23. *QUALITY CONTROL PLOTS FOR EVALUATING CO-RELATIONS*

This chapter describes a simple graphical method for studying the relationships between pairs of person measures, pairs of item calibrations or any other pairs of values intended to display a co-relation. This method replaces the elusive numerals of correlation and regression coefficient(s) with easy to see pictures for describing relationships between pairs of values.

PROBLEMS WITH CORRELATION AND REGRESSION COEFFICIENTS

The traditional approach to investigate a relationship between pairs of values has been to calculate a correlation coefficient and stop there. The single number which results is used to describe the relation between the paired values. But how can any single number do justice to a potentially complex relationship so completely and so fully that no further information is of interest?

The correlation coefficient attempts to summarize in one number all the information contained in all pairs of values under consideration. The single resulting number is only occasionally evaluated in terms of its standard error and even less often dis-attenuated for the measurement error inevitably contained in the estimation of the pairs of values compared.

Dis-attenuation for measurement error should always be done for every coefficient used to quantify a relation between pairs of values. Dis-attenuation for measurement error is called for in every regression analysis. And there are further problems to consider. The usual regression analysis assumes there is no error in the independent variable(s) and that error in the dependent variable away from the modeled relation is entirely random and the only error expected. But usually the independent variables themselves are estimates containing their own error of estimation.

Relationship analysis needs to identify, separate and separately consider:

(1) *modeled error*, the explicit stochastic part of the relational theory implemented by the regression analysis,

(2) *measurement error*, an unavoidable part of all values in the analysis which depend on a prior estimation procedure, and

(3) *model misfit error*, the discrepancy between the general theory modeled and the particular data which is being examined for the extent to which it constrains or contradicts the theory modeled.

These misunderstandings occur whenever simple correlation or regression coefficients are accepted as sufficient summaries of relationships. These single values give only the barest and most incomplete description of the situation. They are based on the presumption that nothing is happening in the data except a simple linear relationship between two exactly known variables which can be captured by one coefficient as a single value. To presume this condition is to specify in advance that all people or items whose pairs of values are used to compute the correlation are nothing more than random examples of a single, simple linear relationship.

To say this again: When we rely upon a correlation coefficient to convey all that is operating in the relationship between a set of data points, we are reducing all the people or items examined to the status of equivalent examples of whatever the single, simple linear relationship is determined to be. Every person or item is reduced to being exchangeable in demonstrating the one relationship presumed in the data.

This conceptual reduction is not only never true but also never useful. Reducing any relationship to a single number denies and conceals all of the interesting individual behavior occurring in the data. The reduction prevents any realization of the diagnostic capacity of the data. To routinely discard this rich potential is not good science.

INVESTIGATING CO-RELATIONSHIPS USING PLOTS

We give each pair of values its own identity when we plot their location. Instead of reducing the data to a single correlation coefficient, the paired instances of the two variables are plotted against each other so that every data point represents a relation between the paired values - a relation that invites further investigation before summarizing.

Every plotted point should be clearly labeled so that unusual points can be examined to determine their specificity. When a point represents two measures on two variables for a particular person, that point is specific to that person. If the person is better in spelling than in arithmetic, their data point is uniquely informative about that aspect of that person. This is quite beyond and far more interesting than any general correlation which may be observed between spelling and arithmetic.

The first step in addressing the problem of co-relation is overlooked when the plot of paired values is not drawn, not labeled and not carefully studied for the particular identities of unusual points.

Some people find it difficult to examine a plot by inspection. They do not derive benefit from a simple examination because the plot is not set up in a way that tells them a story about what might be seen in their data.

LABELS FOR PLOTTED POINTS

In order to interpret plots we need to enable the plotted points to bring out the purpose of the plot and to make the story contained in the points immediately visual. Careful attention to labeling enables us to make visible the idiosyncratic and diagnostic possibilities in the data.

It is essential to label each point with a label that identifies in each plot what each point stands for i.e. male (M) or female (F), black (B) or white (W), married (m) or not (n). We can't investigate whether points are as expected or discover a pattern, if we cannot see what the points stand for. If we discover clusters of points we need to see on the graph what characteristics the clustered points share and do not share. This means we will replot the same points but with differing label sets to bring out the dominant patterns.

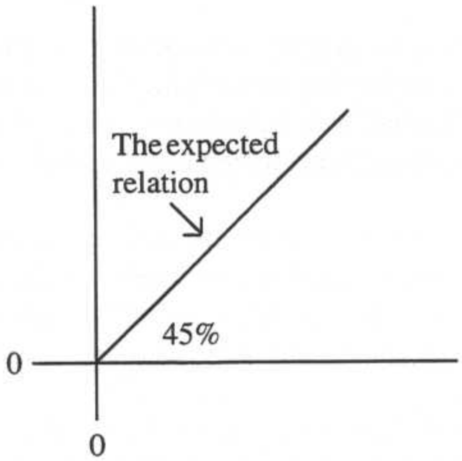
We need ways to label points so that any organization they manifest will be immediately apparent, so that we can see what the points indicate. The labeling of points must be as comprehensive and as versatile as possible. When labeling clutters up the plot or becomes too extreme to show on the plot itself, then a code number can be given each point and an accompanying legend (located next to or on the plot) constructed so that the points can be quickly identified and their pattern understood.

The more comprehensive the legend, the more assistance it will provide in investigating the nature of the plotted points. Graphical notations printed in position on the plot, however, communicate more quickly than text in a legend. Thus it is useful to develop versatility in successively altering point identification and replotting the same data so as to bring out the main patterns contained in the data by making them visible on the graph itself.

IDENTITY LINES

Labeling data points and providing a legend are not enough. We must go further and draw into the plot a line (or curve) that represents the main question to be investigated by these data - the main question which the plot is intended to answer. This "identity (of the question asked) line" should be a smooth, preferably (with data transformed so that it becomes) straight line drawn so that it marks the hypothesized path of the presumed relationship between the two variables.

To expedite the visual interpretation of any plot, it is important to adjust the scales of the horizontal and vertical axes so that the resulting "space" revealed is square. When this adjustment achieves a complete equation, the simplest version of this identity line goes through the origin with a slope of one, proceeding at a 45 degree angle across the plot from lower left to upper right and indicating a positive relationship between two variables on the same scale.



This simplest identity line specifies that the two sets of values are intended to measure the "same" thing from the same origin on the same scale: inches-to-inches, pounds-to-pounds or the logits-to-logits of commonly calibrated items.

If, in a study of item bias, we co-calibrate items to a common scale, we can plot pairs of item calibrations and use the identity line to model "no bias" between the two calibrations. The line shows which item points do not fit the "no bias" hypothesis represented by the identity line and hence which items require further investigation.

Usually the two values plotted originate on somewhat different scales. For pairs of measures, origins and scales are usually expected to be different. Then, a useful representation of the hypothesized relationship may be a different kind of identity line that passes through the means of the two sets of values with a slope equal to the ratio of their standard deviations. Again the best choice of horizontal and vertical scales is one that makes the resulting plot fill out a square.

An appealing and seemingly equivalent approach is to standardize the values of each variable by subtracting a mean M from each value X and dividing the difference by a standard deviation S:

$$z = \frac{X - M}{S}$$

When these standardized z values are plotted, the hypothesis bearing identity line once again goes through the origin with a slope of one.

The shortcoming of this standardization is that it draws our attention away from the metric(s) of the original variables. It is seldom useful to forget what the original metrics stand for. That metric information can be a key to understanding the data plot.

Thus it is usually more informative to retain the original metrics of the variables and not to standardize. That places the hypothesis bearing identity line through the intersection of the means with slope determined by the ratio of the standard deviations.

THE HYPOTHESIS REPRESENTED BY THE IDENTITY LINE

The identity line represents the hypothesis of a perfect relationship. The utility of the identity line is that it guides the eye in examining the data points with respect to the hypothesis. We can see which data points are close to the identity line and which points are far from it and thence indicative of a particular and identifiable digression from the perfect relation hypothesized.

The deviations are the exceptions, the unexpected digressions from the perfect idea indicated by the identity line. The identity line also guides the eye to locations where no data points exist. The data points which follow the identity line confirm our expectations. The data points that deviate contradict our expectations. The data points that are missing show us where we are uninformed.

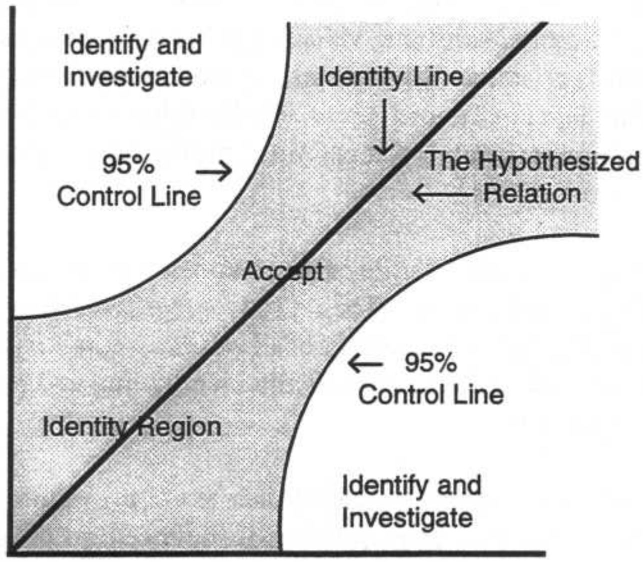
The statistical model used with most correlations is a null hypothesis of zero correlation between the two variables. But when we model a relationship, the relevant null hypothesis is seldom zero but rather a perfect relation as close to one as measurement error allows. This more useful "null hypothesis" of perfect relation is the one relevant to measurement analysis.

CONSTRUCTING QUALITY CONTROL LINES

How can we show the extent of expected error in a plot? How can we make allowance for measurement error visible in the plot? How can we visualize error dis-attenuation? The answer is to draw in quality control lines to guide inspection of the data plot and to provide guidelines for seeing how close, statistically speaking, our estimated points are to the identity line, given their errors of measurement.

These error guidelines are constructed in the same way as the statistics used in industrial quality control. We draw two boundary lines, one above, one below the identity line, to guide inspection of data points. These lines make the statistical boundaries of our hypothesis visible.

We usually construct this pair of boundaries so that they enclose 95% of the data points which measurement error around a perfect relation would produce. These boundary lines enclose a region containing two standard errors of measurement around the identity line in each direction:



Quality control lines enable visual evaluation of the data points. They show us the identity "line" and the identity "region"; the area around the identity line in which it is reasonable for data points to occur, given the measurement error.

Data points which fall within the control lines can be accepted as statistically equivalent to the identity line and hence to the hypothesized relation. These data points do not contradict the hypothesis represented by the identity line.

Data points which lie outside the control lines are, however, instances which contradict the identity line i.e. the hypothesis. Each outlying point is a visible contradiction to the hypothesis and consequently each outlying point needs to be identified and investigated in order to understand and explain what has been observed, in order to discover the meaning of the contradiction.

If, when studying a sample of people who have been measured on two variables, we find that their paired measures follow an identity line, then the paired measures are clearly on a single variable and the two initial variables are empirically co-dimensional, at least for these people.

MEANS AND STANDARD DEVIATIONS

Even when two variables are both conceptually and empirically co-dimensional, there will still be some individuals for whom the relationship does not hold, some exceptions. Inclusion of these deviant values in calculating means and standard deviations for these data, however, disturbs these two commonly used reference statistics.

We want to determine the extent to which the data follow the line which asserts and/or supports our intended hypothesis. To make this determination we begin by evaluating all of the data points in terms of our theory.

Without any theory to guide our observations we are only fishing for something we cannot yet describe. This is not research but blind groping. While there may be times when we find ourselves perplexed by a measurement problem, that confusion is neither optimal nor scientific.

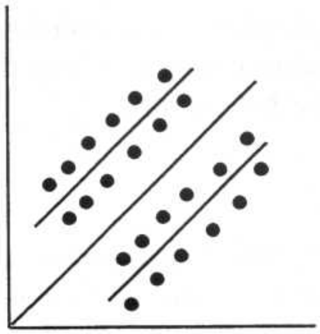
Outliers are contradictory data points. They become unusual in the light of our expectations and so in need of investigation. To include outlying values together with those data points that confirm the hypothesis in computing means and standard deviations is to remain confused by our own data. Means and standard deviations are vulnerable to extreme values. Outliers distort the conclusions we come to. We want robust statistics that are not unduly influenced in central location and dispersion by idiosyncratic, extreme values.

Statistics like the median and interquartile range are sometimes advocated as useful because these statistics are less influenced by extreme values. Their disadvantage is that they lack precision and power. What we want is not the mean and standard deviation of all values, exceptions included, but the mean and standard deviation of just those values which follow the identity line and hence do not contradict the hypothesis of a shared dimensionality.

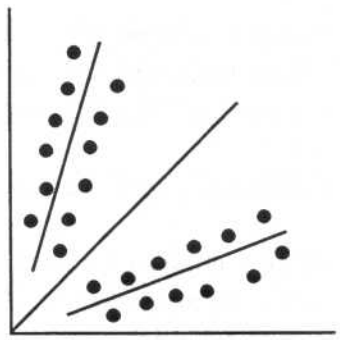
As we survey a plot we need a convenient and consistent way to exclude the outliers. When deviant data points are identified we want to recompute means and standard deviations without including these deviant points and then to re-draw the identity and control lines so that they represent only the points of the subsample of people who confirm the hypothesis of a general relationship and not those of the people who contradict it.

This does not mean that we throw the other data away. On the contrary, it is important to investigate all the data, and most especially the deviant data points. But, it is necessary to determine what criteria is to be used in making the decision concerning deviance. If our hypothesis is depicted by an identity line, then statistically significant deviations are those data points beyond the quality control boundary lines. These values, then, because they are different, do not belong when calculating the summary statistics used to locate the identity and control lines.

Usually when data plots are examined it is easy to see whether the points are following a line. Sometimes we see two groups of points that follow two lines:



or



No statistic can determine which of the two lines we should use as our expectation, our intended hypothesis, and which to consider as deviant, however. We must identify the data points involved and then engage in the hard work of thinking clearly about our intentions and how they emerge in the data plot. That is the only way to determine what reasonable hypothesis they support.

An average, even when successively adjusted for deviations is, nevertheless, only an estimate of central location. No automatic strategy can provide the detective work and speculative inspiration needed for creative analysis. No automatic technique can substitute for patient investigation, visual inspection and careful thought. No automatic process can examine the data points in such a way as to replace intelligent review. The evaluation of unexpected data points requires the combined efforts of data analyst and content specialist so that each can encourage the other to investigate all possible hunches concerning the patterns manifested in the plot.

Two other chapters provide examples of this: Chapter 8 (p.57), Identifying Item Bias and Chapter 9 (p.65), Control Lines for Item Plots.

MEASUREMENT ESSENTIALS

2nd Edition

BENJAMIN WRIGHT

MARK STONE

WIDE RANGE, INC.
Wilmington, Delaware