

마라톤 코스 测定值의 統計的 分析

Statistical Analysis of Marathon Course Measurements

曹 圭 田* 李 榮 鎮**
Cho Kyu-Jon Lee Young-Jin

要 旨

마라톤과 같은 도로 경기 코스는 그 線形이 垂直·水平方向으로 항상 변화한다. 따라서 코스測定에는 檢定된 사이클에 의한 方法이 매우 효과적이며 測定值에는 적절한 統計處理가 수반되어야 한다. 이 論文은 서울 올림픽 마라톤 코스의 設定을 위하여 1986년 5월 18일에 사이클로 测定한 구간의 距離가 提示된 거리보다 짧아지지 않도록 하기 위하여 安全係數의 개념을 도입하고 이의 선택을 위한 統計學的 处理方法을 제시하였다. 距離算定의 결과는 같은 正確度로 再測하더라도 最小測定值보다 짧아지지 않을 信賴度가 99.95%임을 보여주고 있다.

ABSTRACT

The marathon course on road surfaces varies frequently in alignment horizontally and /or vertically. This fact compelled us to use a calibrated bicycle method for the course measurement, and to be required statistical approaches for data processing. This paper deals with the computation of the Seoul Olympic Marathon course lengths measured on May 18, 1986. The concept of safety factor against short course is analyzed for certification, and statistical methods are presented to compute an appropriate safety factor. The results of computation show that the best(lowest)actual measurement provides about 99.95% confidence that the combined lengths of all intervals will not be found short upon an equally accurate remeasurement.

1. 序 言

마라톤과 같은 도로경기의 코스는 水平 및 垂直方向에 따라 수시로 변화하고 있으므로 일반

적으로 수행되는 直線距離의 測定方法과는 큰 차이가 있다. 국제육상경기연맹 (IAAF)의 규정에 의하면 “마라톤 코스는 42195미터~42237미터 이어야 하며 最短코스가 되는 이상적인 선을 따라 测定되어야 한다”고^{2,3} 許容誤差의 限界와

* 京畿大學校 工科大學 助教授
** 漢陽大學校 大學院 博士課程

코스 設定方法을 명시하고 있다.

그러므로 마라톤 코스의 설정에는 “檢定된 사이클에 의한 코스測定法 (the calibrated bicycle method)”이 작업속도와 정확도 면에서 가장 유리한 방법이다.⁽⁴⁾ 이 방법은 보통 그림 1. 에서와 같은 과정으로 수행될 수 있으며 일정구간의 정확한 거리(基線)와 이 구간내에서 사이클의 회전수를 결정하기 위한 檢定作業을 필요로 한다. 그러나 温度, 바람등과 같은 기상조건의 변화, 사이클 주행자, 地表面의 상태 또는 타이어의 壓力 등에 의하여 사이클의 회전수가 변화하므로 코스길이의 算定에는 많은 양의 統計處理를 수반한다.

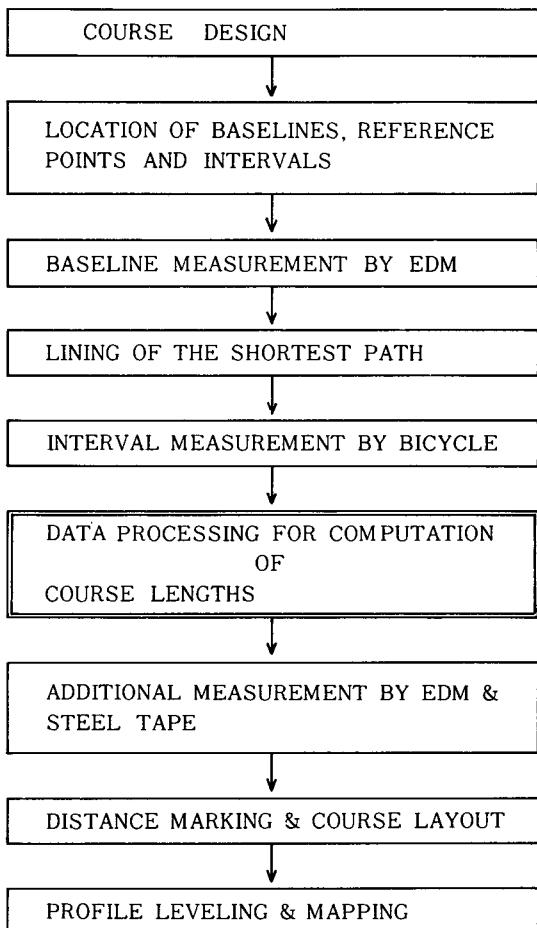


그림 1. 사이클 측정법에 의한 코스설정

이 論文에서는 이와 같은 사이클측정법으로 측정한 구간의 距離算定을 다루고 있으며 回歸分析과 安全係數(safety factor against short course)의 개념을 도입하여 그 結果值를 处理하고 있다.

距離의 算定에 사용된 데이터는⁽¹⁾ 서울 올림픽 마라톤 코스의 設定을 위해 1986년 5월 18일에 测定한 것으로서 코스선상에 설치한 12개의 기선과 33개의 測定點을 이용하였다. 또한 사이클로 주행한 코스의 측정선은 最短距離가 되도록 이상적인 선을 따라 12개 기선과 19개 구간에 백색선으로 도색한 후 총 13대의 사이클로 주행하였다 (그림 2. 참조) 사이클의 앞바퀴 축에는 회전수를 읽을 수 있는 존스 카운터 (Jones Counter)가 부착되었다.

2. COUNTS/KM의 決定

사이클로 코스선상에 설치된 基線을 주행한다면 이때 읽은 카운트 수로 부터 COUNTS/KM (KM 당 카운트 수)를 구할 수 있다. 이와 같이 檢定된 사이클을 이용하여 距離를 알고자 하는 두 점간을 주행하고 읽은 카운트 수를 COUNTS/KM로 나누면 距離를 구할 수 있다. 이 原理가 檢定된 사이클에 의한 코스測定法이다.

그러나 마라톤 코스와 같이 코스가 길고 線形狀態와 條件이 수시로 변화하는 경우에는 한 개의 基線에 의해 檢定한 COUNTS/KM를 전 코스에 똑같이 적용하기가 곤란하다. 따라서 가능한 한 많은 수의 基線을 設置하여 모든 基線에서 COUNTS/KM를 구하고 이로 부터 未知區間의 距離算定에 적용할 COUNTS/KM를 구해야 한다. 이를 위해서 回歸分析 (regression analysis)이 필요하다.

Y_i 를 基線 i 에서의 COUNTS/KM, X_i 를 基線 i 까지의 누적 카운트수(elapsed counts)라 하면 3차 多項式의 線形모델을 다음과 같이 나타낼 수 있다. ^(6, 7, 8, 9)

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i \quad (1)$$

여기서 X_i 의 次数가 多項式의 차수를 나타내며 β_i 는 계수이다. β_i 의 수를 p , 基線의 수를 n 이

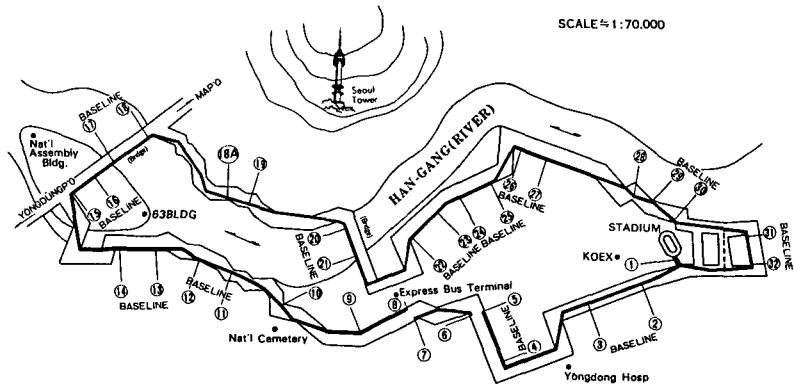


그림 2. 코스 측정의 개요도⁽¹⁾

라고 하면 식(1)은 전 基線에 대해 다음의 매트릭스 형태로 표현 가능하다.

$$Y = X\beta + \varepsilon \quad (2)$$

$(n \times 1) \quad (n \times p) \quad (p \times 1) \quad (n \times 1)$

이 모델식에 最小제곱법을 적용하면 β 의 추정치인 b 를 구할 수 있다.

$$(X^T X) b = X^T Y \quad (3)$$

$$b = (X^T X)^{-1} X^T Y$$

이때의 추정치 $\hat{Y} = Xb$ 로 구할 수 있으므로 残差는 다음 식(4)에 의해 표현된다. 그러므로 이를 이용하여 ANOVA분석이나 t -檢定에 의해 모델식의 適合度를 알 수 있다.

$$e = Y - \hat{Y} \quad (4)$$

식(2)와 식(3)에 의해 추정된 모델식을 이용하여 미지구간에 대한 누적 카운트수 X_h 에서의 추정 COUNTS/KM와 分散은 다음과 같이 주어진다.^[7,8]

$$\hat{Y}_h = X_h b \quad (5)$$

$$\delta_{\hat{Y}_h}^2 = \text{MSE} (1 + X_h (X_h^T X_h)^{-1} X_h^T) \quad (6)$$

여기서 $\text{MSE} = e^T e / (n - p)$ 이며 식(6)은 모델식 자체의 Y 에 대한 分散과 \hat{Y}_h 의 分散이 합해진 형태이다. 그러므로 F 分布를 이용한 Working-Hotelling 방법이나 t 분포에 의한 Bonferroni 방법에 의해 $E\{\hat{Y}_h\}$ 의 信賴領域을 추정할 수 있다.

이 論文에서는 1, 2, 3차 多項式과 두기선 平均法 및 重量平均法을 적용하여 未知區間의 COUNTS/KM를 추정하였다. 두 기선 平均法은

인접한 두 기선의 COUNTS/KM를 평균하여 未知區間에 적용하는 방법이며, 重量平均法은 각 基線의 距離를 重量으로 고려하여 계산하는 방법으로서 사이클마다 동일한 COUNTS/KM 값이 전구간에 적용된다.

3. 安全係數의 導入

앞서 언급한 바와 같이 COUNTS/KM가 결정되면 주행한 각 사이클에 대하여 각 구간의 距離를 計算할 수 있다. 한 구간을 13대의 사이클이 주행하였다면 13개의 距離에 대한 最確值를 결정할 수 있으나 마라톤 코스의 규정을 만족시킬 수 있도록 安全係數(safety factor against short course)를 고려하여야 한다.

이 安全係數가 너무 작으면 檢測距離가 公式 거리보다 短아질 염려가 있고 반대로 安全係數가 너무 크면 코스길이가 불필요하게 길어진다. 이러한 安全係數의 크기는 測定者의 경험과 속력도에 따라 다르므로 이의 적절한 선택이 중요하다.

檢測한 결과가 공식적으로 확정된 距離보다 短아질 確率을 구해 보자. 다시 말해서 공식적으로 인정된 거리를 L , 檢測한 距離를 V 라고 할 때 V 가 L 보다 작을 確率을 구하는 문제이다.

측정치 A_1, A_2, \dots, A_N 이 평균 μ_0 , 표준편차 σ_0 를 갖는 독립확률변수라 하고 檢測值 V 가 평균 μ 표준편차 σ 인 확률분포를 이룬다고 가정하

면 공식거리 L 은 안전계수 S 를 도입하여 다음과 같이 표현할 수 있다.^(5,6)

$$L = \frac{A_1 + A_2 + \dots + A_N}{N} - S \quad (7)$$

만약 공식거리 L 을 구하기 위해 S 를 $K \cdot \hat{\sigma}$ 로 정의하면,

$$L = \bar{A} - K \cdot \hat{\sigma} \quad (8)$$

여기서 \bar{A} 와 $\hat{\sigma}$ 는 측정치 A 의 표본평균 및 표준편차를 나타낸다.

그러므로 檢測值 A 가 L 보다 작지 않을 확률 p 에서 K 값을 구해야 한다. 즉,

$$\text{Prob } \{V\} < \text{Prob } \{L\} \quad (9a)$$

$$\text{Prob } \{V\} < \text{Prob } \left\{ \frac{A_1 + A_2 + \dots + A_N}{N} - S \right\} \quad (9b)$$

$$\text{Prob } \left\{ V - \frac{A_1 + A_2 + \dots + A_N}{N} \right\} < \text{Prob } \{-S\} \quad (9c)$$

따라서 다음과 같은 새로운 확률변수를 도입할 수 있다.

$$X = V - \bar{A} \quad (10)$$

이때 X 는 확률변수 V 의 분포를 결정하는데 t -分布式을 사용할 수 있다.^(6,8)

$$T = \frac{X - \mu_x}{\sigma_x} \quad (11)$$

여기서 X 의 확률분포는 다음의 평균과 표준편차를 갖는다.

$$\mu_x = \mu - \mu_o \quad (12)$$

$$\sigma_x = \sqrt{\sigma^2 + \frac{\sigma_o^2}{N}} \quad (13)$$

L 과 V 의 평균과 표준편차를 같다고 가정하면 $\mu = \mu_o$, $\sigma = \sigma_o$ 이므로

$$\mu_x = 0, \quad \sigma_x = \sqrt{\frac{N+1}{N}} \sigma_o \quad (14)$$

가 성립한다. 식(11)과 식(14)로 부터 다음식이 유도된다. 즉, σ_o 대신에 $\hat{\sigma}$ 을 사용하면,

$$T = \frac{V - \bar{A}}{\sqrt{\frac{N+1}{N}} \hat{\sigma}} = \sqrt{\frac{N}{N+1}} \left(\frac{V - \bar{A}}{\hat{\sigma}} \right) \quad (15)$$

여기서 $(V - \bar{A})/\hat{\sigma}$ 는 $(N-1)$ 의 자유도를 갖는 t -분포를 형성한다. 또한 식(9c)와 식(15)로

부터,

$$F(t) = \text{Prob } (X < -S) = T \left\{ \sqrt{\frac{N}{N+1}} \left(\frac{-S}{\hat{\sigma}} \right) \right\} \quad (16)$$

이때의 신뢰구간은 다음과 같이 된다.

$$1 - C = T \left\{ \sqrt{\frac{N}{N+1}} \left(\frac{-S}{\hat{\sigma}} \right) \right\}$$

$$C = T \left\{ \sqrt{\frac{N}{N+1}} \left(\frac{S}{\hat{\sigma}} \right) \right\} \quad (17)$$

식(17)에서 S 를 구하기 위해 변형하면,

$$S = T^{-1} \left\{ \sqrt{\frac{N+1}{N}} \hat{\sigma} \right\} \quad (18)$$

결과적으로 식(18)에서 사용할 K 값은 다음 식으로 주어진다.

$$K = \sqrt{\frac{N+1}{N}} t_p^{(N-1)} \quad (19)$$

여기서 $t_p^{(N-1)}$ 은 자유도가 $N-1$ 인 t -분포의 $p\%$ 의 영역을 나타낸다.

예를 들어 $N=13$ 이고 $P=99\%$ 인 경우에는 $t_{0.99}^{(12)} = 2.68$ 이므로 $K=2.78$ 의 값을 갖는다.

이상의 경우는 平均을 중심으로 하는 방법을 설명하였으나 中央値(median)을 고려하는 경우에는 좀더 복잡한 것으로 알려져 있으며, 사용의 편리성 때문에 平均을 고려하는 방법이 信賴度문제에 보편적으로 사용되고 있다.⁽⁵⁾

4. 電算處理

마라톤 코스의 設定을 위하여 1986년 5월 18일에 13대의 사이클로 주행하고 33개의 測定點에서 읽은 Raw Jones Counts와 사전에 측정된 각 基線의 距離를 기본 데이터로 하여 코스길이를 計算하였다.

각 구간의 COUNTS/KM에 대한 最確值의 추정과 구간거리의 計算 및 統計處理를 위하여 OLYM 1. BAS, OLYM 2. BAS, OLYM 3. BAS라고 하는 세 가지의 프로그램이 개발 되었으며 距離算定을 위한 흐름도는 그림 3.에서 보여주고 있다.

OLYM 1. BAS는 1개의 주프로그램과 4개의 부프로그램으로, OLYM 2. BAS는 1개의 주프

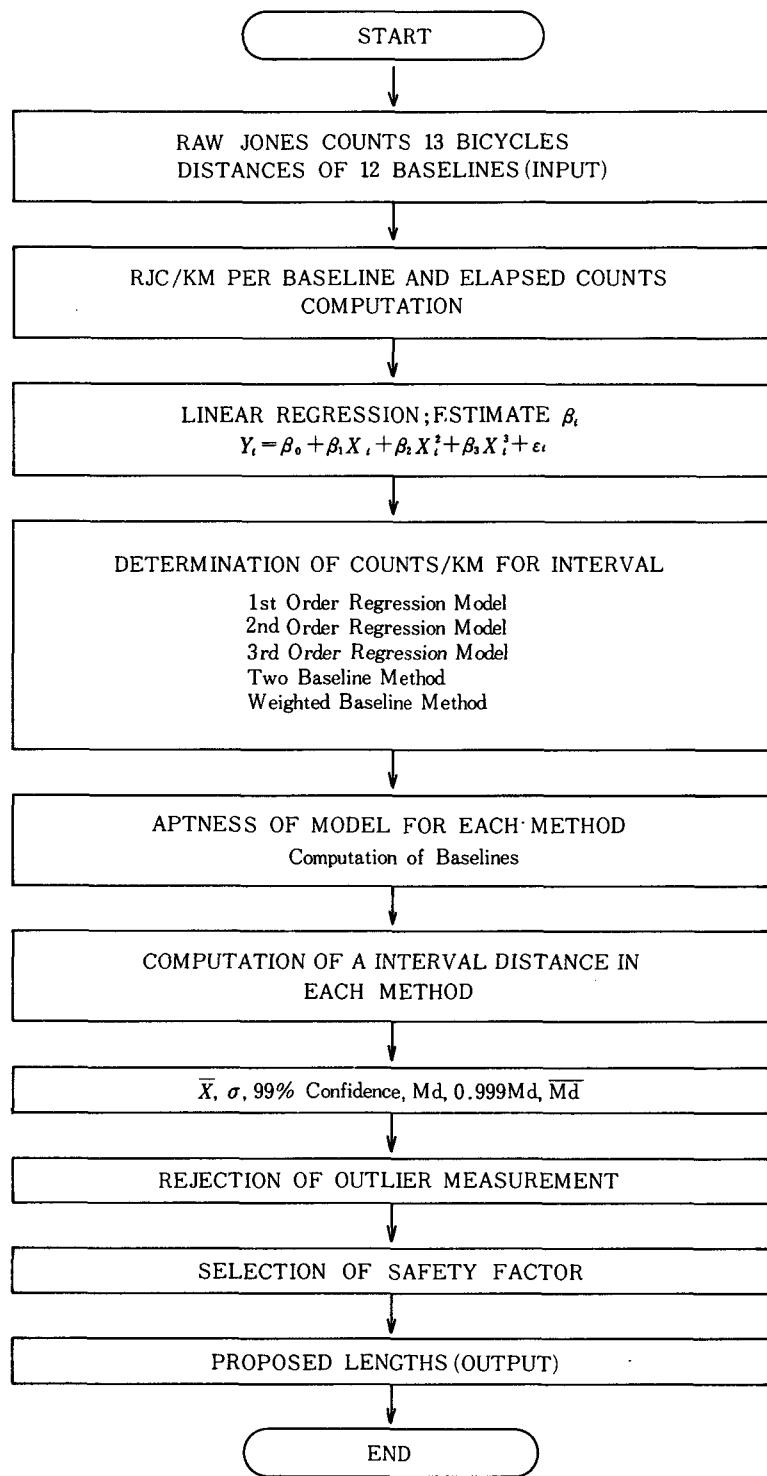


그림 3. 거리계산을 위한 흐름도

로그램과 부프로그램으로 각각 구성되었다. 프로그램 작성은 BASIC언어를 사용하였으며 16비트 소형 컴퓨터를 사용하여 計算하였다.

프로그램 1에서는 基線測定值 파일과 13개의 사이클에 대한 카운트 파일로 부터 사이클별로 각 구간과 기선의 COUNTS/KM 추정치를 구하고 있다. 프로그램 2에서는 프로그램 1에서 얻어진 출력파일을 재편집하여 사이클별 구간거리와 구간별 計算距離를 구하고 각 구간에 대한 통계계산이 실시되며, 프로그램 3에서는 프로그램 1, 2에서 얻어진 출력파일을 각 기선별로 재편집하여 計算된 基線距離와 실제 距離를 비교하여 각 回歸式의 信賴度를 분석하고 있다.

다음 表-1.에서는 사이클별 구간거리의 計算例를 보여주고 있다. 여기서 각 분석방법별 中央값의 평균은 $Md=3086.45$ 이며 다음 장에서 最確值로 사용하고 있다. 또한 4가지 방법에 따른 사이클 13대의 최소값은 3085.85(밀줄침) 이므로 이를 最小測定值 (lowest actual measurement)로 고려할 수 있다.

表-1. 구간거리의 계산예(구간 R3 - R4)

RIDER	LINEAR	2ND-ORDER	3RD-ORDER	TWO-BASE	WGT MEAN
C 1	3087.03	3087.36	3087.34	3087.21	3087.96
C 2	3086.04	3086.41	3086.43	3086.51	3085.71
C 3	3086.26	3086.48	3086.52	3086.79	3076.51
C 4	3086.84	3087.21	3087.26	3087.41	3083.46
C 5	3086.74	3086.72	3086.73	3087.00	3083.01
C 6	3086.12	<u>3086.12</u>	<u>3086.08</u>	<u>3086.00</u>	3082.75
C 7	<u>3085.85</u>	3086.19	3086.21	3086.32	3094.33
C 8	3086.32	3086.43	3086.44	3086.43	3084.75
C 9	3086.15	3086.34	3086.35	3086.42	3094.19
C 10	3086.79	3087.19	3087.15	3087.14	3103.33
C 11	3086.20	3086.47	3086.51	3086.55	3085.26
C 12	3086.69	3086.82	3086.86	3087.47	3094.69
C 13	3086.08	3086.31	3086.27	3086.26	3084.90
MEDIAN	3086.26	3086.47	3086.51	3086.55	3085.27
MEAN	3086.39	3086.62	3086.63	3086.73	3087.76
ST. DEV	0.38	0.41	0.41	0.47	7.04
99%CON	3085.35	3085.49	3085.46	3085.42	3068.17
0.999M	3083.17	3083.38	3083.43	3083.46	3082.18

* $Md=3086.45$

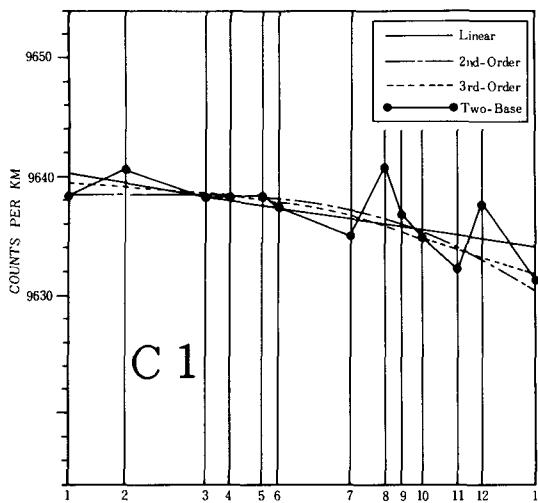


그림 4 (a). 기선에서의 실측 COUNTS/KM와 회귀식의 예

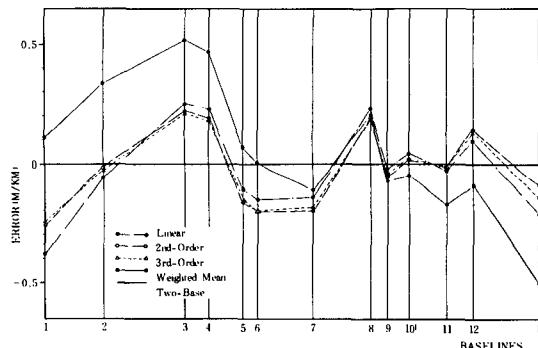


그림 4 (b). 적합도 분석

表-2. 각 방법별 평균오차

	Linear	2nd-order	3rd-order	WGT mean
mean	-0.018	-0.018	-0.018	0.055
S. D.	0.187	0.164	0.162	0.278

表-3. 분석방법별 신뢰영역

Method	Linear	2nd-order	3rd-order	Two-Base	Two-Base	Remarks
\bar{X}_{13}		31166.16	31165.90	31165.88	31166.14	
$t_p^{(12)}$	K\ S. D.	2.87	2.83	2.85	2.94	
p=99%	2.78	31158.18	31158.03	31157.96	31157.97	99% (SUM)
p=99.5%	3.18	31157.03	31156.90	31156.82	31156.79	99.5 (SUM)
p=99.9%	4.08	31154.45	31154.35	31154.25	31154.15	
p=99.95%	4.48	31153.30	31153.22	31153.11	31152.97	
p=99.99%	5.50	31150.38	31150.33	31150.21	31149.97	
p=99.995%	5.92	31149.17	31149.15	31149.01	31148.74	
p=99.999%	6.85	31146.50	31146.51	31146.36	31146.00	

* t_p : 1-Side interference

5. 2 安全係數의 선택

안전계수를 선택하는 방법으로는 다음 세 방법을 고려할 수 있다. 즉, 제안거리에 따라,

- ① $0.999 \times Md.$ (중앙값)을 택할 경우
- ② 信賴度 99%의 값을 택할 경우
- ③ 最小測定值를 택할 경우.

우선 信賴度에 따른 安全係數를 선택하기 위하여 19개 구간의 총거리와 이에 상응하는 평균값(\bar{X}), 표준편차(SD) 및 각 방법별 99%~99.9 99%의 신뢰영역의 계산결과가 表-3.에 제시되었다.

또한 表-4.에는 19개 구간에 대한 99% 신뢰영역의 길이가 해석 방법별로 나타냈다. 여기서 SUM(99%) 중 가장 낮은 값은 31147.67 미터이며, 이 값은 99% (SUM)보다 10.29 미터 작고, 총길이에 대하여 99.997%의 信賴度를 갖는다. (表-3. 참조)

距離를 결정하기 위한 또 다른 방법으로는 각 구간별 最小測定值(best or lowest actual measurement)를 사용하는 방법이 고려될 수 있다. 表-5.는 각 구간에 대한 最小測定值를 나타내고 있으며 최저값을 고딕으로 나타냈다. SUM(BEST)¹중 가장 낮은 두 기선 평균법의 값은 99% (SUM)보다 2.85미터 낮으며 檢測시 검측된 거리보다 작을 확률 99.7% 이상의 信賴度를 갖는다. 또한 SUM(BEST)²은 99.95%의 信賴度를 갖는다.

表- 4. 구간별 99% 신뢰값이($X_{13} - 2.78 S.D.$)

Terminus	Linear	2nd-order	3rd-order	Two-Base
R 1-R 2	1240.46	1240.56	1240.61	1240.51
R 3-R 4	3085.35	3085.49	3085.48	3085.42
R 6-R 7	1430.50	1430.53	1430.49	1430.42
R 8-R 9	1453.59	1453.60	1453.50	1453.42
R 9-R10	1785.97	1785.97	1785.85	1785.74
R 10-R11	1273.16	1273.16	1273.17	1273.02
R 12-R13	1640.79	1640.74	1640.71	1639.91
R 14-R15	2452.12	2451.98	2451.87	2451.74
R 16-R17	397.70	397.67	397.68	397.75
R 18-R18A	2350.04	2349.94	2349.92	2350.31
R 18A-R19	567.40	567.39	567.39	567.48
R 19-R20	2905.04	2904.88	2904.86	2904.80
R 21-R22	2101.67	2101.62	2101.67	2101.44
R 23-R24	848.24	848.21	848.23	848.00
R 25-R26	1335.43	1335.38	1335.40	1335.52
R 27-R28	1855.94	1855.94	1855.91	1855.65
R 28-R29	848.88	848.89	848.89	848.84
R 30-R31	1703.32	1703.33	1703.31	1703.15
R 32-R 1	1874.41	1874.53	1874.53	1874.55
SUM (99%)	31150.01	31149.81	31149.47	31144.67

表- 5. 최소측정치

Terminus	Linear	2nd-order	3rd-order	Two-Base
R 1-R 2	1240.67	1240.83	1240.86	1240.84
R 3-R 4	3085.85	3086.12	3086.08	3086.00
R 6-R 7	1430.73	1430.75	1430.70	1430.69
R 8-R 9	1453.72	1453.73	1453.66	1453.62
R 9-R10	1786.15	1786.15	1786.07	1785.97
R 10-R11	1273.28	1273.28	1273.60	1273.19
R 12-R13	1640.94	1640.91	1640.82	1640.01
R 14-R15	2453.24	2453.23	2542.98	2452.84
R 16-R17	397.89	397.87	397.88	397.96
R 18-R18A	2350.85	2350.69	2350.65	2351.37
R 18A-R19	567.48	567.48	567.48	567.65
R 19-R20	2905.41	2905.17	2905.13	2905.72
R 21-R22	2102.00	2101.99	2102.00	2110.67
R 23-R24	848.35	848.33	848.36	848.21
R 25-R26	1335.64	1335.60	1335.66	1335.75
R 27-R28	1856.20	1856.20	1856.22	1856.00
R 28-R29	849.10	849.10	849.13	849.01
R 30-R31	1703.54	1703.53	1703.55	1703.46
R 32-R 1	1875.49	1875.00	1875.99	1875.14
SUM(BEST) ²	31156.53	31156.53	31155.82	31155.10

$$\text{SUM (BEST)}^2 = 31152.92$$

그림5.는 이상의 방법들에 대하여 각종 결과치의 상호관계를 그래프로 표시한 것이다. 이 그림에서는 99% (SUM)의 표시선 아래에 있는 모든 값들이 99.7%의 신뢰도를 갖고 있다. 다시 말해서 공인된 거리가 檢測距離보다 클 확률이 99.7%이다. 특히 여기서 주목할 사항은 중앙값 (M_d)의 0.999 배값이 99.999%의 信賴度를 갖는다는 사실이다.

결론적으로 모든 것을 종합해 볼때 99.95%의 信賴度를 갖는 最小測定值(SUM(BEST)²)를 제안거리(proposed length)로 채택하는 것이 타당할 것이다.

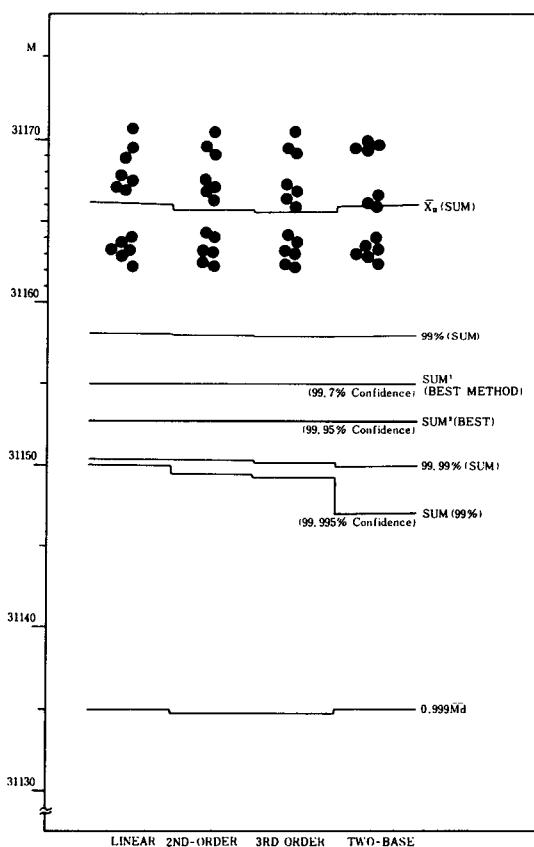


그림5. 각 신뢰길이간의 관계

6. 結論

서울 올림픽 마라톤 코스의 사이클測定區間에 대한 統計分析 결과 다음과 같은 結論을 얻었다.

1. 最小測定值를 사이클로 测定한 구간의 距離로 확정한다면 檢測에서 이보다 더 짧아지지 않을 信賴度는 99.95%이다.
2. 1986년 5월 18일의 测定과 같은 正確度로 再測定을 실시하더라도 再測定값이 最小測定值보다 작을 確率은 약 1/2,000이다.

謝辭

본 研究는 韓國測地學會에서 수행한 「마라톤 및 경보코스 설정에 관한 조사연구 및 측정」⁽¹⁾의 일부이며, 이를 가능케 한 서울올림픽대회 조직위원회와 韓國測地學會에 謝意를 표한다.

参考文獻

1. 서울올림픽대회 조직위원회·한국측지학회, MARATHON/RACE WALKS COURSE: Study and Measurement, report of the 1988 Olympic Marathon Measurement, 1986.
2. IAAF, Official Handbook 1985/1986.
3. IAAFF, Guidelines for the Conduct of Road Races, IAAF, 1984.
4. TAC Road Running Technical Committee, Road Race Course Measurement and Certification Procedures, The Athletics Congress of the USA, 1985.
5. Los Angeles Olympic Organizing Committee, MARATHON/RACE WALKS, report of the 1984 Olympic Marathon Measurements, 1983.
6. Hogg, R. V. and A. T. Craig, Introduction to Mathematical Statistics (4th ed.), Macmillan, 1983.
7. Neter, J. and W. Wasserman, Applied Linear Statistical Models, Richard D. Irwin, Inc., 1974.
8. Kleinbaum, D. G. and L. L. Kupper, Applied Regression Analysis and Other Multivariable Methods, Duxbury Press, 1978.
9. Taylor, J. R., An Introduction to Error Analysis; the study of uncertainties in physical measurements, University Science Books, 1982.