A Tale of Two Logratios (is this a counter-example to using ILRs?) *Michael Greenacre, 12 February 2018*

Here's a set of data from a (fictitious) survey of drinking habits of people who drink wine, beer and spirits daily, in approximately the same total quantity but in different compositional amounts. The units are not relevant, but all are on the same scale.

	wine	beer	spirits
[1]	1.2	1.8	0.6
[2]	1.8	1.4	0.3
[3]	2.8	0.3	0.6
[4]	2.6	0.5	0.5
[5]	2.5	0.7	0.7
[6]	3.0	0.1	0.3
[7]	1.1	1.9	0.3
[8]	1.9	1.2	0.3
[9]	1.4	1.5	0.4
[10]	1.7	1.3	0.2
[11]	1.3	1.7	0.3
[12]	2.4	0.8	0.3
[13]	2.1	1.1	0.5
[14]	1.0	2.1	0.2
[15]	2.0	1.1	0.6
[16]	1.6	1.5	0.4
[17]	2.7	0.4	0.1
[18]	2.3	0.9	0.4
[19]	1.5	1.7	0.6
[20]	2.2	1.0	0.5
[21]	2.9	0.1	0.4

Researchers are interested in relationships between these variables, and they have heard about CODA and the ILR transformation. In particular they are interested in how the high alcohol spirits are related to the lower alcohol wine & beer.

After closing the data to make it compositional, they form the ILR balance in the form of the logratio of spirits divided by the geometric mean of wine & beer, and then the "orthogonal" ILR balance in the form of the simple logratio of beer to wine.

They plot these two against each other (Fig. 1). There is a positive relationship with a correlation of 0.53 (p=0.01). The conclusion is that as the ratio of wine to beer increases, so the ratio of spirits to the group wine&beer of lower alcohol drinks increases.

Now there is an ongoing debate about using amalgamations rather than geometric means to simplify the interpretation. So, as a check, the ILR balance on the y-axis is plotted against the logratio of spirits divided by the sum (i.e. amalgamation) of wine and beer (Fig. 2), what I call an *amalgamation balance*. The two look very similar, apart from some differences at the upper end, so the researchers feel justified in their conclusions.

But just to check even more , they plot the amalgamation balance against log(wine/beer) and get Fig. 3, a surprise! There is no relationship: correlation = 0 (p=0.99).

So what caused the relationship in Fig. 1?

Figure 1

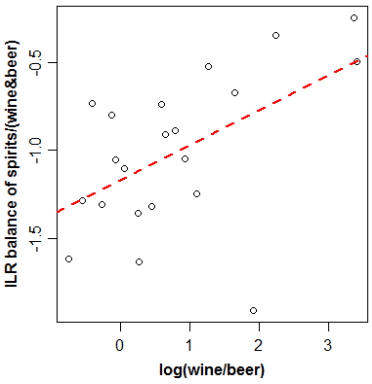


Figure 2

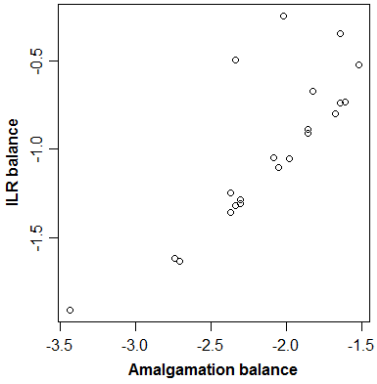
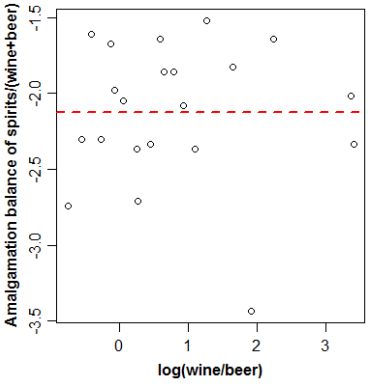
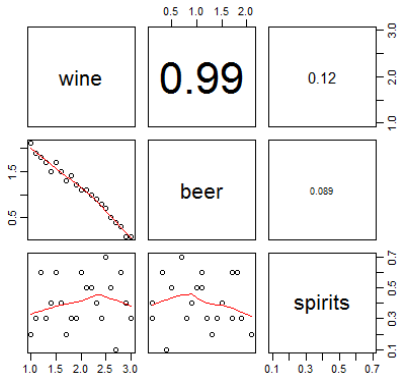


Figure 3



The truth about these data is that there is only one relationship between the variables, as can be seen in the pairwise scatterplots below (correlations in upper triangle):



Wine and beer have a negative relationship, and in fact the sum of wine and beer consumption is practically a constant (approx. 3.0-3.2). In Fig. 3, since wine+beer ≈ constant, it is just the varying spirits that could be related to log(wine/beer), but it isn't.

The ILR in Fig. 1 creates a relationship, because the ILR varies as the wine and beer components of the geometric mean of the denominator change, even though their sum is almost a constant (I illustrate how much the ILR can vary in my contribution **`5-rejoinder-to-vera-and-juanjo.pdf`** to this debate, see page 2 of that document). Thus, the scatterplot of the two ILR balances in Fig. 1 is showing that one relationship between wine and beer (the changes in the geometric mean in the denominator of the ILR balance) is correlated with another relationship between wine and beer (the logratio of wine to beer)!

The true state of affairs is revealed in the **logratio analysis (LRA)** below, namely the logratio of wine to beer (which accounts for about 93% of the variance), with nothing much else to interpret: