MAC IIci: Miller Asymptotics

Chapter 5: Regression

Section ?: Asymptotic Relationship Between a CC and its Associated Slope Estimates in Multiple Linear Regression

The asymptotic null distribution of a CC can be used to determine the asymptotic distribution of linear combinations of the corresponding slope estimates in a multiple linear regression. The normal distribution will be used because, in general, only asymptotic distributions for CC have been developed for the normal distribution; however, the process is general and could be used whenever other asymptotic distributions have been derived for other distributions. For example, for nonparametric CC the limiting distributions hold over a class of distributions. It is known, for example, that the GDCC has the same population value and limiting distribution for the bivariate Cauchy as for the bivariate normal.

The method is first developed for Pearson's r and then extended to NPCC where GDCC is used as an example. The full rank multivariate normal model is used with covariance matrix partitioned as follows into the response and regressor variables.

$\sum_{p+1xp+1} = \begin{pmatrix} s_1^2 & s_{12}' \\ s_{12} & \Sigma_{22} \end{pmatrix}$ where s_1^2 is the variance of the response variate and Σ_{22} is the covariance matrix of the regressor variates. Let Y be the response variate and X the column vector of regressor variates. Let H represent the distribution function of this multivariate normal distribution and $b(H)$ the parameters in the regression of Y on X. Then it is known that $b(H)$ = $\Sigma_{22}^{-1} s_{12}$. Let m = E(Y) and m_x = E(X), a p dimensional vector. The regression model is $E(Y|X = x) = m + (x - m_X)'b(H)$. Let the parameter r($X_i$,Y) be the i^{th} element of $s_{12}, i = 1,2,\cdots,p$. For a sample of size n, let the columns of data be $(y, x_1, x_2,\cdots,x_p)$ and $r(x_i, y)$ the sample CC of x_i and y. The notation is now set for the development using a Taylor Series.

Let l be be a p dimensional column vector of constants and consider the correlation parameter as a function of b; $f(b) = r(X'l, Y - X'b)$. It is a continuous differentiable function of b. In order to relate the null distribution of CC r to linear combinations of the estimated slopes, $f(b)$ will be expanded into a truncated multivariate Taylor Series about $b(H)$. Then the resulting equation will be approximated by data at $\hat{b}$, the estimate slopes. Finally, using the asymptotic null distribution of r, the asymptotic distribution of any linear combination of $\hat{b}$ will be found.

For convenience and without loss of generality let $m = 0$ and $m_x = 0$. We start by determining $f(b)$ in an explicit form and then taking its partial derivatives with respect to b.

$$r(X'l, Y - X'b) = \frac{E(X'l(Y - X'b))}{\sqrt{V(X'l)V(Y - X'b)}}.$$

$$EX'l(Y - X'b) = l'E(XY) - l'E(XX')b = l's_{12} - l'\Sigma_{22}b.$$

$V(X'l) = l'\Sigma_{22}l = a(l), say.$
$V(Y - X'b) = s_1^2 + b'\Sigma_{22}b - 2b's_{12} = b(b), say$
Now $b(H) = \Sigma_{22}^{-1}s_{12}$ so that $r(X'l, Y - X'b(H)) = l's_{12} - l'\Sigma_{22}\Sigma_{22}^{-1}s_{12} = 0$, and $b(b(H)) = s_1^2 - s_{12}'\Sigma_{22}^{-1}s_{12} = s_{res}^2$ where res stands for residuals. We now expand $r(X'l, Y - X'b) = \dfrac{l's_{12} - l'\Sigma_{22}b}{\sqrt{a(l)b(b)}}$ into a Taylor Series.

$$\frac{\partial f(b)}{\partial b} = \frac{\sqrt{a(l)b(b)}(-l'\Sigma_{22}) - (l's_{12} - l'\Sigma_{22}b)\dfrac{\partial\sqrt{a(l)b(b)}}{\partial b}}{a(l)b(b)}.$$

$$\left.\frac{\partial f(b)}{\partial b}\right|_{b=b(H)} = \frac{-l'\Sigma_{22}}{\sqrt{a(l)b(b)}} = \frac{-l'\Sigma_{22}}{\sqrt{l'\Sigma_{22}l}\,s_{res}}.$$

We now are ready to expand $f(b) = r(X'l, Y - X'b)$ into a Taylor Series around $b(H)$ with just the first partials being used.

$$f(b) = r(X'l, Y - X'b) = r(X'l, Y - X'b(H)) - \frac{l'\Sigma_{22}}{\sqrt{l'\Sigma_{22}l}\,s_{res}}(b - b(H)).$$

We now approximate the terms in this series by replacing X by $(x_1, x_2, \cdots, x_p)$ = x, Y by y and evaluate b at $\hat{b}$ where $r(x_i, y - x\hat{b}) = 0, i = 1, 2, \cdots, p$. Because of the linearity of the covariance function, the p equations imply that $r(xl, y - x\hat{b}) = 0$. Thus, the Taylor Series becomes $f(\hat{b}) = r(xl, y - x\hat{b}) = 0 = r(xl, y - xb(H)) - \dfrac{l'\Sigma_{22}}{\sqrt{l'\Sigma_{22}l}\,s_{res}}(\hat{b} - b(H))$. Now the first term on the right hand side has a null distribution since outcomes xl and $y - xb(H)$ come independent random variables $X'l$ and $Y - X'b$. It is also known (reference here) that $\sqrt{n}r(xl, y - xb(H))$ has an asymptotic N(0,1) distribution function. Consequently, $\dfrac{\sqrt{n}l'\Sigma_{22}}{\sqrt{l'\Sigma_{22}l}\,s_{res}}(\hat{b} - b(H))$ has the same asymptotic distribution.

In order to relate this result to standard methods, transform from vector l to vector k where $l = \Sigma_{22}^{-1}k$; thus, $\Sigma_{22}l = k$ and $l'\Sigma_{22} = k'$. Then the quadratic form equality is $l'\Sigma_{22}l = k'\Sigma_{22}^{-1}k$. It follows that $\dfrac{\sqrt{n}k'(\hat{b} - b(H))}{\sqrt{k'\Sigma_{22}^{-1}k}\,s_{res}}$ has an asymptotic N(0,1) distribution. In terms of $\hat{b} - b(H)$, $k'(\hat{b} - b(H)$ is approximately $N(0, \dfrac{(k'\Sigma_{22}^{-1}k)s_{res}^2}{n})$.

This result is now related to the classical least squares or normal theory fixed x multiple linear regression model.

$y = xb + e$ where $e \sim N(0, s^2 I)$ independent.

Let $x^* = (x_1^*, x_2^*, \cdots, x_p^*)$ where the * indicates that the data have been centered at the means. Then the sum of squares matrix is $x^{*\prime}x^*$ with $\hat{b} = (x^{*\prime}x^*)^{-1}x^{*\prime}y$ and

$V(\hat{b}) = s^2(x^{*\prime}x^*)^{-1}$. The distribution of $\hat{b} - b$ is multivariate normal: $MN(0, V(\hat{b}))$. Thus, the distribution of $k'(\hat{b} - b)$ is $N(0, s^2 k'(x^{*\prime}x^*)^{-1}k)$.

We now connect the two notations between CC and classical methods. $(k'\frac{\Sigma_{22}^{-1}}{n}k)s_{res}^2 = k'(x^{*\prime}x^*)^{-1}ks^2$. In this development n was used rather than the usual n-p which would give an unbiased estimate of variance.

(exercise - look at second derivative in the Taylor series)

ASYMPTOTICS FOR r_{gd}

It is known (Gideon, Hollister) that for joint normal random variables W_1, W_2 the population value of $r_{gd}(W_1,W_2)$ is $\frac{2}{p}\sin^{-1} r_{W_1,W_2}$ where r_{W_1,W_2} is the bivariate normal correlation parameter between W_1 and W_2. Thus, for $X'l$ and $Y - X'b$

$f_1(b) = r_{gd}(X'l, Y - X'b) = \frac{2}{p}\sin^{-1} r_{X'l,Y-X'b}$. For normal random variables $r(X'l, Y - X'b) = r_{X'l,Y-X'b}$ and so the results of the previous section can be used. The Taylor series for $f_1(b)$ is

$$f_1(b) = r_{gd}(X'l, Y - X'b) = r_{gd}(X'l, Y - X'b(H)) + \frac{\P}{\P b} r_{gd}(X'l, Y - X'b)\Big|_{b=b(H)}(b - b(H))$$

.

For the partial derivatives $\frac{\P}{\P b} r_{gd}(X'l, Y - X'b) = \frac{2}{p}\frac{1}{\sqrt{1 - r_{X'l,Y-X'b}^2}}\frac{\P}{\P b} r_{X'l,Y-X'l}\Big|_{b=b(H)}$.

At $b = b(H)$, $X'l$ and $Y - X'b$ are independent random variables, so $r_{X'l,Y-X'b(H)} = 0$, and the last term is, as before, $\frac{-l'\Sigma_{22}}{\sqrt{l'\Sigma_{22}l}\,s_{res}}$. The Taylor series becomes

$$f_1(b) = r_{gd}(X'l, Y - X'b) = r_{gd}(X'l, Y - X'b(H)) - \frac{2}{p}\frac{l'\Sigma_{22}}{\sqrt{l'\Sigma_{22}l}\,s_{res}}(b - b(H)).$$

Now solve the associated "normal equations" for data x and y.

$$r_{gd}(x_i, y - x\hat{b}_{gd}) = 0, i = 1, 2, \cdots, p.$$

$\hat{b}_{gd}$ is a solution vector with i^{th} individual component $\hat{b}_{i,gd}$. The CC r_{gd} does not have the same linearity properties that r has and so it is not necessarily true that $r_{gd}(xl, y - x\hat{b}_{gd}) = 0$; however, computer simulations have shown that $r_{gd}(x\hat{b}_{gd}, y - x\hat{b}_{gd})$ is zero or close to zero. We shall again approximate the Taylor series above by their sample counterparts at $\hat{b}_{gd}$ and assume that the left hand side is zero. The simulations in the example below indicate that the asymptotic distribution theory is still good. Again $r_{gd}(X'l, Y - X'b(H))$ has the null distribution and its sample equivalent multiplied by $\sqrt{n}$ is $\sqrt{n}r_{gd}(xl, y - x\hat{b}_{gd})$ and this has an approximate N(0,1) distribution (Gideon, Pyke, Prentice). It now follows from the Taylor series that $\frac{2}{p}\sqrt{n}\frac{l'\Sigma_{22}}{\sqrt{l'\Sigma_{22}l}\, s_{res}}(\hat{b}_{gd} - b(H))$ has an approximate N(0,1) distribution. Solving this for the centered slopes we have $l'\Sigma_{22}(\hat{b}_{gd} - b(H)) \sim N(0, \frac{p^2 l'\Sigma_{22} l s^2_{res}}{4n})$.

If again we let $l'\Sigma_{22} = k'$,

$k'(\hat{b}_{gd} - b(H)) \sim N(0, \frac{p^2 k'\Sigma_{22}^{-1} k s^2_{res}}{4n})$.

As a special case we let k be a vector of 0's except for a 1 in the i^{th} position, the above result gives the asymptotic distribution of $\hat{b}_{i,gd} - b_i(H)$ as $N(0, \frac{p^2 s^{ii} s^2_{res}}{4n})$ where $s^{ii} = (i,i)$ element of Σ_{22}^{-1}.

An Example

In order to ascertain if the asymptotic results are accurate, some simulations were run for samples of size n = 10. They showed that the distribution of the chosen linear combinations of the estimated b's was as good for r_{gd} as for r; ie, very similar distributions. These results are now outlined.

Let Z be $N(0,I_3)$, $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, and Y = AZ. Let Y be written as $Y = \begin{pmatrix} y \\ x_1 \\ x_2 \end{pmatrix}$.

Then $Y \sim N(0, \Sigma_y)$ where $\Sigma_y = AI_3A' = \begin{pmatrix} (3) & (2 \;\; 2) \\ \begin{pmatrix} 2 \\ 2 \end{pmatrix} & \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} s_1^2 & s_{12} \\ s'_{12} & \Sigma_{22} \end{pmatrix}$. Then

$\Sigma_{22}^{-1} = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, $b(H) = \Sigma_{22}^{-1}s'_{12} = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix}$ and $s^2_{res} = s_1^2 - s_{12}\Sigma_{22}^{-1}s'_{12} = 1/3$, so that

$Y|X = x \sim N(\frac{2}{3}x_1 + \frac{2}{3}x_2, \frac{1}{3})$.

The general results from the previous work give the following approximate distribution for $k = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$; $k'\hat{b} = (\hat{b}_1 + \hat{b}_2), k'\Sigma_{22}^{-1}k = 2/3$. So the asymptotic distributions for the estimated slopes from the two regressions are first for r and then for r_{gd}, $(\hat{b}_1 + \hat{b}_2 - 4/3) \sim N(0, 2/9n)$

$(\hat{b}_{1,gd} + \hat{b}_{2,gd} - 4/3) \sim N(0, \frac{p^2}{18n})$.

The population parameters for the CC are

$r_{y|x_1,x_2} = \frac{\sqrt{s'_{12} b(H)}}{s_1} = \frac{2\sqrt{2}}{3} = .9428$ and

$r_{gd,y|x_1,x_2} = \frac{2}{p}\sin^{-1} r_{y|x_1 x_2} = \frac{2}{p}\sin^{-1}(.9428) = .7836$.

One hundred simulations were run with $W_1 = \hat{b}_1 + \hat{b}_2$ and $W_2 = \hat{b}_{1,gd} + \hat{b}_{2,gd}$ recorded each time. A second plot of W_1 -vs- W_2 and a second qqplot of W_1 -vs- W_2 reveal linearity; thus, the distributions are nearly the same except for the scale factor. More simulations need to be run on other cases.

Section: An example of simple and multiple regression with the 1992 Atlanta Braves team record of 175 games.

Three regression are run with the response variable , y, being the length of a game in hours. The first regression , I, will use the first three of the following four regressor variables.

x_1, the total number of runs by both teams in a game,

x_2, the total number of hits by both teams in a game,

x_3, the total number of runners by both teams left on base in a game

x_4, the total number of pitchers used in a game by both teams.

Thus, the interest is in determing how various conditions in a game affect the length of the game. The second regression, II, will be a simple linear regression of time, y, on x_4. The third regression, III, will use all four of the regressor variables. The main purpose is to use the asymptotic distributions of the slopes to compare least squares (LS) or Pearson regression to the NPCC Greatest Deviation(GD). The residual standard deviations are compared and the GD one is less than that of LS! Also the multiple CC's are computed and one partial CC is computed. QQ plots on the residuals are discussed.

Although time is a continuous random variable all the regresssor variables are discrete; so at best for the classical analysis only an approximate multivariate normal distribution would model the data. The advantage is using GD would be the usual NP advantage, the distributions of the statistics hold under a wider class of data distributions. For example, it is known that GD has the same population value for a bivariate Cauchy distribution as for a bivariate Normal. All classical inference is based on the normal distribution or central limit theorems which give asymptotic results.

For any CC and in particular for the NPCC r_{gd} the regression equations for example I are $r_{gd}(x_i, y - b_1 x_1 - b_2 x_2 - b_3 x_3) = 0, i = 1,2,3$. Thus, the regressor variables are uncorrelated with the regression residuals. The intercept of the regression is obtained by taking the median of these residuals

$$b_0 = \text{median}(y - b_1 x_1 - b_2 x_2 - b_3 x_3).$$

The residual SD is obtained by the methods of Chapter 3; ie, a simple linear regression of the sorted residuals on the ordered N(0,1) quantiles. Let quan and res represent these ordered vectors. Then the estimated SD is s taken from the solution to

$$r_{gd}(quan, res - s * quan) = 0 \qquad \text{(a)}$$

For regression example I, Splus and some C routines have been developed that do multiple regression with GD, the lm command, linear models, was used for LS.

GD: $\hat{y} = 1.8374 + 0.04908x_1 - 0.01022x_2 + 0.05479x_3$

$s = \hat{s} = 0.2518$ or (.2518)60 = 15.1 minutes

LS: $\hat{y} = 1.7179 + 0.04459x_1 - 0.01079x_2 + 0.06910x_3$

$s = \hat{s} = 0.2919$ on 171 degrees of freedom (.2919)60 = 17.5 minutes

Note that $\hat{s}_{gd} < \hat{s}_{LS}$.

Before proceeding with more of the regression analysis , the r_{gd} and classical LS normal quantile plots on the regression residuals are compared. The quantile plot related to equation (a) above shows the "regression" line going through the center of about $\frac{175-16}{175} = 90.9\%$ of the data while the normal quantile plot connected with LS goes directly through $\frac{175-39}{175} = 77.7\,\%$ of the data. That is, about 16 games have GD residuals markedly away for the straight line plotted through the quantiles whereas for LS, the number of departures is 39. As is claimed for GD regression it fits the bulk of the data very well. It does this, by making outliers more extreme than LS does, which as is well-known, is more sensitive to outliers. The three most extreme outliers are games 28, 86, and 147 which are extra innings games of length 16, 10 and 12 respectively. There are total of 18 extra innings games and these could all be consider non-standard data or outliers even thought all four regressor variables relate well to the time of these games. Deleting the most serious "outliers" would make LS more like the GD regression.

The asymptotics of the work earlier in this chapter is now illustrated for this example. The results are given and then all the calculations behind them are listed.

Table

standard errors of the regression coefficients, z scores, and P-values

slopes	GD SE	LS SE	GD z score	GD p-value	LS z score	LS p-value
b_1	0.01318	0.0094	3.72	.0002	4.77	.0000
b_2	0.01324	0.0101	-0.77	.44	-1.07	.2867
b_3	0.00998	0.0077	5.49	.0000	9.02	.0000

The calculation of the estimated standard errors of the slopes is now given. From the text the asymptotic distributions are

$$\text{LS:} \quad N(b_i, \frac{s^{ii} s_{res}}{n}) \qquad \text{and GD:} \quad N(b_i, \frac{s^{ii} s_{res} p^2}{4n})$$

where n=175 and for GD s^2_{res} is the square of the slope of the GD regression line of the sorted residuals, $y - \hat{y}$, on N(0,1) quantiles. For LS the lm, linear models, from Splus was used, although the s^2_{res} coming from a QQ plot with Pearson's CC was close to the LS result. Let Σ_{22} be the 3x3 covariance matrix of the regressor variables, and s^{ii} denote the ith diagonal element.

For the GD case Σ_{22} was obtained by using the GD estimates of the SD's ($\hat{s}_i, i = 1,2,3$ obtained similar to equation (a)) and the GD correlation matrix where each GD correlation was transformed to an estimate of a bivariate normal (or bivariate Cauchy) correlation by $\hat{r} = \sin(\frac{p r_{gd}}{2})$, label this 3x3 matrix Σ_{gd} . With this in mind,

$$\Sigma_{22} = \begin{pmatrix} \hat{s}_1 & 0 & 0 \\ 0 & \hat{s}_2 & 0 \\ 0 & 0 & \hat{s}_3 \end{pmatrix} \Sigma_{gd} \begin{pmatrix} \hat{s}_1 & 0 & 0 \\ 0 & \hat{s}_2 & 0 \\ 0 & 0 & \hat{s}_3 \end{pmatrix}.$$

The SD's and correlations needed for all of these calculations are now given.

Basic statistics for our data are given in the following tables.

Table

Pearson's CC

	y	x_1	x_2	x_3	x_4
y	1	0.4835	0.6053	0.6745	0.7201
x_1		1	0.7686	0.2308	0.6025
x_2			1	0.6117	0.6279
x_3				1	0.4764
x_4					1

Table

GDCC, upper triangular is r_{gd}, lower half is $(\sin p r_{gd}/2)$

	y	x_1	x_2	x_3	x_4
y	1	0.3736	0.4138	0.4023	0.4885
x_1	(0.5537)	1	0.5690	0.1839	0.4023
x_2	(0.6052)	(0.7794)	1	0.4080	0.3678
x_3	(0.5907)	(0.2849)	(0.5979)	1	0.2529
x_4	(0.6942)	(0.5907)	(0.5461)	(0.3869)	1

Table

Estimates of the standard deviations of the regression variables

LS is the classical least squares, GD is using a GD fitting on the quantile plot

	y	x_1	x_2	x_3	x_4
LS	0.4420	4.1994	4.7855	4.1457	2.1885
GD	0.4003	3.8825	4.16199	4.0078	2.1468

From these tables using just the x_1,x_2,x_3 part the covariance matrix can be formed, Σ_{22}, and inverted to obtain for the LS or Pearson and GD the following

For Pearson the calculation gives $\Sigma_{22}^{-1} = \begin{pmatrix} 0.1785 & -0.1570 & 0.0691 \\ -0.1570 & 0.2079 & -0.1101 \\ 0.0691 & -0.1101 & 0.1198 \end{pmatrix}$.

For GD the calculation gives $\Sigma_{22}^{-1} = \begin{pmatrix} 0.1943 & -0.1548 & 0.0530 \\ -0.1548 & 0.1962 & -0.0925 \\ 0.0530 & -0.0925 & 0.1114 \end{pmatrix}$

Recall that s^{ii} is the notation for the (i,i) element in Σ_{22}^{-1}. From the asymptotic results connecting slope estimates and correlation, for Pearson's r

$\hat{s}_{\hat{b}_i} = \sqrt{\dfrac{s^{ii} s_{res}^2}{n}} = \dfrac{0.2919}{\sqrt{175}} \sqrt{s^{ii}} = (0.02207)\sqrt{s^{ii}}$, and this equals for

(i=1), 0.009324; (i=2), 0.010063; (i=3), 0.0077. Note the close agreement with this and the standard errors direct from the Splus linear models (lm) command.

For GD $\hat{s}_{\hat{b}_i} = \sqrt{\dfrac{p^2 s^{ii} s_{res}^2}{4n}} = \sqrt{\dfrac{p^2(0.2518)^2}{4(175)}} \sqrt{s^{ii}} = (0.02990)\sqrt{s^{ii}}$, and this equals for

$(i=1), (0.02990)\sqrt{.1943} = 0.01318$

$(i=2), (0.02990)\sqrt{.1962} = 0.01324$

$(i=3), (0.02990)\sqrt{.1114} = 0.00998$

Note that the inference on the significance of the slopes using P-values is essentially the same whether LS or GD regression is used. The estimated values $\hat{b}_i$ are somewhat different and the smallest standard error of the regression is GD not LS (15.1 minutes versus 17.5 minutes). The GD inference does very well for the bulk of the data but does not fare as well on the "outliers" as does LS. However, in inference like this, one probably wants knowledge for the standard games and not to be swayed by a few unusual games. The larger z-scores for GD are the price paid for NP inference, the $p^2/4$ term in the SD, but the inference is valid over a larger class of distributions.

Subsection: example of Partial Correlation

In order to more fully compare LS and GD regression, the partial correlation of Y and X_2 was computed deleting the effects of X_1 and X_2. The variable X_2 was chosed because the Pearson and GD correlations with Y are nearly the same and positive but in the Case I regression the coefficient of X_2 is negative. Recall that $r(y,x_2) = 0.6053$ and $r_{gd}(y,x_2) = 0.4138$ with $\sin(pr_{gd}/2) = 0.6052$. For each CC the regression of Y and X_2 on X_1 and X_3 must be computed in order to obtain residuals and then the correlation of these Y and X_2 residuals give the partial correlations. The regression are for LS: $\hat{y} = 1.60807 + 0.03645x_1 + 0.06339x_3$ and $\hat{x}_2 = 3.4463 + 0.7553x_1 + 0.5296x_3$. Therefore for LS the partial correlation is $r(y - \hat{y}, x_2 - \hat{x}_2) = -0.08146$.

For GD $\hat{y} = 1.7982 + 0.04030x_1 + 0.05083x_3$. Because X_2 is a discrete quantity the multiple regression Splus-C routine for the regression of X_2 on X_1 and X_3 would not converge. Therefore an approximation was obtained by continuizing X_2. Random normal increments with $m = 0$ and $s = 0.01$ were added to the 175 X_2 values. Call these x_{22}. Then the GD regression of x_{22} on x_1 and x_2 did converge. The result is $\hat{x}_{22} = 3.5831 + 0.7632x_1 + 0.5022x_3$. Note that this result is similar to the LS output. To check how close the approximation is, we note that $r_{gd}(x_1, x_{22} - \hat{x}_{22}) = 0$ and that also $r_{gd}(x_1, x_2 - \hat{x}_{22}) = 0$. However, although $r_{gd}(x_3, x_{22} - \hat{x}_{22}) = 0$ we have only that $r_{gd}(x_3, x_2 - \hat{x}_{22}) = -0.0057$. This accuracy is sufficient for this example. The partial correlation is given for both y and x_2 and y with x_{22}. $r(y - \hat{y}, x_2 - \hat{x}_{22}) = -0.04022$ with $\sin(p(-0.04022)/2) = -0.06639$ and $r(y - \hat{y}, x_{22} - \hat{x}_{22}) = -0.05747$ with $\sin(p(-0.05747)/2) = -0.09015$. Thus, it is seen that the Pearson and GD partial CC of Y and X_2 removing the effects of X_1 and X_3 are very similar.

Section: The multiple CC of the regression.

The multiple CC of Y on X_1, X_2, and X_3 is

LS: $r(y,\hat{y}) = \sqrt{0.5713} = 0.7558$,

GD: $r(y,\hat{y}) = 0.5805$ and $\sin(pr_{gd}/2) = 0.7906$.

This result is in agreement with the SD of the residuals in that GD gives a larger multiple CC than does Pearson indicating a slightly closer relationship.

Section: Cases II and III

The regression of Y on just X_4 and then on all four regressor variables is now given so that the LS and GD methods can be compared. First for case II the CC's are $r(y,x_4) = 0.7201$ and $r_{gd}(y,x_4) = 0.4485$ with $\sin(pr_{gd}/2) = 0.6942$. This factor X_4 has a higher correlation with Y for both CC's than for each of the other three regressor variables. Hence, possibly an important predictor variable has been left out of the regression equation. Here are all the Case II and III regression equations:

LS: $\hat{y} = 1.9167 + 0.1454x_4$ with $\hat{s}_{res} = 0.3076$,
GD: $\hat{y} = 1.9000 + 0.1500x_4$ with $\hat{s}_{res} = 0.2769$.

LS: $\hat{y} = 1.5753 + 0.0217x_1 - 0.0127x_2 + 0.0533x_3 + 0.0897x_4$
with $r(y,\hat{y}) = \sqrt{0.6332} = 0.8205$ and $\hat{s}_{res} = 0.2556$ on 170 degrees of freedom,
GD: $\hat{y} = 1.7473 + 0.0457x_1 - 0.0310x_2 + 0.0567x_3 + 0.0718x_4$
with $r_{gd}(y,\hat{y}) = 0.5805$, $\sin(pr_{gd}/2) = 0.7906$, and $\hat{s}_{res} = 0.2280$.

In the four variable regression LS has a slightly higher multiple CC but a higher residual SE, somewhat a contradiction. The two slopes with the biggest difference between the two regressions are for X_1 and X_2. The X_2 coefficient for GD is over twice as large as for LS, and if the value of the X_2 coefficient for GD had been the LS coefficient, it would have been very significant as the P-value would have been much lower than the 0.1514 as given below. The X_3 difference is even greater. From the "lm" command in Splus the P-values for variables 1,2,3,4 are respectively, 0.0142, 0.1514, 0.0000, and 0.0000

The normal quantile plots on the residuals for the four variable regression now reveal fewer unusual games, 5 for LS, and 6 for GD, with all the remaining points lying very close to the straight line through the points. However, the distance from the line to the unusual points is much greater for GD than for LS. This explains the main difference in the regression output. GD obtains a smaller residual SE by not weighting the very few unusual points as much as LS does. Whether or not the difference in the coefficients and the (0.2280)60=13.6 minutes is meaningfull to a data analysis compared to (0.2556)60=15.3 minutes depends on qualitative thought. In the current problem on the length of major league baseball games, with the idea that they are too long, the GD analysis would be more appropriate for this analyst.

For the simple linear regression of Case II the asymptotic inference on the slope is now given. The SE of the slope coefficient is calculated. Because in this case Σ_{22} is a 1x1 matrix, the inverse is $s^{11} = \frac{1}{s_4^2}$ where s_4^2 is the estimate of the variance of X_4 by whatever method is being used. The formulas for LS and GD are

Pearson or LS: $\hat{s}_{\hat{b}} = \sqrt{\frac{s_{res}^2}{ns_4^2}} = \sqrt{\frac{(0.3086)^2}{175(4.7894)}} = 0.0107$

GD: $\hat{s}_{\hat{b}} = \sqrt{\frac{p^2 s_{res}^2}{4ns_4^2}} = \sqrt{\frac{p^2(0.2768)^2}{(4)175(4.6088)}} = 0.0153.$

As in the other examples the SE of the slope is higher for the NPCC than for the classical case. However, the real question is which one is more reliable over a variety of data sets. Direct from the Splus "lm" command $\hat{s}_{\hat{b}} = 0.0107$.