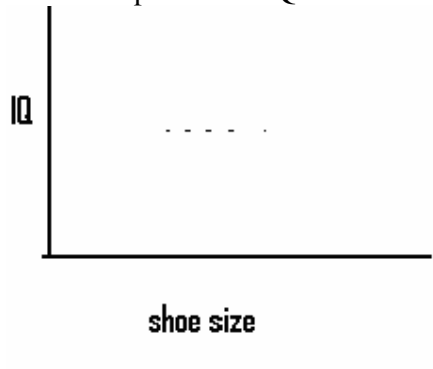
On the table, find a line of constant age. That's any line on the table parallel to the shoe size axis. All the points along such a line correspond to a single age. In fact, any point in the vertical plane including that line corresponds to a single age. Take any such plane and look at the data points in it. What relationship do they show between shoe size and IQ?



I have drawn three large data points, to make them salient, that lie along one line of constant age. One line of constant age is the bottom edge of the paper - all points

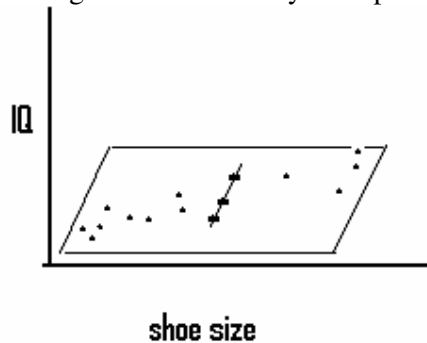
along this edge have a constant age (as would any points along any line in the paper parallel to this edge). As you move from left to right shoe size increases. For those three points, how does increasing shoe size relate to changes in IQ?

If you look at your paper, still held in the way specified, you will see that the relationship between IQ and shoe size *in a plane of constant age* is flat.



Wherever you place your plane of constant age, the plane of our paper intersects it with a horizontal straight line. This is because your paper plane has a horizontal slope along the table edge corresponding to shoe size. Make sure you can see that; this is the essence of multiple regression. Taking all the children of one precise age, IQ does not vary with shoe size.

Now take a plane of constant shoe size, i.e. a vertical plane containing a line parallel to the age axis. Within any such plane, what is the relationship between IQ and age?



I have shown three large points on one line of constant shoe size (the line will be parallel to the sloping edge of your paper). For these points, as you move from front to back age increases, even though shoe size is constant. What happens to IQ?

You will see that IQ increases with age, even holding shoe size constant. Your paper plane has a slope along the table edge corresponding to age. So your paper will always intersect a plane of constant shoe size with a line with that slope.

So the slopes of the plane tell us different information than just the correlations between the variables. We have seen that all the variables are positively correlated with each other. But when we consider the two-dimensional plane fitting the points in a three dimensional space, the slope of the plane along the shoe size axis is zero. That is because, as we have seen, the slope of the plane represents the ability of shoe size to predict IQ (or whatever is our dependent variable) when age has been held constant. In general, the plane slopes tell us the ability of each predictor variable to predict IQ when all other predictors have been held constant.

Simple regression fits a line to a two-dimensional plot. With two predictor variables, multiple regression fits a two-dimensional plane to a three dimensional plot. In fact, multiple regression can deal with any number,  $n$ , of predictor variables, and

conceptually it is fitting a  $n$ -dimensional solid to the  $(n+1)$  dimensional plot. But since our concrete imagination is limited to three dimensions, we always talk of multiple regression as fitting a plane, no matter how many predictor variables there are. And the same principle applies to the  $n$ -dimensional case as to the three dimensional case we are considering: The slope of this plane along the axis of one predictor represents the unique ability of that predictor to predict the dependent variable when all other predictors are held constant.

You see how this gets us closer to the true relationships between variables. If we had the scatter plot of IQ against shoe size we might try to come up with a theory of why shoe size is related to IQ. If we think that age might be the true underlying variable mediating the relationship, we could perform a multiple regression with both age and shoe size in. Finding a positive slope for age but not shoe size would support our theory.

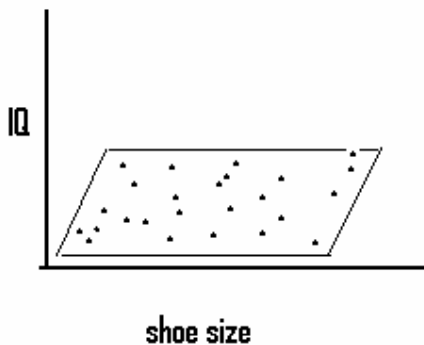
Exercises:

- 1) Orient your paper so as to show that age does not predict IQ when shoe size is held constant, but shoe size predicts IQ when age is held constant.
- 2) Orient your paper so that *both* predictors have independent effects on IQ.
- 3) Orient your paper so that neither has any effect on IQ.
- 4) Orient your paper to show that age is negatively related to IQ when shoe size is held constant, and shoe size has no independent effect on IQ.

Make sure you can do all these exercises; if not, come to see me.

Let's return to the original orientation of the plane, with a slope for age but not shoe size. We obtained a correlation between shoe size and IQ even though the plane had a zero slope for shoe size *because* shoe size and age were correlated. That meant that as shoe size increased, age went up on average, and it was because age was going up that IQ went up. In other words, we got that result because we drew our points on our piece of paper going from one diagonal to the other.

Now I want you to add points randomly all over your piece of paper. Now orient your plane as before, and it will look like this:

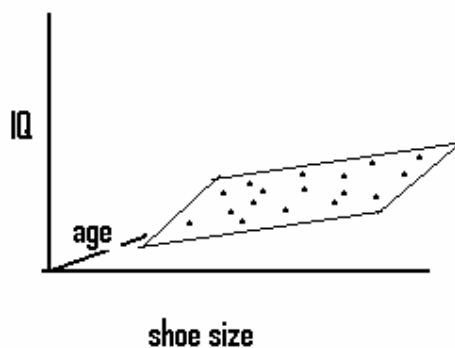


Now what happens if you shine your bright light from above? The points will project on the table surface all over the place. There will be no correlation between shoe size and age. If you look at the projection of the points in the IQ-age plane, there will be a positive correlation between age and IQ. If you look at the projection in the shoe-size IQ plane, there will be no correlation between shoe size and IQ.

In summary, if your predictor variables do not correlate with each other, the correlations of each predictor with IQ reflect the slopes for that predictor. There will be a correlation if there is a slope, and vice versa.

You might wonder if it is worth performing a multiple regression if the predictors do not correlate with each other at all. After all, if the predictors don't correlate with each other, the correlations of the predictors with the dependent variable reflect the predictor's slopes, so why bother with the multiple regression? There is a reason for still being interested in multiple regression in this situation.

Imagine the points scattered all over your piece of paper, so no correlation between predictors. Now orient your plane to show some slope for shoe size and some for age; both variables can uniquely predict IQ.



Imagine all your points lie perfectly in the plane. Now shine your light to get the correlations. Project the points onto the age-IQ plane. See the scatter of points going from the bottom diagonal to the top: A correlation. But the points do not lie perfectly on a straight line, they are scattered about the line. If there were zero slope for shoe size, what would the projection be like onto the age-IQ plane? All the points lie in the plane, and you would be seeing the plane edge on. The projection would be a perfect straight line. The correlation between age and IQ would be 1.0. As you tilt the plane along the shoe size axis, you no longer see the plane edge on: You get a scatter of points projected onto the IQ-age plane. The more you tilt along the shoe size axis, the greater the scatter. So for a constant underlying relationship between age and IQ, a constant age slope, the correlation between age and IQ can get lower and lower as the shoe size slope increases. The scatter of the points about the line is the noise through which you are trying to see the signal of the population simple regression slope. There will be a certain amount of noise when the noise is so great that the correlation you obtain in your sample is non-significant. There is a population correlation, but you do not have the power to detect it. Now consider the full three dimensional plot with shoe size, age and IQ. The noise through which you try to detect the population plane slopes is the scatter of the points above and below the plane. The height of a point above or below the plane is the residual for that point, the error in prediction. In our example, by assumption, there was no scatter above and below the plane, all the points lay in the plane. (The residuals were all zero.) No noise, you can easily detect the multiple regression slopes.

So even when the predictors are uncorrelated with each other, multiple regression can increase your sensitivity in picking up population relations. As more of the variability in IQ is accounted for, there is less to appear as noise.

### Summary

Multiple regression involves using two or more predictors to predict a dependent variable. Multiple regression fits a plane in order to make the best predictions of the dependent variable based on accurate knowledge of the predictor variables. The

slopes of the plane correspond to the unique ability of each predictor to predict the dependent variable when all other predictors are held constant.

If ever you find yourself wondering what a multiple regression slope means, just think of shoe size, IQ, and age.

Post script: a note on interaction.

Thus far, we have not dealt with interactions between predictors. Do not be tempted to think of the slope as somehow reflecting the interaction between shoe size and age. It is not. The slope for shoe size is the effect of shoe size on IQ with age held constant; this is not the interaction between shoe size and age. An interaction between shoe size and age would mean the effect of shoe size on IQ is different for different levels of age (interaction: the effect of one IV is different at different levels of another IV). In our example, the slope for shoe size (i.e. effect of shoe size on IQ) was identical (i.e. 0) for each level of age; and the slope for age was identical (some positive value) for each level of shoe size. This arose because the plane was flat, ensuring these constant slopes. So no interaction.

There is a way of testing interactions with multiple regression but we do not cover it in this course. It may well come up in your third year project if you use multiple regression in your project, and then your supervisor will show you how to test for interactions with multiple regression.