Dependencies in Interval-valued Symbolic Data

Lynne Billard
University of Georgia
lynne@stat.uga.edu

Tribute to Professor Edwin Diday:
Paris, France; 5 September 2007

Naturally occurring Symbolic Data -- **Mushrooms**

w_u	Species	Pileus Cap Width	Stipe Length	Stipe Thickness	Edibility
w_1	*arorae*	[3.0, 8.0]	[4.0, 9.0]	[0.50, 2.50]	U
w_2	*arvenis*	[6.0, 21.0]	[4.0, 14.0]	[1.00, 3.50]	Y
w_3	*benesi*	[4.0, 8.0]	[5.0, 11.0]	[1.00, 2.00]	Y
w_4	*bernardii*	[7.0, 6.0]	[4.0, 7.0]	[3.00, 4.50]	Y
w_5	*bisporus*	[5.0, 12.0]	[2.0, 5.0]	[1.50, 2.50]	Y
w_6	*bitorquis*	[5.0, 15.0]	[4.0, 10.0]	[2.00, 4.00]	Y
w_7	*califorinus*	[4.0, 11.0]	[3.0, 7.0]	[0.40, 1.00]	T
..	....	....	....	....	..

Patient Records – Single Hospital, **Cardiology**

Patient	**Hospital**	**Age**	**Smoker ….**
Patient 1	Fontaines	74	heavy
Patient 2	Fontaines	78	light
Patient 3	Beaune	69	no
Patient 4	Beaune	73	heavy
Patient 5	Beaune	80	light
Patient 6	Fontaines	70	heavy
Patient 7	Fontaines	82	heavy
⋮	⋮	⋮	⋮

Patient Records by Hospital -- aggregate over patients

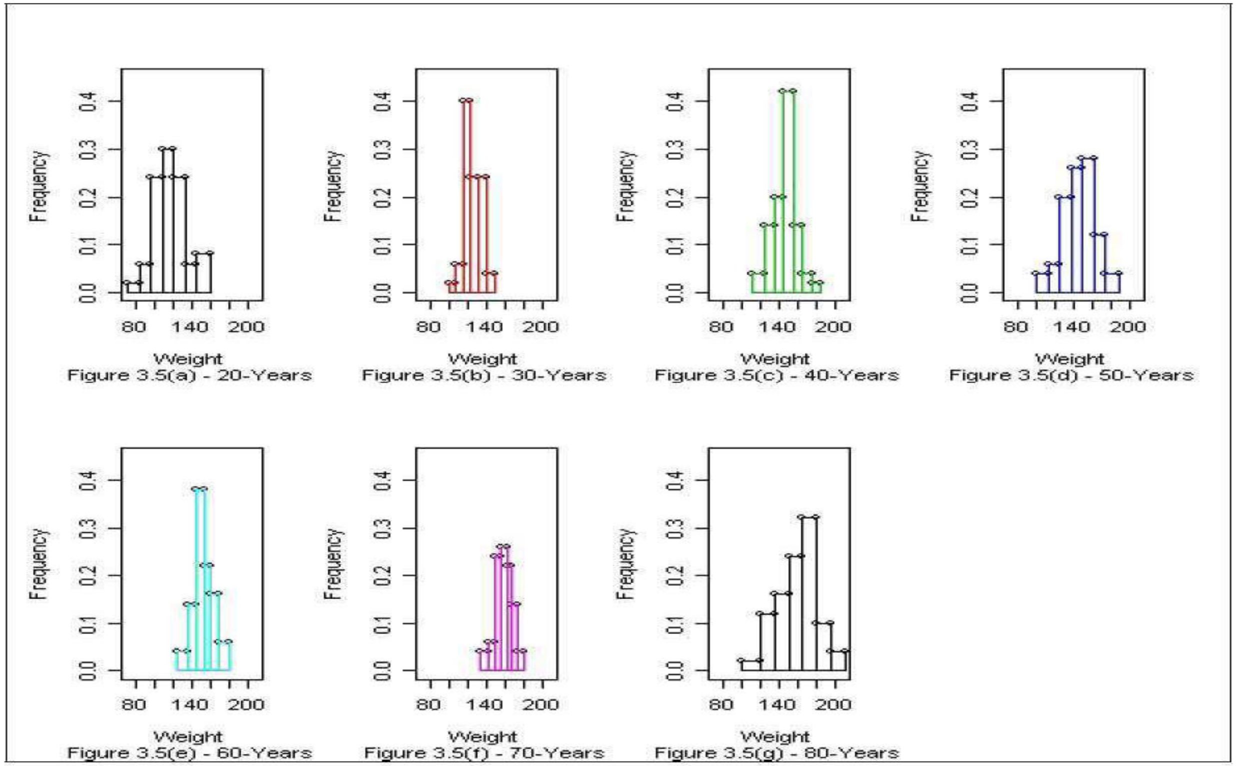
Result: Symbolic Data

Patient	Hospital	Age	Smoker
Patient 1	Fontaines	74	heavy
Patient 2	Fontaines	78	light
Patient 3	Beaune	69	no
Patient 4	Beaune	73	heavy
Patient 5	Beaune	80	light
Patient 6	Fontaines	70	heavy
Patient 7	Fontaines	82	heavy
⋮	⋮	⋮	⋮

Hospital	Age	Smoker
Fontaines	[70, 82]	{light ¼, heavy ¾}
Beaune	[69, 80]	{no, light, heavy}
⋮	⋮	⋮

Histogram-valued Data --

Weight by Age Distribution:



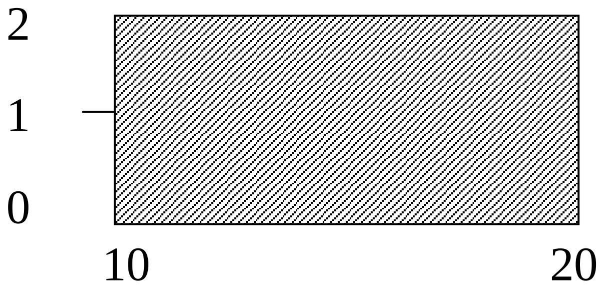
Logical dependency rule

E.g. Y_1 = age Y_2 = # children

Classical: $\boldsymbol{Y_a} = \mathbf{(10, 0)}$, $\boldsymbol{Y_b} = \mathbf{(20, 2)}$, $\boldsymbol{Y_c} = \mathbf{(18, 1)}$

Aggregation →

Symbolic: $\boldsymbol{\xi} = \mathbf{(10\ ,\ 20)} \times \mathbf{(0, 1, 2)}$



I.e., ξ implies classical Y_d = (10, 2) is possible

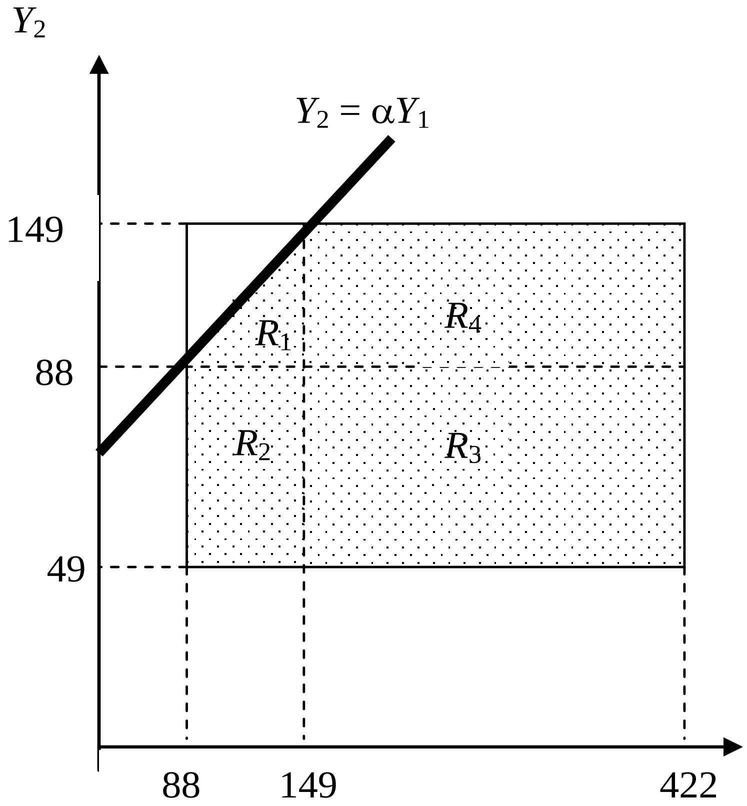
Need **rule** ν: {If $Y_1 < 15$, then $Y_2 = 0$}

Interval-valued data

u Team	Y_1 # At-Bats	Y_2 # Hits	u Team	Y_1 # At-Bats	Y_2 # Hits
1	(289, 538)	(75, 162)	11	(212, 492)	(57, 151)
2	**(88, 422)**	**(49, 149)**	12	(177, 245)	(189, 238)
3	(189, 223)	(201, 254)	13	(342, 614)	(121, 206)
4	(184, 476)	(46, 148)	14	(120, 439)	(35, 102)
5	(283, 447)	(86, 115)	15	(80, 468)	(55, 115)
6	(24, 26)	(133, 141)	16	(75, 110)	(75, 110)
7	(168, 445)	(37, 135)	17	(116, 557)	(95, 163)
8	(123, 148)	(137, 148)	18	(197, 507)	(52, 53)
9	(256, 510)	(78, 124)	19	(167, 203)	(48, 232)
10	(101, 126)	(101, 132)			

$\xi(2)$: **$Y_2 = 149$ not possible when $Y_1 < 149$**

Observation $\xi(2)$



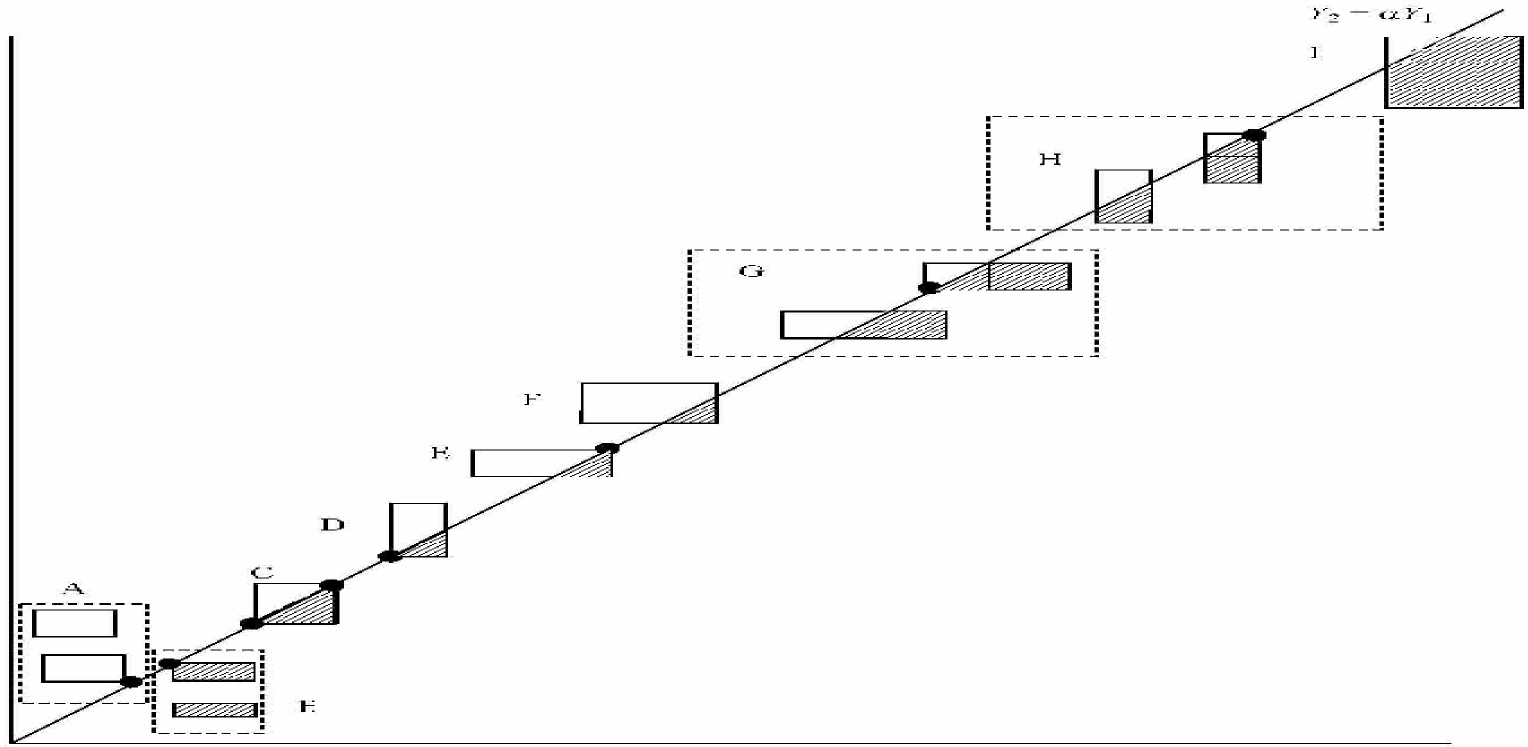


Figure 4.5 – Patterns for Virtual V – shaded regions

Dependencies between Variables – Interval-valued Variables

E.g., Regression Analysis

Dependent variable: $\mathbf{Y} = (Y_1, \cdots, Y_q)$, e.g., $q=1$

Predictor/regression variable: $\mathbf{X} = (X_1, \cdots, X_p)$

Multiple regression model:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + e$$

Error: $e \sim E(e)=0$, $Var(E) = \sigma^2$, $Cov(e_i, e_k)= 0$, $i \neq k$.

Multiple Regression Model: $Y = \beta_o + \beta_1 X_1 + \cdots + \beta_p X_p + e$

In vector terms,

$$\mathbf{Y} = \mathbf{X}\,\beta + \mathbf{e}$$

Observation matrix: $\mathbf{Y}' = (Y_1, \cdots, Y_n)$

Design matrix: $$\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix}$$

Regression coefficient matrix: $\beta' = (\beta_o, \beta_1, \cdots, \beta_p)$

Error matrix: $\mathbf{e}' = (e_1, \cdots, e_n)$

Model: **Y** = **X** β + **e**

Least squares estimator of β is

$$\hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$$

When p=1,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{Cov(X,Y)}{Var(X)},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}\bar{X}$$

where

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Model: $Y = \beta_o + \beta_1 X_1 + \cdots + \beta_p X_p + e$

Or, write as

$$Y - \bar{Y} = \beta_1(X_1 - \bar{X}_1) + \ldots + \beta_p(X_p - \bar{X}_p) + e$$

$$\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}, \quad j = 1, \ldots, p.$$

$$\beta_0 \equiv \bar{Y} - (\beta_1 \bar{X}_1 + \ldots + \beta_p \bar{X}_p)$$

Then,

$$Y - \bar{Y} = \beta_1(X_1 - \bar{X}_1) + \ldots + \beta_p(X_p - \bar{X}_p) + e$$

Least squares estimator of β is

$$\hat{\boldsymbol{\beta}} = [(\boldsymbol{X} - \bar{\boldsymbol{X}})'(\boldsymbol{X} - \bar{\boldsymbol{X}})]^{-1}(\boldsymbol{X} - \bar{\boldsymbol{X}})'(\boldsymbol{Y} - \bar{\boldsymbol{Y}})$$

where

$$(\boldsymbol{X} - \bar{\boldsymbol{X}})'(\boldsymbol{X} - \bar{\boldsymbol{X}}) =$$

$$= \begin{pmatrix} \Sigma(X_1 - \bar{X}_1)^2 & \cdots & \Sigma(X_1 - \bar{X}_1)(X_p - \bar{X}_p) \\ \vdots & & \vdots \\ \Sigma(X_p - \bar{X}_p)(X_1 - \bar{X}_1) & \cdots & \Sigma(X_p - \bar{X}_p)^2 \end{pmatrix}$$

$$= \left(\sum_i (X_{j_1} - \bar{X}_{j_1})(X_{j_2} - \bar{X}_{j_2}) \right), \qquad j_1, j_2 = 1, \cdots, p$$

$$(\boldsymbol{X} - \bar{\boldsymbol{X}})'(\boldsymbol{Y} - \bar{\boldsymbol{Y}}) = \left(\sum_i (X_j - \bar{X}_j)(Y - \bar{Y}) \right), j = 1, \cdots, p$$

Interval-valued data:

$$Y_{uj} = [a_{uj}, b_{uj}],\ j = 1, \ldots, p,\ u \in E = \{w_1, \ldots, w_u, \ldots w_m\}$$

Bertrand and Goupil (2000):

Symbolic sample mean is

$$\bar{Y}_j = \frac{1}{2m} \sum_{u \in E} (b_{uj} + a_{uj}),$$

Symbolic sample variance is

$$S_j^2 = \frac{1}{3m} \sum_{u \in E} (b_{uj}^2 + b_{uj} a_{uj} + a_{uj}^2) - \frac{1}{4m^2} [\sum_{u \in E} (b_{uj} + a_{uj})]^2$$

Notice, e.g., m = 1, Y = Weight

$Y_1 = [132, 138] \rightarrow$ $\bar{Y}_1 = 135,\ S_1^2 = 3$

$Y_2 = [129, 141] \rightarrow$ $\bar{Y}_1 = 135,\ S_2^2 = 12$

Can rewrite

$$S_j^2 = \frac{1}{3m} \sum_{u \in E} [(a_{uj} - \bar{Y}_j)^2 + (a_{uj} - \bar{Y}_j)(b_{uj} - \bar{Y}_j) + (b_{uj} - \bar{Y}_j)^2]$$

Then, by analogy, for j = 1,2, for interval-valued variables Y_1 and Y_2,

empirical covariance function $Cov(Y_1, Y_2)$ is

$$Cov(Y_1, Y_2) = \frac{1}{3m} \sum_{u \in E} G_1 G_2 [Q_1 Q_2]^{1/2}$$

$$Q_j = (a_{uj} - \bar{Y}_j)^2 + (a_{uj} - \bar{Y}_j)(b_{uj} - \bar{Y}_j) + (b_{uj} - \bar{Y}_j)^2$$

$$G_j = \begin{cases} -1, & \text{if} \quad \bar{Y}_{uj} \le Y_j, \\ 1, & \text{if} \quad \bar{Y}_{uj} > Y_j, \end{cases}$$

$$\bar{Y}_{uj} = (a_{uj} + b_{uj})/2.$$

Notice, special cases: (i) $Cov(Y_1, Y_1) \equiv S_1^2$

(ii) If $a_{uj} = b_{uj} = y_j$, for all u, i.e., classical data,

$$Cov(Y_1, Y_2) = \frac{1}{m} \Sigma (y_1 - \bar{Y}_1)(y_2 - \bar{Y}_2)$$

Back to Bertrand and Goupil (2000)

Sample variance is

$$S_j^2 = \frac{1}{3m} \sum_{u \in E} (b_{uj}^2 + b_{uj} a_{uj} + a_{uj}^2) - \frac{1}{4m^2} [\sum_{u \in E} (a_{uj} + b_{uj})]^2$$

This is total variance.

Take Total Sum of Squares = Total $SS_j = mS_j^2$

Then, we can show

$$\text{Total } SS_j = \text{Within Objects } SS_j + \text{Between Objects } SS_j$$

where

$$S_j^2 = \frac{1}{3m} \sum_{u \in E} [(a_{uj} - \bar{Y}_j)^2 + (a_{uj} - \bar{Y}_j)(b_{uj} - \bar{Y}_j) + (b_{uj} - \bar{Y}_j)^2]$$

Within Objects $SS_j = \frac{1}{3} \sum_{u \in E} [(a_{uj} - \bar{Y}_{uj})^2 + (a_{uj} - \bar{Y}_{uj})(b_{uj} - \bar{Y}_{uj}) + (b_{uj} - \bar{Y}_{uj})^2]$

Between Objects $SS_j = \sum_{u \in E} [(a_{uj} + b_{uj})/2 - \bar{Y}_j]^2$

with

$$\bar{Y}_{uj} = (a_{uj} + b_{uj})/2, \quad \bar{Y}_j = \frac{1}{2m} \sum_{u \in E} (a_{uj} + b_{uj}).$$

Classical data: $a_{uj} = b_{uj} = \bar{Y}_{uj}$

$\rightarrow$ Within Objects $SS_j = 0$

So, for Y_j, we have Sum of Squares SS,

$$\text{Total } SS_j = \text{Within Objects } SS_j + \text{Between Objects } SS_j$$

Likewise,

for (Y_i, Y_j), we have Sum of Products SP

$$\text{Total } SP_{ij} = \text{Within Objects } SP_{ij} + \text{Between Objects } SP_{ij}$$

Can rewrite

$$S_j^2 = \frac{1}{3m} \sum_{u \in E} [(a_{uj} - \bar{Y}_j)^2 + (a_{uj} - \bar{Y}_j)(b_{uj} - \bar{Y}_j) + (b_{uj} - \bar{Y}_j)^2]$$

Then, by analogy, for j = 1,2, for interval-valued variables Y_1 and Y_2,

empirical covariance function $Cov(Y_1, Y_2)$ is

$$Cov(Y_1, Y_2) = \frac{1}{3m} \sum_{u \in E} G_1 G_2 [Q_1 Q_2]^{1/2}$$

$$Q_j = (a_{uj} - \bar{Y}_j)^2 + (a_{uj} - \bar{Y}_j)(b_{uj} - \bar{Y}_j) + (b_{uj} - \bar{Y}_j)^2$$

$$G_j = \begin{cases} -1, & \text{if} \quad \bar{Y}_{uj} \le Y_j, \\ 1, & \text{if} \quad \bar{Y}_{uj} > Y_j, \end{cases}$$

$$\bar{Y}_{uj} = (a_{uj} + b_{uj}) / 2.$$

Can rewrite

$$S_j^2 = \frac{1}{3m}\sum_{u \in E} [(a_{uj} - \bar{Y}_j)^2 + (a_{uj} - \bar{Y}_j)(b_{uj} - \bar{Y}_j) + (b_{uj} - \bar{Y}_j)^2]$$

Then, by analogy, for $j = 1,2$, for interval-valued variables Y_1 and Y_2,

empirical covariance function $Cov(Y_1, Y_2)$ is

$$Cov(Y_1, Y_2) = \frac{1}{3m}\sum_{u \in E} G_1 G_2 [Q_1 Q_2]^{1/2}$$

$$Q_j = (a_{uj} - \bar{Y}_j)^2 + (a_{uj} - \bar{Y}_j)(b_{uj} - \bar{Y}_j) + (b_{uj} - \bar{Y}_j)^2$$

$$G_j = \begin{cases} -1, & \text{if} \quad \bar{Y}_{uj} \le Y_j, \\ 1, & \text{if} \quad \bar{Y}_{uj} > Y_j, \end{cases}$$

$$\bar{Y}_{uj} = (a_{uj} + b_{uj})/2.$$

(Total)SP part can be replaced by

$$\text{Total SP} = \frac{1}{6}\sum_{u} [2(a - \bar{Y})(c - \bar{X}) + (a - \bar{Y})(d - \bar{X}) + (b - \bar{Y})(c - \bar{X}) + 2(b - \bar{Y})(d - \bar{X})]$$

How is this obtained?

Recall that for a Uniform distribution,

$$Y \sim \bigcup(a,b),\ Var(Y) = \frac{(b-a)^2}{12}$$

By analogy, we can show, for *u=1,…,m* observations,

$$\text{Within SP} = \frac{1}{12}\sum_{u=1}^{m}(a_u - b_u)(c_u - d_u)$$

$$\text{Between SP} = \sum_{u=1}^{m}\left(\frac{a_u + b_u}{2} - \bar{Y}_1\right)\left(\frac{c_u + d_u}{2} - \bar{Y}_2\right)$$

where

$$Y_{u1} = [a_u, b_u], \qquad Y_{u2} = [c_u, d_u]$$

$$\bar{Y}_1 = \frac{1}{m}\sum_{u=1}^{m}\left(\frac{a_u + b_u}{2}\right), \qquad \bar{Y}_2 = \frac{1}{m}\sum_{u=1}^{m}\left(\frac{c_u + d_u}{2}\right)$$

$$\text{Within SP} = \frac{1}{12}\sum_{u=1}^{m}(a_u - b_u)(c_u - d_u)$$

$$\text{Between SP} = \sum_{u=1}^{m}\left(\frac{a_u + b_u}{2} - \bar{Y}_1\right)\left(\frac{c_u + d_u}{2} - \bar{Y}_2\right)$$

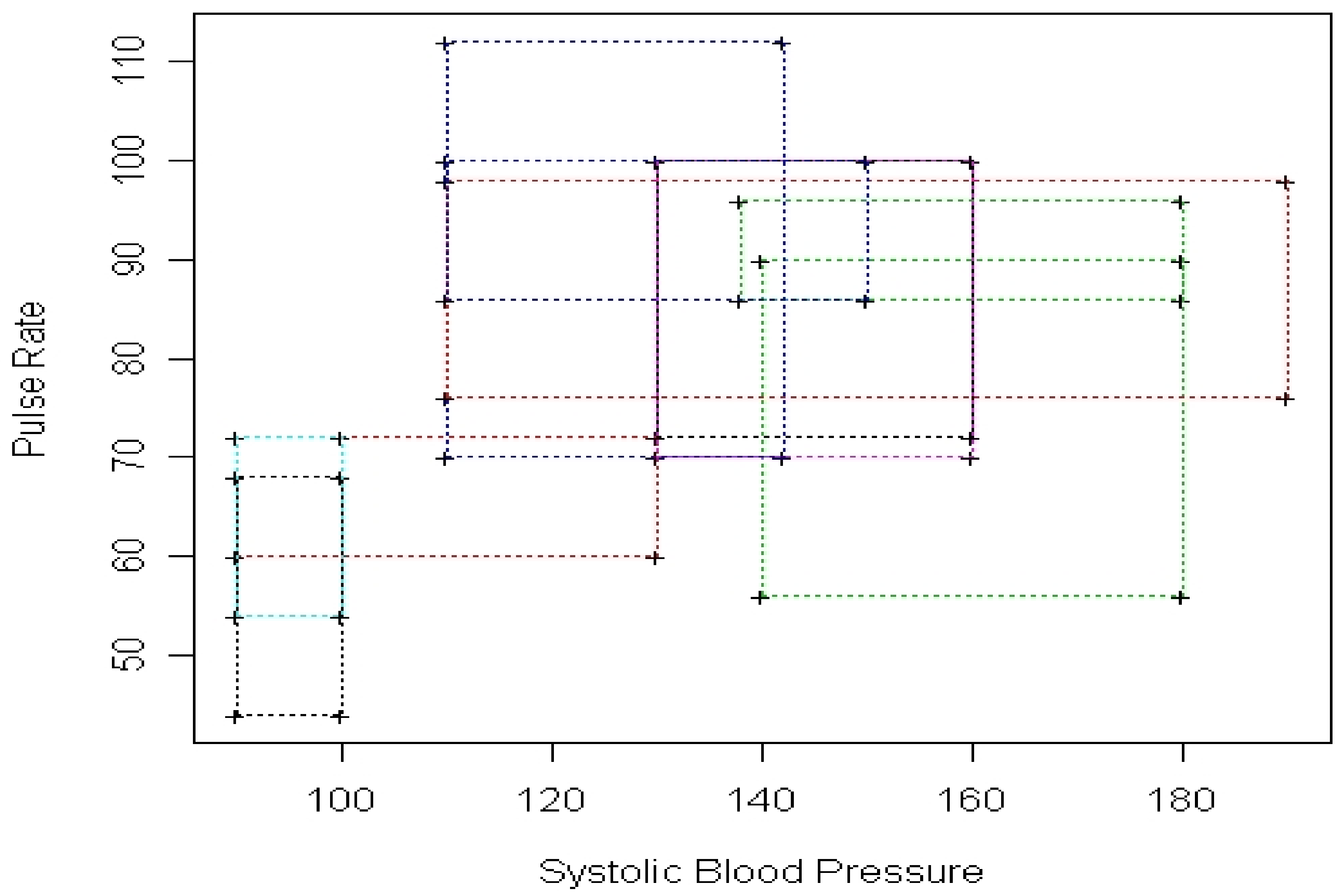
Hence, from

Total SP = Within SP + Between SP

$$= \frac{1}{6}\sum_{u=1}^{m}\left[2(a_u - \bar{Y}_1)(c - \bar{Y}_2) + (a - \bar{Y}_1)(d - \bar{Y}_2)\right.$$
$$\left. + (b - \bar{Y}_1)(c - \bar{Y}_2) + 2(b - \bar{Y}_1)(d - \bar{Y}_2)\right]$$

u	*Y* Pulse Rate	*X1* Systolic Pressure	*X2* Diastolic Pressure
1	[44, 68]	[90, 110]	[50, 70]
2	[60, 72]	[90, 130]	[70, 90]
3	[56, 90]	[140, 180]	[90, 100]
4	[70, 112]	[110, 142]	[80, 108]
5	[54, 72]	[90, 100]	[50, 70]
6	[70, 100]	[134, 142]	[80, 110]
7	[72, 100]	[130, 160]	[76, 90]
8	[76, 98]	[110, 190]	[70, 110]
9	[86, 96]	[138, 180]	[90, 110]
10	[86, 100]	[110, 150]	[78, 100]
11	[63, 75]	**[60, 100]**	**[140, 150]**

Rule: X2 = Diastolic Pressure < Systolic Pressure = X1


Pulse Rate
110
100
90
80
70
60
50
100
120
140
160
180
Systolic Blood Pressure

The regression equation becomes,

for

Y = Pulse Rate, X1 = Systolic Pressure

Y = 25.228 + 0.410X1

$\bar{Y} = 79.1$ $\quad\quad$ $\bar{X} = 131.5$

Std Devn(Y) = 14.692   Std Devn(X1) = 26.013

Cov(Y, X1) = 277.217   rho(Y, X1) = 0.725

Y = Pulse Rate, X1 = Systolic Pressure

Y = 25.228 + 0.410X1

Prediction

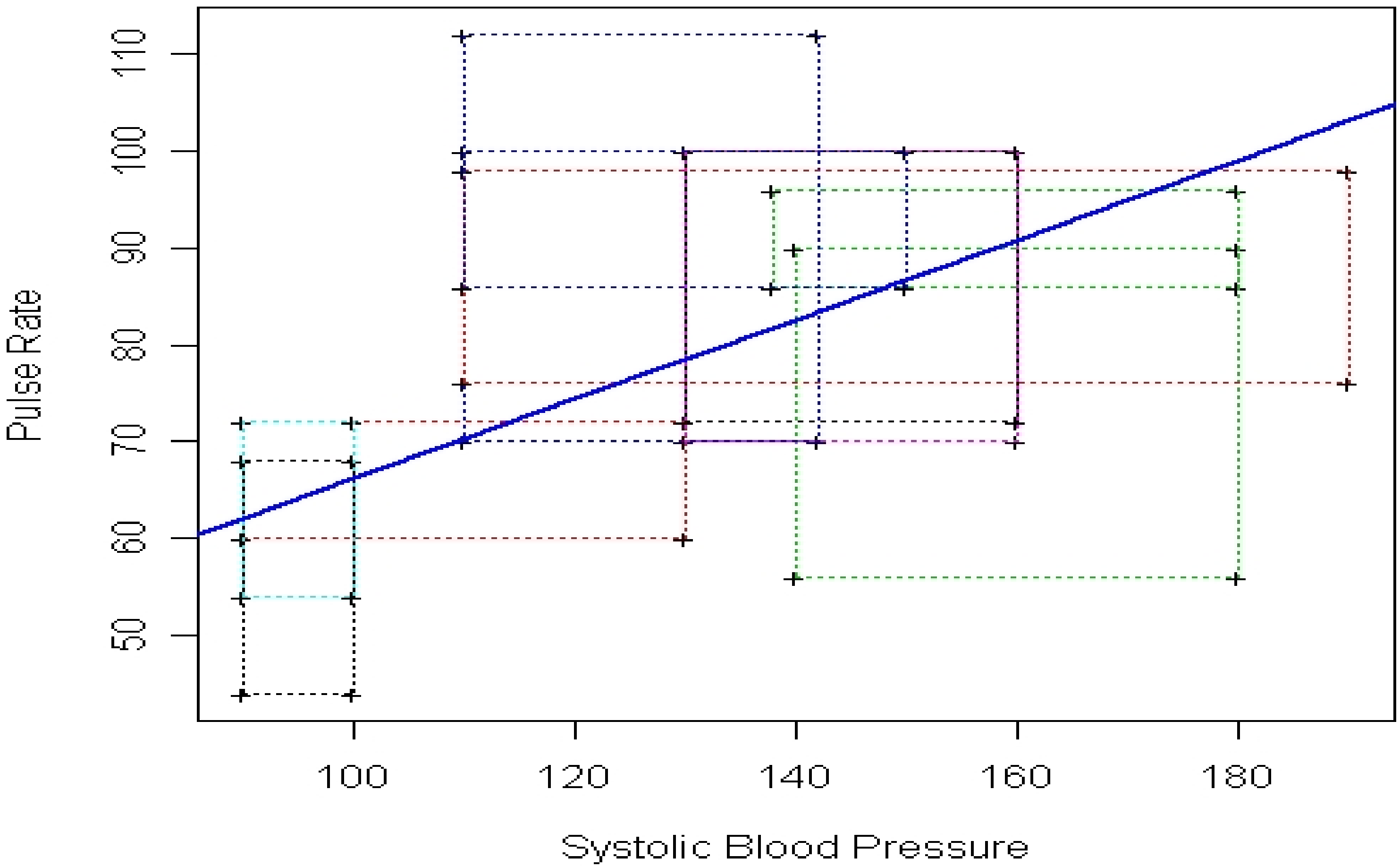
$$\widehat{Y}_u = [\widehat{a}_{u1}, \widehat{b}_{u1}]$$

with

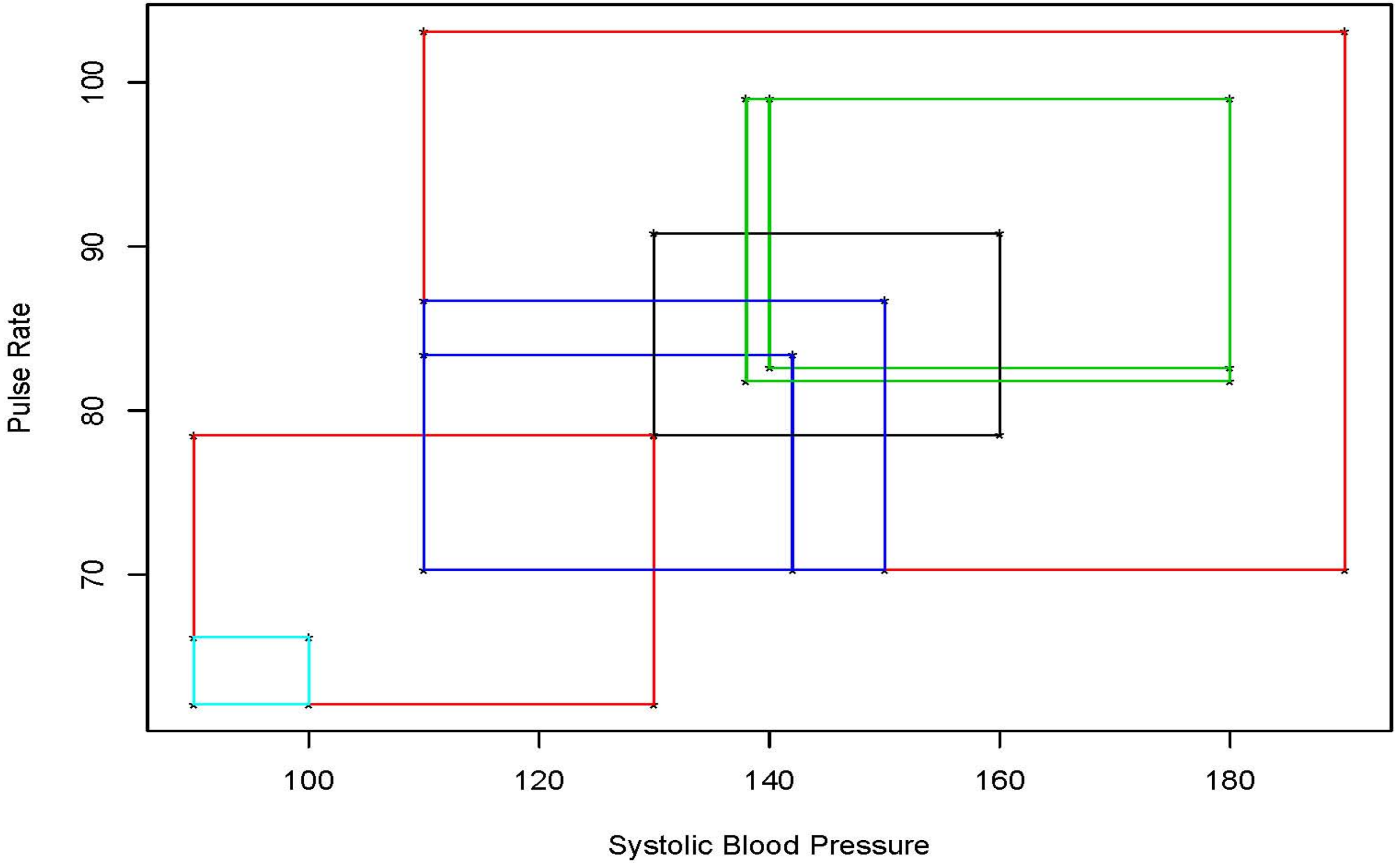
$$\widehat{a}_{u1} = 25.228 + 0.410 a_{u2}$$

$$\widehat{b}_{u1} = 25.228 + 0.410 b_{u2}$$

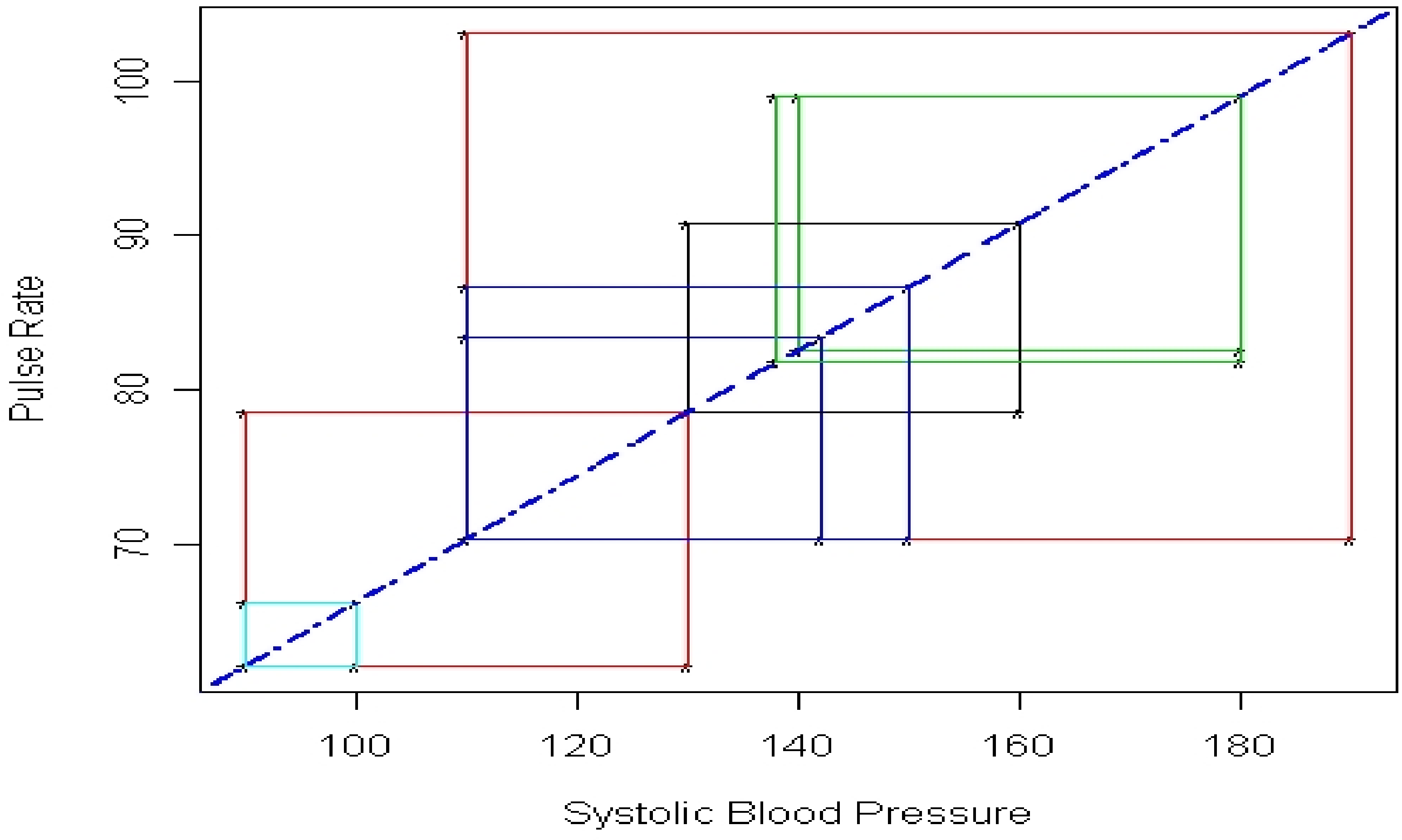
Symbolic Prediction Equation



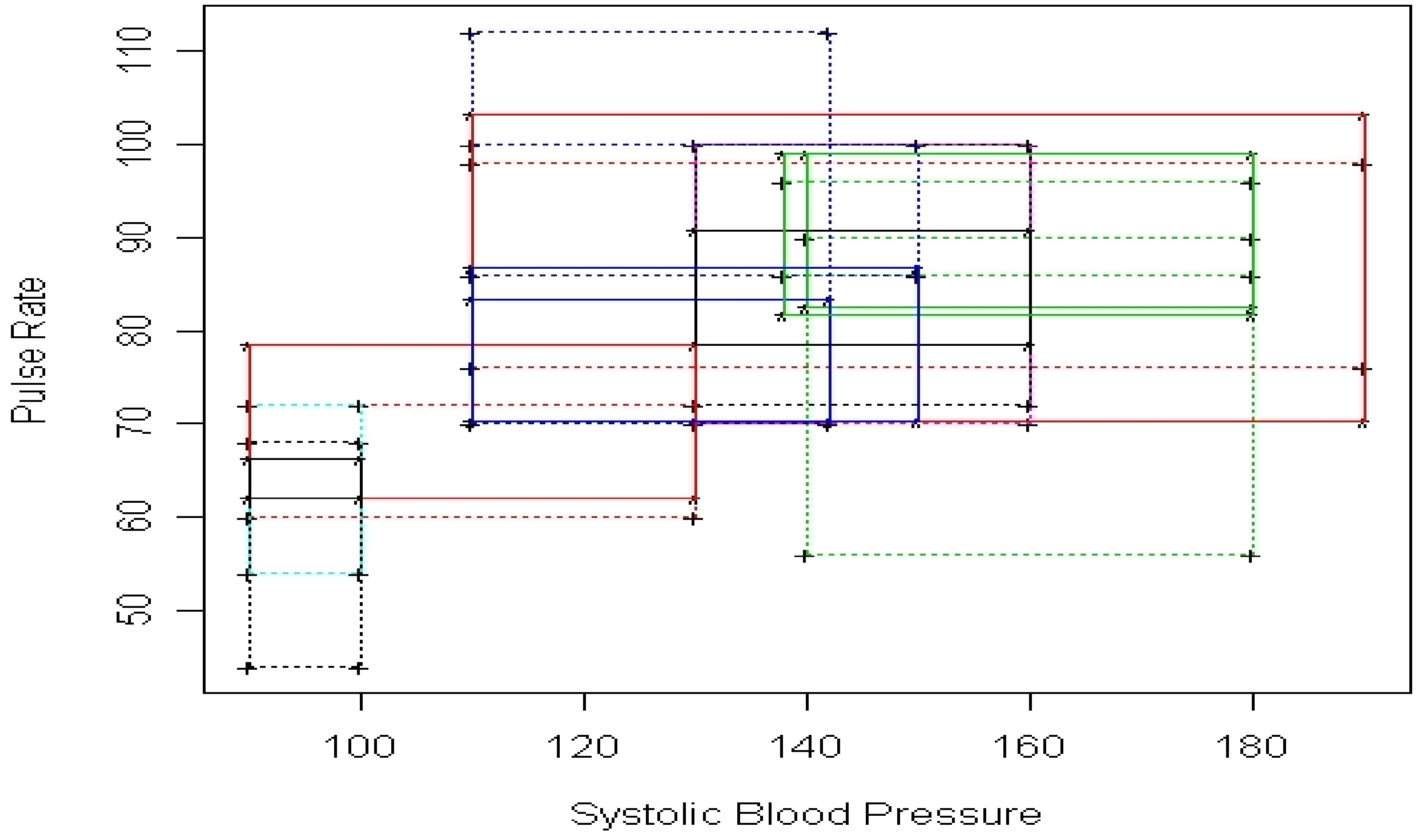
Symbolic Prediction Intervals



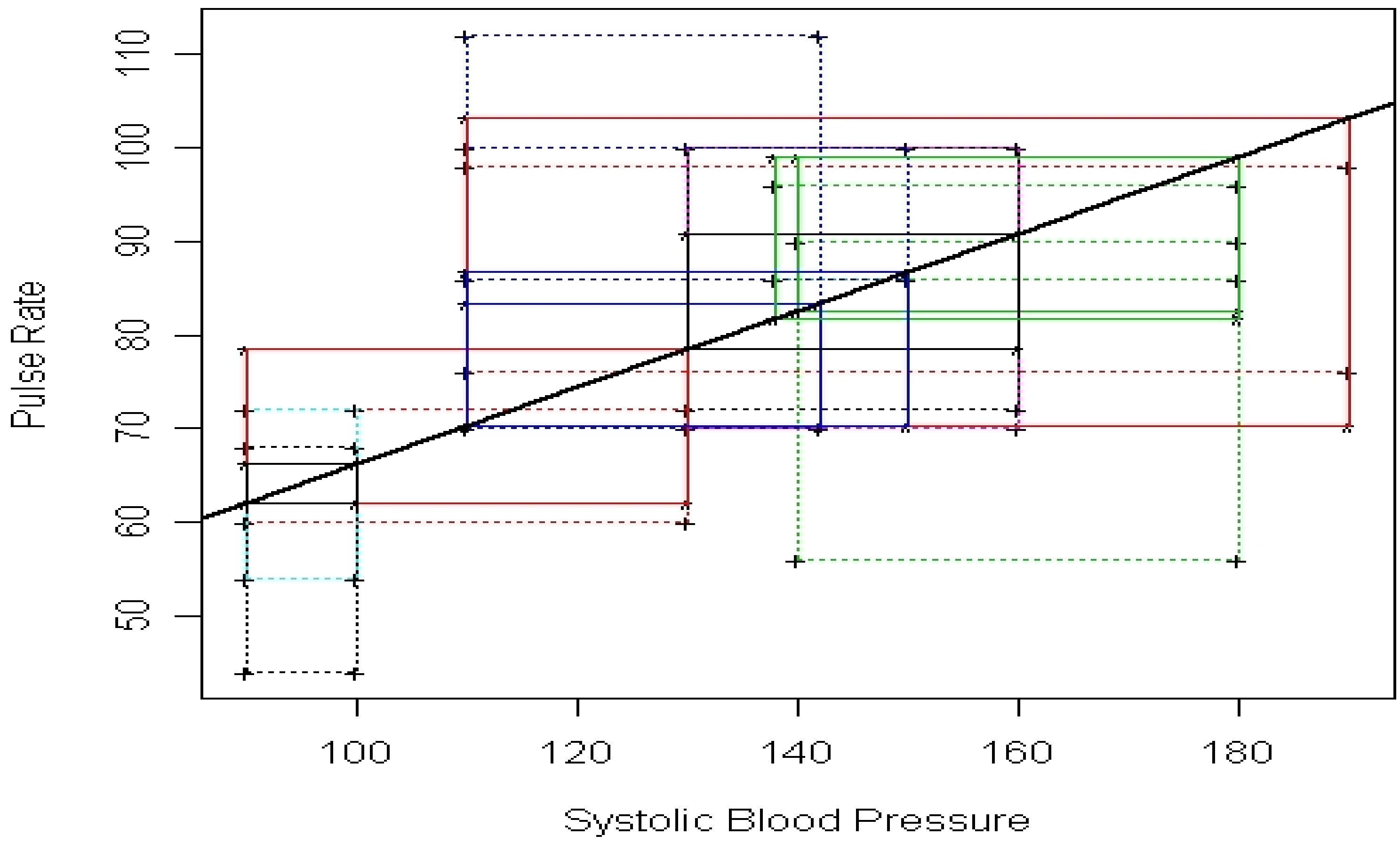
Symbolic Prediction Intervals and Equation



Original Intervals …… Prediction Intervals -------



Data Intervals: …….   Prediction Intervals: ------



Predicted Pulse Rates and Residuals

	Observed (Y, X1)		Predicted Y		Residuals	
u	Pulse Rate	Systolic	$\hat{a}_u$	$\hat{b}_u$	Res_a	Res_b
1	[44,68]	[90,100]	[62.099,	66.195]	[-18.099,	1.805]
2	[60,72]	[90,130]	[62.099,	78.485]	[-2.099,	-6.486]
3	[56,90]	[140,180]	[82.582,	98.969]	[-26.582,	-8.969]
4	[70,112]	[110,142]	[70.292,	83.402]	[-0.292,	28.599]
5	[54,72]	[90,100]	[62.099,	66.195]	[-8.099,	5.805]
6	[70,100]	[130,160]	[78.486,	90.776]	[-8.486,	9.224]
7	[72,100]	[130,160]	[78.486,	90.776]	[-6.486,	9.224]
8	[76,98]	[110,190]	[70.292,	103.066]	[5.708,	-5.066]
9	[86,96]	[138,180]	[81.763,	98.969]	[4.237,	2.969]
10	[86,100]	[110,150]	[70.292,	86.679]	[15.708,	13.321]

Y_u = Pulse Rate = $[a_u, b_u]$

$\hat{Y}_u$ = Predicted Pulse Rate = $[\hat{a}_u, \hat{b}_u]$

Residual = $[Res_a, Res_b]$

Sum of Residuals for Symbolic Fit

Sum of Min Residuals Σ_u Res_{au} = -44.488

Sum of Max Residuals Σ_u Res_{bu}= 44.488

Sum of Squared Residuals for Symbolic Fit

Sum of Min Squared Residuals = 1515.592

Sum of Max Squared Residuals = 1359.434

Classical Regression on Midpoints

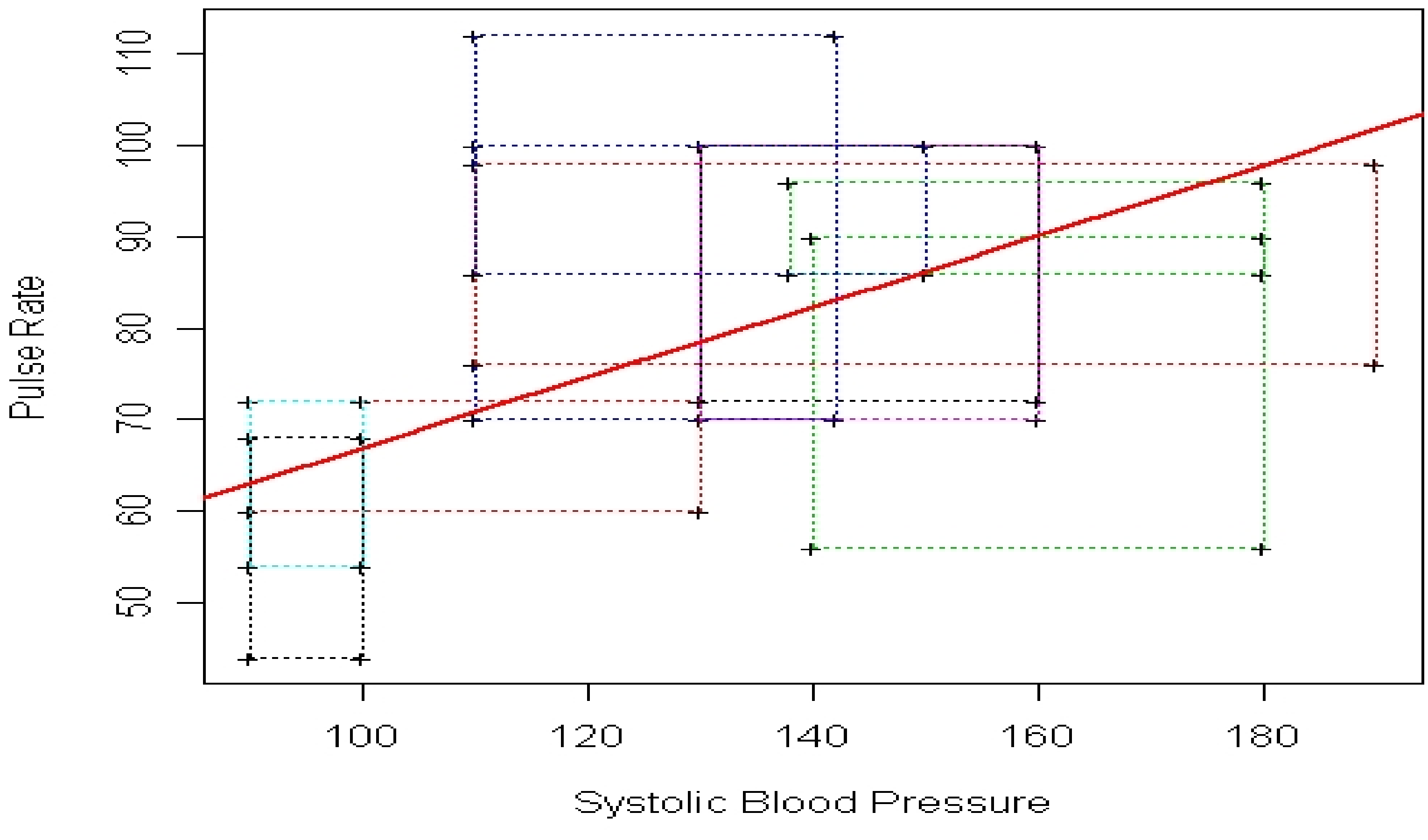
$$Y_u^c = (a_{u1}+b_{u1})/2, \qquad X_{ju}^c = (a_{uj}+b_{uj})/2, \qquad j = 1, 2$$

$$\rightarrow Y^c = 28.322 + 0.386 X_1$$

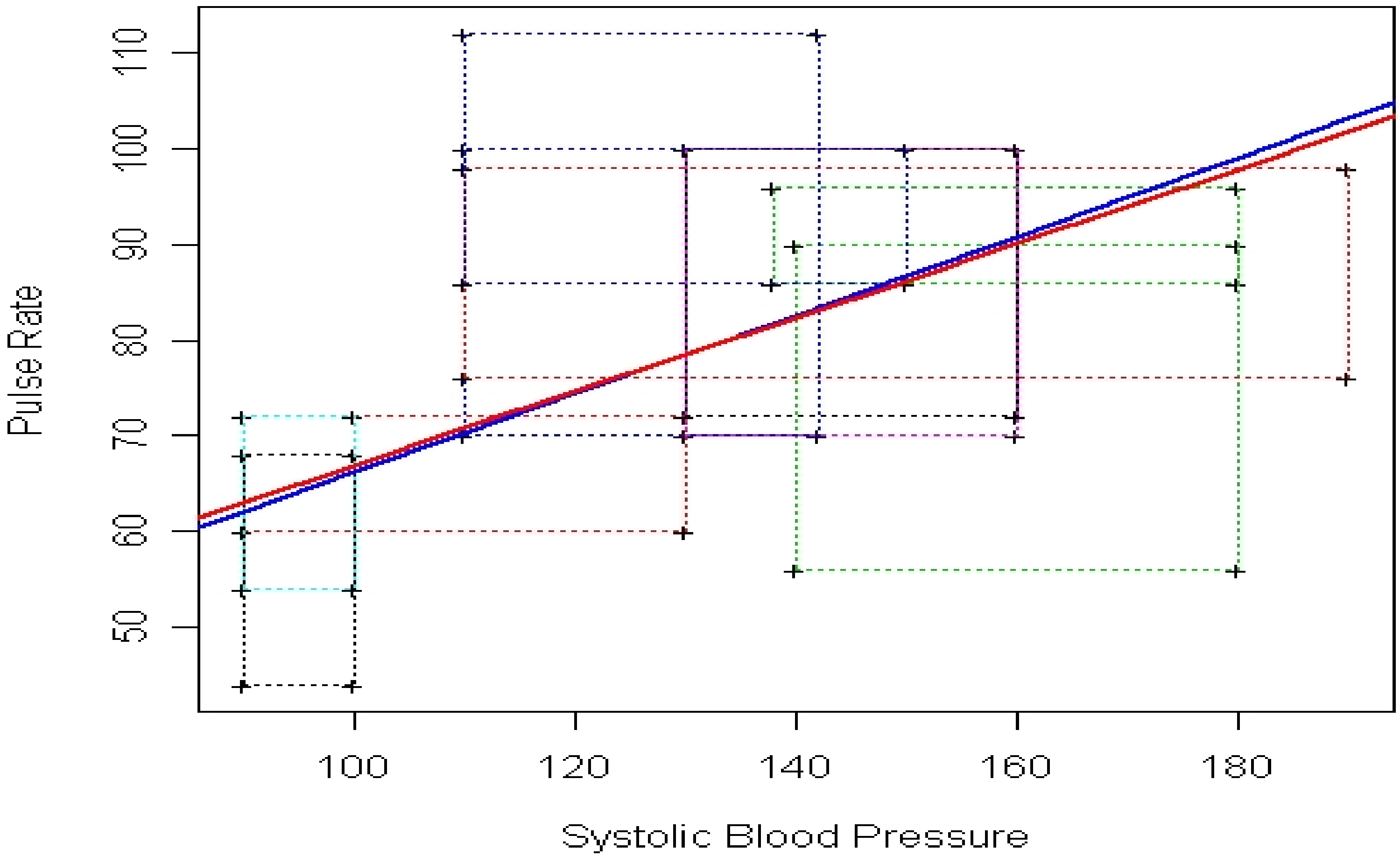
$$\hat{Y}^c = [\hat{a}^c, \hat{b}^c]$$

$$\hat{a}_u^c = 28.322 + 0.386 a_{u2}$$
$$\hat{b}_u = 28.322 + 0.386 b_{u2}$$

Classical Regression through Midpoints



Symbolic Regression ---- Classical regression ----



Comparison of Regression Fits

Sum of Residuals for Symbolic Fit

Sum of Min Residuals = -44.488

Sum of Max Residuals = 44.48

Sum of Squared Residuals for Symbolic Fit

Sum of Min Squared Residuals = 1515.592

Sum of Max Squared Residuals = 1359.434

Sum of Residuals for Classical Fit

Sum of Min Residuals = -48.652

Sum of Max Residuals = 48.652

Sum of Squared Residuals for Classical Fit

Sum of Min Squared Residuals = 1544.889

Sum of Max Squared Residuals = 1364.639

Centers and Range Regression

DeCarvalho, Lima Neto, Tenorio, Freire, ... (2004, 2005, …)

Midpoint: $Y^c = (a + b)/2$, $X^c = (c + d)/2$

Range: $Y^r = (b – a)/2$, $X^r = (d - c)/2$

$$\hat{Y}^c = 28.322 + 0.386 X^c$$
$$\hat{Y}^r = 25.444 - 0.05875 X^r$$

Centers and Range Regression

DeCarvalho, Lima Neto, Tenorio, Freire, ... (2004, 2005, …)

Midpoint: $Y^c = (a + b)/2$, $X^c = (c + d)/2$

Range: $Y^r = (b – a)/2$, $X^r = (d - c)/2$

$$\hat{Y}^c = 28.322 + 0.386 X^c$$
$$\hat{Y}^r = 25.444 - 0.05875 X^r$$

$$\hat{Y}^c = 31.788 + 0.3300 X_1^c + 0.111 X_1^r$$
$$\hat{Y}^r = 7.866 + 0.170 X_1^c + -0.194 X_1^r$$

Centers and Range Regression --Predictions

Obs	Y	Single $[\widehat{Y}_a, \widehat{Y}_b]$	Multiple $[\widehat{Y}_a, \widehat{Y}_b]$
1	[44,68]	[52.572,77.439]	[53.195,75.299]
2	[60,72]	[59.230,82.365]	[63.089,81.937]
3	[56,90]	[78.537,101.672]	[75.334,102.695]
4	[70,112]	[65.178,88.774]	[65.349,88.470]
5	[54,72]	[52.572,77.439]	[53.195,75.299]
6	[70,100]	[72.457,96.168]	[69.587,96.331]
7	[72,100]	[72.457,96.168]	[69.587,96.331]
8	[76,98]	[75.831,96.655]	[81.180,99.092]
9	[86,96]	[78.209,101.228]	[75.504,102.308]
10	[86,100]	[66.953,90.087]	[67.987,90.241]

Symbolic Principal Components -- BATS

Y1=Head, Y2=Tail, Y3=Height, Y4=Forearm

Obs	[Y1a,Y1b]	[Y2a,Y2b]	[Y3a,Y3b]	[Y4a,Y4b]
1	[33, 52]	[26, 33]	[4, 7]	[27, 32]
2	[38, 50]	[30, 40]	[7, 8]	[32, 37]
3	[43, 48]	[34, 39]	[6, 7]	[31, 38]
4	[44, 48]	[34, 44]	[7, 8]	[31, 36]
5	[41, 51]	[30, 39]	[8, 11]	[33, 41]
6	[40, 45]	[39, 44]	[9, 9]	[36, 42]
7	[45, 53]	[35, 38]	[10, 12]	[39, 44]
8	[44, 58]	[41, 54]	[6, 8]	[35, 41]
9	[47, 53]	[43, 53]	[7, 9]	[37, 41]
10	[50, 69]	[30, 43]	[11, 13]	[51, 61]
11	[65, 80]	[48, 60]	[12, 16]	[55, 68]
12	[82, 87]	[46, 57]	[11, 12]	[58, 63]

Symbolic Principal Components -- BATS

Y1=Head, Y2=Tail,Y3=Height,Y4=Forearm

Obs	PC1a	PC1b	PC2a	PC2b	PC3a	PC3b
1	45.276	62.471	11.935	22.006	-28.931	-10.135
2	53.826	67.716	13.788	24.556	-24.948	-11.019
3	57.185	66.275	17.708	24.377	-22.581	-15.398
4	58.198	67.908	17.736	27.816	-21.739	-13.517
5	56.421	71.418	11.433	23.055	-25.695	-12.063
6	61.999	70.061	19.368	25.247	-17.330	-10.843
7	64.941	74.123	14.485	19.875	-24.414	-15.651
8	62.968	80.264	22.096	36.217	-27.290	-10.011
9	66.990	77.698	23.402	33.956	-22.355	-12.302
10	72.282	94.342	6.237	21.763	-39.804	-18.374
11	90.753	112.874	18.529	34.738	-40.761	-21.056
12	99.870	110.547	21.800	32.763	-46.392	-37.047

Symbolic Principal Component Analysis -- BATS