Few Crystal balls are Crystal Clear, Eyeballing Regression
Rob T. Wittebrood, Wittebrood Petrophysical Services, Hycal Energy Research Laboratories

Well log analysis, and other techniques, are increasingly done using computer programs. These programs often have many routines and subroutines available, acting like crystal balls, to predict various reservoir parameters. Most of us realize that crystal balls are only to be used by the initiated few. Most users find these routines far from clear. Many appear opaque and some even may seem to be black to the casual user.

Statistical analysis can easily be abused, often even unintentionally. However, most users do not realize that the majority of our statistical results are in fact no better than a well-educated guess, especially in the oil industry. For instance, most often, no one is able to prove that a regression line is better or worse than a plain eyeballed line, in spite of all the crystal balls that have been used.

There are many reasons why this is the case. These are rarely revealed in standard statistical texts or in school. Before regression can be done, you should have enough knowledge about: the shape of the joint probability distribution, the presence of lurking variables, the generation process of the data, the sampling procedures, the presence of outliers etc. Even when the data is fully cooperative and satisfies all conditions for a regression, the regression problem can not be solved. *We always lack enough knowledge to calculate the proper regression line.* This fact is hardly known. In order to calculate a regression line we have to include at least one other assumption, which we have to make ourselves, and is usually based on the potential use of the regression results. The freedom in this assumption leads to the large variety of regression techniques that are available i.e. standard regression, weighted regression, inverse regression, reduced regression, bivariate regression, robust regression and others. All these different techniques lead to different regression lines. The accuracy and the validity of the additional assumption is generally uncertain, making the regression line somewhat of a guess rather than a presumed truth.

We shall discuss all the prerequisite conditions and the various common assumptions using *examples from the field of petrophysics*. We will show that in nearly all cases the regression lines are actually uncertain to some extent. So uncertain in fact, that an eyeballed line would do equally well. We do not want to dissuade you from using regression in your professional future. We want to convince you that you should accept your neighbors line like your own. In fact we are convinced that in many cases an eyeballed line is a better fit than a regression line based on limited knowledge about the data.

However, a reliable regression line can be calculated when the behavior of the data is well understood. It is imperative that you know the distributions, dependencies and source of your data and their deviations. A causal understanding of the expected relation is also essential. Regression is not magic, it just is hard work.

(19; :)'I gq/ʋtkcf "); : <'Cduʋtcevʋ<'Rngpct{ .'Qtcn 'Rquʋgtu. 'Eqtg''Y qtmʋj qr