신뢰성응용연구 신뢰성기술
제11권, 제3호. pp.213-224, 2011

상수도용 Pipeline의 누수고장 자료 분석

나명환·함상민

전남대학교 통계학과

Data Analysis of First Leak Time of Water Pipeline

Myung Hwan Na·Sang Min Ham

Department of Statistics, Chonnam National University

Abstract

In this paper, we analyze statistically the data set of first leak time of water pipeline. We classify first the leak time data by pipe type, location, diameter of pipe and, length of pipe. We perform the analysis of variance to indicate that there are significant difference of mean of the time between levels of the factor and also compare the distribution of levels using the multiple box-plot. When there are the difference of the mean, we perform the least significant test to find out what levels of the facor has a different mean.

Key words; The first leak time, Pipeline, Analysis of Variance, Box-plot

논문접수일 : 2011년 04월 13일 논문수정일 : 2011년 09월 16일 게재확정일 : 2011년 09월 17일

1. 서 론

본 논문에서는 상수도를 공급하는 외국의 회사에서 매립되어 있는 상수도용 pipeline이 파괴되어 새는 경우 pipe의 직경, 길이 종류, 매설 위치에 따라서 처음으로 새는 시간 (first leak time)자료를 이용하여 여러 가지 통계적 분석을 하고자 한다. 여기서 첫 번째 누수 시간만을 고려하는 이유는 어느 특정 부위에 누수가 발생하면 그 위치를 수리하면서 수리해야 할 부분을 포함한 일정 부분을 sealing을 함으로써 재차 누수가 발생하지 않는다고 판단되어 첫 번째 고장시간만을 고려하여 분석을 한다. 초기누수현상이 많이 발생하는데 이 부분에 대한 보다 정밀한 분석을 통해 자료를 제거하고 분석을 하든지 아니면 초기누수현상이 여러 경제적인 면이나 사회적인 또는 기업차원의 이미지에 영향을 미친다고 판단이 되면 보다 정밀한 분석을 통해 그 원인을 밝혀내고, 차후에 이러한 초기 누수현상이 없도록 계획을 세우는 것이 중요하겠다.

본 자료분석을 통해 pipe의 type에 따라서 또는 직경 및 길이에 따라서 처음누수시간의 차이가 있는지를 확인하고 향후 새로운 상수도용 pipeline을 설계할 때 이점을 감안하여 pipeline 매설 계획을 수립하는 것을 목적으로 하고 있다. 또한 매설지역에 따라 누수시간에 차이가 있는 경우 그 매설지역에 대한 토질 및 환경에 대한 특이성을 분석하고자 하며, 그 지역의 환경과 맞는 형태의 pipeline을 설계하고자 한다.

이와 같이 pipe의 type, 매설지역, 직경, 길이에 따라 자료를 구분하여 각 요인에 따라 차이가 있는지를 알아보기 위해 분산분석을 실시하였다. 분산분석에 대한 자세한 이론은 배종성 외(2011)과 Mickey(2004)를 참조바란다. 분산분석을 이용한 품질관리 및 신뢰성 자료 분석을 한 연구는 매우 다양하다 할 수 있으며 그 응용논문은 김형욱(1982), 윤중범(1988), 전세창과 손환규(2011) 등을 참조하기 바란다. 분산분석 결과에 유의한 차이가 있는 경우 이를 바탕으로 최소유의차 검정을 실시하였으며, 또한 다중상자그림을 그려 비교 분석하였다. 최소유의차 검정은 박성현(2009)을 참조바란다. 모든 통계적 분석은 통계 패키지 Minitab을 이용하여 분석하였으며 Minitab에 대한 사용설명은 박성현, 김종욱(2010)을 참조바란다. 이렇게 비교분석 자료를 바탕으로 분석 결과를 보면 type, 매설 지역, 직경, 길이에 따라 누수고장시간에 유의한 차이가 있는 것을 알 수 있다.

2절에서는 분산분석에 대하여 간단히 설명을 하였으며, 각 자료의 분석결과는 3절에 수록하였으며 4절에는 결론과 제언을 하였다.

2. 자료분석을 위한 통계적 방법

Pipeline의 누수고장자료를 분석하기 위하여 일원배치법을 이용하였다. 반복수가 다른 일원배치법의 모형은 다음과 같다.

$$x_{ij} = \mu_i + \epsilon_{ij}\,, i = 1,2,...,a, j = 1,2,...,n_i$$

여기서 x_{ij}는 i번째 수준의 j번째 측정치의 값, μ_i는 i번째 수준의 모평균, ϵ_{ij}는 실험의 오차항에 해당되는 확률변수로 평균이 0이고 분산이 σ^2인 정규분포 $N(0,\sigma^2)$을 따르고 서로 독립이다. 총 실험 횟수를 $N=\sum_{i=1}^{a} n_i$이라 한다.

일원배치 분산분석의 귀무가설과 대립가설은 다음과 같다.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \qquad H_1 : \text{적어도 하나의 } \mu_i \text{는 다르다.}$$

이제 분산분석법의 과정을 소개해보자.

각 측정치의 편차 $(x_{ij}-\overline{x})$를 다음과 같이 분해할 수 있다.

$$x_{ij} - \overline{x} = (\overline{x_i} - \overline{x}) + (x_{ij} - \overline{x_i}).$$

여기서 $\overline{x_i}$는 i번째 수준에서 측정치 평균, $\overline{x}$는 측정치 전체 평균을 나타낸다.

즉 전체 편차는 수준(집단)간 편차와 수준(집단)내 편차의 합으로 나누어진다.

$$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(\overline{x_i}-\overline{x})(x_{ij}-\overline{x_i}) = \sum_{i=1}^{a}(\overline{x_i}-\overline{x})(\sum_{j=1}^{n_i}(x_{ij}-\overline{x_i})) = \sum_{i=1}^{a}(\overline{x_i}-\overline{x}).0 = 0$$

이 되기 때문에 양변을 제곱하여 합하면 다음과 같다.

$$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\overline{x})^2 = \sum_{i=1}^{a} n_i(\overline{x_i}-\overline{x})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(x_{ij}-\overline{x_i})^2$$

총편차제곱합(SST) = 집단간 편차제곱합(SS_A) + 집단내 편차제곱합(SSE)

SS_A와 SSE를 각각의 자유도 인 $a-1$과 $N-a$로 나누어 주면 이를 각 집단간 편차제곱합의 평균(MS_A)과 집단내 편차 제곱의 평균(MSE)이라 한다.

여기서 SS_A 및 SSE의 자유도가 $a-1$과 $N-a$인 것은 표본의 분산

$$s^2 = \sum_{i=1}^{a}(x_i-\overline{x})^2/(a-1)$$

로 계산되는 것과 마찬가지로 생각하면 된다.

귀무가설 H_0가 참이면

$$F = \frac{MS_A}{MSE} \sim F(a-1,\ N-a)$$

가 성립한다. H_1이 참일 때 F통계량은 1보다 큰 값을 가지려는 경향이 있다.
따라서 $F > F_\alpha(a-1,\ N-a)$이면 H_0를 기각하고, 그렇지 않으면 H_0를 기각하지 못한다.

<표 1> 분산분석표

요인	제곱합	자유도	평균제곱	F
인자 A	SS_A	$a-1$	MS_A	MS_A/MSE
오차	SSE	$N-a$	MSE	
합계	SST	$N-1$		

이와 같이 F값을 얻기 위한 과정을 <표 1>과 같이 작성한 것을 분산분석표(ANOVA table)라 한다.

3. 실증분석

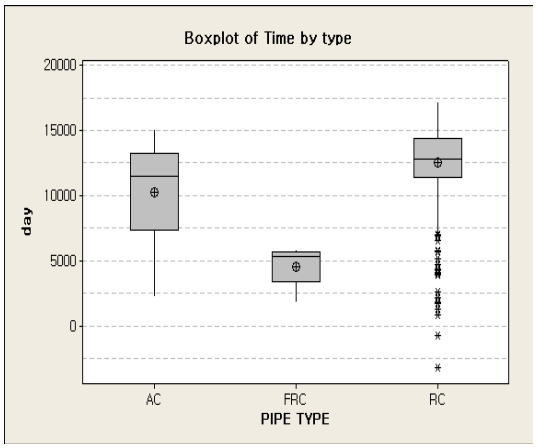
3.1 Pipeline 자료의 type별 분석

본 장에서는 상수도를 보급하는 회사에서 매립된 pipe가 매립위치에 따라 첫 번째로 pipe가 새는(leak)시간을 기록하여 얻은 자료를 이용하여 자료분석을 하였다. 얻어진 자료는 파이프의 type별로 FRC, AC, RC의 3종류이며, 각 type별로 첫 번째 고장시간과 자료의 수는 다음 <표 2>과 같다.

<표 2> Pipe Type별 first leak time(단위: day)

Pipe Type	Frist Leak Time	자료수
FRC	5341 … 5642	14
AC	11793 … 5387	56
RC	8704 … 9297	1061

먼저, 누수시간의 분포를 파악하기 위하여 누수시간에 대한 다중 상자그림을 그린 결과가 <그림 1>이다.



<그림 1> pipe의 type별 다중 상자그림

<그림 1>을 보면 3가지 type에 대한 분포에 차이를 보이고 있으며, 그 차이는 AC와 RC는 적은 반면 RC와 FRC 에서는 큰 차이를 보이고 있다. 또한 특이하게도 RC type에서 유독 많은 이상치를 보이고 있으며, RC, AC, FRC 순으로 frist leak time이 큰 것을 알 수 있다. Pipe의 type별로 frist leak time에 통계적으로 유의한 차이가 있는지 알아보기 위하여 분산분석을 실시하여 정리한 것이 <표 3>이다.

<표 3> Pipe의 Type별 ANOVA Table

요인	제곱합	자유도	평균제곱	F value	P value
모형	1,172,057,563	2	586,028,781	89.62	<.0001
오차	7,317,296,938	1,119	6,539,139		
Total	489,354,501	1,121			

<표 3>의 분산분석 결과를 보면, type별로 누수시간의 유의한 차이를 나타내는 통계량값은 89.62이며, 이 통계량의 P값(P-value)이 0.001보다 작다. 따라서 유의수준 1%에서 pipe의 type별로 frist leak time에 유의 한 차이가 있음을 알 수가 있다. 또한, 어떤 type에서 유의한 차이가 있는지 알아보기 위해 최소유의차 검정(LSD)을 실시한 결과 RC, AC, FRC 모두 서로 차이를 보이고 있으며 RC와 FRC 간의 차이가 가장 심한 것으로 보이고 있다.

<표 4> LSD 검정결과

type Comparison	Difference Between Means	95% Confidence Limits	
RC - AC	2359.0	1670.9	3047.0
RC - FRC	8080.1	6730.2	9429.9
FRC - AC	5721.1	-7220.4	-4221.9

따라서 pipe의 type에 따른 누수시간은 서로 유의한 차이가 있으면 그 크기 순서는 FRC, AC, RC순임을 알 수 있다. 이후 분석은 각 타입별로 frist leak time에 대한 분석 방법이 같지 때문에 가장 많은 자료를 가지고 있으며, 가장 높은 frist leak time을 가지는 RC Type으로 분석을 실시하였다.

3.2 Pipeline의 매설 위치에 따른 분석

RC Type은 총 8 곳의 위치에 매설 돼 있어 편의상 A, B, …, H로 구분하여서 자료를 분석하였다. 정리된 자료는 <표 5>와 같다.

<표 5> RC type의 pipeline 매설 위치에 따른 frist leak time

(단위: day)

위치	Frist Leak Time	자료수
A지역	9958 9304 9297 8733 8968 … 9723 8479 11481 12381 8478	18
B지역	12755 15655 12389 14998 … 13940 12900 13282 13429 12816	52
C지역	4310 3966 4695 12197 11857 … 11424 15453 15753 11456 15872	136
D지역	14002 15083 15508 12186 … 12093 13289 14019 15001 15006	379
E지역	13627 13527 10129 11976 … 11991 11647 12444 12041 1056 4100	81
F지역	13636 13536 12142 12342 … 15047 13648 15502 12843 13233	290
G지역	12441 10315 10911 12080 … 12078 12150 12577 12427 12890	101
H지역	9906 12799 12364 12393	4

지역 별로 차이가 있는지 알기 위하여 분산 분석을 하여 <표 6>과 같은 결과를 얻었다. P-value가 0.001보다 작은 값으로 유의 수준 1%에서 지역별로 차이가 있음을 알 수가 있다.

<표 6> Location에 의한 ANOVA Table

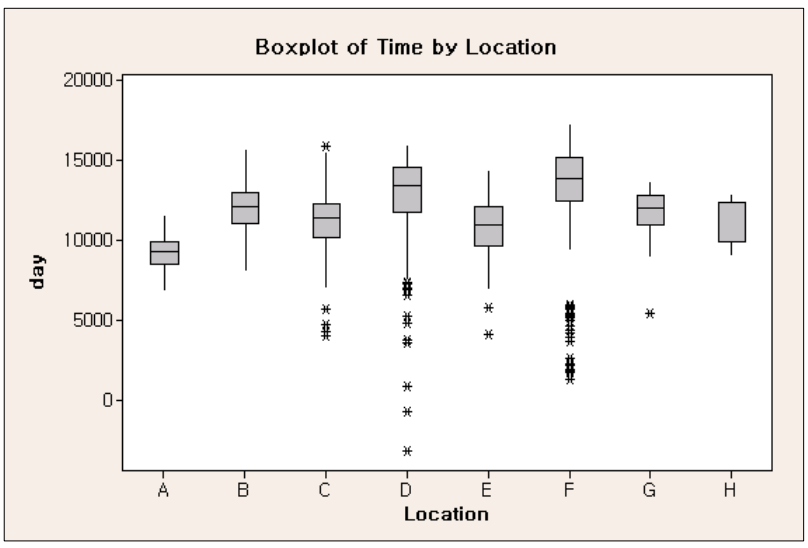
요인	제곱합	자유도	평균제곱	F value	P value
모형	1316643307	6	219440551	43.34	<.0001
오차	316172449	1050	5063021		
Total	6632815756	1056			

또한 지역별로 어떠한 차이가 있는지 보기 위해 LSD검정을 실시하였다.

<표 7> LSD 검정결과

lot Comparison	Difference Between Means	95% Confidence Limits	
F - D	546.2	201.7	890.7
F - B	1638.9	974.0	2303.8
F - G	1814.3	1304.2	2324.4
⋮			
A - G	-2631.3	-3760.9	-1501.7
A - C	-1793.2	-2900.6	-685.8
A - E	-1575.4	-2726.0	-424.9

그 결과를 보면 B-C-E 이 3지역 간과 B-G-H 이 3 지역 간은 차이가 없는 반면 나머지 지역들은 서로 차이가 있는 것을 알 수가 있었다. 어떠한 차이가 있는지 알아보기 위하여 다중상자그림을 그려보면 B-C-E 지역과 B-G-H 지역의 frist leak time은 서로 비슷한 것을 알 수 있었고 B지역이 이 두 군들의 중간임을 알 수 있었다. 또한 F 지역이 가장 높은 frist leak time임을 알 수 있고 A지역이 가장 낮은frist leak time 임을 알 수 있었으며 특이점으로는 D 지역에서 아주 낮은 frist leak time의 이상치가 발견되고 있어 차후 원인을 밝혀볼 필요가 있다. 이러한 차이는 다중상자그림에서 더욱더 쉽게 알아 볼 수 있는데 지역 D인경우가 다른 지역보다 좀 더 높은 frist leak time을 가지고 있는 것을 알 수 있다.



<그림 2> Pipe의 매립지역별 다중 상자그림

3.3 Pipe의 직경에 따른 분석

직경에 따른 pipe 자료는 <표 8>과 같다.

<표 8> Pipe의 직경에 따른 frist leak time

(단위: day)

Pipe Dia	Frist Leak Time	자료수
250	15032 14702 … 10679 11050 13597	11
300	13495 9211 … 12617 12427 12890	420
375	12720 9836 … 10911 12150 12577	234
450	12709 13565 … 13559 11109 12211	189
525	11977 11004 … 9723 8479 11481	108
600	11859 12364 … 12393 14549 15748	40
675	11763 12187 … 11813 10693 10686	14

이 자료 역시 분산 분석을 실시하여 <표 9>과 같은 결과를 얻을 수 있었다. P-value가 역시 0.0001 보다 작은 값으로 이 자료 또한 유의 수준 1%에서 직경에 따라서 차이가 있다는 것을 알 수 있다.

<표 9> Pipe의 직경자료의 ANOVA Table

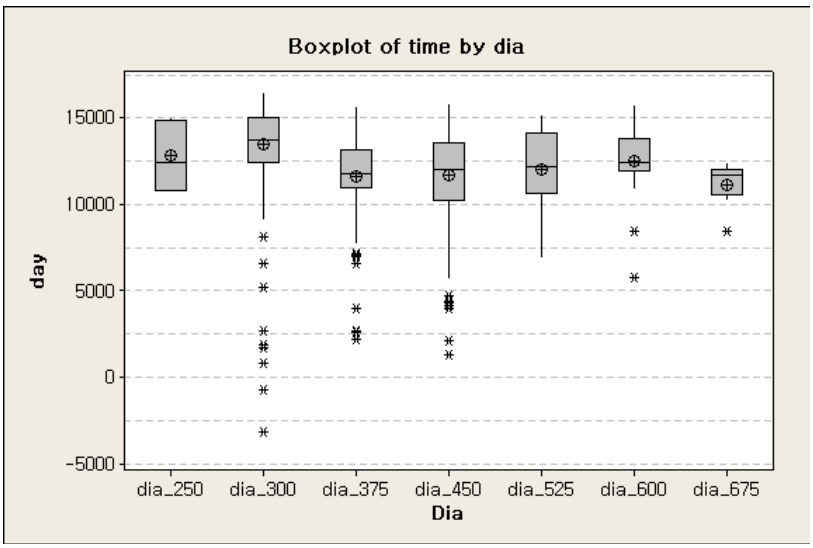
요인	제곱합	자유도	평균제곱	F value	P value
모형	739576708	6	123262785	22.03	<.0001
오차	5644660552	1009	5594312		
Total	6384237260	1015			

차이를 보기 위하여 LSD 검정을 한 결과 직경이 300 인 경우가 다른 직경들의 경우보다 frist leak time이 큰 차이를 보이고 있었으며 직경이 675인 경우도 약간의 차이를 보이고 있었다.

<표 10> LSD 검정 결과

diameter Comparison	Difference Between Means	95% Confidence Limits	
dia_300 - dia_450	1758.4	1351.9	2165.0
dia_300 - dia_375	1833.8	1455.2	2212.4
dia_300 - dia_675	2352.9	1091.9	3613.8
	⋮		
dia_675 - dia_525	-906.0	-2224.4	412.4
dia_675 - dia_450	-594.4	-1880.0	691.1
dia_675 - dia_375	-519.1	-1796.1	757.9

이러한 차이는 다중상자그림에서 더욱더 쉽게 알아 볼 수 잇는데 직경이 300인경우가 다른 직경의 경우보다 좀 더 높은 frist leak time 을 가지 고 잇는 것을 알 수 있지만 많은 이상치들이 낮은 frist leak time에서 많이 나타나고 있어 원인을 밝혀볼 필요가 있다 또한 이러한 이상치들은 375와 450 인 경우에도 나타나고 있다.



<그림 3> Pipe의 직경별 다중 상자그림

3.4 Pipe의 길이에 따른 분석

길이에 따른 파이프 자료는 총 100부터 1000까지 10개의 구간으로 <표 11>과 같은 자료이다. Pipe의 길이에 따라 차이가 있는가 알아보기 위해 분산분석을 하였다 <표 12>이 분산분석의 결과이며 P-value 값이 0.0001보다 작아 1% 유의 수준에서 길이에 따라서 frist leak time의 차이가 있음을 알 수 있다.

<표 11> Pipe 길이에 따른 frist lake time

(단위 : day)

pipe length	Frist Leak Time	자료수
100	14193 14662 13980 14375 … 16085 16161 15508 12427 12890	120
200	12364 12393 15940 12186 … 12093 11626 12617 12441 10315	62
	⋮	
900	15502 13135 13494 23494 … 9518 11042 11763 13214 12219	96
1000	16100 13589 15778 16512 … 12909 15835 12835 13721 13518	60

<표 12> Pipe length ANOVA Table

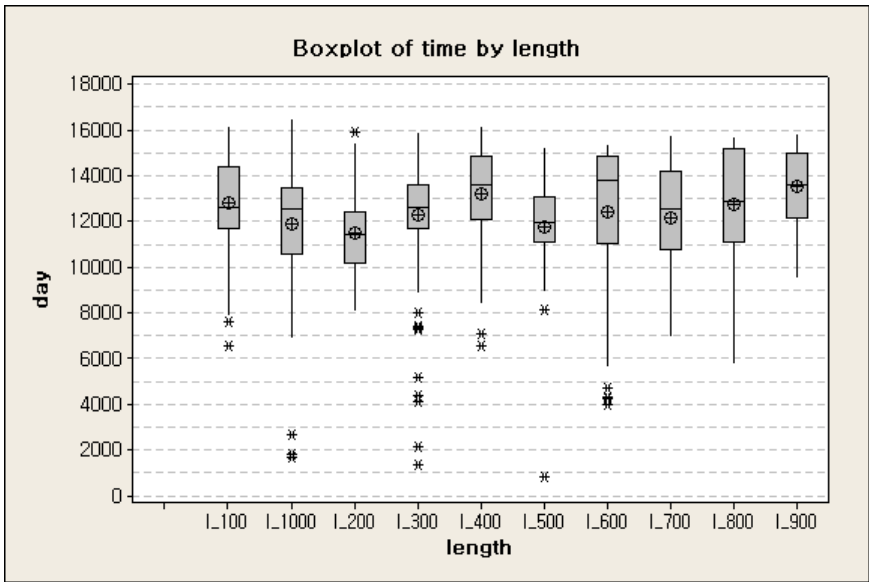
요인	제곱합	자유도	평균제곱	F value	P value
모형	3040204384	8	33780049	6.49	<0.0001
오차	180280759	803	5205829		
Total	4484301196	812			

또한 pipe 길이에 대하여 frist leak time이 어떠한 차이가 있는지 LSD 검정을 하였다. 검정 결과 길이가 900 인 경우와 400 인 경우 frist leak time 이 다른 길이들의 frist leak time 과 뚜렷한 차이를 보이고 있으며 전반 적으로 걸쳐 모든 길이에 대하여 서로 조금씩 frist leak time이 차이를 보이고 있다.

<표 13> LSD 검정 결과

lenth Comparison	Difference Between Means	95% Confidence Limits		
l_900 - l_400	282.8	-362.0	927.5	
l_900 - l_100	695.5	82.2	1308.8	***
l_900 - l_800	747.5	-146.3	1641.3	
		⋮		
l_200 - l_800	-1291.0	-2246.7	-335.2	***
l_200 - l_600	-904.4	-1719.0	-89.9	***
l_200 - l_300	-817.8	-1494.0	-141.6	***

아래의 다중상자그림을 보면 pipe 길이에 대하여 일정 한 추세가 없이 고루 분포 하고 있으며 길이가 300인 경우 낮은 frist leak time에서 많은 이상치를 보이고 있다. 4번에 걸친 다중상자그림을 보면 낮은 frist leak time의 이상치를 보이고 있어 이에 대한 연구가 필요하다.



<그림 4> Pipe의 길이별 다중 상자그림

4. 결론 및 제언

본 논문에서는 상수도용 pipeline의 frist leak time의 통계적 분석을 하였다. 그 결과를 요약하면 type별로 분산분석을 한 결과 RC, FRC, AC는 서로 간에 frist leak time이 차이가 있었으며, 특히 FRC같은 경우는 다른 두 type에 비해 현저히 낮은 frist leak time을 보이고 있어 FRC는 품질이 의심되며 차후 RC나 AC 등 다른 type 으로 교체가 필요하다고 볼 수 있다. 이후 RC type이 가장 많은 자료수와 높은 frist leak time을 보이고 있어 RC type 만을 가지고 분석을 하였다.

RC type의 분석은 매설지역, pipe의 직경과 길이로 구분하여서 분석한 결과 전체적으로는 분산분석의 결과가 0.0001보다 낮은 P-value로 매설지역, 직경, 길이 별로 모두 다른 frist leak time을 보이고 있는 것을 알 수 있었다. 세부적으로 매설지역으로 구분한 자료의 결과를 보면 2개의 동질적인 군과 나머지 서로 이질적으로 구분된 지역이 존재 하고 있었다. 그 폭도 frist leak time의 평균이 1000일에서 15000일까지 무려 5000일이 넘을 정도로 차이의 폭이 커서 지역에 따른 영향이 크다고 할 수 있다. 또 D 지역과 F 지역에서 아주 낮은 frist leak time의 이상치가 보이고 있어 원인 규명이 필요한 부분이다. 다음으로 직경에 의한 분석은 매설 지역보다는 적은 frist leak time의 차이를 보이고 있지만 직경 마다 차이가 있는 것은 확실 했으며, 직경 300, 375, 450에서 확연이 낮은 frist leak time의 이상치가 있으며 300의 경우에는 아주 긴 꼬리의 이상치가 보였다. 길이로 구분을 한 자료의 분석에서 또한 위와 비슷한 결과로 frist leak time의 차이와 이상치들이 발견됐다. 따라서 전체적인 매설지역과, 직경, 길이간의 상호 작용이나 이원배치 분산분석을 위한 실험 계획을 통한 분석이 필요하다고 본다. 또한 D, F 지역과 직경 300, 375, 450, 길이 300, 600에서 낮은 frist leak time의

이상치들이 보이고 있어 이들 간의 상호 작용이나 이원배치 분산분석 또한 별도로 해볼 필요가 있겠다. 그리도 RC_type의 자료들을 분석해본결과 어떤 요인에 의한 추세 등은 찾을 수가 없었다.

참고문헌

[1] 김형욱(1982), 방적제품의 품질개선에 대한 연구, 전남대학교 석사학위논문.

[2] 박성현(2009), 현대실험계획법, 민영사.

[3] 박성현, 김종욱(2010), (Minitab을 활용한) 현대실험계획법: 공학실험의 설계와 분석을 위한 필수 지침서, 민영사.

[4] 배종성 외 7(2011), SPSS를 이용한 통계학, 경문사.

[5] 윤중범(1988) 발수가공 데이터의 분산분석, 품질경영학회지, 16, 1, 43-48,

[6] 전세창, 손환규(2011), 보험회사의지속가능경영 전략에 관한연구, 한국통계학회논문집, 18, 1, 119-130

[7] Mickey, Ruth M.(2004), Applied statistics: analysis of variance and regression. 3rd ed., Wiley.