REGRESSION

CORRELATION & REGRESSION

- Data: We have measurements made on each subject, one is the response variable Y, the other predictor X (maybe more). There are two types of analyses:
- Correlation: is concerned with the association between them, measuring the strength of the relationship
- Regression: To predict response from predictor. Is a woman's Weight Gain during pregnancy predictive of her newborn's Birth Weight?

REGRESSION MODEL

- Let Y be the Dependent Variable and X the Independent Variable. For a particular value x of X, the values of Y is assumed to follow certain distribution.
- For example, among the mothers who gained 37 lbs during their pregnancies and gave birth to baby boys, the boys' birth weights may not be all the same but form certain distribution – with mean f(x).

$$Y_{|X=x} = f(x) + \varepsilon$$

$$\varepsilon \text{ is the "random" part}$$

NORMAL LINEAR REGRESSION MODEL

- Let Y be the Dependent Variable and X the Independent Variable. The Dependent Variable Y for the sub-population with X=x, is assumed to be "normally distributed". The Normal Error Regression Model describes the Mean of that Normal Distribution as a function of value X=x.
- If we focus only on linear relationship, the above function represents the equation of a straight line- with two parameters, a Slope and an Intercept.

REGRESSION MODEL

- Model: $Y = \beta_0 + \beta_1 x + \varepsilon$ where β_0 & β_1 are two parameters called regression coefficients, the Intercept and the Slope, respectively. The term ε is the "error" representing the random fluctuation of y-values around their mean, $\beta_0 + \beta_1 x$,
- The "error" is an important characteristic of a statistical relationship.

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

$$\varepsilon \in N(0, \sigma^2)$$

The "normal" assumption can sometimes be weakened to $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$

REGRESSION COEFFICIENTS

- The error ε - with variance σ^2 -would tell how spread the dots are around the regression line.
- The regression coefficients, β_0 and β_1, determine the position of the line. As parameters, both β_0 and β_1 are unknown; but they can be “estimated” by statistics from data.

SUM OF SQUARED ERRORS

- When X=x, the Mean of Y is $\beta_0 + \beta_1 x$.
- Let b_0 and b_1 are estimates of β_0 and β_1; then $(b_0 + b_1 x)$ is an estimate of y. The error of that estimate is $[y - (b_0 + b_1 x)]$ so that $Q = \Sigma\ [y - (b_0 + b_1 x)]^2$ represents "the sum of squared errors"
- The method of least squares requires that we find "good estimates" of β_0 and β_1 the values of b_0 and b_1 so as to minimize this "sum of squared deviations".

METHOD OF LEAST SQUARES

$$\text{Data}: \ \{(x_i, y_i)\}_{i=1}^{n}$$

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\delta Q}{\delta \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{\delta Q}{\delta \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{\delta Q}{\delta \beta_0} = \frac{\delta Q}{\delta \beta_1} = 0$$

called "Normal Equations" (page 17):

$$\sum \mathrm{y_i} = nb_0 + b_1 \sum x_i$$

$$\sum \mathrm{x_i y_i} = b_0 \sum x_i + b_1 \sum x_i^2$$

Results: Point estimators/estimates

RESULTS

- The “Least Squares Estimates” are:

$$\mathbf{b_1} = \frac{\sum(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{y}-\bar{\mathbf{y}})}{\sum(\mathbf{x}-\bar{\mathbf{x}})^2},\quad \mathrm{b}_0 = \bar{y} - b_1\,\bar{x}$$

- Given the estimates “b_0” of the Intercept and “b_1”^ of the Slope, Estimate of y (for the mean or a “new” value x of X) is $\hat{Y} = b_0 + b_1x$; this is called “fitted value”.

SAMPLING DISTRIBUTION

Under the "Normal Error Regression Model":

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

$$\varepsilon \in N(0, \sigma^2)$$

The sampling distribution of the estimated slope

b_1 is Normal with Mean and Variance:

$$\mathrm{E}(\mathrm{b}_1) = \beta_1$$

$$\sigma^2(\mathrm{b}_1) = \frac{\sigma^2}{\sum (x - \bar{x})^2}$$

$$\widehat{Var}(b_1) = \frac{MSE}{\sum (x - \bar{x})^2}$$

$$SE(b_1) = \sqrt{\frac{MSE}{\sum (x - \bar{x})^2}}$$

COEFFICIENT OF CORRELATION

The Correlation Coefficient measures the strength of the relationship:

$$\boxed{r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{[\sum (x-\bar{x})^2][\sum (y-\bar{y})^2]}}}$$

WE could “explain” the denominator as necessary for standardization: to obtain a statistic in [-1,1].

There are many different ways to express the coefficient of correlation n; one of which is often mentioned as the “short-cut” formula:

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{[\sum x^2 - \frac{(\sum x)^2}{n}][\sum y^2 - \frac{(\sum y)^2}{n}]}}$$

$$= \frac{s_{xy}}{s_x s_y}$$

s_{xy} is the “sample covariance” of X and Y

Another very useful formula is to express the coefficient of correlation r as the “Average Product” in “standard units” – where s_x and s_y are the (sample) standard deviations of X and Y, respectively:

$$r = \frac{1}{n}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$$

The following is an important and very interesting characteristic:

$$r(cX + d, aY + b) = r(X, Y)$$

we can prove by showing that (u = c*x + d) is the same as x in standard units & (v = a*y + b) is the same as y in standard units; e.g.

$$\frac{u - \bar{u}}{s_u} = \frac{x - \bar{x}}{s_x}$$

"Slope" is the more important parameter and the followings are few alternative expressions:

$$\boxed{b_1 = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}} \longrightarrow \quad b_1 = \frac{s_{xy}}{s_x^2}$$

$$= r\frac{s_y}{s_x}$$

$$b_1^2 = r^2\frac{s_y^2}{s_x^2}$$

$$b_1 = \frac{s_{xy}}{s_x^2}$$

$$= r\frac{s_y}{s_x}$$

From this simple result, we can see that "b_1" and "r" are of the same sign – and both are equal to zero at the same time

The classic work of Pearson (Biometrika, 1909) and Fisher (Biometrika, 1915; Metron, 1921) led to the (Pearson's, product moment) coefficient of correlation r which generated a steady stream of development of measures of association and correlation coefficients appropriate in different contexts (latest: Kraemer, SMMR, 2006).

At the beginning of the 20th century, correlation analysis was one of the most common statistical analyses, especially in health psychology and epidemiology. Conversely, the use of coefficient of correlation r has had major influence on medical policy decision making.

We end up having “a correlation analysis” and “a regression analysis” that go hand-in-hand: the slope “b_1” and the coefficient of correlation “r” are of the same sign – and both are equal to zero at the same time. But, in general, it does not have to be that way

A measure of the strength of an association, say θ, needs only satisfying the following conditions: (1) It is an unit-free measure that range from -1 to +1, (2) If X and Y are independent, θ = 0, and (3) If one can perfectly predict Y from X, or X from Y, then θ = 1.

Besides the (Pearson’s) coefficient of correlation r, we have (i) Spearman’s rho and (ii) Kendall’s tau

Most of the times, we have more than just two measurements; may be only one “Response” but many Predictors – at least many “Potential Predictors”. Then we could ask why predicting the response from just one predictor? Or, if there are more than one predictors, then which predictor? Or which potential predictor would be the most valuable one? etc…

THE DRAWBACKS OF SLR

- The effect of some factor X on the dependent variable Y may be influenced by the presence of other factors through effect modifications, i.e. interactions; SLR does not allow us to study "effect modifications"
- Even without interactions, information provided by different factors may be "redundant" (overlapsed). There is a difference between "contribution" and "marginal contribution".

SLR does not allow us to investigate "marginal contribution" of a factor in the prediction of a response (where the effect of some factor on a dependent variable may be influenced by the presence of other factors, not by effect modification, but because of redundancies). Marginal contribution of a factor is its contribution on top of or in addition to the contributions by other factors.

In addition, not all relationships are “linear”; high blood pressure (hypertension) is not good but very low blood pressure (hypotension) is also bad;

Simple Linear Regression does not allow us to study, say, quadratic relationships – or higher-order polynomials.

A POSSIBLE SOLUTION

In order to provide more comprehensive prediction of the dependent variable Y – say the outcome of certain treatment, it is very desirable to:

(1) Consider a large number of factors – with available data - and

(2) Sort out which ones are most closely related to that outcome.

THE MULTIPLE REGRESSION MODEL

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \varepsilon$$

$$\varepsilon \in N(0, \sigma^2)$$

The “fixed” part is the Mean and the “random” part is the error ε

THE MEAN IN A MLR MODEL

- Suppose we want to consider k independent variables simultaneously, the simple linear model can be easily generalized and expressed as:

  Mean of Y = $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$
- The β's are the (k+1) unknown parameters; β_0 is the intercept and β_i's are the slopes, one slope for each independent variables; x_i is the value of the ith independent variable (i = 1 to k) – considered as "fixed" or "designed".

GENERAL LINEAR MULTIPLE REGRESSION MODEL

- The linear multiple regression model can be generalized and expressed as:
  Mean of Y = $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$
  plus a normally-distributed error term.
- However, the k X's do not need to represent k different predictors; some term could be the product of two predictors, some term could be the quadratic power of another predictor.
- The terms involved could be continuous, binary, or categorical (more than 2 categories)

TESTING HYPOTHESES

- Once we have fitted a multiple linear regression model and obtained estimates for the various parameters of interest, we want to **answer questions about the contributions of factor or factors** to the prediction of the dependent variable Y. There are three types of tests:

(1) An **overall** test

(2) Test for the value of a **single factor**

(3) Test for contribution of a **group of factors**

OVERALL TEST

- The question is: “ Taken collectively, does the entire set of explanatory or independent variables contribute significantly to the prediction of the Dependent Variable Y?”.
- The **Null Hypothesis** for this test may stated as: **“All k independent variables, considered together, do not explain the variation in the values of Y"**. In other words, we can write:

$$H_o : \beta_1 = \beta_2 = ... = \beta_k = 0$$

This “Over Test” often precedes others, making sure that we have “reason” to look for effects of individual factors – and to control for type I errors; similar to performing F-test in One-way ANOVA before looking for possible pairwise differences.

TEST FOR SINGLE FACTOR

- The **question** is: “Does the addition of one particular factor of interest add significantly to the prediction of Dependent Variable **over and above that achieved by other factors in the model**?”.
- The **Null Hypothesis** for this test may stated as: **"Factor X_i does not have any value added to the explain the variation in Y-values over and above that achieved by other factors** ". In other words, we can write:

$$H_0 : \beta_i = 0$$

Logically, we could look for some individual effect – even without the overall test – if it’s the “primary aim” of the research project. For example, this factor represents “**treatment assignment**” in a **clinical trial** (say, =1 if treated & 0 if placebo)

TEST FOR A GROUP OF VARABLES

- The **<u>question</u>** is: “Does the addition of <u>a group of factors</u> add significantly to the prediction of Y **over and above that achieved by other factors**?
- The **Null Hypothesis** for this test may stated as: "Factors $\{X_{i+1}, X_{i+2}, \ldots, X_{i+m}\}$, considered together as a group, do not have any value added to the prediction of the Mean of Y over and above that achieved by other factors ". In other words,

$$H_0 : \beta_{i+1} = \beta_{i+2} = \ldots = \beta_{i+m} = 0$$

TEST FOR A GROUP OF VARABLES

- This "**multiple contribution**" test is often used to test whether a **similar group of variables**, such as demographic characteristics, is important for the prediction of the mean of Y; these variables have **some trait in common**.
- Other groupings: **Psychological** and **Social** aspects of health in **QofL research**

Another application: **collection of powers and/or product terms**. It is of interest to assess powers & interaction effects collectively before considering individual interaction terms in a model. It reduces the total number of tests & helps to provide **better control of overall Type I error rates** which may be inflated due to **multiple testing**.

ANOVA IN REGRESSION

- The variation in Y is conventionally measured in terms of the deviations $(Y_i - \bar{Y})$'s; the total variation, denoted by SST, is the **sum of squared deviations**: $SST = \Sigma(Y_i - \bar{Y})^2$. For example, SST = 0 when all observations are the same; SST is the numerator of the sample variance of Y, the greater SST the greater the variation among Y-values.
- When we use the regression approach, the variation in Y is decomposed into **two components**: $(Y_i - \bar{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$

Three Basic SUMS OF SQUARES

- In the decomposition: $(Y_i - \overline{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \overline{Y})$
- The first term reflects the variation **around** the regression mean (fitted value); the part than <u>cannot</u> be explained by the regression itself with the sum of squared deviations: $\mathbf{SSE = \Sigma(Y_i - \hat{Y}_i)^2}$.
- The difference between the above two sums of squares, **SSR** = SST - SSE = $\mathbf{\Sigma(\hat{Y}_i - \overline{Y})^2}$, is called the **regression sum of squares**; SSR measures of the variation in Y associated with the regression model.
- Three basic sums of squares are: SST = SSR + SSE (this result can be easily proved – and we did).

“ANOVA” TABLE

- The breakdowns of the total sum of squares and its associated degree of freedom are displayed in the form of an “analysis of variance table” (ANOVA table) for regression analysis as follows:

Source of Variation	SS	df	MS	F Statistic	p-value
Regression	SSR	k	MSR	MSR/MSE	
Error	SSE	n-k-1	**MSE**		
Total	SST	n-1			

- **MSE**, the “error mean square”, **serves as an estimate of the constant variance σ^2** as stipulated by the regression model.

In using “Multiple Regression”, the emphasis is in “**Marginal Contribution**”. For example, if we use the following two-factor” model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\varepsilon \in N(0, \sigma^2)$$

The results would tell us:

(a)The effects of X_1 on Y **<u>after adjusted for</u>** X_2

(b)The effects of X_2 on Y **<u>after adjusted for</u>** X_1

SEQUENTIAL PROCESS

- SSR measures of the variation in Y associated with the regression model with k variables; Regression helps to "reduce" SST **by the amount** SSR to what left unexplained in SSE.
- Instead of a model with k independent variables, let consider the sequential process to include one or some variables at a time.
- By doing this we can focus on the "**marginal contribution**" of the variable or variables to be added – as **further reduction** in SSE.

“EXTRA” SUMS OF SQUARES

- An extra sum of squares measures the (marginal) **reduction** in the error sum of squares when one or several independent variables are added to the regression model, given that the other variables are already in the model.
- Equivalently, an extra sum of squares measures the (marginal) **increase** in the regression sum of squares when one or several independent variables are added to the regression model, given that the other variables are already in the model.

MARGINAL CONTRIBUTION

- The changes from the model containing only X_1 to the model containing both X_1 and X_2 represent the “marginal contribution” by X_2.
- The marginal contribution represent the part (of SSE) that is further explained by X_2 (in addition to what already explained by X_1).
- **$SSR(X_2|X_1) = SSR(X_1,X_2) - SSR(X_1)$** is the **“extra sum of squares due to the addition of X_2 to the model that already includes X_1.**

The ANOVA Table was constructed to show the decomposition of SST into SSR(X_1,X_2) and SSE(X_1,X_2); **SSR(X_1,X_2) represents the combined contribution by X_1 and X_2**.

In the sequential process, we can further decompose **SSR(X_1,X_2)** to show the contribution of X_1 (which enters first), **SSR(X_1)** ,and the marginal contribution of X_2, **SSR($X_2|X_1$)** ; this involves addition of one single factor, so this SSR($X_2|X_1$) is associated with one degree of freedom.

(Amended) ANOVA TABLE

- The breakdowns of the total sum of squares and its associated degree of freedom could be displayed as:
- (amended) ANOVA Table:

Source of Variation	SS	df	MS
Regression	SSR(X1,X2)	2	MSR(X1,X2)
Due to X1	SSR(X1)	1	MSR(X1)
Addition of X2: X2|X1	SSR(X2|X1)	1	MSR(X2|X1)
Error	SSE(X1,X2)	n-3	MSE(X1,X2)
Total	SST	n-1	

TEST FOR SINGLE FACTOR

- The **question** is: “Does the addition of one particular factor of interest add significantly to the prediction of Dependent Variable **over and above that achieved by other factors**?”.
- The **Null Hypothesis** for this test may stated as: "Factor X_i does not have any value added to the explain the variation in Y-values over and above that achieved by other factors ". In other words,

$$H_0 : \beta_i = 0$$

TEST FOR SINGLE FACTOR #1

- The Null Hypothesis is $H_0 : \beta_i = 0$
- Regardless of the number of variables in the model, one simple approach is using

$$\mathbf{t} = \frac{\mathbf{b_i}}{\mathbf{SE(b_i)}}$$

- Refer it to the percentiles of the **t-distribution** with df is the error degree of freedom, where b_i is the corresponding estimated regression coefficient and $SE(b_i)$ is the standard error of β_i , both of which are provided by any computer package – in **one "run".**

TEST FOR SINGLE FACTOR #2

- Suppose we have a multiple regression model with 2 independent variables (the **Full Model**):

$$\mathbf{Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon}$$

$$\mathbf{\varepsilon \in N(0, \sigma^2)}$$

& Suppose we are interested in the Null Hypothesis:

$$H_0 : \beta_2 = 0$$

- We can compare the Full Model to the **Reduced Model**:

$$\mathbf{Y = \beta_0 + \beta_1 x_1 + \varepsilon}$$

$$\mathbf{\varepsilon \in N(0, \sigma^2)}$$

TEST FOR SINGLE FACTOR #1a

- Test Statistic

$$F^* = \mathbf{MSR(X_2|X_1)/MSE(X_1,X_2)}$$

- Rejection Rule

Reject H_0 if $F^* > F_\alpha$

where F_α is based on an F distribution with **one (1) degree of freedom (numerator) and (n-3) degrees of freedom (denominator).**

We can perform an F test or a t-test.

Actually, **the two tests are equivalent**; numerical value of F* is equal to the square of the calculated value of "t": **$F^* = t^2$**. The t-test needs one computer run and the F test requires two computer runs – one with all variables in and one with the variable under investigated set aside.

TEST FOR A GROUP OF VARABLES

- The **question** is: "Does the addition of a group of factors add significantly to the prediction of Y **over and above that achieved by other factors**?
- The **Null Hypothesis** for this test may stated as: "Factors $\{X_{i+1}, X_{i+2},\ldots, X_{i+m}\}$, considered together as a group, do not have any value added to the prediction of the Mean of Y that other factors are already included in the model". In other words,

$$H_0 : \beta_{i+1} = \beta_{i+2} = \ldots = \beta_{i+m} = 0$$

The “ANOVA Approach”, through the use of “extra sums of squares”, was not really useful for testing concerning single factors because we could more easily use the point estimate and standard error of the regression coefficient to form a t-test. In fact, we do not have to do anything; all results are obtained “automatically” from any computer package. However, the story is very **different for tests concerning a group of several variable**: **NO EASY WAY OUT**; **In the sequential process, we can add in more than one independent variables at a time.**

(Amended) ANOVA TABLE

- The breakdowns of the total sum of squares and its associated degree of freedom could be displayed as:
- (amended) ANOVA Table:

Source of Variation	SS	df	MS
Regression	$SSR(X_1,X_2,X_3)$	1	$MSR(X_1,X_2,X_3)$
Due to X1	$SSR(X_1)$	1	$MSR(X_1)$
Addition of X2,X3|X1	$SSR(X_2,X_3\|X_1)$	2	$MSR(X_2,X_3\|X_1)$
Error	$SSE(X_1,X_2,X_3)$	n-4	$MSE(X_1,X_2,X_3)$
Total	SST	n-1	

In the decomposition of the sums of squares; the "extra sums of squares" are not only useful for testing for the marginal contribution of individual variable or group of variables; they can also be used to "**measure**" these contributions. Unlike the statistical tests of significant, **the measures are not affected by the sample size**. Let start with the coefficient of multiple determination as a measure of the **proportionate reduction in the variation of Y** achieved by the regression model.

COEFFICIENT OF DETERMINATION

- The ratio, called the **coefficient of multiple determination**, defined as:

$$\mathbf{R^2 = \frac{SSR}{SST}}$$

- It represents the portion of total variation in y-values attributable to difference in values of independent variables or covariates.

COEFFICIENT OF PARTIAL DETERMINATION

- Suppose we have a multiple regression model with 2 independent variables (the Full Model) and suppose we are interested in the marginal contribution of X_2:

$$\mathbf{Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon}$$

$$\boldsymbol{\varepsilon \in N(0, \sigma^2)}$$

- The **coefficient of partial determination**, between Y and X_2 measures the marginal reduction in the variation of Y associated with the addition of X_2, when X_1 is already in the model:

$$R^2_{Y2|1} = \frac{SSR(X_2 \mid X_1)}{SSE(X_1)}$$

A limitation of the usual residual plots (say, residuals against values of a predictor variable): **they may not show the nature of the "additional contribution"** of a predictor variable to those by other variables already in the model.

For example, we consider a multiple regression model with 2 independent variables X_1 and X_2; is the relationship between Y and X_1 linear?

"**Added-variable plots**" (also called "partial regression plots" or "adjusted variable plots") are **refined residual plots** that provide graphic information about the marginal importance of a predictor variable given the other variables already in the model.

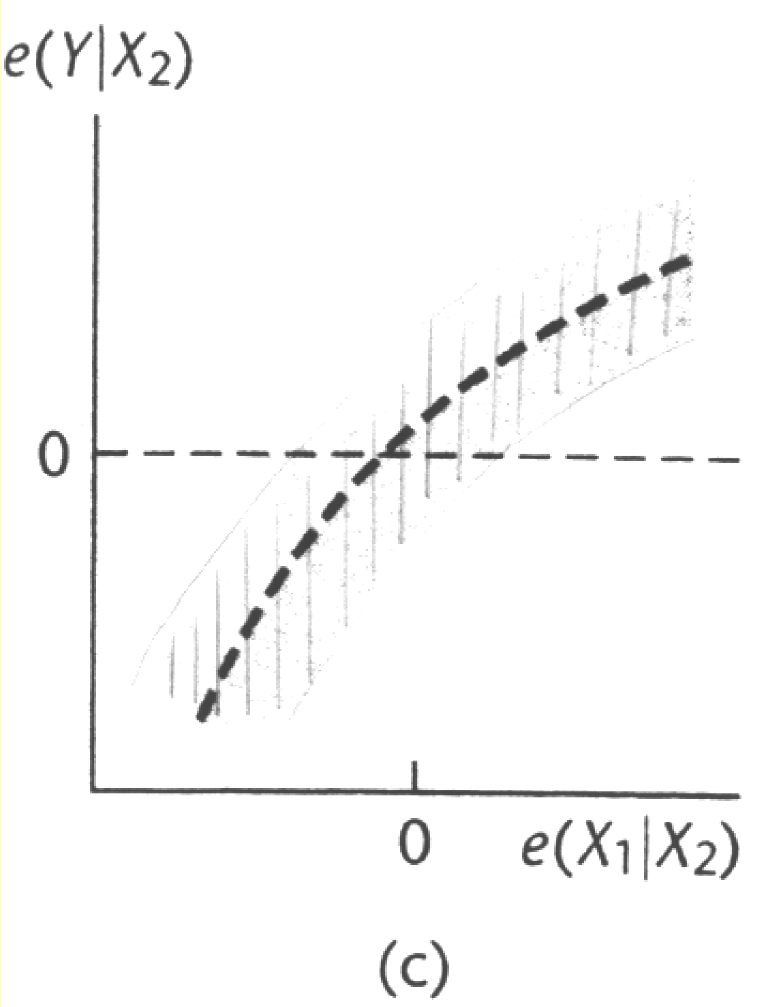
ADDED-VARIABLE PLOTS

In an added-variable plot, both the response variable **Y** and the predictor variable under investigation (say, $\mathbf{X_1}$) are **both regressed against the other predictor variables** already in the regression model and the **residuals are obtained for each**. These two sets of residuals reflect the part of each (Y and X_1) that is not linearly associated with the other predictor variables.

The plot of one set of residuals against the other set would **show the marginal contribution of the candidate predictor** in reducing the residual variability as well as the information about the nature of its marginal contribution.

A SIMPLE & SPECIFIC EXAMPLE

Consider the case in which we already have a regression model of Y on predictor variable X_2 and is now considering **if we should add X_1** into the model (if we do, we would have a multiple regression model of Y on (X_1,X_2)). In order to decide, we investigate 2 simple linear regression models: (a) A **regression of Y on X_2** and obtained residuals, and (b) A **regression of X_1 on X_2** and **obtain 2 sets of residuals**:



The "curvilinear band" indicates that, like the case (b) previously, X_1 should be added to the model already containing X_2. However, it further suggests that the **inclusion of X_1 is justified but some power terms – or some type of data transformation** – are needed.

The fact that an added-variable plot may suggest “the nature of the **functional form**” in which a predictor variable should be added to the regression model is more important **than that the variable possible inclusion** (which can be resolved without graphical help). The added-variable plots play the role that we use scatter diagrams for in simple linear regression; they would tell if data transformation or if certain polynomial model is desirable.

COEFFICIENTS OF
PARTIAL CORRELATION

The square root of a coefficient of partial determination is called a **coefficient of partial correlation**. It is given the **same sign as that of the corresponding regression coefficient** in the fitted regression model. Coefficients of partial correlation are frequently used in practice, although they do not have a clear meaning as coefficients of partial determination nor the (single) coefficient of correlation.

AN INTERESTING RESULT

Let both the response variable Y and the predictor under investigation (say, X_1) be both regressed against the other predictor variables already in the regression model and the residuals are obtained for each. These **two sets of residuals** reflect the part of each (Y and X_1) that is not linearly associated with the other predictor variables.

Result: The (simple) correlation **between the above two sets of residuals** is equal to the Coefficient of Partial Correlation between Y and X1. This has been verified numerically & **I found a proof in a book on "Linear Models" by Ronald Christensen.**

The importance of this result has been over-looked. It is important because it shows that we could study “marginal contributions” using non-parametric methods – such as Spearman and Kendall.