Albert Ludwigs University Freiburg
Department of Computer Science
BIOINFORMATICS GROUP FREIBURG

MASTER THESIS CALCULATIONS

5TH SEPTEMBER 2018

Calculations

AUTHOR:

ALEXANDER MATTHEIS

SUPERVISORS:

DR. MARTIN RADEN

PD DR. HANS-PETER KAHLE

CROSS-DATING
ABSOLUTE TO NORMALIZED COORDINATES

given

$$\mathrm{Y} = \langle 5, 3, 2 \rangle$$

calculation

$$\mathrm{mean}(\mathrm{Y}) = \frac{10}{3}$$

$$\sigma(\mathrm{Y}) = \sqrt{\frac{\sum_{i=1}^{3}\left(y_i - \mathrm{mean}(\mathrm{Y})\right)^2}{n-1}}$$

$$= \sqrt{\frac{\left(5 - \frac{10}{3}\right)^2 + \left(3 - \frac{10}{3}\right)^2 + \left(2 - \frac{10}{3}\right)^2}{2}} \approx 1.527525$$

$$y_1^{std} = \frac{5 - \frac{10}{3}}{1.527525} \approx 1.09109$$

$$y_2^{std} = \frac{3 - \frac{10}{3}}{1.527525} \approx -0.2182179$$

$$y_3^{std} = \frac{2 - \frac{10}{3}}{1.527525} \approx -0.8728717$$

Cross-Dating
Double Weighting - Single Column

given

$sampleLength = 3$

$bestYears = \langle 1992, 1502, 1493, 1801, 1723 \rangle$ (sorted by rank)

calculation

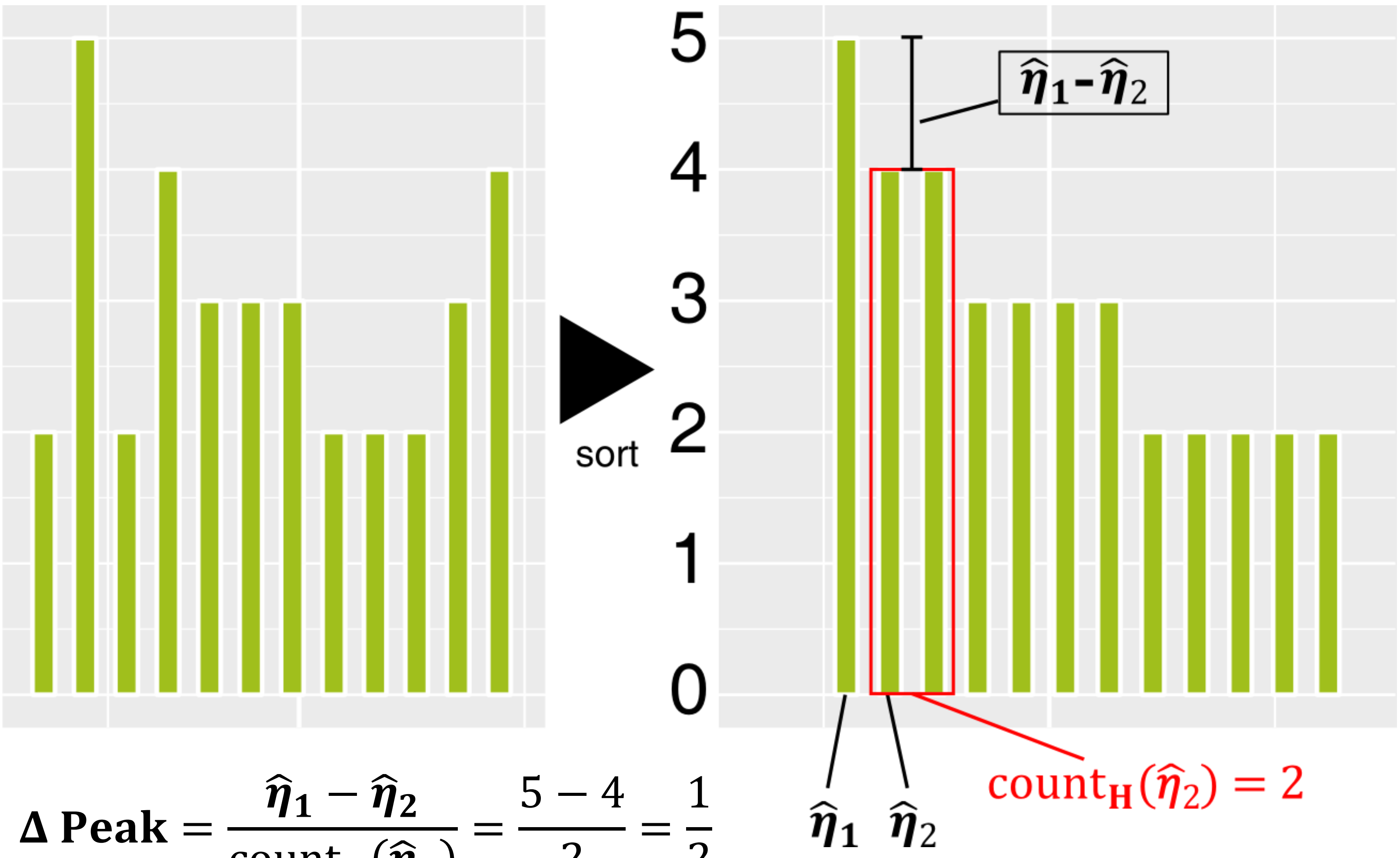
$output = \langle 1992, 1992, 1992, 1502, 1502, 1493, 1801, 1723 \rangle$

1992 has rank 1 $\Rightarrow$ 3 times in list
1502 has rank 2 $\Rightarrow$ 2 times …
1493 has rank 3 $\Rightarrow$ once …

$1801, 1723$ no ranks lower 1 $\Rightarrow$ once …

Cross-Dating

ΔPeaks



$$\Delta\,\textbf{Peak} = \frac{\hat{\eta}_1 - \hat{\eta}_2}{\text{count}_\text{H}(\hat{\eta}_2)} = \frac{5-4}{2} = \frac{1}{2}$$

Cross-Dating
Empirical Distribution

given

$\mathbf{S}_{\mathrm{S}}^{\mathrm{C}} = \{\varsigma_1 = 2, \varsigma_2 = 15, \varsigma_3 = 5, \varsigma_4 = 52, \varsigma_5 = 3\}$

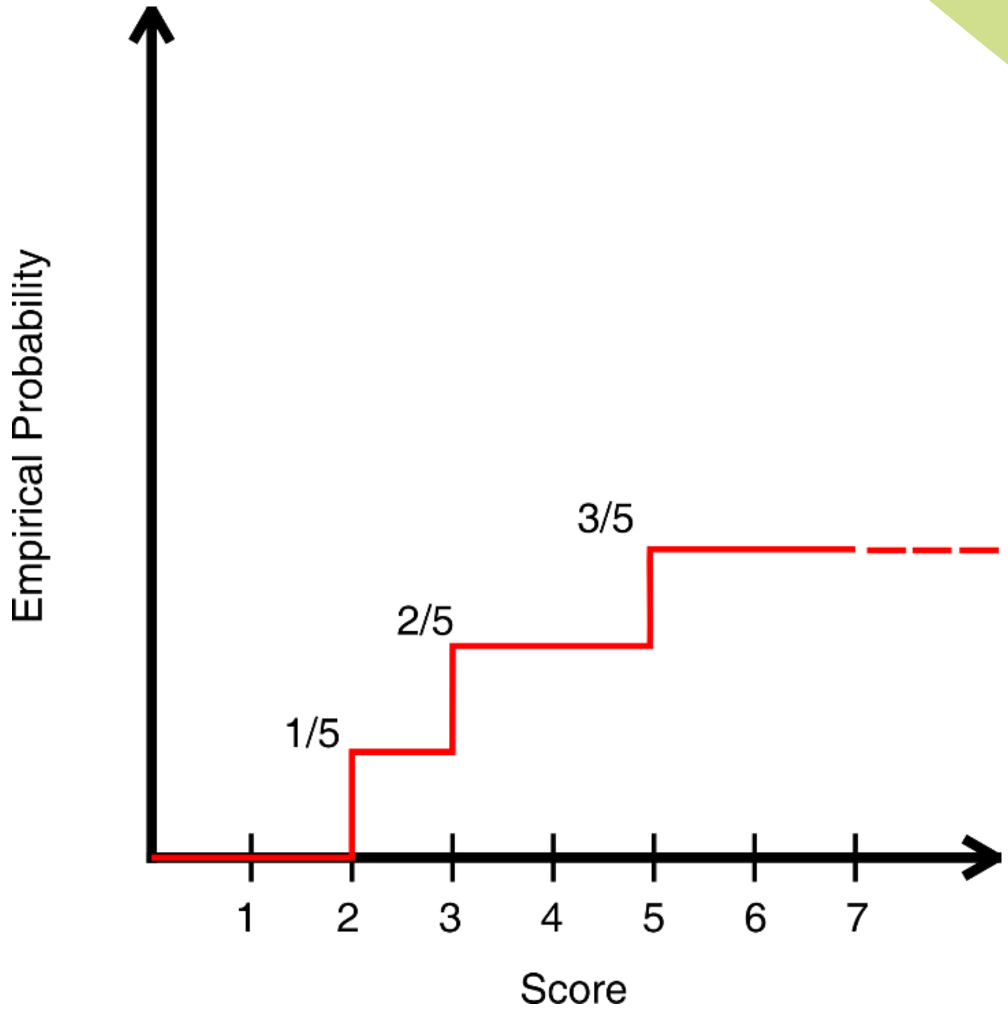
calculation

$$\hat{F}(\varsigma = 2) = \frac{1}{5} \cdot (1) = \frac{1}{5}$$

$$\hat{F}(\varsigma = 3) = \frac{1}{5} \cdot (1 + 0 + 0 + 0 + 1) = \frac{2}{5}$$

$$\hat{F}(\varsigma = 5) = \frac{1}{5} \cdot (1 + 0 + 1 + 0 + 1) = \frac{3}{5}$$

…



CROSS-DATING
GAPS IN SAMPLES

given

sample: ABCD[–gap–]XYZ

calculation

first sample:

chronology

1	2	3	4	5	6	7	8	9	10
A	B	C	D						

Shift	Profil A	Profil B	Profil C	Profil D
1	0.5	0.6	0.7	0.8
2				
3				

Cross-Dating
Gaps in Samples

given

sample: ABCD[–gap–]XYZ

calculation

first sample:

chronology

1	2	3	4	5	6	7	8	9	10
	A	B	C	D					

Shift	Profil A	Profil B	Profil C	Profil D
1	0.5	0.6	0.7	0.8
2	0.9	1.0	1.1	1.2
3				

CROSS-DATING
GAPS IN SAMPLES

given

sample: ABCD[–gap–]XYZ

calculation

first sample:

chronology

1	2	3	4	5	6	7	8	9	10
		A	B	C	D				

Shift	Profil A	Profil B	Profil C	Profil D
1	0.5	0.6	0.7	0.8
2	0.9	1.0	1.1	1.2
3	1.3	1.4	1.5	1.6

Cross-Dating
Gaps in Samples

given

sample: ABCD[−gap−]XYZ

calculation

second sample:

chronology

1	2	3	4	5	6	7	8	9	10
					X	Y	Z		

Shift	Profil A	Profil B	Profil C	Profil D
1	0.5	0.6	0.7	0.8
2	0.9	1.0	1.1	1.2
3	1.3	1.4	1.5	1.6

Shift	Profil X	Profil Y	Profil Z
1	1.6	1.7	1.8
2			
3			

CROSS-DATING
GAPS IN SAMPLES

given

sample: ABCD[–gap–]XYZ

calculation

second sample:

chronology

1	2	3	4	5	6	7	8	9	10
						X	Y	Z	

Shift	Profil A	Profil B	Profil C	Profil D
1	0.5	0.6	0.7	0.8
2	0.9	1.0	1.1	1.2
3	1.3	1.4	1.5	1.6

Shift	Profil X	Profil Y	Profil Z
1	1.6	1.7	1.8
2	1.9	2.0	2.1
3			

Cross-Dating
Gaps in Samples

given

sample: ABCD[–gap–]XYZ

calculation

second sample:

chronology

1	2	3	4	5	6	7	8	9	10
							X	Y	Z

Shift	Profil A	Profil B	Profil C	Profil D
1	0.5	0.6	0.7	0.8
2	0.9	1.0	1.1	1.2
3	1.3	1.4	1.5	1.6

Shift	Profil X	Profil Y	Profil Z
1	1.6	1.7	1.8
2	1.9	2.0	2.1
3	2.2	2.3	2.4

Cross-Dating
Gaps in Samples

calculation

GAP-SIZE = 3

chronology

1	2	3	4	5	6	7	8	9	10
A	B	C	D				X	Y	Z

Shift	Profil A	Profil B	Profil C	Profil D
1	0.5	0.6	0.7	0.8
2	0.9	1.0	1.1	1.2
3	1.3	1.4	1.5	1.6

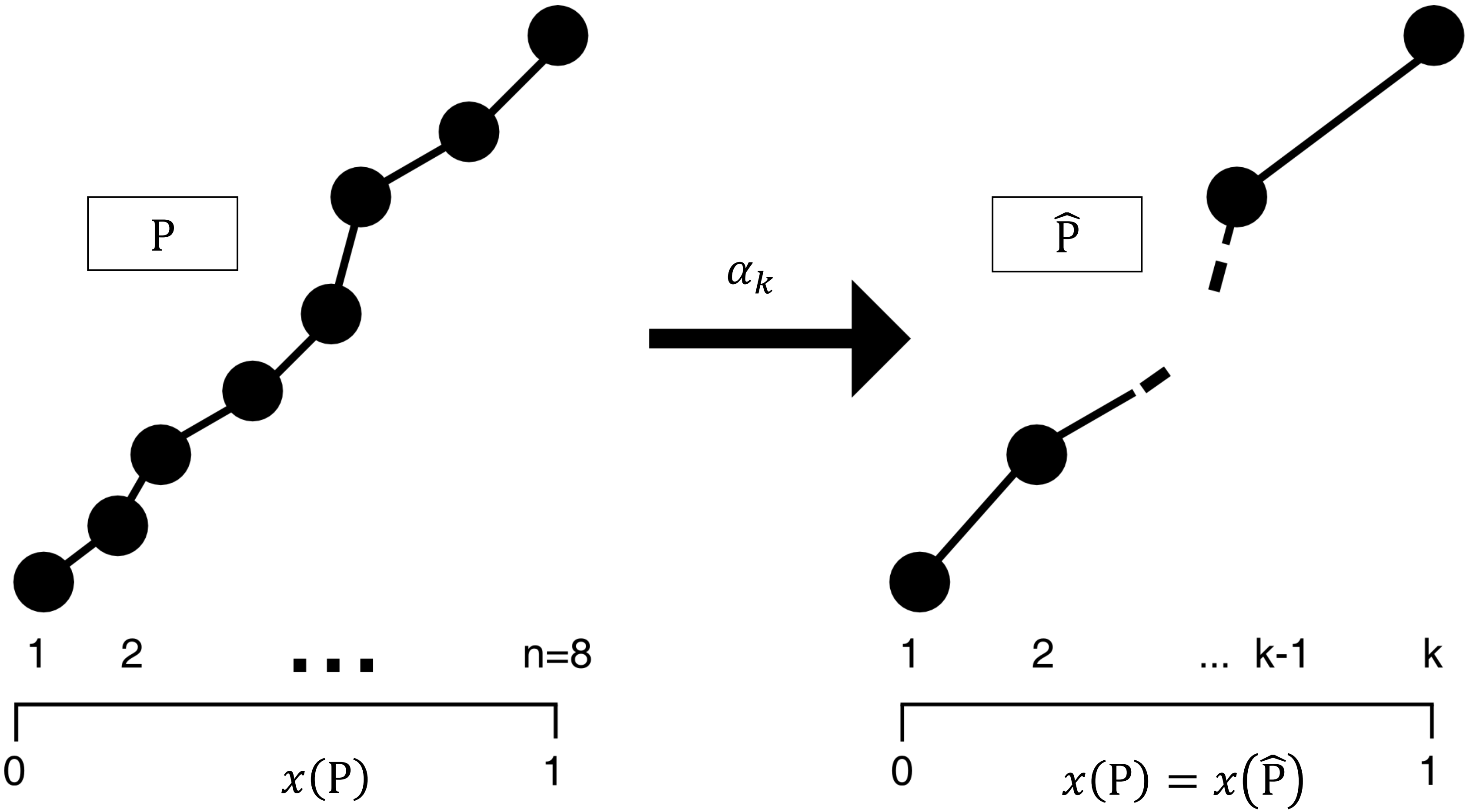
Shift	Profil X	Profil Y	Profil Z
1	1.6	1.7	1.8
2	1.9	2.0	2.1
3	2.2	2.3	2.4

Score = 0.5 + 0.6 + 0.7 + 0.8 + 0 + 0 + 0 + 2.2 + 2.3 + 2.4

Cross-Dating
Interpolation

given

$\alpha_k(\mathrm{P}) = \widehat{\mathrm{P}}$



CROSS-DATING
KENDALL

given

$V^a = \cdots$

$V^b = \cdots$

for $i < j$

$concordant$:	$v_i^a < v_j^a \wedge v_i^b < v_j^b$	|	$v_i^a > v_j^a \wedge v_i^b > v_j^b$	(sort order agreement)
$discordant$:	$v_i^a < v_j^a \wedge v_i^b > v_j^b$	|	$v_i^a > v_j^a \wedge v_i^b < v_j^b$	(sort order disagreement)
T_a:	$v_i^a = v_j^a \wedge v_i^b \neq v_j^b$			(tie in a)
T_b:	$v_i^a \neq v_j^a \wedge v_i^b = v_j^b$			(tie in b)

calculation (values taken from [3])

$r(v_i^a)$	1	2	3.5	3.5	5	6	7	8
$r(v_i^b)$	2	3	5	1	4	7.5	7.5	6

Cross-Dating
KENDALL

calculation (values taken from [3])

$r(v_i^a)$	1	2	3.5	3.5	5	6	7	8
$r(v_i^b)$	2	3	5	1	4	7.5	7.5	6

it holds

if	$r(v_i^b) - r(v_j^b) < 0$
then	**concordant** because $r(v_i^a) - r(v_j^a) < 0$ holds
	(ranks $r(v_i^a)$ are sorted in ascending order)
else if	$r(v_i^b) - r(v_j^b) > 0$
then	**discordant**
else…	

CROSS-DATING
KENDALL

calculation (values taken from [3])

$r(v_i^a)$	1	2	3.5	3.5	5	6	7	8
$r(v_i^b)$	2	3	5	1	4	7.5	7.5	6

1-2	**1-3.5**	**1-3.5**	**1-5**	**1-6**	**1-7**	**1-8**
2-3	2-5	2-1	2-4	2-7.5	2-7.5	2-6
+	**+**	**-**	**+**	**+**	**+**	**+**

CROSS-DATING
KENDALL

calculation (values taken from [3])

$r(v_i^a)$	1	2	3.5	3.5	5	6	7	8
$r(v_i^b)$	2	3	5	1	4	7.5	7.5	6

1-2	**1-3.5**	**1-3.5**	**1-5**	**1-6**	**1-7**	**1-8**
2-3	2-5	2-1	2-4	2-7.5	2-7.5	2-6
+	+	-	+	+	+	+

2-3.5	**2-3.5**	**2-5**	**2-6**	**2-7**	**2-8**
3-5	**3-1**	**3-4**	**3-7.5**	**3-7.5**	**3-6**
+	-	+	+	+	+

Cross-Dating
Kendall

calculation (values taken from [3])

$r(v_i^a)$	1	2	3.5	3.5	5	6	7	8
$r(v_i^b)$	2	3	5	1	4	7.5	7.5	6

1-2	**1-3.5**	**1-3.5**	**1-5**	**1-6**	**1-7**	**1-8**
2-3	2-5	2-1	2-4	2-7.5	2-7.5	2-6
+	+	-	+	+	+	+

2-3.5	**2-3.5**	**2-5**	**2-6**	**2-7**	**2-8**
3-5	**3-1**	**3-4**	**3-7.5**	**3-7.5**	**3-6**
+	-	+	+	+	+

3.5-3.5	**3.5-5**	**3.5-6**	**3.5-7**	**3.5-8**
5-1	**5-4**	**5-7.5**	**5-7.5**	**5-6**
a	+	-	-	-

CROSS-DATING
KENDALL

calculation (values taken from [3])

1-2	1-3.5	1-3.5	1-5	1-6	1-7	1-8
2-3	2-5	2-1	2-4	2-7.5	2-7.5	2-6
+	+	-	+	+	+	+

2-3.5	2-3.5	2-5	2-6	2-7	2-8
3-5	3-1	3-4	3-7.5	3-7.5	3-6
+	-	+	+	+	+

3.5-3.5	3.5-5	3.5-6	3.5-7	3.5-8
5-1	5-4	5-7.5	5-7.5	5-6
a	-	+	+	+

⋮

CROSS-DATING

KENDALL

⋮

calculation

3.5-5	3.5-6	3.5-7	3.5-8
1-4	1-7.5	1-7.5	1-6
+	+	+	+

5-6	5-7	5-8
4-7.5	4-7.5	4-6
+	+	+

6-7	6-8
7.5-7.5	7.5-6
b	-

7-8
7.5-6
-

N_{conc}: 21 N_{disc}: 5 N_{ties}^{a}: 1 N_{ties}^{b}: 1

$$\tau = \frac{N_{conc} - N_{disc}}{\sqrt{\left(N_{conc} + N_{disc} + N_{ties}^{a}\right) \cdot \left(N_{conc} + N_{disc} + N_{ties}^{b}\right)}} = \frac{21 - 5}{\sqrt{27 \cdot 27}}$$

$$\approx 0.5925926$$

Cross-Dating
Mean-Distance

given

$Y^{a} = \langle 2, 3, 5 \rangle$

$Y^{b} = \langle 3, 1, 2 \rangle$

calculation

$$\mathrm{dist}_{avg}^{y}\left(\mathrm{P}^{\mathrm{a}}, \mathrm{P}^{\mathrm{b}}\right) = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| y_i^a - y_i^b \right| = \frac{|2-3| + |3-1| + |5-2|}{3} = \frac{1+2+3}{3} = 2$$

Cross-Dating
Median

given

$\mathbf{R} = \{r_1 = 1,\ r_2 = 6,\ r_3 = 7,\ r_4 = 23,\ r_5 = 65,\ r_6 = 76\}$ with $r_1 \leq r_2 \leq \cdots \leq r_u$

calculation

$u = 6$

$$\text{median}(\mathbf{R}) = \begin{cases} r_{\frac{u+1}{2}} & , \quad u\ odd \\ \frac{1}{2}\left(r_{\frac{u}{2}} + r_{\frac{u}{2}+1}\right) & , \quad u\ even \end{cases}$$

$$= \frac{1}{2}\left(r_{\frac{6}{2}} + r_{\frac{6}{2}+1}\right)$$

$$= \frac{1}{2}(r_3 + r_4)$$

$$= \frac{1}{2}(7 + 23) = 15$$

Cross-Dating
Normalization of Coordinates

given

$Y = \langle 2 \quad 3 \quad 6 \rangle$

calculation

$$y_1^{std} = \frac{2 - \mathrm{mean}(Y)}{\sigma(Y)}$$

$$= \frac{2 - \frac{11}{3}}{\sqrt{\frac{1}{2}} \cdot \sqrt{\left(2 - \frac{11}{3}\right)^2 + \left(3 - \frac{11}{3}\right)^2 + \left(6 - \frac{11}{3}\right)^2}}$$

$$= \frac{-\frac{5}{3}}{\sqrt{\frac{1}{2}} \cdot \sqrt{\left(-\frac{5}{3}\right)^2 + \left(-\frac{2}{3}\right)^2 + \left(\frac{7}{3}\right)^2}}$$

$$\approx -0.801$$

Cross-Dating
POWERSET TABLE

given

year	score1	score2	score3
1992	0.1	0.2	0.3
1993	0.4	0.5	0.6
1994	0.7	0.8	0.9

calculation

year	score1	score2	score3	score1 + score2	score1 + score3	score2 + score3	score1 + score2 + score3
1992	0.1	0.2	0.3	0.3	0.4	0.5	0.6
1993	0.4	0.5	0.6	0.9	1.0	1.1	1.5
1994	0.7	0.8	0.9	1.5	1.6	1.7	2.4

Cross-Dating
Spearman Rank Correlation Coefficient

given

$\mathrm{V}^{\mathrm{a}} = \langle 4, 2, 3 \rangle$

$\mathrm{V}^{\mathrm{b}} = \langle 2, 5, 6 \rangle$

calculation

$$\varrho\left(\mathrm{V}^{\mathrm{a}}, \mathrm{V}^{\mathrm{b}}\right)$$

$$= \rho\left(r(\mathrm{V}^{\mathrm{a}}), r\left(\mathrm{V}^{\mathrm{b}}\right)\right)$$

$$= \rho\left((3{,}1{,}2), (1, 2, 3)\right)$$

$$= \frac{(1, -1, 0) \cdot (-1, 0, 1)}{2} = \frac{-1 + 0 + 0}{2} = -0.5$$

Cross-Dating
Tukey's Biweight Robust Mean

given

$\mathbf{V}_{\leq} = \{2, 3, 5\}_{\leq}$

$c_{tun} = 9$

calculation

$\tilde{v} = 3$

$\mathrm{MAD}_v = \mathrm{median}\{|2-3|, |3-3|, |5-3|\} = \mathrm{median}\{0, 1, 2\} = 1$

$$\zeta_1 = \frac{v - \tilde{v}}{c_{tun} \cdot \mathrm{MAD}_v + \varepsilon} = \frac{2-3}{9 \cdot 1 + 0.0001} = \frac{-1}{9.0001} \approx -0.111$$

$$\zeta_2 = \frac{3-3}{9 \cdot 1 + 0.0001} = 0$$

$$\zeta_3 = \frac{5-3}{9 \cdot 1 + 0.0001} = \frac{2}{9.0001} \approx 0.222$$

Cross-Dating
Tukey's Biweight Robust Mean

calculation

$$w(\zeta_1) = \begin{cases} (1-\zeta_1^2)^2 & ,|\zeta_1| \leq 1 \\ 0 & ,|\zeta_1| > 1 \end{cases} = \left(1 - \left(\frac{-1}{9.0001}\right)^2\right)^2 \approx 0.975$$

$$w(\zeta_2) = (1-0^2)^2 = 1$$

$$w(\zeta_3) = \left(1 - \left(\frac{2}{9.0001}\right)^2\right)^2 \approx 0.904$$

$$\bar{y} = \frac{w(\zeta_1)\cdot 2 + w(\zeta_2)\cdot 3 + w(\zeta_3)\cdot 5}{w(\zeta_1) + w(\zeta_2) + w(\zeta_3)} = 3.288936942$$

Sources

- [1] **Mattheis Alexander** «Cross-Dating of Intra-annual Wood Density Series».

  Master Thesis. 2018
- [2] **Affymetrix** «Statistical algorithms description document».

  Technical Paper. 2002, pp. 22–23
- [3] **Walz Guido**, ed. Lexikon der Mathematik: Band 3. Springer, 2017,
  pp. 98-99.doi: 10.1007/978-3-662-53502-8