A SIMPLE LINEAR REGRESSION MODEL WITH A NON-CONSTANT VARIANCE

Howard S. Kaplon, Towson State College

In the classical theory of regression assumptions of normality are generally made and these in turn lead to the assumption of constancy of variance. In many situations the assumption of constancy of variance is not satisfied. This paper is concerned with the study of such a model in simple linear regression. Computational programs have been obtained to give estimates of parameters in case the standard deviation is assumed to be linear.

Simple linear regression of Y on X is usually defined by the equation $E[Y|X=x] = \alpha+\beta x$ and $Var[Y|X=x] = \sigma^2$ for all x. Here α and β are the regression coefficients and σ^2 is the assumed constant variance. The statistical problem of the investigator here is to estimate values of α and β with $\hat{\alpha}$ and $\hat{\beta}$ computed from a random sample. These estimates are of use in predicting values of the dependent statistical variable Y for observed values of the independent mathematical variable X. In addition, since in most situations the investigator would want to calculate a confidence band for his predicting equation, an estimator $\hat{\sigma}^2$ of σ^2 is also needed. The three estimators are usually derived using the method of maximum likelihood.

Let $(X_1,Y_1), (X_2,Y_2), \ldots, (X_n,Y_n)$ be a random sample taken from the population of concern (assumed to be normal and $Var[Y|X=x]$ is constant for all x). To estimate the parameters α, β and σ^2 the likelihood function

$$L=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i-(\alpha+\beta x_i)]^2\right\} \quad (1)$$

is set up and we take the three partial derivatives

$$\frac{\partial(\log L)}{\partial\alpha}, \frac{\partial(\log L)}{\partial\beta} \text{ and } \frac{\partial(\log L)}{\partial\sigma^2}$$

and set them equal to 0. This procedure yields

$$\sum_{i=1}^{n}(y_i-\alpha-\beta x_i) = 0, \quad (2)$$

$$\sum_{i=1}^{n}x_i(y_i-\alpha-\beta x_i) = 0 \quad (3)$$

$$\text{and } n\sigma^2 = \sum_{i=1}^{n}(y_i-\alpha-\beta x_i). \quad (4)$$

Upon solving the normal equations (2) and (3) for $\hat{\alpha}$ and $\hat{\beta}$ we obtain the desired estimators

$$\hat{\alpha} = \frac{(\Sigma y_i)(\Sigma x_i^2)-(\Sigma x_i)(\Sigma x_i y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} \quad (5)$$

$$\text{and } \hat{\beta} = \frac{n\Sigma x_i y_i-(\Sigma x_i)(\Sigma x_i y_i)}{n\Sigma x_i^2-(\Sigma x_i)^2} \quad (6)$$

then by using these values (5) and (6) in equation (4) $\hat{\sigma}^2$ can easily be found.

As an example of the above model consider the situation where an engineer wishes to determine the effect of heat and cold on the expansion and contraction of a certain metal (with a specified temperature) to be used in the construction of a bridge. He takes a sample of known length of this metal at a known temperature (the specified temperature) and heats it to a new, known temperature, and he measures the new length of the sample. He repeats this experiment as often as he feels is necessary, each time heating or cooling a sample of the same length from the same temperature. However, due to the measuring device, the time elapsed between the time when the metal has reached the desired temperature and the time it is measured, and human error, he does not expect his measurements to be exactly correct. Further he assumes that the errors in the measurements are normally distributed and independent of each other. By allowing the lengths of the pieces of metal to be represented by the random variable Y and the temperature by the variable X, the preceding model seems to fit this situation. The engineer can compute $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$, set up prediction equation $\hat{Y} = \hat{\alpha} + \hat{\beta}X$, calculate a desired confidence band, and construct the bridge accordingly.

However, consider an example of the of the public health official who needs to predict the hours of health care that will be needed per year by the adults at each age over 21 years of age. He takes a stratified random sample and obtains a pair of values for each individual: X - age and Y - hours of medical care needed per year. The official might very well choose the following as his model: Let $Y_i = \alpha + \beta X_i + e_i$ where e_i are normal distributed independent errors with a mean of 0 and at least two of the x_i are distinct. In this model the assumption that all σ_i^2 are equal is not a logical one. It would seem that as age increases, not only would the average need for medical care increase, but so would the variation of this need. Also, just because of the increase in the values of the random variable with the increase in the age - X, the variances σ_i^2 tend to increase. For these reasons the official could pick his model for regression as

$$E[Y_i|X_i=x_i] = \alpha+\beta x_i$$

and $\mathrm{Var}[Y_i|X_i=x_i] = \sigma_i^2$

where $\sigma_i = \gamma+\delta x_i$.

Now as before, the statistical problem is to estimate the regression coefficients and the variances, i.e. to find $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$. However, with this model the computation of these estimators is not as simple as in the previous model. It is the purpose of this paper to offer a method of deriving these estimators. The general theory used here would seem to generalize to models with non-linear regression and standard deviation equations, as well as multiple regression models, but the computational procedures are bound to be more complicated.

Given a set of data $\{(X_1,Y_1), (X_2,Y_2), \ldots, (X_n,Y_n)\}$ that fulfill the requirements of our new model, the density function of Y_i is given by

$$f(y_i;\alpha,\beta,\gamma,\delta)=\frac{\exp\left\{-\frac{1}{2}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^2}\right\}}{2\pi(\gamma+\delta x_i)} \quad (7)$$

The method of estimation will again be that of maximum likelihood. The likelihood function now becomes

$$L(\alpha,\beta,\gamma,\delta)=\frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^2}\right\}}{(2\pi)^{\frac{n}{2}}\prod_{i=1}^{n}(\gamma+\delta x_i)} \quad (8)$$

$$\text{and } \log(L)=-\frac{n}{2}\log 2\pi-\sum_{i=1}^{n}\log(\gamma+\delta x_i) - \frac{1}{2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^2}. \quad (9)$$

Taking the first partial derivative of (9) with respect to each of α, β, γ and δ and setting these equal to 0 we have the following system of four equations in the four unknowns α, β, γ and δ:

$$\frac{\partial\log(L)}{\partial\alpha}=\sum_{i=1}^{n}\frac{[Y_i-(\alpha+\beta x_i)]}{(\gamma+\delta x_i)^2}=0$$

$$\frac{\partial\log(L)}{\partial\beta}=\sum_{i=1}^{n}\frac{x_i\,[y_i-(\alpha+\beta x_i)]}{(\gamma+\delta x_i)^2}=0$$

$$\frac{\partial\log(L)}{\partial\gamma}=-\sum_{i=1}^{n}\frac{1}{\gamma+\delta x_i}+\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^3}=0 \quad (10)$$

$$\frac{\partial\log(L)}{\partial\delta}=-\sum_{i=1}^{n}\frac{x_i}{\gamma+\delta x_i}+\sum_{i=1}^{n}\frac{x_i[y_i-(\alpha+\beta x_i)]^2}{(\gamma+\delta x_i)^3}=0.$$

$\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ are the solutions to these four equations and are our desired estimators of α, β, γ and δ respectively.

Before continuing, let's consider two special cases: one where $\delta = 0$ and one where $\gamma = 0$. If $\delta = 0$ ($\gamma \neq 0$) then the system of equations (10) reduces to equations (2), (3) and (4), the case with a constant variance, and a closed form solution exists. If $\gamma = 0$ ($\delta \neq 0$) then system (10) becomes

$$\frac{\partial\log(L)}{\partial\alpha}=\frac{1}{\delta^2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{x_i^2}=0$$

$$\frac{\partial\log(L)}{\partial\beta}=\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]}{x_i}=0 \quad (11)$$

$$\frac{\partial\log(L)}{\partial\delta}=-\frac{n}{\delta}+\frac{1}{\delta^3}\sum_{i=1}^{n}\frac{[y_i-(\alpha+\beta x_i)]^2}{x_i^2}=0.$$

Upon solving the first two of these for $\hat{\alpha}$ and $\hat{\beta}$ we obtain the closed form solution

$$\hat{\alpha}=\frac{n\sum\frac{y_i}{x_i^2}-\sum\frac{1}{x_i}\sum\frac{y_i}{x_i}}{n\sum\frac{1}{x_i^2}-\left(\sum\frac{1}{x_i}\right)^2} \quad (12)$$

$$\text{and } \hat{\beta}=\frac{\sum\frac{1}{x_i^2}\sum\frac{y_i}{x_i}-\sum\frac{1}{x_i}\sum\frac{y_i}{x_i^2}}{n\sum\frac{1}{x^2}-\left(\sum\frac{1}{x_i}\right)^2}. \quad (13)$$

Then by using these values from (12) and (13) in the third equation of system (11) we get

$$\hat{\delta}=\left[\frac{1}{n}\sum\frac{[y_i-(\hat{\alpha}+\hat{\beta}x_i)]^2}{x_i^2}\right]^{\frac{1}{2}}. \quad (14)$$

However if both $\gamma \neq 0$ and $\delta \neq 0$, then a simple closed form solution of the system of equations (10) does not exist. An approach to this problem can be taken through the Newton-Raphson iteration method of approximating the solution set for α, β, γ and δ. This method is based on the Taylor series expansion. Consider the system of equations (10) as the following system

$$\begin{aligned} f_1(\alpha, \beta, \gamma, \delta) &= 0\\ f_2(\alpha, \beta, \gamma, \delta) &= 0\\ f_3(\alpha, \beta, \gamma, \delta) &= 0\\ f_4(\alpha, \beta, \gamma, \delta) &= 0 \end{aligned} \quad (15)$$

Expanding these four functions in a Taylor series we get for $i = 1, 2, 3, 4$

$$\begin{aligned}0=f_i(\alpha,\beta,\gamma,\delta)=&f_i(\alpha_k,\beta_k,\gamma_k,\delta_k) + \\ &(\alpha-\alpha_k)f_{i\alpha}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \\ &(\beta-\beta_k)f_{i\beta}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \\ &(\gamma-\gamma_k)f_{i\gamma}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \\ &(\delta-\delta_k)f_{i\delta}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \dots\end{aligned} \quad (16)$$

where $f_{i\alpha}$, $f_{i\beta}$, $f_{i\gamma}$ and $f_{i\delta}$ are respective first partial derivatives of f_i. By replacing α, β, γ, δ with α_{k+1}, β_{k+1}, γ_{k+1}, δ_{k+1}, neglecting the non-linear terms in $(\alpha_{k+1}-\alpha_k)$, $(\beta_{k+1}-\beta_k)$, $(\gamma_{k+1}-\gamma_k)$, $(\delta_{k+1}-\delta_k)$ of the expansion and dropping the middle term of the double equality (16), these equations for $i = 1, 2, 3, 4$ become

$$\begin{aligned}-f_i(\alpha_k,\beta_k,\gamma_k,\delta_k)=&\Delta\alpha_k f_{i\alpha}(\alpha_k,\beta_k,\gamma_k,\delta_k)+ \\ &\Delta\beta_k f_{i\beta}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \\ &\Delta\gamma_k f_{i\gamma}(\alpha_k,\beta_k,\gamma_k,\delta_k) + \\ &\Delta\delta_k f_{i\delta}(\alpha_k,\beta_k,\gamma_k,\delta_k)\end{aligned} \quad (17)$$

where $\Delta\alpha_k=\alpha_{k+1}-\alpha_k$, $\Delta\beta_k=\beta_{k+1}-\beta_k$, $\Delta\gamma_k=\gamma_{k+1}-\gamma_k$, $\Delta\delta_k=\delta_{k+1}-\delta_k$.

We solve these four linear equations for $\Delta\alpha_k$, $\Delta\beta_k$, $\Delta\gamma_k$ and $\Delta\delta_k$; and knowing the values of α_k, β_k, γ_k and δ_k, we now have values for α_{k+1}, β_{k+1}, γ_{k+1} and δ_{k+1} which are better estimates of actual solutions α, β, γ and δ, than α_k, β_k, γ_k and δ_k are. Then increase k by one and continue this process until the maximum of $\Delta\alpha_k$, $\Delta\beta_k$, $\Delta\gamma_k$ and $\Delta\delta_k$ tends toward zero. We stop the iteration process when $\max[\Delta\alpha_k, \Delta\beta_k, \Delta\gamma_k, \Delta\delta_k] < \varepsilon$ where ε is a predetermined small number, and for this value of k we use $\alpha^*=\Delta\alpha_k+\alpha_k$, $\beta^*=\Delta\beta_k+\beta_k$, $\gamma^*=\Delta\gamma_k+\gamma_k$ and $\delta^*=\Delta\delta_k+\delta_k$ to serve as values for $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$. In order to complete this process, one needs to have initial estimates α_0, β_0, γ_0 and δ_0. Each step in this process, that is, the system (17), is solved easiest by setting up the matrix system

$$\begin{bmatrix} f_{1\alpha} & f_{1\beta} & f_{1\gamma} & f_{1\delta} \\ f_{2\alpha} & f_{2\beta} & f_{2\gamma} & f_{2\delta} \\ f_{3\alpha} & f_{3\beta} & f_{3\gamma} & f_{3\delta} \\ f_{4\alpha} & f_{4\beta} & f_{4\gamma} & f_{4\delta} \end{bmatrix} \cdot \begin{bmatrix} \Delta\alpha_k \\ \Delta\beta_k \\ \Delta\gamma_k \\ \Delta\delta_k \end{bmatrix} = \begin{bmatrix} -f_1 \\ -f_2 \\ -f_3 \\ -f_4 \end{bmatrix} \quad (18)$$

It seems easier to estimate α_0 and β_0 separately from γ_0 and δ_0 instead of estimating all four in one step. When σ does not depend on x, the method of maximum likelihood for obtaining the estimates of α and β does not involve the variance, σ^2. Our new model assumes a linear change in the standard deviation of normal populations, but since the normal distribution is symetric, it seems reasonable to use the same method even in our case where $\sigma = \gamma + \delta x$.

To find the initial estimates, α_0 and β_0, we solve the normal equations (2) and (3) as before. Since we originally assumed that there are at least two values of x_i that are different, we see that the normal equations have a unique solution. Hence the values we use for α_0 and β_0 are the right sides of equations (5) and (6) respectively.

Because of our assumption, $\sigma = \gamma + \delta x$, the original method of estimating σ_0 will not work. However, the fact that σ is a linear function of x is very helpful. When x is small, we naturally expect σ to take on different values than when x is large. In the case when we have several values of y for each value of x, one way of finding initial estimates for γ and δ is to estimate the population of the distribution of y's for each x, and then a least-squares line is fitted to the points of the estimated standard deviations for the various x's. The coefficients of this straight line would then serve as the initial estimates γ_0 and δ_0. This method of obtaining initial estimates would not work if there were only one value of y for each value of x since there would be no way of calculating the standard deviation with only one value. (If only a few x's had multiple y's, this method would produce poor and erratic results.) In that case, and in general, we can group some of smaller x values together and assume that the standard deviation of the y distributions, for each of the grouped x values, is the same. We form two more groups collecting middle values of x and larger values of x together. The standard deviations, s, for the y's and the means, $\bar{x}$, are obtained for each of these three groups. The graph of $(\bar{x},s)$, with these three points on it, gives a least-squares line. This line may be regarded as an estimate of the equation $\sigma = \gamma + \delta x$ thus allowing us to use the coefficients of this calculated line as our values for γ_0 and δ_0. By using standard least-squares techniques our estimates are

$$\delta_0 = \frac{\sum_{i=1}^{n}(\bar{x}_i-\bar{\bar{x}})(s_i-\bar{s})}{\sum_{i=1}^{n}(x_i-\bar{\bar{x}})^2} \quad (19)$$

and $\gamma_0 = \bar{s} - \delta_0\bar{\bar{x}}$ (20)

where $\bar{\bar{x}} = \frac{1}{3}(\bar{x}_1+\bar{x}_2+\bar{x}_3)$ and $\bar{s} = \frac{1}{3}(s_1+s_2+s_3)$, the means and standard deviations of the three groups.

The importance of good estimates α_0, β_0, γ_0 and δ_0 is that if these initial values are close to the values of $\alpha*$, $\beta*$, $\gamma*$ and $\delta*$, then the Newton-Raphson iteration produces solutions that converge to $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ quickly.

There is one major limitation to the Newton-Raphson iteration method of solving the system of equations (17) or equivalently (18). When the Jacobian determinant

$$J=\begin{vmatrix} f_{1\alpha} & f_{1\beta} & f_{1\gamma} & f_{1\delta} \\ f_{2\alpha} & f_{2\beta} & f_{2\gamma} & f_{2\delta} \\ f_{3\alpha} & f_{3\beta} & f_{3\gamma} & f_{3\delta} \\ f_{4\alpha} & f_{4\beta} & f_{4\gamma} & f_{4\delta} \end{vmatrix}$$

vanishes at or near any of the points $(\alpha_k, \beta_k, \gamma_k, \delta_k)$ in the process, slow convergence, or especially, divergence of the iteration may be expected. This can easily be seen in the case of one variable. In this case equations (16) become

$$0=f(\alpha)=f(\alpha_k)+(\alpha-\alpha_k)f'(\alpha_k)+\frac{(\alpha-\alpha_k)^2 f''(\alpha_k)}{2}+\cdots \qquad 22)$$

and thus equations (17) reduce to

$$-f(\alpha_k)=(\alpha_{k+1}-\alpha_k)f'(\alpha_k). \qquad (23)$$

Upon solving for α_{k+1} we obtain

$$\alpha_{k+1}=\alpha_k+\frac{f(\alpha_k)}{f'(\alpha_k)} \qquad (24)$$

where now the Jacobian is $J=\left|f'(\alpha_k)\right|$. If at any time this determinant, $\left|f'(\alpha_k)\right|$, vanishes, we see that the resulting solution for α_{k+1} in equation (24) does not make sense. When we have four variables instead of one, the situation is more involved, but the idea is essentially the same. If the value of J vanishes at or (in the case of many variables) near the point $(\alpha_k, \beta_k, \gamma_k, \delta_k)$, then the resulting solution for $(\alpha_{k+1}, \beta_{k+1}, \gamma_{k+1}, \delta_{k+1})$ does not make sense.

If this situation occurs, as an alternative to using the unobtainable $\alpha*$, $\beta*$, $\gamma*$ and $\delta*$ for $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$, it appears that α_0, β_0, γ_0 and δ_0 may serve as acceptable substitutes.

In conclusion, consider two examples. For both of these examples the values of the parameters α, β, γ and δ were set, and eight different values of X were chosen. For each value of X, five values of Y were derived using a table of normal, zero-one, random deviates - z - and letting $y=(\gamma+\delta x)(z)+(\alpha+\beta x)$.

Example 1:

TABLE I

x.	y				
1.0	-3.64	11.96	11.00	-1.48	11.60
2.0	15.76	44.88	4.28	-17.28	21.22
3.0	9.68	-3.28	45.36	26.32	-14.00
4.0	24.04	16.10	2.60	-1.72	13.22
5.0	24.20	23.80	33.20	8.40	27.40
6.0	16.84	-0.54	30.48	4.52	13.98
7.0	73.52	41.81	29.36	31.04	67.28
8.0	9.50	51.36	24.06	3.24	14.70

The preset values of the parameters were $\alpha = 5.0$, $\beta = 3.0$, $\gamma = 10.0$ and $\delta = 2.0$. With $\varepsilon = 0.0001$ the values computed by an IBM 7094 computer are $\hat{\alpha} = \alpha* = 3.16$, $\hat{\beta} = \beta* = 3.42$, $\hat{\gamma} = \gamma* = 12.46$ and $\hat{\delta} = \delta* = 1.14$.

Example 2:

TABLE II

x.	y				
1.0	-1.58	11.93	-6.13	6.33	-5.57
2.0	19.07	18.89	0.02	6.38	7.01
5.0	27.05	39.05	51.35	33.05	15.80
6.0	9.63	15.07	25.61	16.09	16.77
9.0	39.12	-27.35	33.60	36.36	54.30
10.0	62.75	-1.75	30.75	-35.75	35.25
12.0	50.47	108.96	88.46	53.08	39.45
13.0	3.73	56.81	4.66	68.28	57.12

The preset values of the parameters were $\alpha = 2.0$, $\beta = 3.0$, $\gamma = 5.0$ and $\delta = 2.0$. With $\varepsilon = 0.0001$ the computed values are $\hat{\alpha} = \alpha* = 0.488$, $\hat{\beta} = \beta* = 3.83$, $\hat{\gamma} = \gamma* = 4.26$ and $\delta = \delta* = 2.52$.

When the values of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$ are substituted for α, β, γ and δ in the left sides of equations (15), the maximum value of $f_i(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta})$ for i = 1, 2, 3, 4 is 0.00000021 in example 1 and 0.00000014 in example 2.

And finally, in support of the adequacy of α_0, β_0, γ_0 and δ_0 in case the Newton-Raphson iteration method diverges, the values of these initial estimates in example 1 are $\alpha_0 = 3.0$, $\beta_0 = 3.46$, $\gamma_0 = 11.22$ and $\delta_0 = 1.13$.