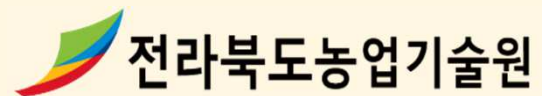


2020년 빅데이터 분석 활용 경진대회(20. 12. 18)

스마트팜 빅데이터 분석



발표자 : 서재원

CONTENTS



분석 개요



데이터 탐색



분석 결과



활용





분석 개요

- ❖ 스마트팜 설명
- ❖ 분석 목적 및 배경

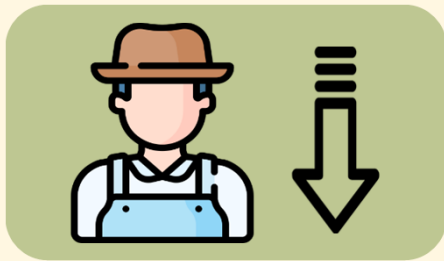


01. 분석 개요

분석 배경 및 목적



Smart Farm



농사기술에 사물인터넷(IoT)와 빅데이터등을 활용해 시공간의 제약 없이 최적의 생육환경을 자동 제어하는 '지능화된 농장을 말한다.

일손절감과 생산량 증가라는 장점이 있다.

분석 목적

환경, 생육 등 수집된 빅데이터 분석을 통한 작물 생육 예측





02

데이터 탐색

- ❖ 분석 데이터
- ❖ 분석 전체 프로세스
- ❖ 데이터 모니터링



02. 분석 과정

* 분석 데이터



방울토마토



완숙토마토



생육데이터

초장, 생장길이, 엽장,
엽폭, 수확과 등



환경데이터

내부온도, 내부습도,
일사량 등



양액데이터

배양액, 급수량, EC,
PH 등





02. 분석 과정

* 분석 농장 정보



방울토마토

아리울 농장

전라북도 김제시 청하면

번영 농장

전라북도 김제시

하늘 농장

전라북도

화천 농장

전라북도



완숙토마토

로즈밸리 농장

전라북도

장수파프리카 농장

전라북도

민수네토마토 농장

전라북도

오마이토마토 농장

전라북도

햇살 농장

전라북도

등 11개 농장





02. 분석 과정

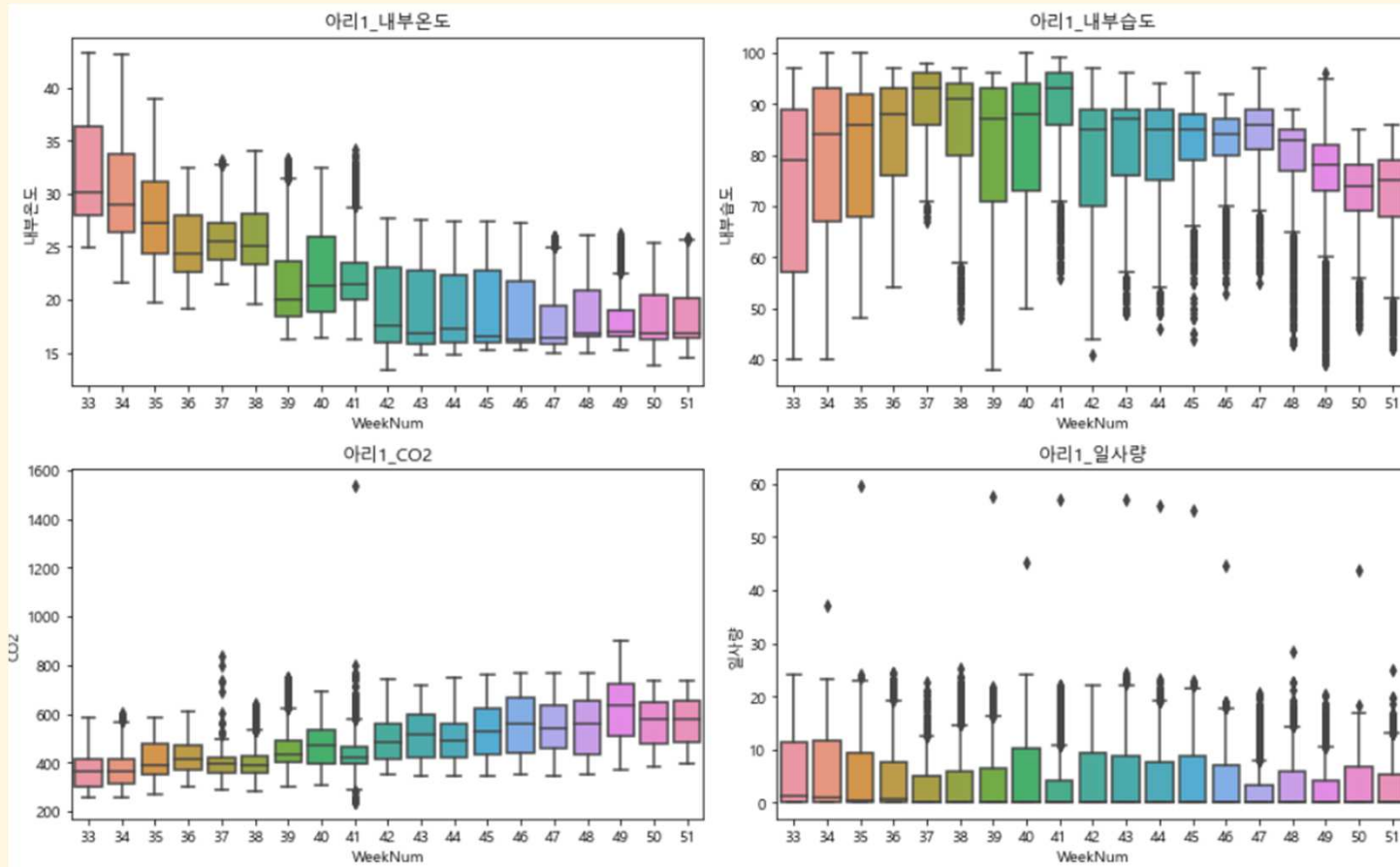
* 분석 전체 프로세스





02. 분석 과정 - 데이터 모니터링

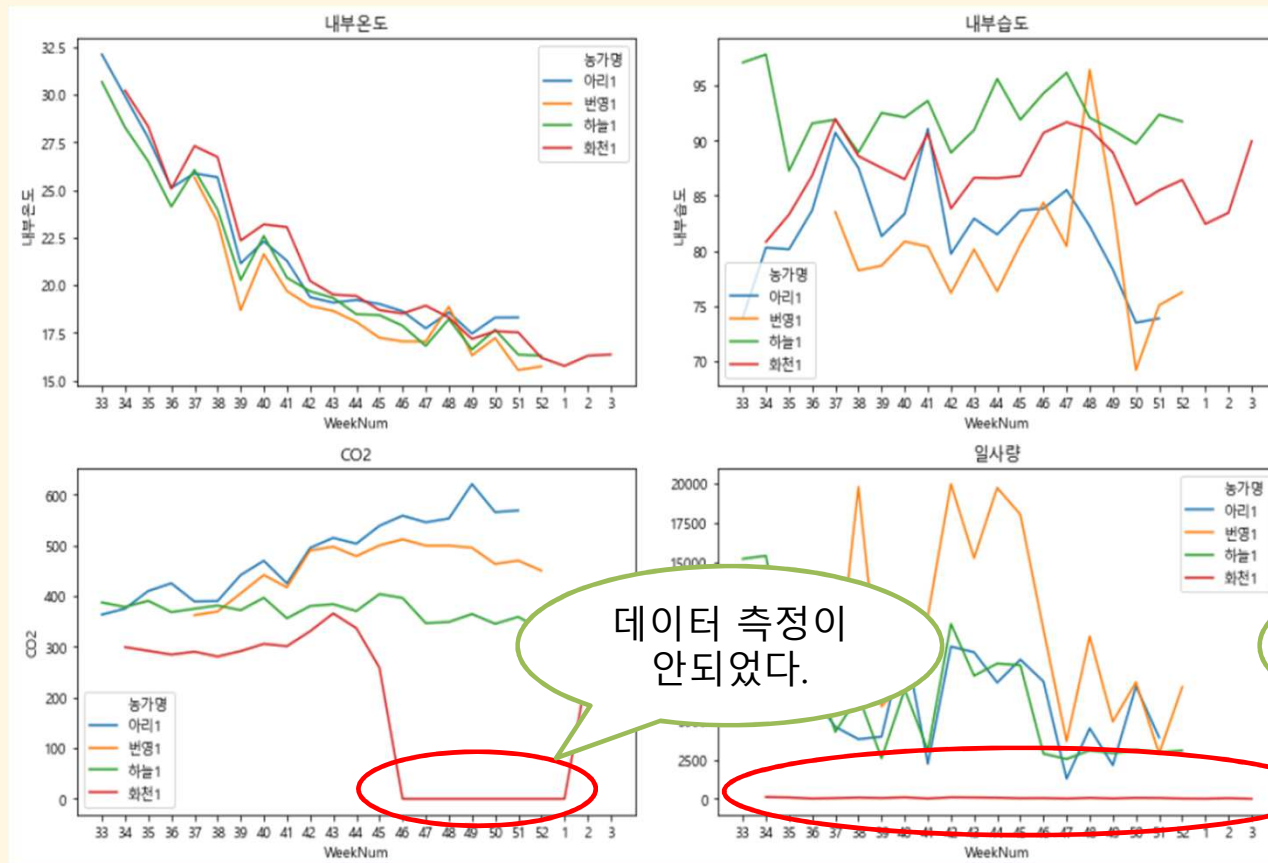
* 주차별 환경 데이터 시각화





02. 분석 과정 - 데이터 모니터링

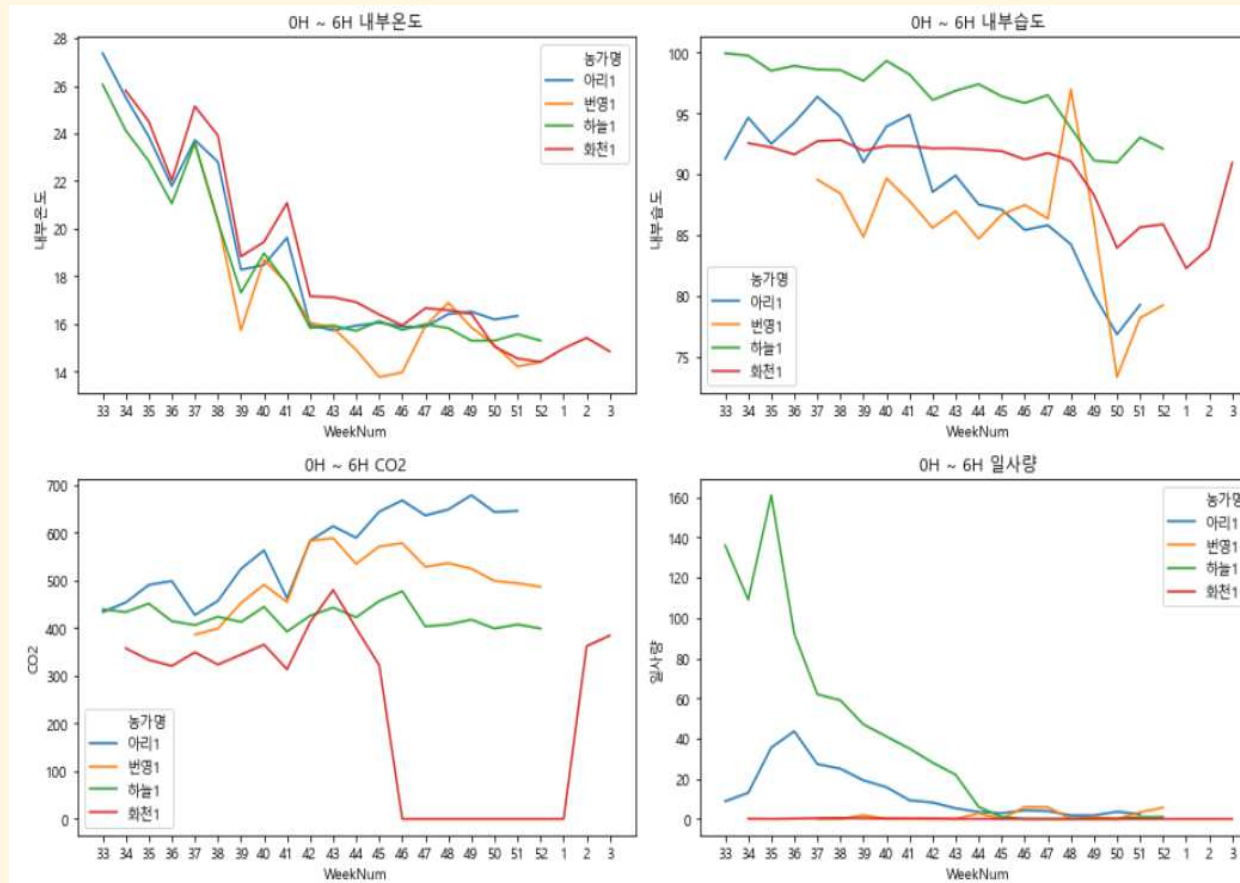
농가별 주차별 환경 현황





02. 분석 과정 - 데이터 모니터링

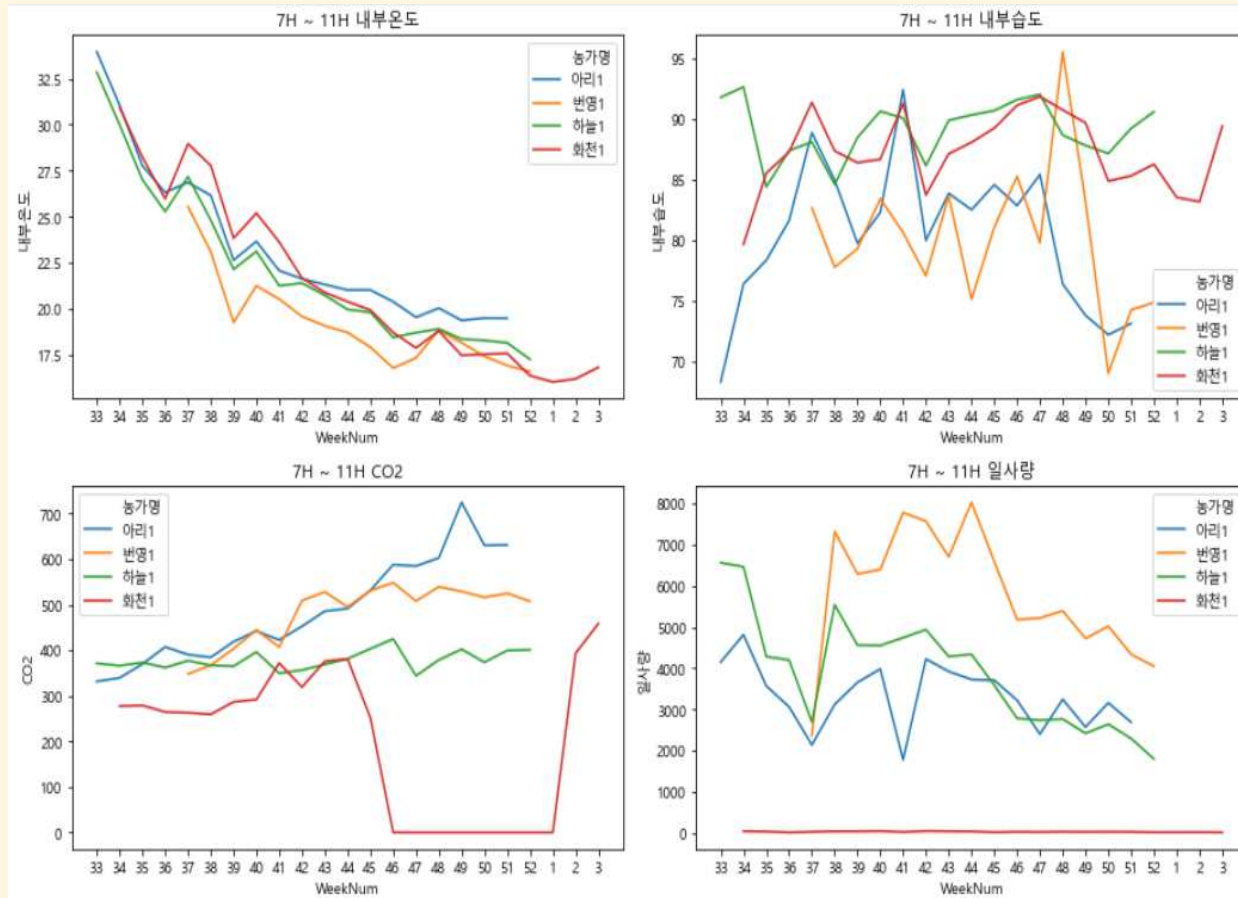
시간 구간별 환경 비교 (0시~06시)





02. 분석 과정 - 데이터 모니터링

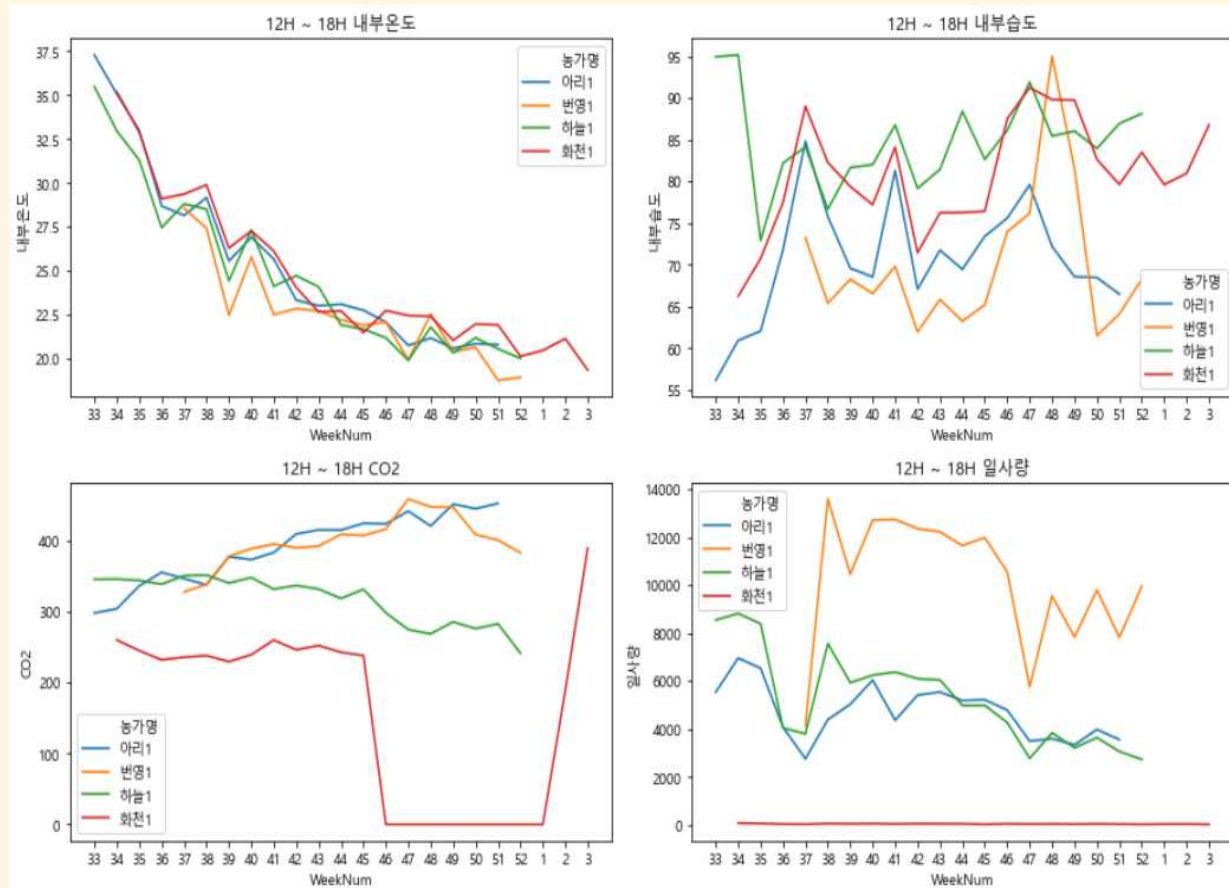
시간 구간별 환경 비교 (07시~11시)





02. 분석 과정 - 데이터 모니터링

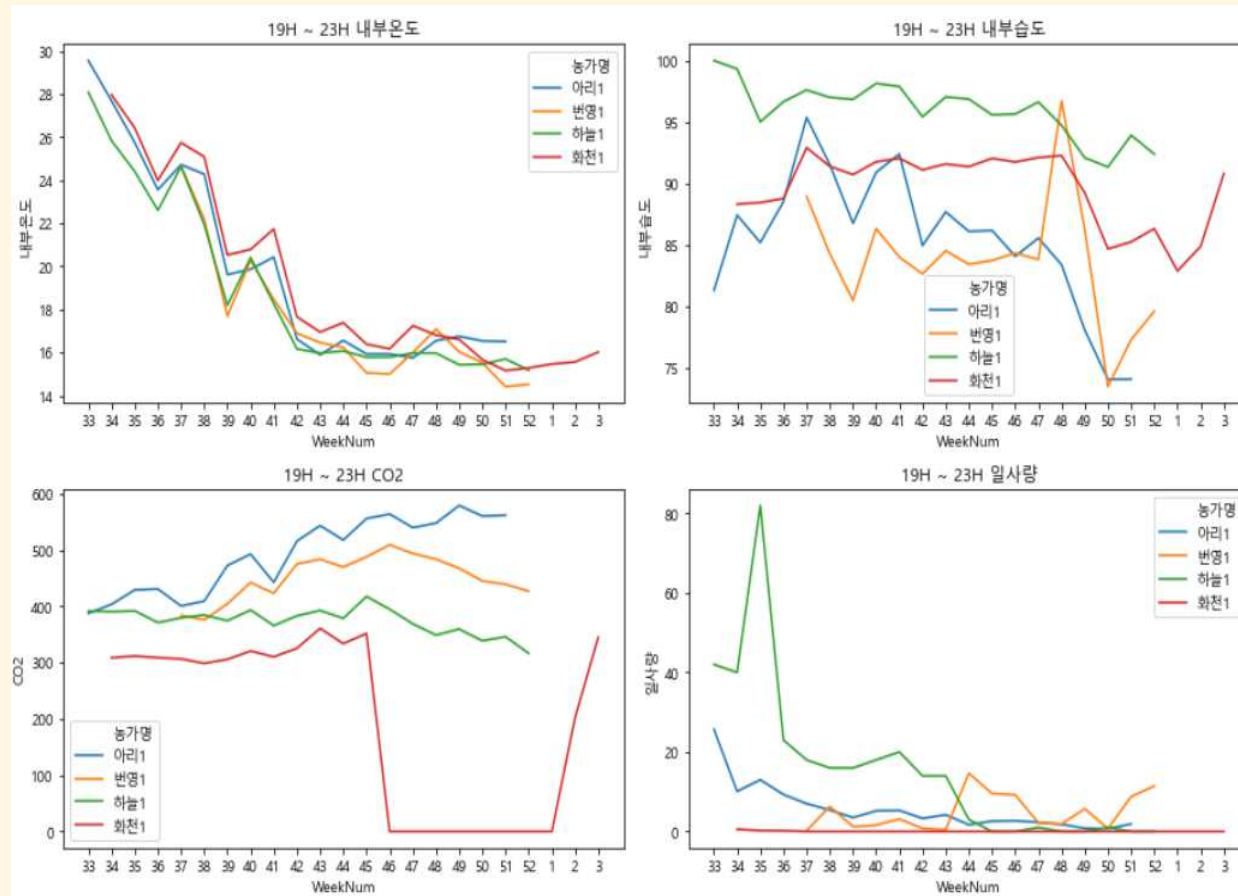
시간 구간별 환경 비교 (12시~18시)





02. 분석 과정 - 데이터 모니터링

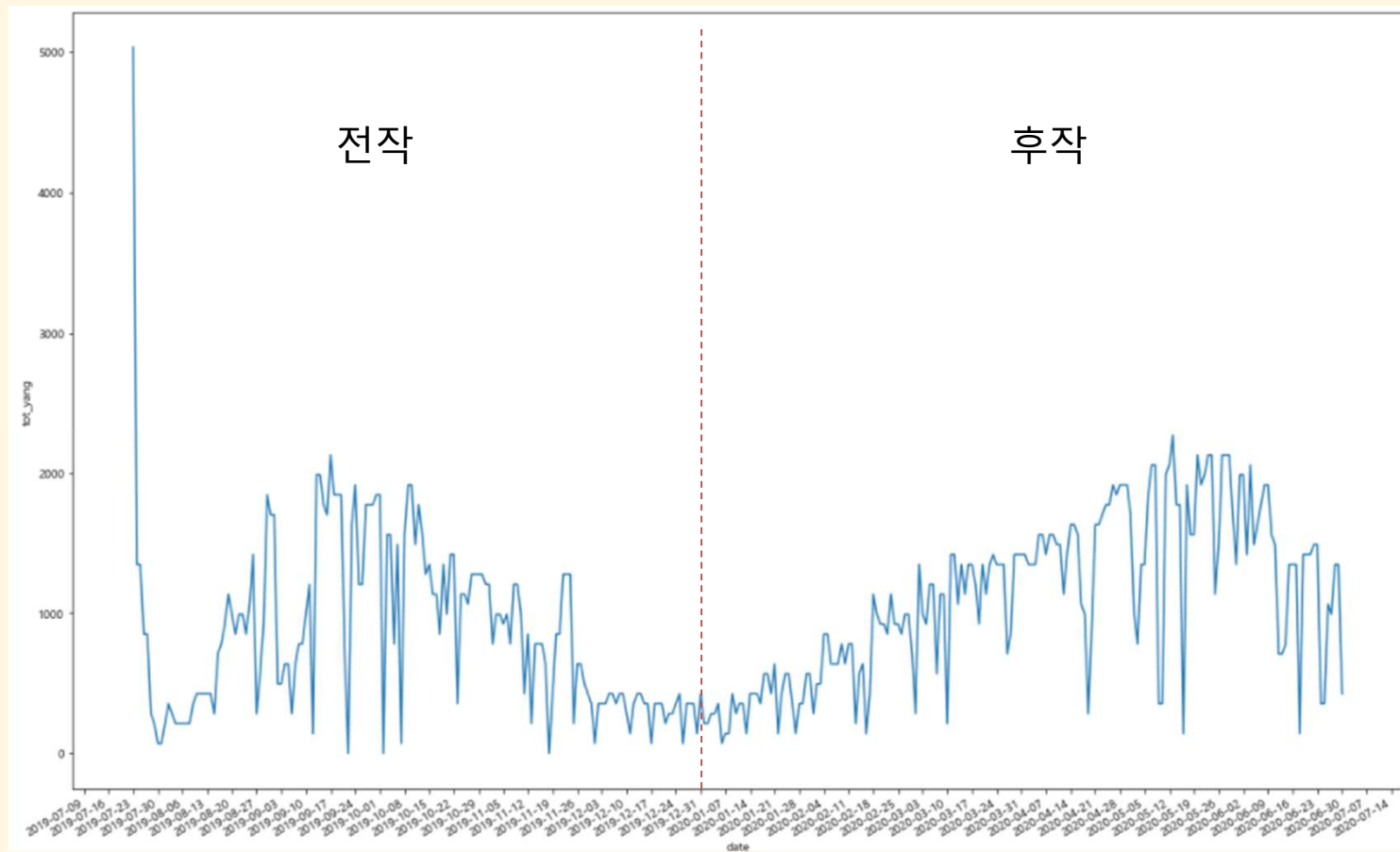
시간 구간별 환경 비교 (19시~23시)





02. 분석 과정 - 데이터 모니터링

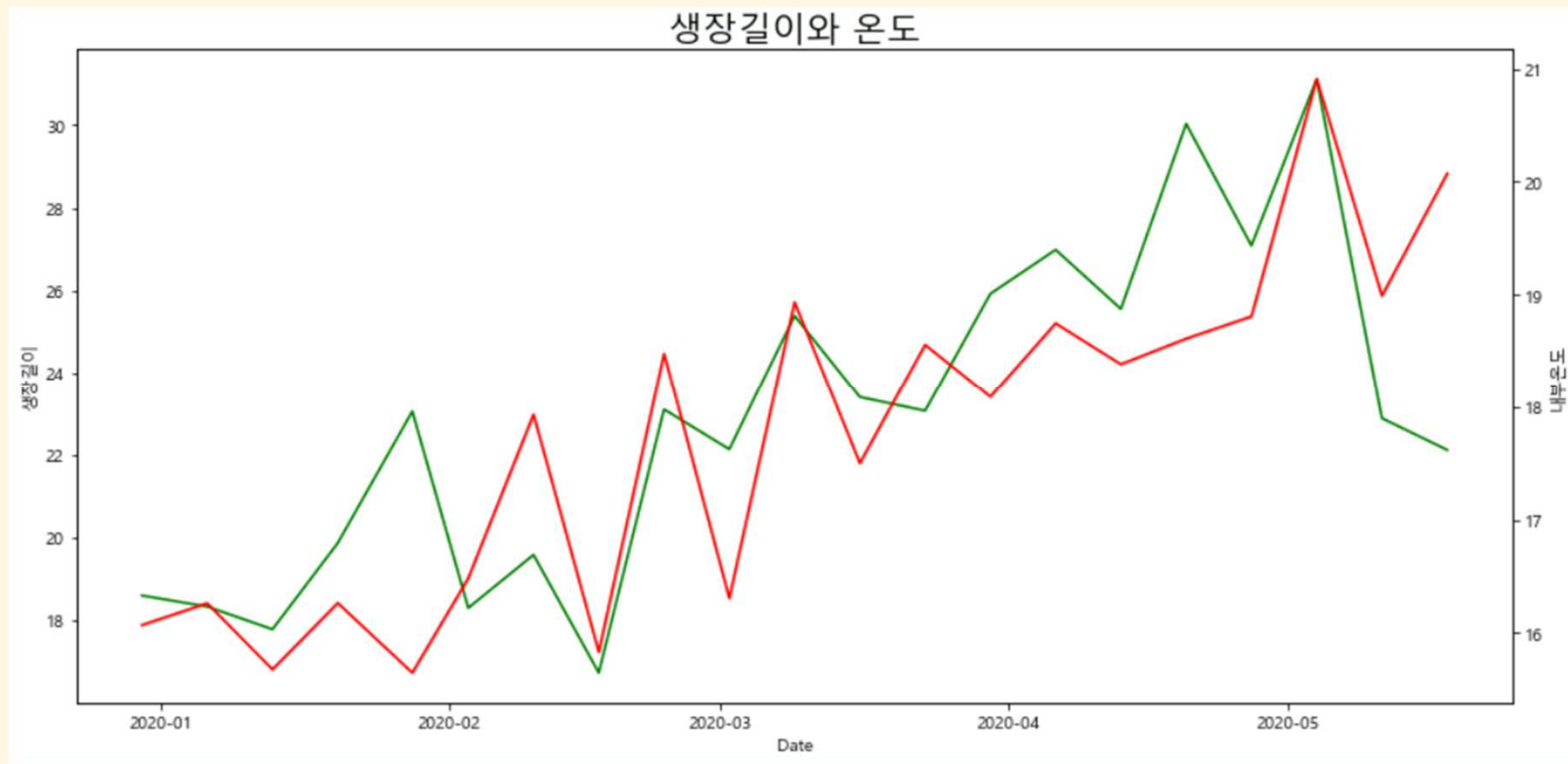
양액 관수량





02. 분석 과정 - 데이터 모니터링

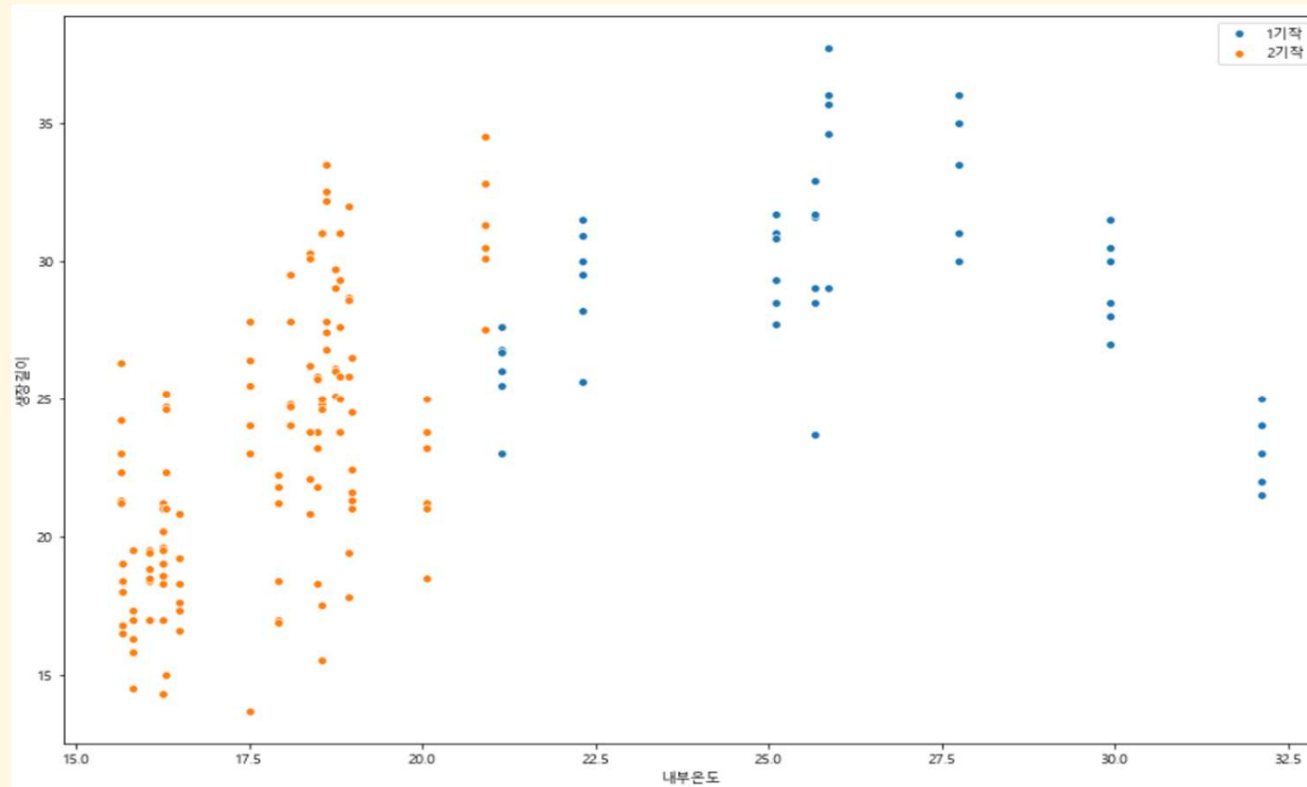
생장길이와 내부온도의 관계





02. 분석 과정 - 데이터 모니터링

1기작과 2기작 온도에 따른 생장길이



1기작 :
2019년 8~12월
2기작 :
2020년 1~6월

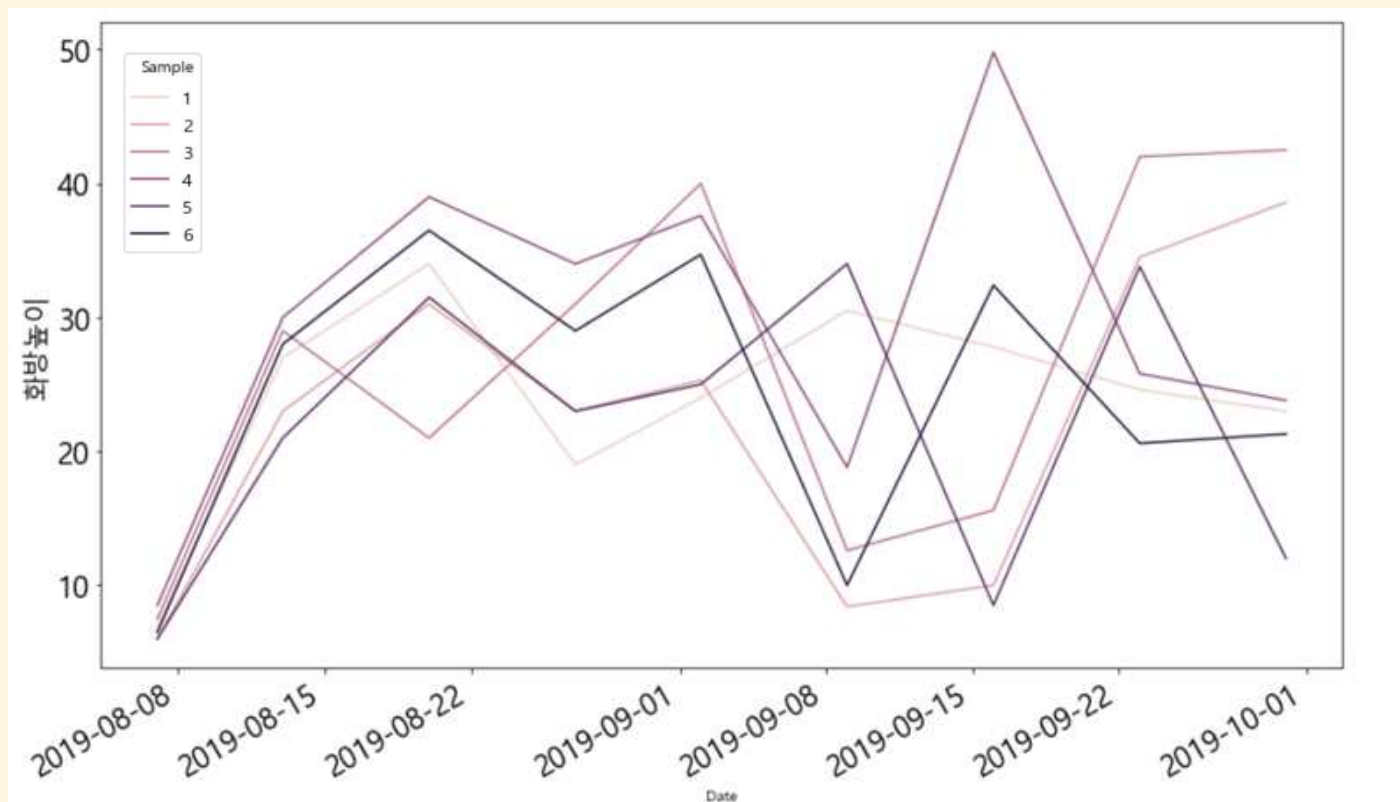
온도가 높은 **1기작** 때 1주일 생장 길이가 더 긴 것을 알 수 있다.





02. 분석 과정 - 데이터 모니터링

샘플별 화방높이





03

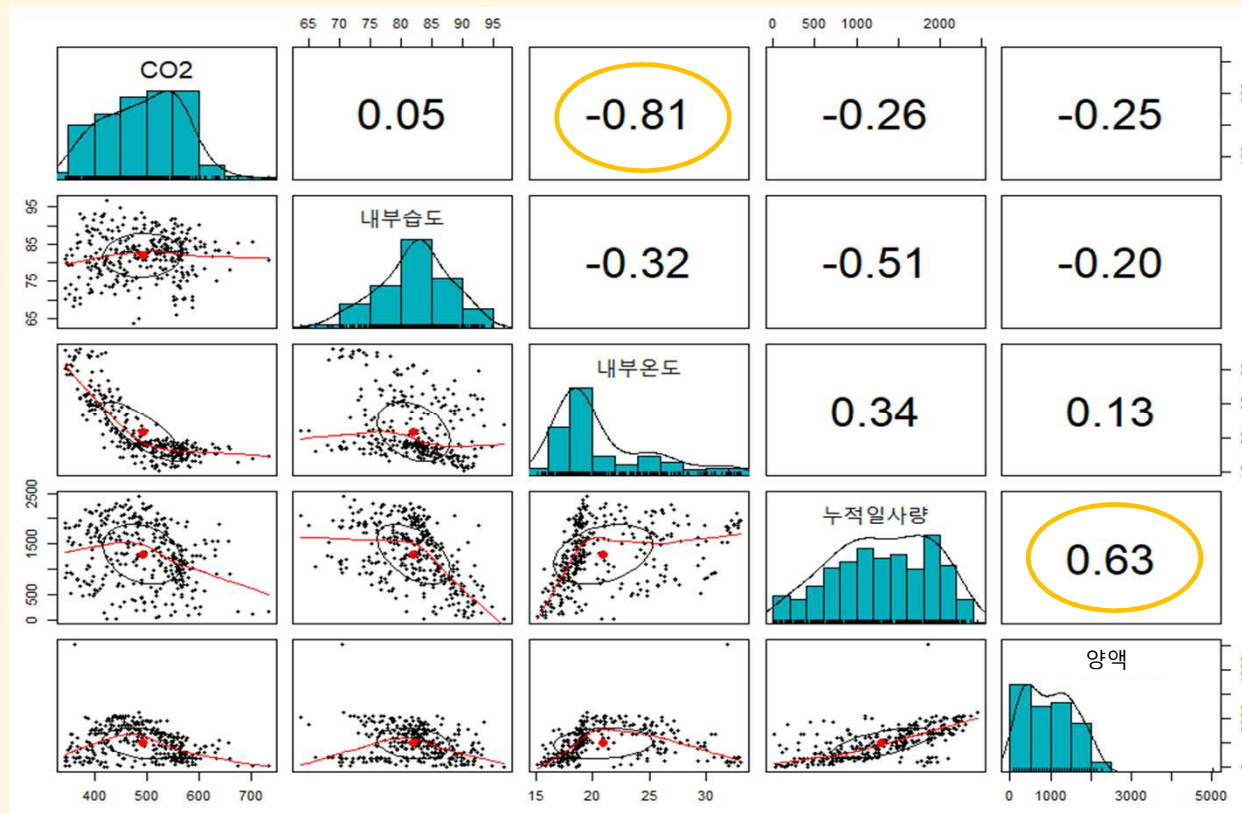
분석 및 결과

- ❖ 상관분석
- ❖ 생장길이 예측
- ❖ 수확과 검정



03. 분석 및 결과

외부요인 상관관계



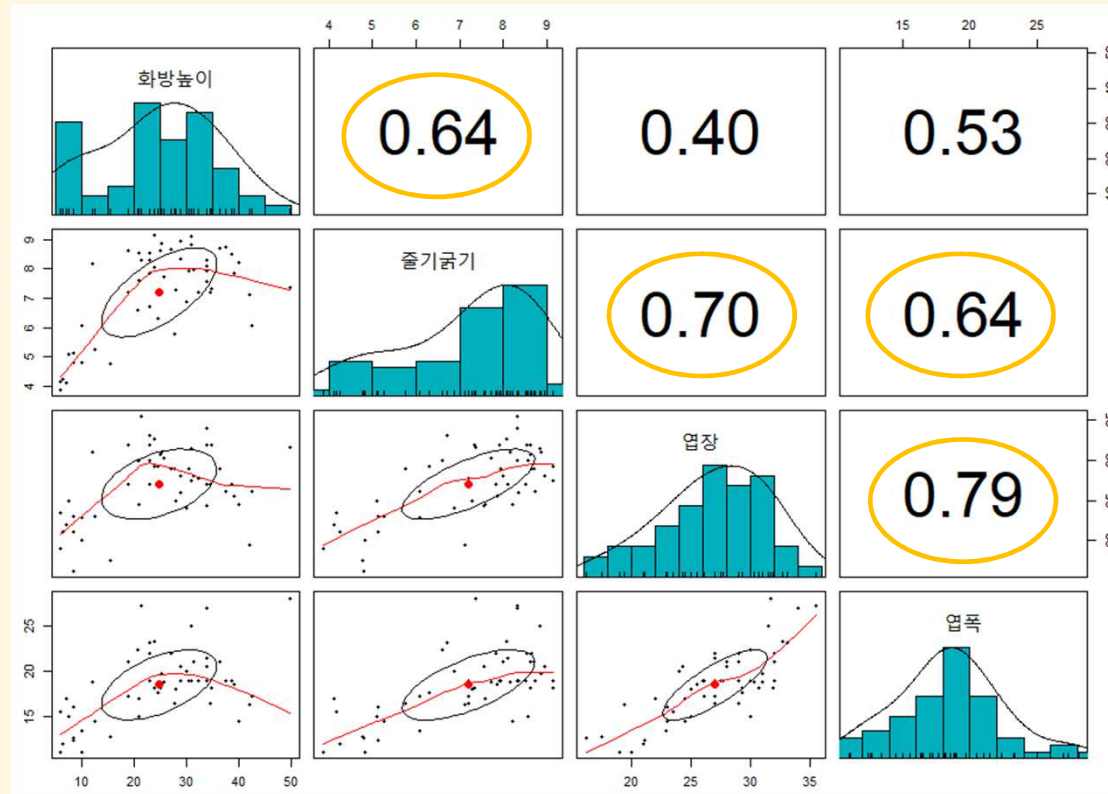
CO2~내부습도는 **음**의 상관관계를,
누적일사량~양액은 **양**의 상관관계를 갖는다.





03. 분석 및 결과

생육 상관관계



높은 상관 관계를 갖는다.

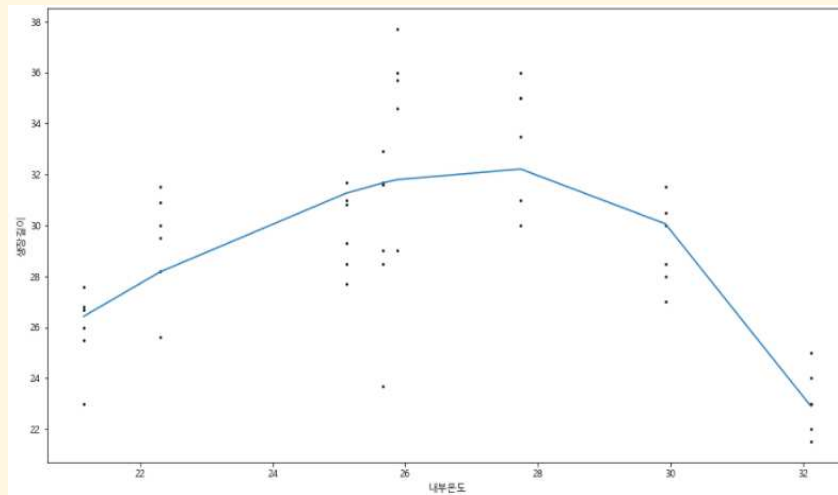




03. 분석 및 결과

다항회귀분석

1주 전 내부온도와 생장길이의 관계



약 $24^{\circ}\text{C} \sim 28^{\circ}\text{C}$ 에서
일주일간 생장 길이가 가장 길다.

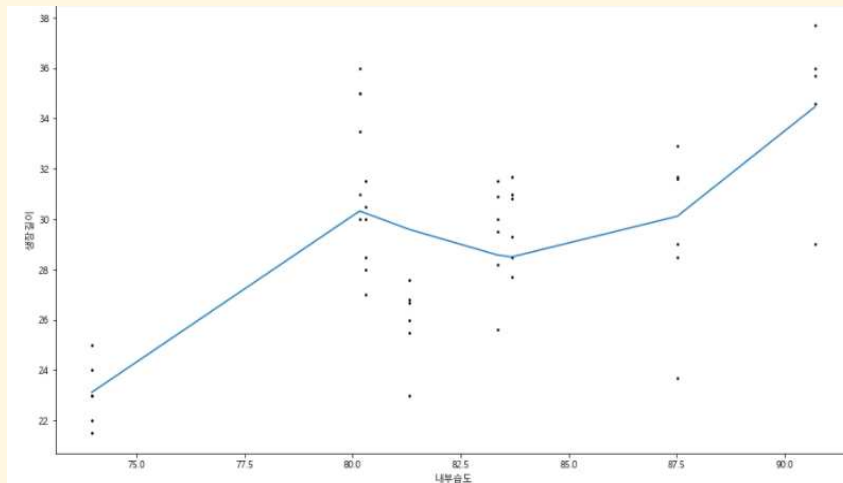




03. 분석 및 결과

다항회귀분석

1주 전 내부습도와 생장길이의 관계



습도가 높아질수록 생장길이가
길어지는 경향을 보인다.

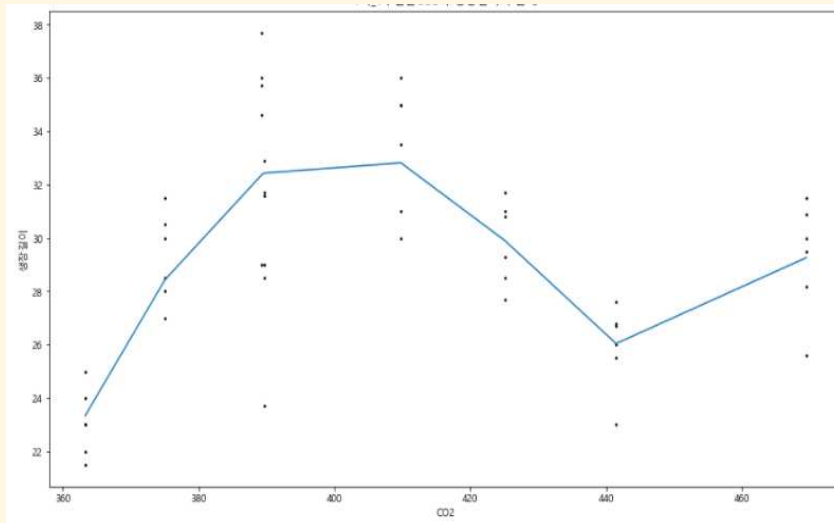




03. 분석 및 결과

다항회귀분석

1주 전 CO2와 생장길이의 관계



390~410에서
생장길이가 가장 길다.

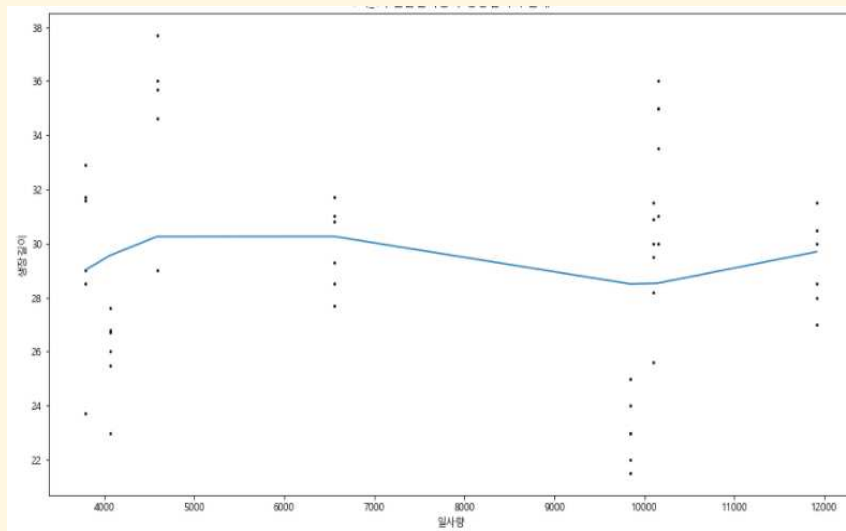




03. 분석 및 결과

다항회귀분석

1주 전 일사량과 생장길이의 관계



일사량은 생장길이에
영향을 미치지 않는다.





03. 분석 및 결과

생장길이 예측을 위한 데이터셋 구성

5주차

6주차

7주차

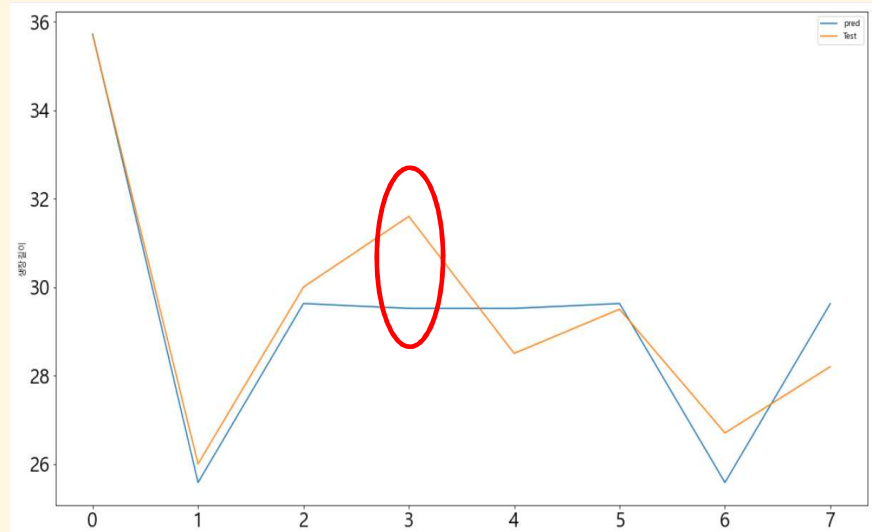
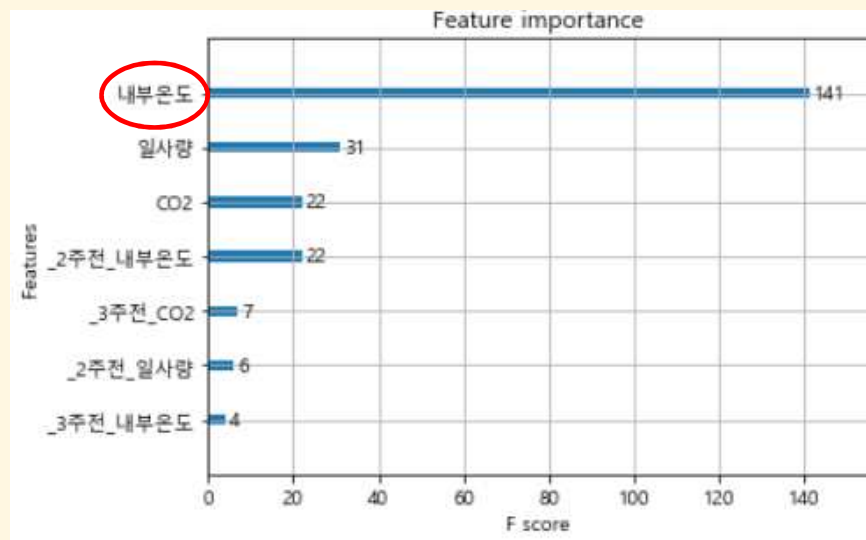
| 생장 길 이 | | 내부온도 | 내부습도 | CO2 | 일사 량 | 2주전_내 | 2주전_내 | 2주전_ | 2주전_ | 3주전_내 | 3주전_내 | 3주전_ | 3주전_ | 4주전_내 | 4주전_내 | 4주전_ | 4주전_ | tot_yang | 2주 전 | 3주 전 | 4주 전 |
|--------------|------|-----------|-----------|------------|---------|-----------|-----------|------------|---------|-----------|-----------|------------|---------|-----------|-----------|------------|---------|----------|-----------|-----------|-----------|
| | | | | | | 부온도 | 부습도 | _CO2 | 일사량 | 부온도 | 부습도 | _CO2 | 일사량 | 부온도 | 부습도 | _CO2 | 일사량 | | _tot_yang | _tot_yang | _tot_yang |
| 24 | 31.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 32.113484 | 73.962384 | 363.309606 | 9842.1 | 7019.1 | 7160.9 | 4679.4 | 2056.1 |
| 25 | 28.5 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 32.113484 | 73.962384 | 363.309606 | 9842.1 | 7019.1 | 7160.9 | 4679.4 | 2056.1 |
| 26 | 31.7 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 32.113484 | 73.962384 | 363.309606 | 9842.1 | 7019.1 | 7160.9 | 4679.4 | 2056.1 |
| 27 | 29.3 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 32.113484 | 73.962384 | 363.309606 | 9842.1 | 7019.1 | 7160.9 | 4679.4 | 2056.1 |
| 28 | 27.7 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 32.113484 | 73.962384 | 363.309606 | 9842.1 | 7019.1 | 7160.9 | 4679.4 | 2056.1 |
| 29 | 30.8 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 32.113484 | 73.962384 | 363.309606 | 9842.1 | 7019.1 | 7160.9 | 4679.4 | 2056.1 |
| 30 | 37.7 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 3970.4 | 7019.1 | 7160.9 | 4679.4 |
| 31 | 34.6 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 3970.4 | 7019.1 | 7160.9 | 4679.4 |
| 32 | 35.7 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 3970.4 | 7019.1 | 7160.9 | 4679.4 |
| 33 | 39.4 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 3970.4 | 7019.1 | 7160.9 | 4679.4 |
| 34 | 36.0 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 3970.4 | 7019.1 | 7160.9 | 4679.4 |
| 35 | 29.0 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 29.921131 | 80.302083 | 375.036072 | 11912.0 | 3970.4 | 7019.1 | 7160.9 | 4679.4 |
| 36 | 28.5 | 25.667758 | 87.516195 | 389.678028 | 3791.2 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 8862.5 | 3970.4 | 7019.1 | 7160.9 |
| 37 | 31.6 | 25.667758 | 87.516195 | 389.678028 | 3791.2 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 8862.5 | 3970.4 | 7019.1 | 7160.9 |
| 38 | 32.9 | 25.667758 | 87.516195 | 389.678028 | 3791.2 | 25.873094 | 90.706793 | 389.284788 | 4588.0 | 25.109201 | 83.681134 | 425.112847 | 6559.8 | 27.740427 | 80.161706 | 409.793155 | 10149.5 | 8862.5 | 3970.4 | 7019.1 | 7160.9 |





03. 분석 및 결과

XGBOOST



가장 높은 영향력을 미치는 변수는 측정 1주일전의 내부온도이다.
성능 : 가장 큰 오차는 약 2cm이다.



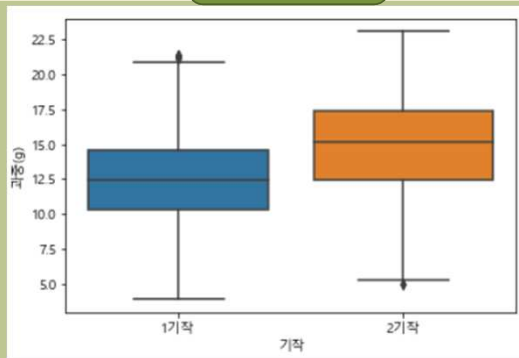


03. 분석 및 결과

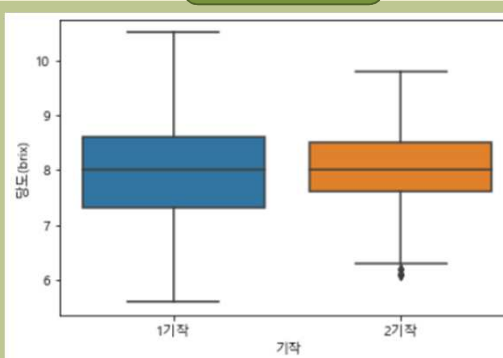
1기작 vs 2기작 수확과 : 이(理)분산 독립 표본 T검정

귀무가설 : 1기작과 2기작의 과중/당도/산도는 각각 차이가 없다.

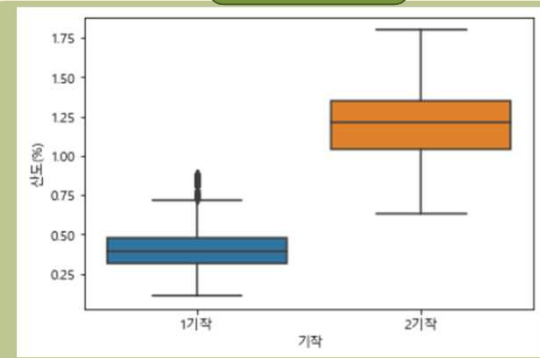
과중



당도



산도



| | |
|---------|----------------|
| n | ≥30 |
| 등분산 | 불만족 |
| p-value | 1.799e-65 (기각) |
| 결론 | 평균의 차이 有 |

| | |
|---------|-------------|
| n | ≥30 |
| 등분산 | 불만족 |
| p-value | 0.2945 (채택) |
| 결론 | 평균의 차이 無 |

| | |
|---------|-----------------|
| n | ≥30 |
| 등분산 | 불만족 |
| p-value | 2.598e-260 (기각) |
| 결론 | 평균의 차이 有 |

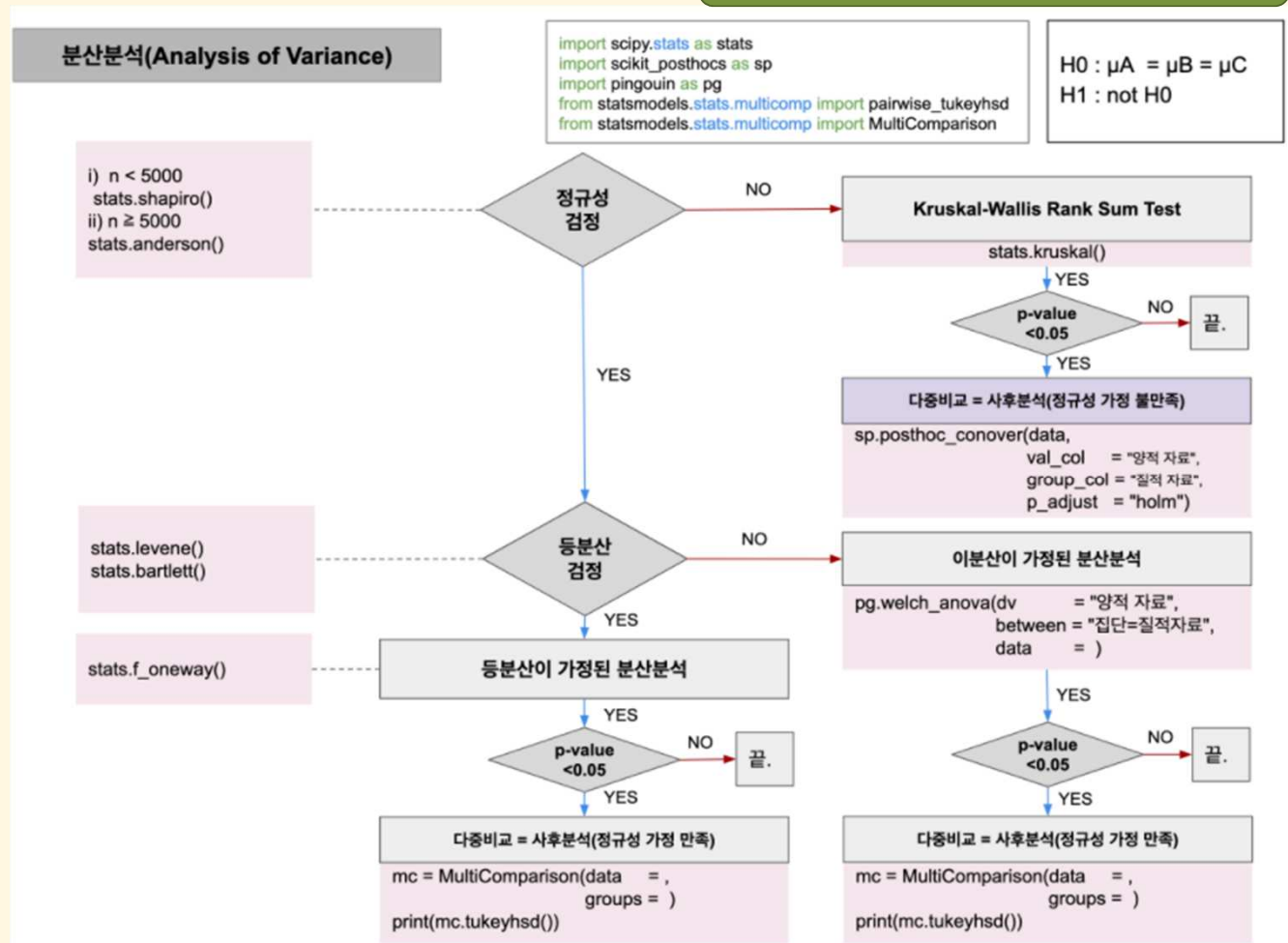




03. 분석 및 결과

* 분산분석 로드맵

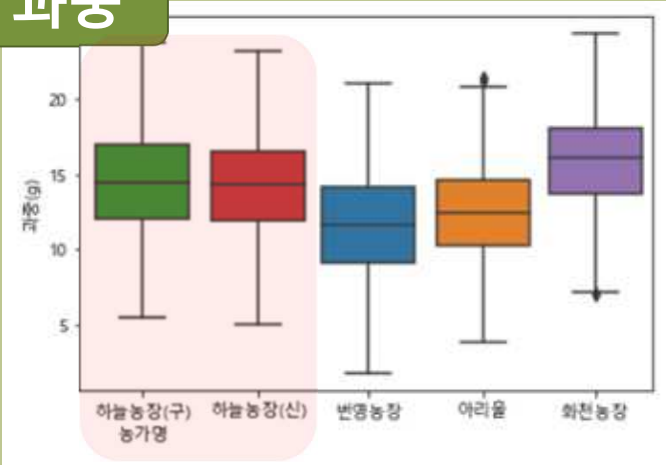
집단이 3개 이상일 때





03. 분석 및 결과

과중



| | |
|---------|----------------|
| n | ≥30 |
| 등분산 | 불만족 |
| p-value | 5.574e-24 (기각) |
| 결론 | 평균의 차이 有 |

Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---------|---------|----------|--------|---------|--------|--------|
| 번영농장 | 아리울 | 1.0122 | 0.001 | 0.6055 | 1.4189 | True |
| 번영농장 | 하늘농장(구) | 3.0212 | 0.001 | 2.5992 | 3.4431 | True |
| 번영농장 | 하늘농장(신) | 2.7742 | 0.001 | 2.3573 | 3.1912 | True |
| 번영농장 | 화천농장 | 4.3331 | 0.001 | 3.9594 | 4.7068 | True |
| 아리울 | 하늘농장(구) | 2.009 | 0.001 | 1.5727 | 2.4452 | True |
| 아리울 | 하늘농장(신) | 1.762 | 0.001 | 1.3306 | 2.1935 | True |
| 아리울 | 화천농장 | 3.3209 | 0.001 | 2.9311 | 3.7107 | True |
| 하늘농장(구) | 하늘농장(신) | -0.2469 | 0.5477 | -0.6928 | 0.1989 | False |
| 하늘농장(구) | 화천농장 | 1.3119 | 0.001 | 0.9062 | 1.7176 | True |
| 하늘농장(신) | 화천농장 | 1.5588 | 0.001 | 1.1584 | 1.9593 | True |

귀무가설 : 각 농가간 과중의 평균차이가 없다.
대립가설 : 각 농가간 과중의 평균차이가 있다.

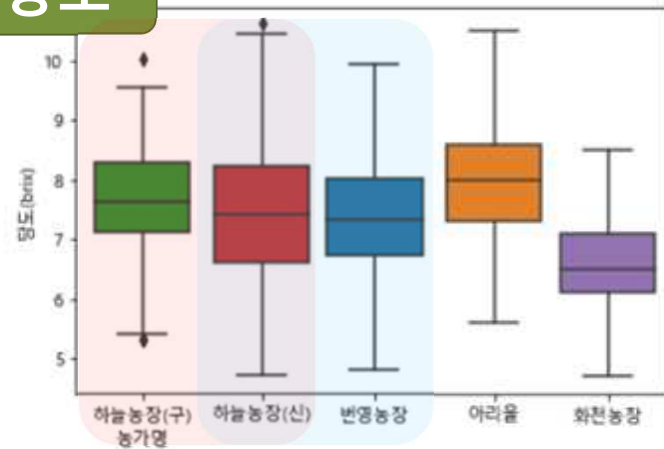
Reject = False일 경우 귀무가설을 기각하지 못한다는 뜻이므로 평균차이가 없다는 것을 말한다.





03. 분석 및 결과

당도



| | |
|---------|----------------|
| n | ≥30 |
| 등분산 | 불만족 |
| p-value | 7.688e-84 (기각) |
| 결론 | 평균의 차이 有 |

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---------|---------|----------|--------|---------|---------|--------|
| 번영농장 | 아리울 | 0.6522 | 0.001 | 0.4479 | 0.8564 | True |
| 번영농장 | 하늘농장(구) | 0.3195 | 0.001 | 0.0975 | 0.5415 | True |
| 번영농장 | 하늘농장(신) | 0.1438 | 0.3884 | -0.0772 | 0.3648 | False |
| 번영농장 | 화천농장 | -0.7681 | 0.001 | -0.96 | -0.5763 | True |
| 아리울 | 하늘농장(구) | -0.3327 | 0.001 | -0.5551 | -0.1102 | True |
| 아리울 | 하늘농장(신) | -0.5084 | 0.001 | -0.7298 | -0.287 | True |
| 아리울 | 화천농장 | -1.4203 | 0.001 | -1.6127 | -1.228 | True |
| 하늘농장(구) | 하늘농장(신) | -0.1757 | 0.2581 | -0.4136 | 0.0622 | False |
| 하늘농장(구) | 화천농장 | -1.0876 | 0.001 | -1.2988 | -0.8765 | True |
| 하늘농장(신) | 화천농장 | -0.9119 | 0.001 | -1.1219 | -0.7019 | True |

귀무가설 : 각 농가간 과중의 평균차이가 없다.
대립가설 : 각 농가간 과중의 평균차이가 있다.

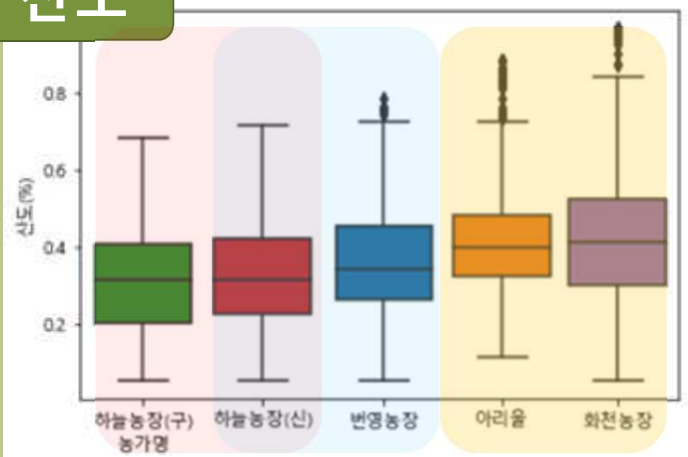
Reject = False일 경우 귀무가설을 기각하지 못한다는 뜻이므로 평균차이가 없다는 것을 말한다.





03. 분석 및 결과

산도



| | |
|---------|----------------|
| n | ≥30 |
| 등분산 | 불만족 |
| p-value | 5.574e-24 (기각) |
| 결론 | 평균의 차이 有 |

Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---------|---------|----------|--------|---------|---------|--------|
| 번영농장 | 아리울 | 0.0688 | 0.001 | 0.0333 | 0.1044 | True |
| 번영농장 | 하늘농장(구) | -0.0432 | 0.0151 | -0.0808 | -0.0056 | True |
| 번영농장 | 하늘농장(신) | -0.029 | 0.2139 | -0.0665 | 0.0084 | False |
| 번영농장 | 화천농장 | 0.0682 | 0.001 | 0.0351 | 0.1013 | True |
| 아리울 | 하늘농장(구) | -0.112 | 0.001 | -0.1504 | -0.0736 | True |
| 아리울 | 하늘농장(신) | -0.0979 | 0.001 | -0.1362 | -0.0596 | True |
| 아리울 | 화천농장 | -0.0006 | 0.9 | -0.0347 | 0.0334 | False |
| 하늘농장(구) | 하늘농장(신) | 0.0141 | 0.8606 | -0.0261 | 0.0544 | False |
| 하늘농장(구) | 화천농장 | 0.1114 | 0.001 | 0.0752 | 0.1476 | True |
| 하늘농장(신) | 화천농장 | 0.0972 | 0.001 | 0.0612 | 0.1333 | True |

귀무가설 : 각 농가간 과중의 평균차이가 없다.
대립가설 : 각 농가간 과중의 평균차이가 있다.

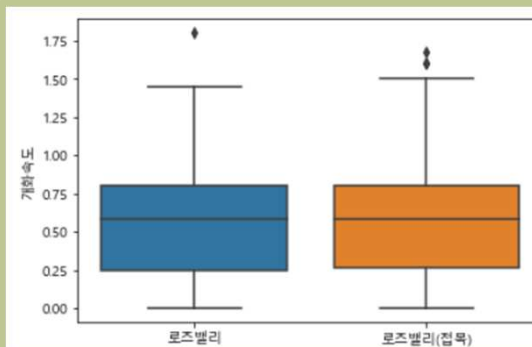
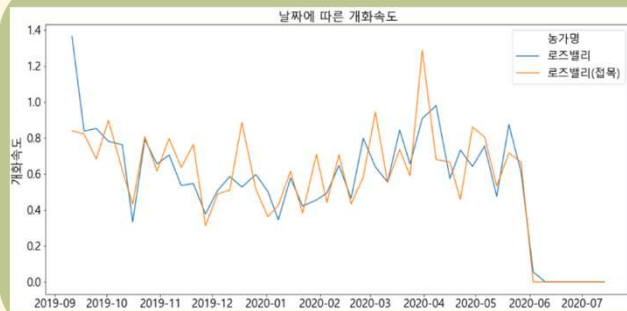
Reject = False일 경우 귀무가설을 기각하지 못한다는 뜻이므로 평균차이가 없다는 것을 말한다.



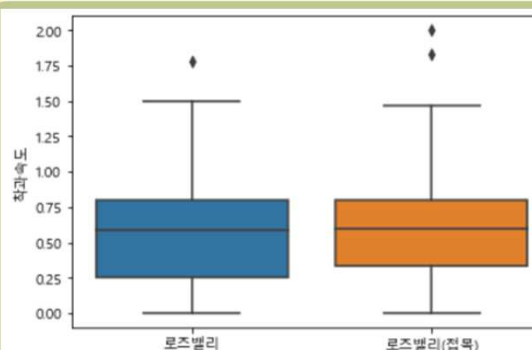


03. 분석 및 결과 수확과

일반묘 vs 접목묘의 개화속도와 착과속도 차이 검정



| | |
|---------|-------------|
| n | ≥30 |
| 등분산 | 만족 |
| p-value | 0.9895 (채택) |
| 결론 | 평균 차이 無 |



| | |
|---------|-------------|
| n | ≥30 |
| 등분산 | 만족 |
| p-value | 0.8420 (채택) |
| 결론 | 평균 차이 無 |





04

활용

- ❖ 활용 방안
- ❖ 개선 방안



04. 활용

활용 방안 및 개선 방안

1. 활용방안

- 생육 예측 모델에 일기예보 자료와 습도, CO2를 넣음으로써 생육을 예측할 수 있다.
- 농민이 의도하는 생육을 위한 환경을 테스트해 볼 수 있다.
- 선도 농가와 환경 비교를 통해 환경 관리

2. 개선 방안

생육데이터

- 데이터 표준화
- 측정 값 입력 정확도 향상

환경데이터

- 더 정확한 센서 사용
- 데이터 표준화
- 측정주기 통일



감사합니다



Q & A

