SM

짝맞춤 조립을 위한 최적 분할 방식과 시뮬레이션을 이용한 검증

2023 제조 빅데이터 분석 경진대회

TEAM 박광진

목차

분석 대상

"tube와 shaft의 치수가 주어졌을 때, 최적의 ball size 제안"

✓ 짝맞춤 조립방식

- 여러 개의 핵심 부품을 조립하여 완성되는 제품을 생산할 때 많이 사용하는 제조 기법
- 각 부품의 치수 정밀도를 낮추어 부품의 가격을 떨어뜨리는 대신 조립 시에 부품을 잘 조합하여 제품의 품질을 유지시킴

✓✓ 치수분할방식

: 짝맞춤 조립방식의 대표적인 유형으로, 생산된 **부품 치수에 의해 각기 여러 범주로 분류**한 후, 대응되는 범주에 속한 부품들끼리 짝을 맞추어 조립하는 방식

- **등간격분할방식**(equal width partitioning)
  : 각 범주의 **치수 폭**이 동일하도록 하는 방식
- **등면적분할방식**(equal area partitioning)
  : 각 범주에 속할 **기대 부품 수**를 동일하도록 하는 방식

현상 및 문제점

- 기존 짝맞춤 조립표

→ **정확도가 보장되지 않고, 조립표를 확인하는 시간이 길어** 거의 사용되지 않음

✓ 짝맞춤 조립표의 부품별 grade 기준이 **실제 부품 치수와 맞지 않음**

```
MultiIndex: 1296 entries
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Tgrade  826 non-null    object
 1   Sgrade  793 non-null    object
 2   Bbush   190 non-null    float64
```

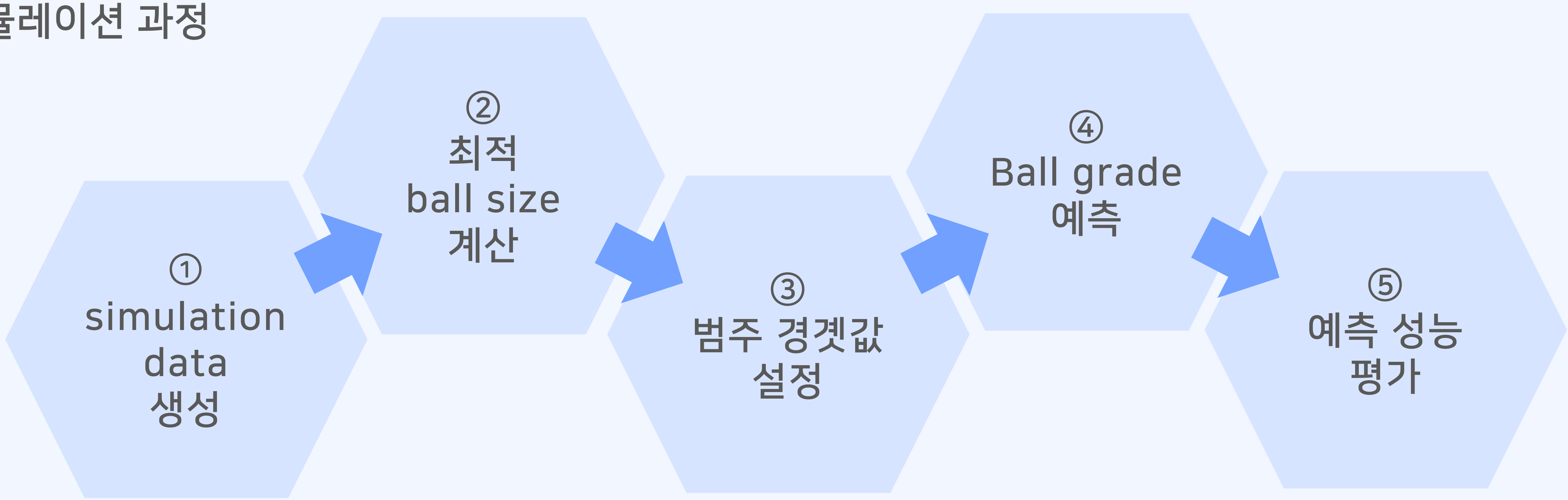
⇒ 기존 짝맞춤 조립표의 분류 방식이 적절하다고 할 수 없으며

최적의 분류 방식을 찾고 이를 검증할 필요가 있다

데이터 분석 방법 및 제반 가정

- tube와 shaft의 치수는 over-ball-diameter(OBD) 방식으로 측정된 **중간 부분의 치수(M 위치의 평균값)**로 정의

시뮬레이션 과정



데이터 분석 방법 및 제반 가정

- tube와 shaft의 치수 분포가 **서로 독립**이며 **정규분포**를 따른다고 가정
  → 100,000개의 난수를 무작위로 추출
- **random seed**를 임의의 수(1906)로 **고정**하여 같은 데이터에 의해 정확도를 계산

데이터 분석 방법 및 제반 가정

- 최적 ball size는 **기존 짝맞춤 조립표에 의해 얻어진 선형회귀식이 정확하다는 가정** 하에 계산
- 베어링 볼은 **0.002mm 간격**으로 생산되며 **소수점 아래 셋째 자리가 짝수라 가정**하고
  ball grade가 소수점 아래 셋째 자리에서 짝수를 갖도록 변환(홀수인 경우 올림, 짝수인 경우 버림)
- 예측한 ball grade가 최적 ball grade로부터 **±0.002mm 이내에 포함되는 경우 양품**이라고 간주

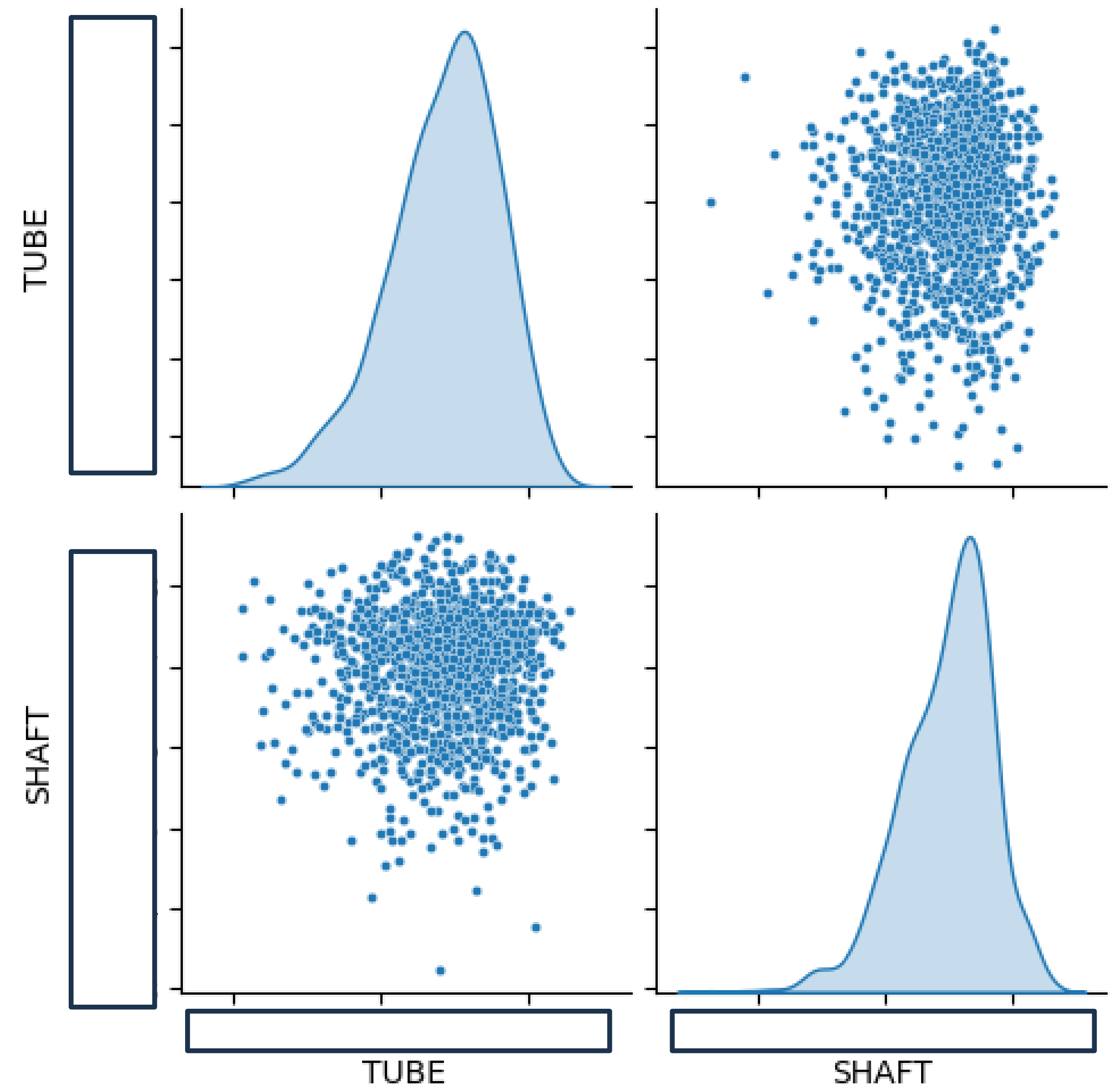
데이터 분석 방법 및 제반 가정

- 예측 ball grade와 최적 ball size의 **MSE**
- 최적 ball grade로 분류된 비율인 **1st accuracy**
- 양품으로 분류된 비율인 **2nd accuracy**

- 각 부품별 범주 개수는 2~20개로 제한
- 2nd accuracy가 90%를 넘으면서 **범주의 개수가 가장 적은** 분류 방식을 **최적 분류 방법**이라 정의

데이터 분석 절차와 내용

① simulation data 생성



Tube와 Shaft의 치수 분포

- 시뮬레이션 데이터는 **실제 제품 치수 데이터의 평균과 표준편차**를 반영
  - Tube Size $\sim N(\quad,\quad)$
  - Shaft Size $\sim N(\quad,\quad)$

→ 난수 **100,000개를 무작위로 추출**하여 생성

데이터 분석 절차와 내용

② 최적 ball size 계산

- shaft 치수가 **0.004mm 증가**하면
  0.002-0.003mm 작은 ball size 사용,
  tube 치수가 **0.01mm 증가**하면
  0.002-0.003mm 큰 ball size 사용

→ **ball size가 tube, shaft 치수와 선형 관계**를 갖는다고 판단

→ 기존 짝맞춤 조립표를 만족하는 **임의의 데이터를 생성**하여
선형 회귀를 통해 그 관계를 모델링

SM

데이터 분석 절차와 내용

② 최적 ball size 계산

✓ 데이터는 **기존 짝맞춤 조립표에서의 최솟값과 최댓값**을 갖고, 범위 내에서 **균일하게 분포**되도록 하여 **난수 100,000개 생성**

$$\text{Tube Size} \sim U(\boxed{}, \boxed{})$$

$$\text{Shaft Size} \sim U(\boxed{}, \boxed{})$$

✓ 짝맞춤 조립표에 의해 **ball size가 지정되지 못한 경우는 제거**하여 선형회귀 모형에 적합

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.948
Model:	OLS	**Adj. R-squared:**	0.948
Method:	Least Squares	**F-statistic:**	4.474e+05
Date:	Thu, 14 Sep 2023	**Prob (F-statistic):**	0.00
Time:	10:34:28	**Log-Likelihood:**	2.5651e+05
No. Observations:	49051	**AIC:**	-5.130e+05
Df Residuals:	49048	**BIC:**	-5.130e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	12.0385	0.015	790.348	0.000	12.009	12.068
Tmean	0.2072	0.001	300.404	0.000	0.206	0.209
Smean	-0.4534	0.000	-933.245	0.000	-0.454	-0.452

Omnibus:	912.151	**Durbin-Watson:**	1.996
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	579.440
Skew:	0.128	**Prob(JB):**	1.50e-126
Kurtosis:	2.533	**Cond. No.**	7.78e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 7.78e+04. This might indicate that there are strong multicollinearity or other numerical problems.

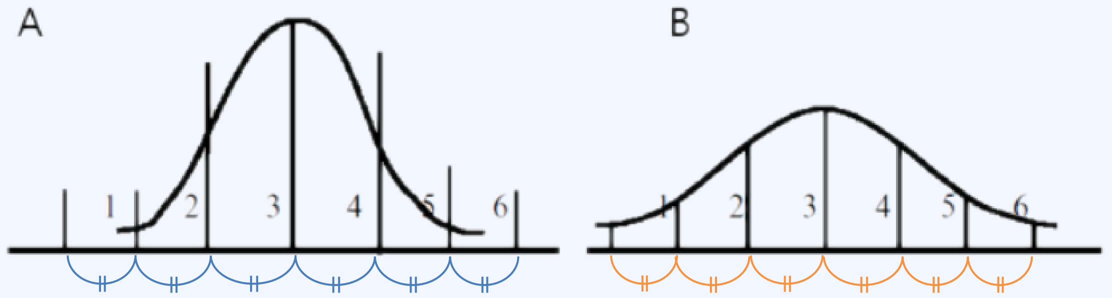
$$\Rightarrow \boldsymbol{BallSize = 12.0385 + 0.2072 \times TubeSize - 0.4534 \times ShaftSize}$$

SM

데이터 분석 절차와 내용

- 등간격분할방식

: 부품 A와 B를 **치수** 기준으로 분류할 때 각 범주의 **치수 폭이 동일하도록** 분할하는 방식



③ 범주 경곗값 설정

→ 부품의 분포는 **정규분포**를 따른다고 가정

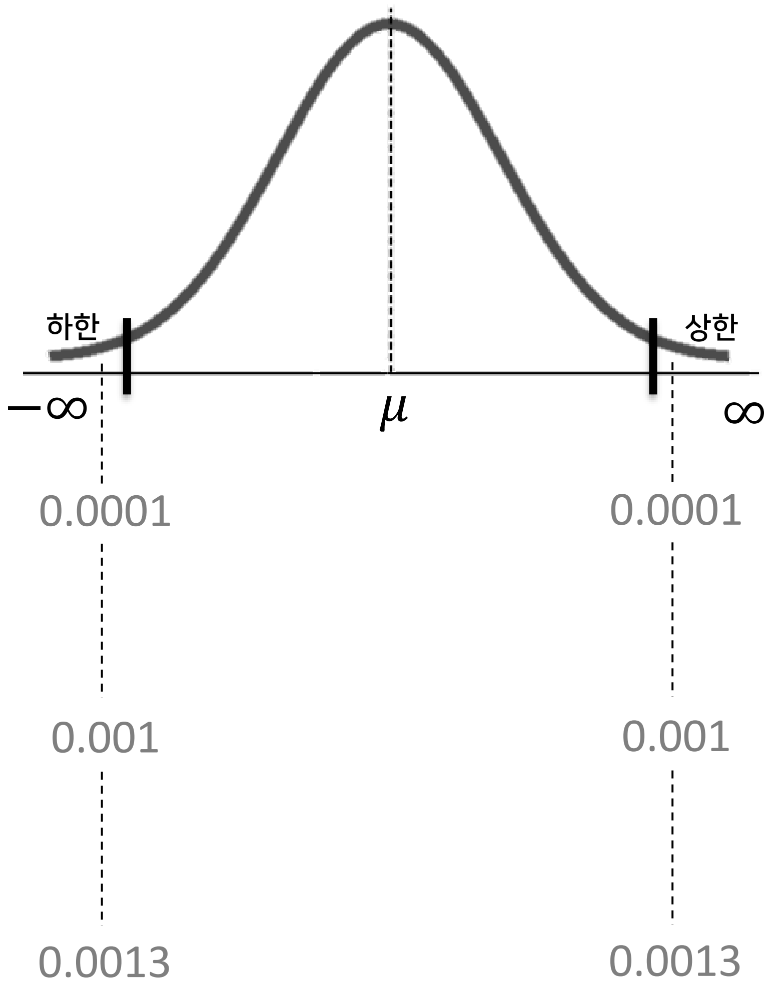
→ 정규분포는 값의 범위가 (-∞, ∞)이므로 **상·하한 지점**을 지정

✔ 0.0001 → $F^{-1}(0.9999)$ & $F^{-1}(0.0001)$

✔ 0.001 → $F^{-1}(0.999)$ & $F^{-1}(0.001)$

✔ 0.0013 → $\mu \pm 3 \times \sigma^2$

*F^{-1}은 역누적분포함수



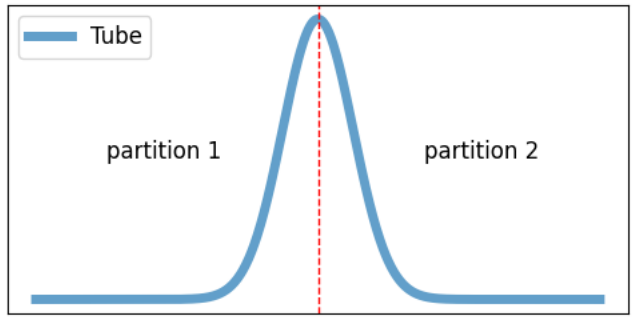
데이터 분석 절차와 내용

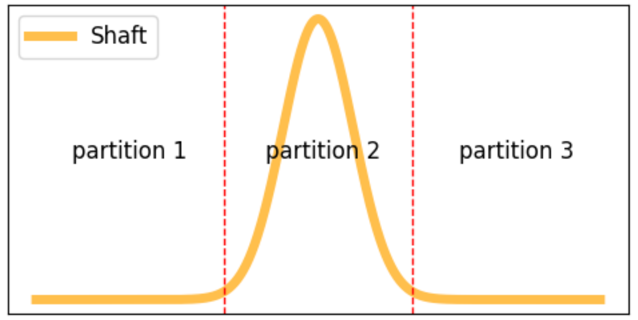
- 등간격분할방식

④ ball grade 예측

- 각 범주에 해당하는 **대푯값(평균값, 중앙값) 지정**
- tube와 shaft 치수의 대푯값을 회귀식에 넣어 **ball size를 예측**
- 이를 **0.002mm 간격**으로 변환

Ex. 튜브 범주 2개, 샤프트 범주 3개, 대푯값으로 중앙값 지정





Tube partition 1의 중앙값 X_1과 Shaft partition 1의 중앙값 Y_1을 회귀식에 대입

$(Ball\ Size = 12.0385 + 0.2072 \times Tube\ size\ - 0.4534 \times Shaft\ size)$

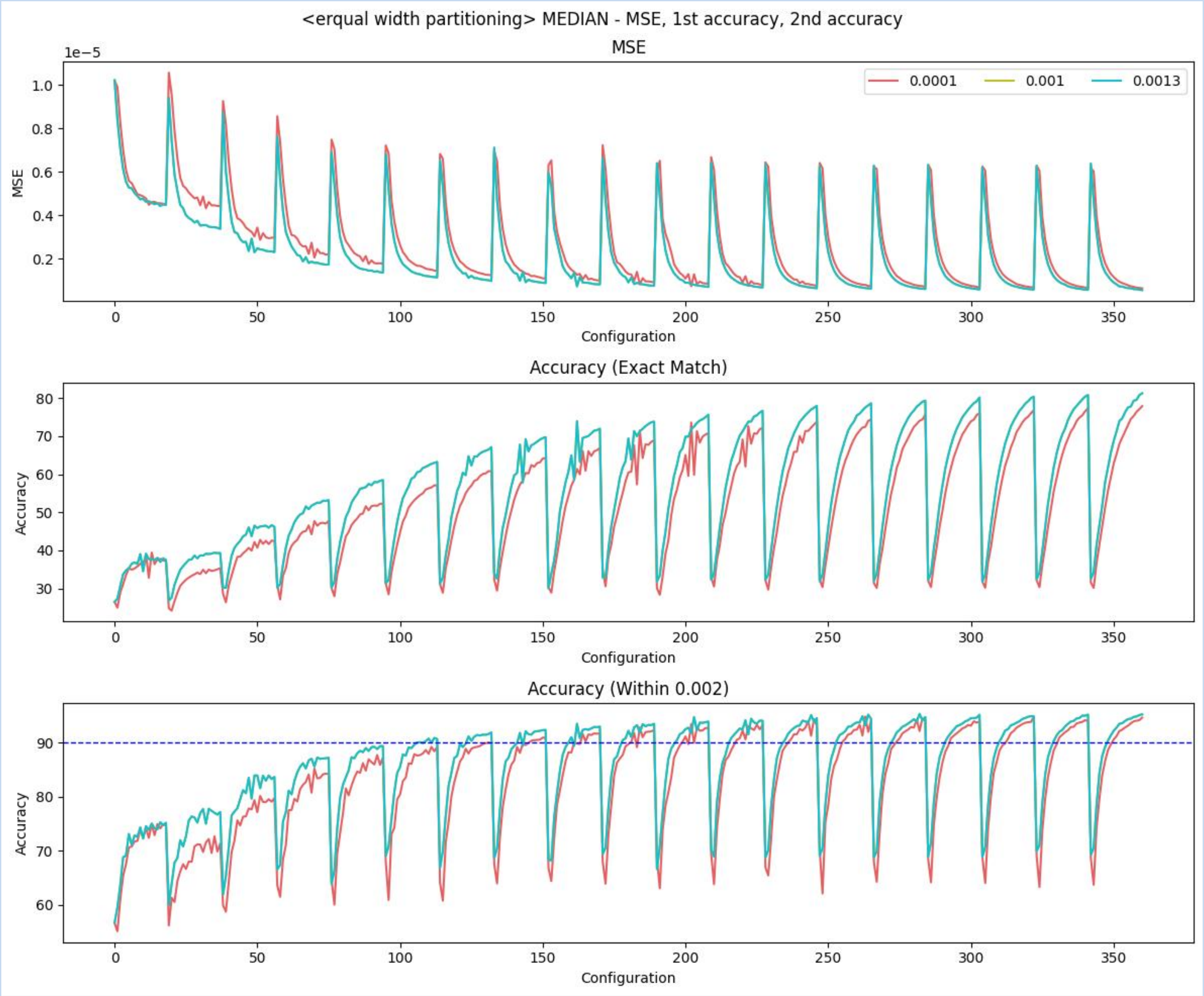
→ 나머지 partition에 대해서 반복

→ 예측한 Ball Size를 0.002mm 간격에 맞게 변환해 ball grade를 구함

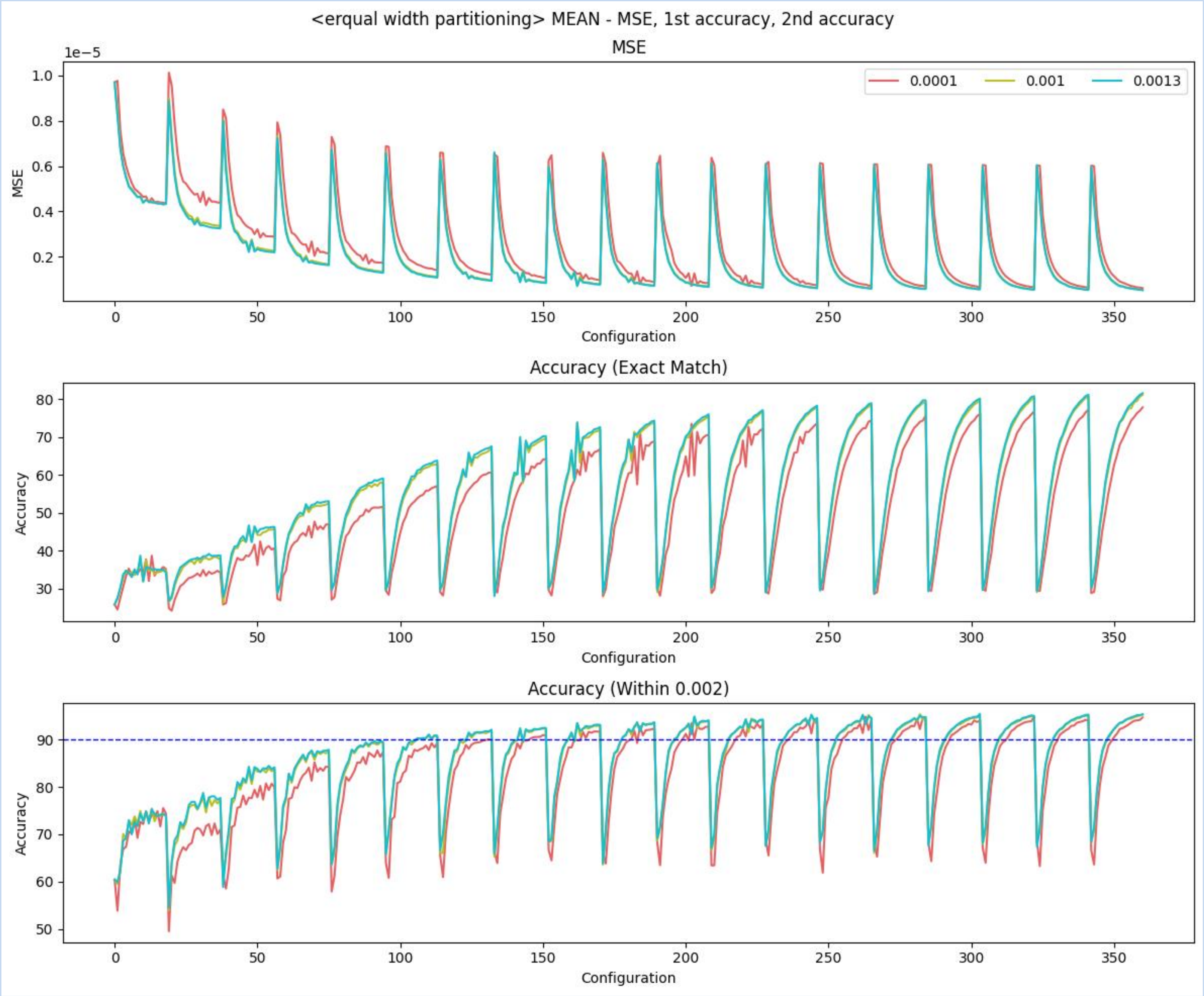
데이터 분석 절차와 내용

- 등간격분할방식

⑤ 예측 성능 평가



대푯값 중앙값



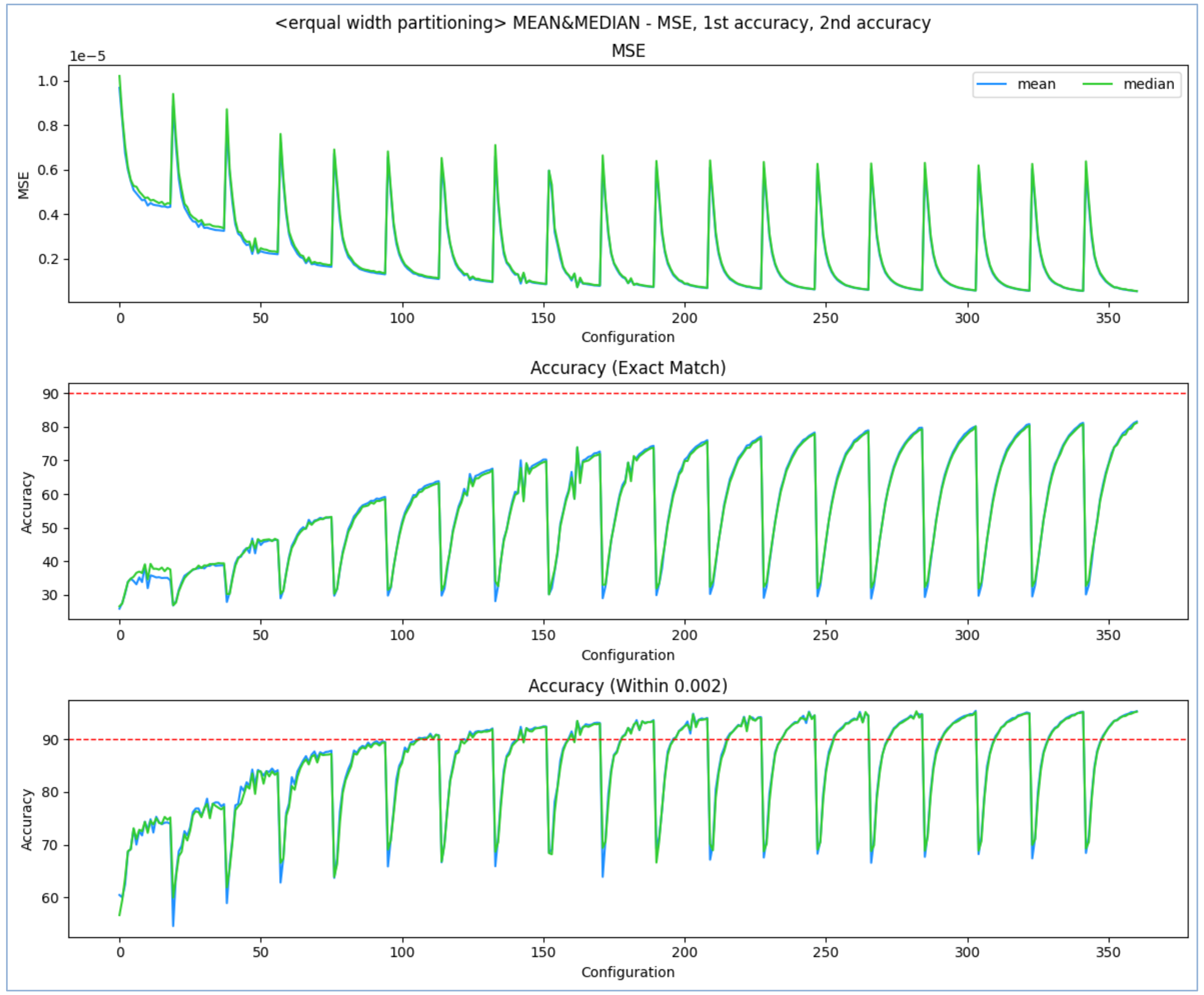
대푯값 평균값

→ 두 대푯값 모두 구간 밖에 속할 확률이 순서대로 0.0013, 0.001, 0.0001일 때 높은 정확도를 보임.

데이터 분석 절차와 내용

- 등간격분할방식

⑤ **예측 성능 평가** 가장 높은 정확도를 보였던 0.0013에서 **중앙값과 평균값 비교**



- 큰 차이를 보이지는 않지만 **범주의 개수에 따라** 약간의 차이를 보임.
  - 범주의 수가 **적을 때** MSE, 1st accuracy, 2nd accuracy 모두 **중앙값**이 약간 더 좋은 성능을 냄.
  - 범주의 수가 **많을 때** 1st accuracy, 2nd accuracy는 **평균값**이 약간 더 좋은 성능을 냄.

- 범주의 개수가 더 증가하면 중앙값과 평균값 간의 차이가 거의 없어짐.

⇒ tube 범주 10개, shaft 범주 8개일 때 2nd accuracy가 90% 이상을 만족하고 범주의 개수가 가장 작았음 (대푯값 : 평균값)

데이터 분석 절차와 내용

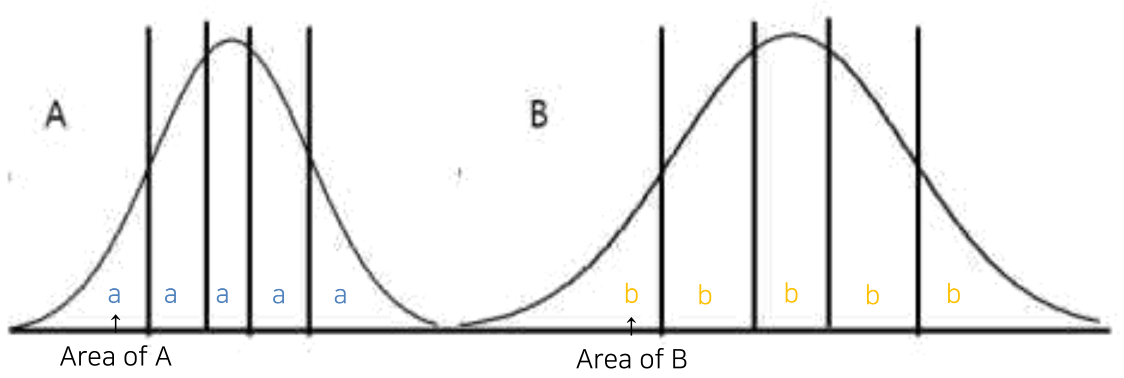
- 등면적분할방식

: 부품 A와 B를 **확률** 기준으로 분류할 때 각 범주의 **면적이 동일하도록** 분할하는 방식

③ 범주 경곗값 설정

✓ 등면적분할방식은 두 부품이 각 범주에 속할 확률이 같게 되도록, 즉 **구간별 적분값(면적)**이 같게 되도록 분할하는 방식

→ 각 구간의 면적이 **1/n**이 되도록 하는 경곗값을 구함



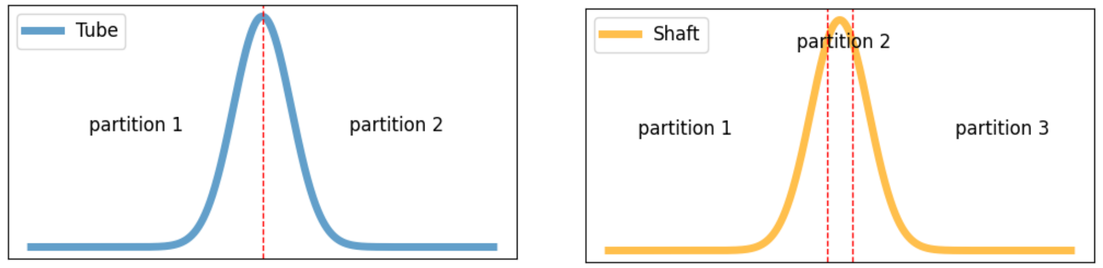
✓ 등면적분할방식은 등간격분할방식에서
③ **범주 경곗값 설정** 에서만 차이가 있고, 나머지 절차는 동일

④ ball grade 예측

- 범주를 분할하고 각 범주에 해당하는 **대푯값(평균값, 중앙값) 지정**

→ 튜브와 샤프트의 대푯값을 회귀식에 넣어 **ball size를 예측**

→ 이를 **0.002mm 간격으로 변환**



✓ Tube partition 1의 중앙값 X_1과 Shaft partition 1의 중앙값 Y_1을 구해 회귀식에 대입

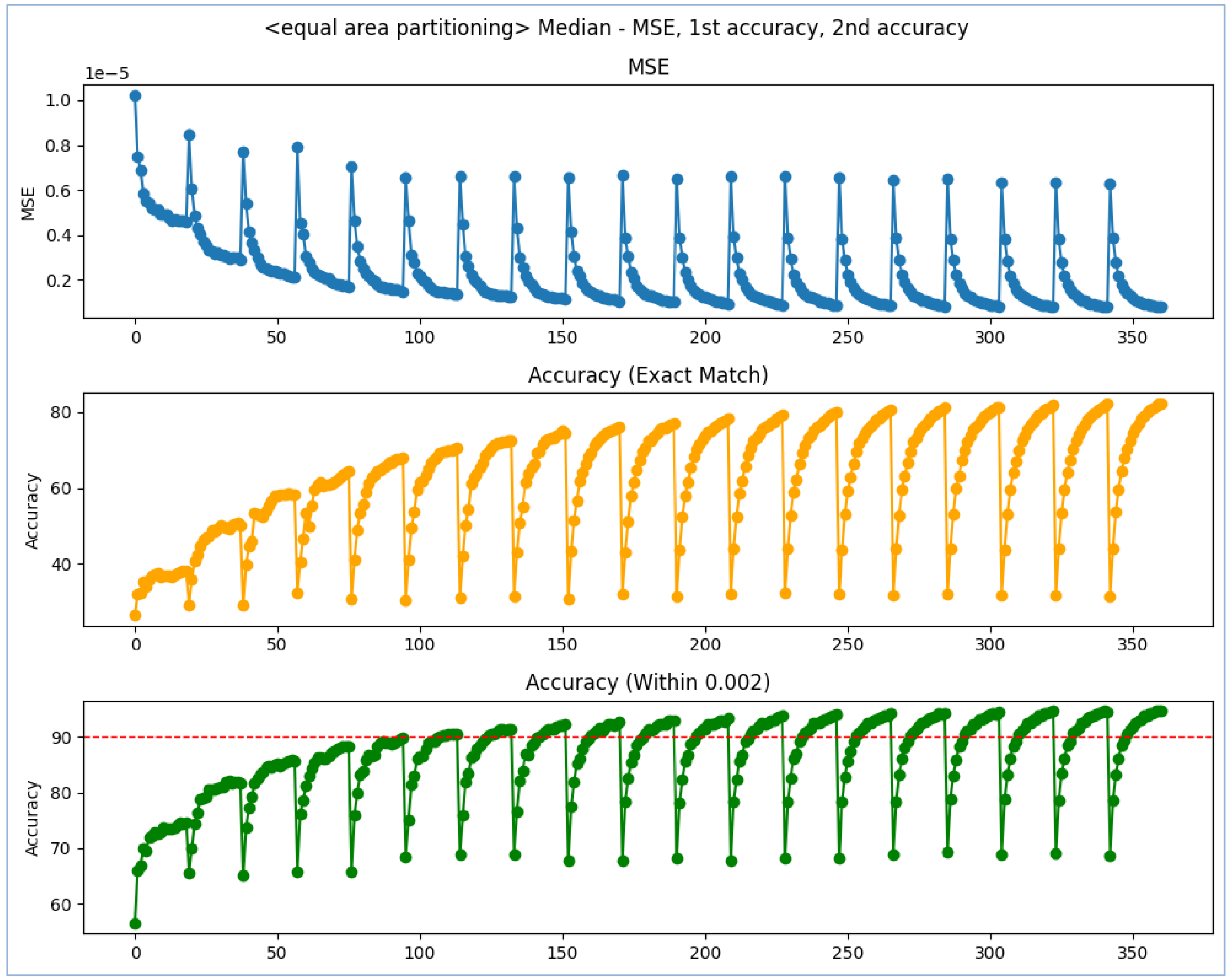
✓ 예측한 Ball Size를 0.002mm 간격에 맞게 변환해 ball grade를 구함

SM

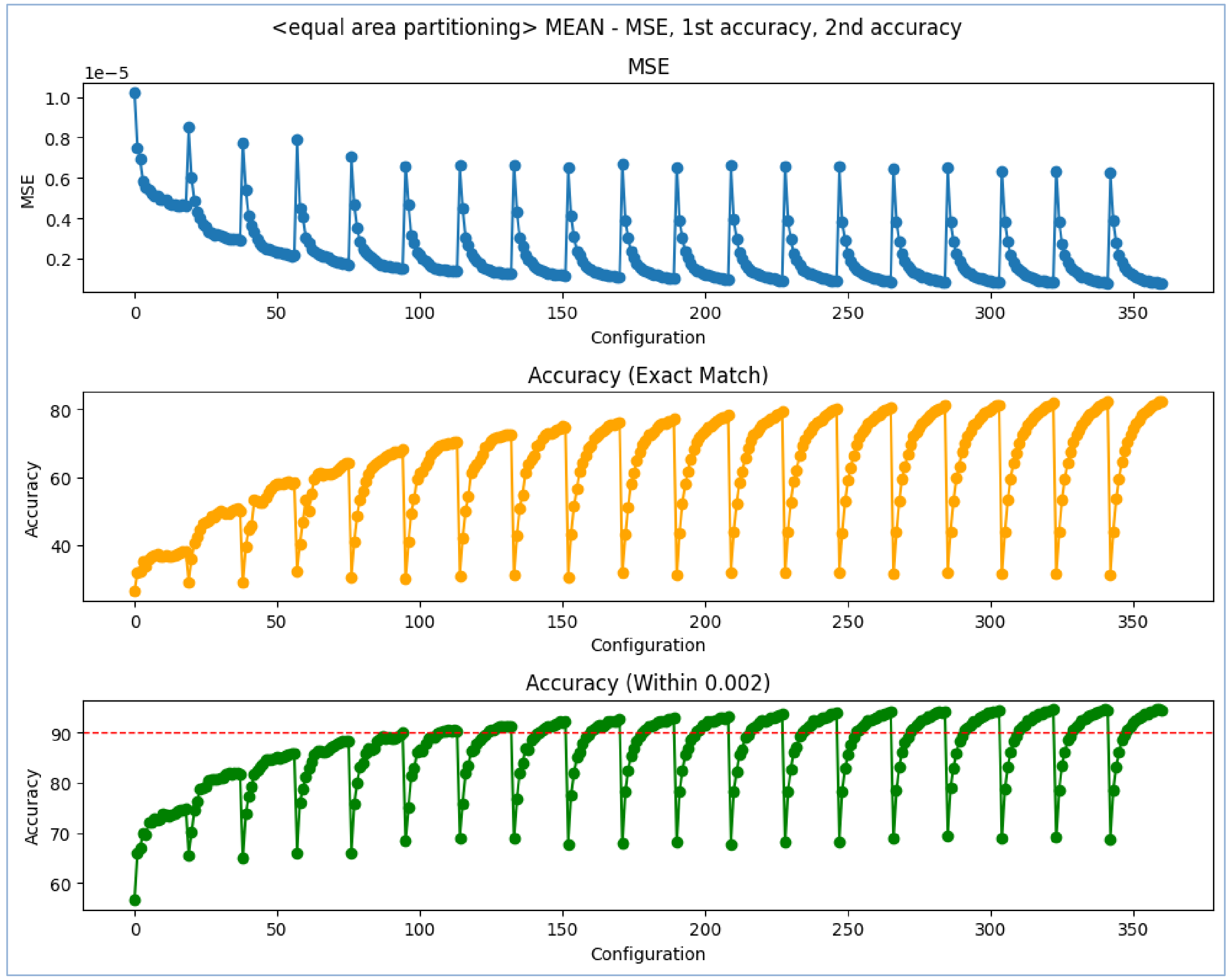
데이터 분석 절차와 내용

- 등면적분할방식

⑤ 예측 성능 평가 결과



대푯값 중앙값



대푯값 평균

→ 두 대푯값 중 **중앙값**을 사용했을 때 더 적은 범주 개수(**tube 범주 9개, shaft 범주 11개**)로 2nd accuracy가 90%를 넘김

데이터 분석 결과

∴ 상·하한 지점을 $\mu \pm 3 \times \sigma^2$(구간 밖 면적 각각 0.0013)로 하고 대푯값을 평균값으로 사용하는

등간격분할방식에 의해 tube 10개, shaft 8개의 범주로 분류하는 것이 최적

(이때의 1st accuracy는 0.61553, 2nd accuracy 는 0.90049)

- 신뢰구간

정규분포 근사를 이용한 신뢰구간

분류 과정은 이항 분포(정답1,오답0)를 따르고,

표본의 크기가 크므로 **정규분포 근사**를 이용해

분류 정확도의 신뢰구간을 구할 수 있음

$$accuracy \pm z\sqrt{\frac{accuracy(1-accuracy)}{n}}$$

✓ 1st accuracy : (0.6125, 0.6185)

✓ 2nd accuracy : (0.8986, 0.9023)

몬테카를로 시뮬레이션을 이용한 신뢰구간

몬테카를로 시뮬레이션을 이용하여

Random seed를 바꿔가며 10만 개의 표본을 생성하고

정확도를 구하는 과정을 **100번 반복 시행**

✓ 1st accuracy : (0.6112, 0.6165)

✓ 2nd accuracy : (0.8967, 0.9004)

SM

데이터 분석 결과

- 짝맞춤 조립표

GRADE 기준표(TUBE)

GRADE	SIZE	MARKING
A		A
B		B
C		C
D		D
E		E
F		F
G		G
H		H
I		I
J		J

GRADE 기준표(SHAFT)

GRADE	SIZE	MARKING
A		A
B		B
C		C
D		D
E		E
F		F
G		G
H		H

개선된 짝맞춤 조립표

Tgroup	Sgroup	2nd (-0.002)	1st	2nd (+0.002)
A	A			
	B			
	C			
	D			
	E			
	F			
	G			
	H			
J	A			
	B			
	C			
	D			
	E			
	F			
	G			
	H			

공정 개선 방향 및 기대효과

- 기존 공정 과정에 큰 변화 없이 **(수집한 데이터를 이용한) 짝맞춤 조립표의 분할 방식 변경**만으로 정확도를 향상시켜 **비용 측면에서 효율적인 디지털 전환**이 가능함
- 같은 범주의 tube와 shaft에서는 모두 동일한 ball size가 대응되기 때문에 **작업자가 짝맞춤표를 확인하는 시간을 절약**할 수 있음
- 향상된 정확도로 신규 작업자도 **적은 횟수로 최적의 조합**을 찾아낼 수 있음

⇒ 인적 오류에 대한 위험을 줄이고, 결과적으로 부품 조립 정확도를 높일 수 있음

분석 한계 및 향후 개선 방향

- **최적 ball size가 tube 치수, shaft 치수와 선형 관계**를 갖고 짝맞춤 조립표에 의해 **얻어진 선형 관계가 적절**하다는 가정
  → 기계에 의한 제품 검사 결과(ex. 슬라이딩 부의 힘 크기) 등의 **정보가 보충**된다면 최적 ball size에 대한 정의를 보완할 수 있고, 노이즈가 제거되어 **질적으로 향상된 데이터를 이용해 더 정확한 예측과 시뮬레이션이 가능**할 것

- 최적 분류 방법을 결정하기 위해 **2nd accuracy가 90%가 넘으면서 부품 범주의 수를 최소로 하는 것을 기준**으로 설정
  → 범주 수가 증가함에 따라 증가하게 되는 **분류 비용과 정확도와 품질의 향상이 가져올 이익**을 안다면, 이를 함수 형태(ex. Taguchi 손실함수)로 정의하여 **총이익을 최대화 시키도록 최적 분류 방법을 결정**할 수 있을 것

⇒ 향후 추가적인 데이터 수집 및 분석 방법을 활용하여 더 정확하고 효율적인 공정으로의 개선이 가능함

참고 문헌

- Pugh, G.A. 1986. “Partitioning for selective assembly.” Computers and Industrial Engineering Conference Proceedings, 175-179.
- 권혁무, 이영준, 이민구, 홍성훈. (2017). 선택조립방식의 효율성에 대한 시뮬레이션 검토. 품질경영학회지, 45(4), 829-846.

SM

감사합니다

TEAM 박광진