

ALL_DATA(범주형_시간추가_결측치열모두제거) (1).CSV

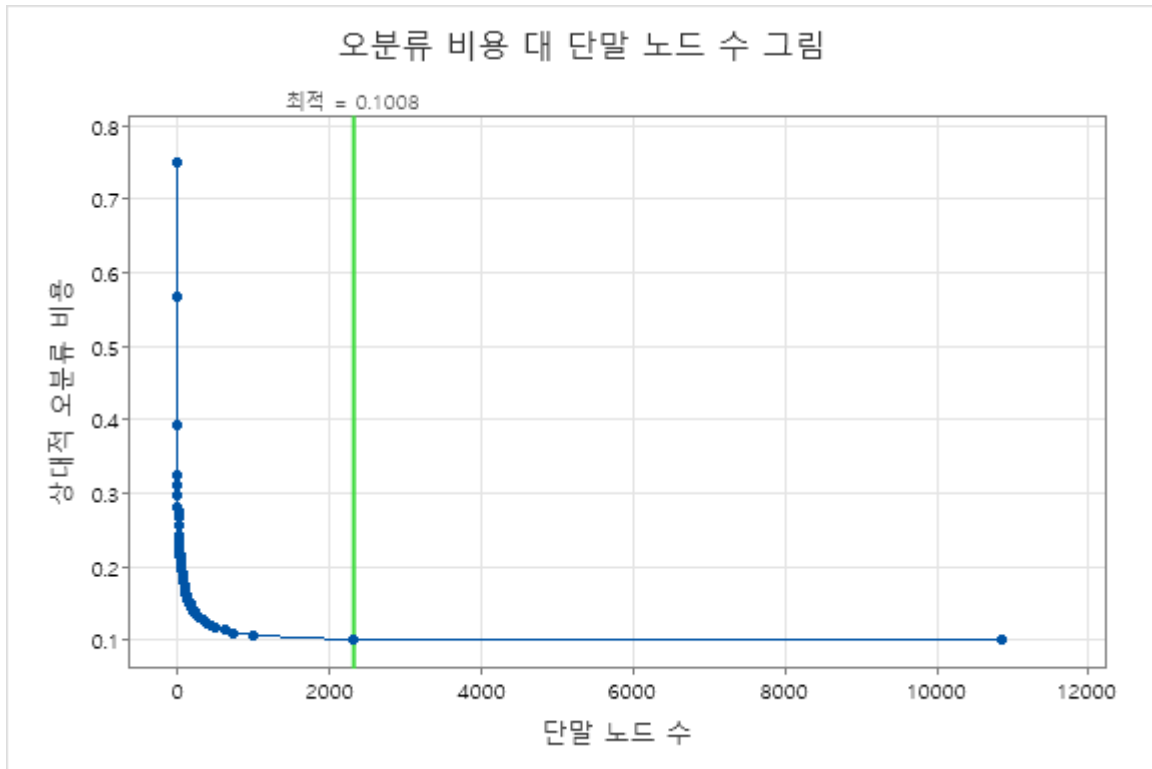
2318 노드 CART® 분류: 전력거래량_범주 대 기온(°C), 풍속(m/s), 풍향(16방위), 습도(%), 증기압(hPa), 이슬점온도(°C), 현지기압(hPa), 해면기압(hPa), 전운량(10분위), 중하층운량(10분위), 시정(10m), 지면온도(°C), 년도, 요일, 월, 지역, 시간, 일

방법

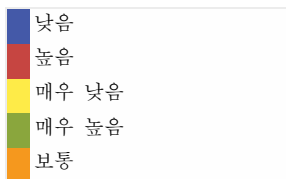
사전 확률	모든 등급에 대해 같음
노드 분할	지니
최적 트리	최소 오분류 비용의 1 표준 오차 이내
모형 검증	70/30% 교육/검정 집합
사용된 행	973458
사용되지 않은 행	5358

다항 반응 정보

변수	등급	교육		검정	
		카운트	%	카운트	%
전력거래량_범주	낮음	182296	26.75	78251	26.80
	높음	55001	8.07	23594	8.08
	매우 낮음	269750	39.59	115672	39.61
	매우 높음	6470	0.95	2772	0.95
	보통	167916	24.64	71736	24.57
	모두	681433	100.00	292025	100.00



최적 트리 다이어그램



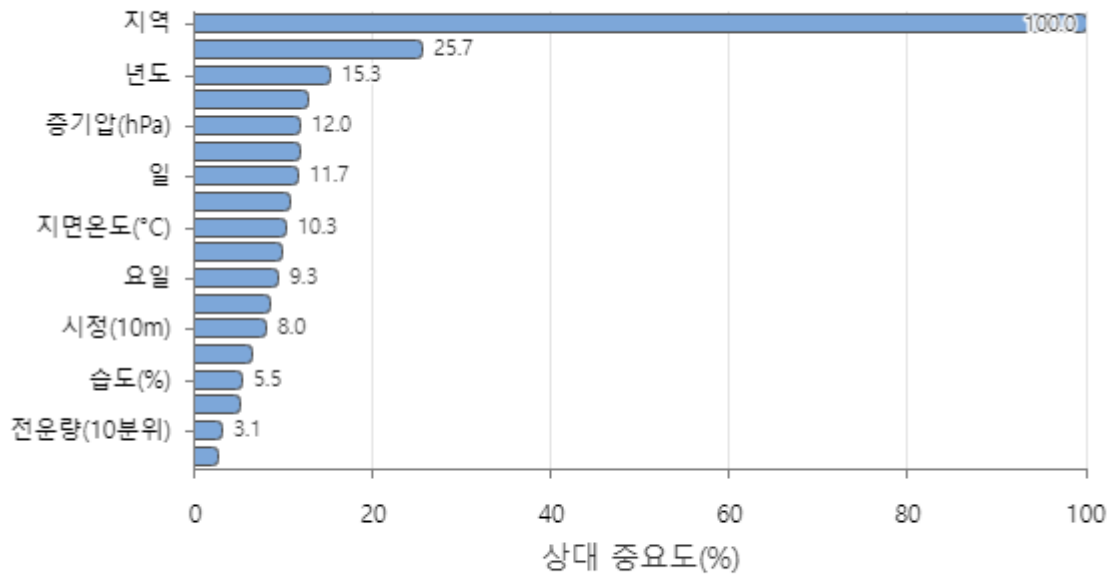
모형 요약

전체 예측 변수 18
 중요 예측 변수 18
 단말 노드 수 2318
 최소 단말 노드 크기 3
 오분류 비용(교육) 0.0525
 오분류 비용(검정) 0.1008

전력거래량_범

주 = 수준	ROC 곡선 아래 면적(교육)	ROC 곡선 아래 면적(검정)
대 ≠ 수준		
낮음	0.9899	0.9882
높음	0.9966	0.9902
매우 낮음	0.9987	0.9986
매우 높음	0.9996	0.9371
보통	0.9898	0.9863

상대 변수 중요도



변수 중요도는 예측 변수에 분할을 적용할 때 모형 개선을 측정합니다. 상대 중요도는 상위 예측 변수를 기준으로 한 % 개선으로 정의됩니다.

오차 행렬

예측된 등급(교육)

실제 등급	카운트	낮음	높음	매우 낮음	매우 높음	보통	정답률(%)
낮음	182296	171536	128	1496	1	9135	94.1
높음	55001	0	52536	0	1224	1241	95.5
매우 낮음	269750	6380	4	263354	0	12	97.6
매우 높음	6470	0	6	0	6464	0	99.9
보통	167916	6184	7414	7	70	154241	91.9
모두	681433	184100	60088	264857	7759	164629	95.1

예측된 등급(검정)

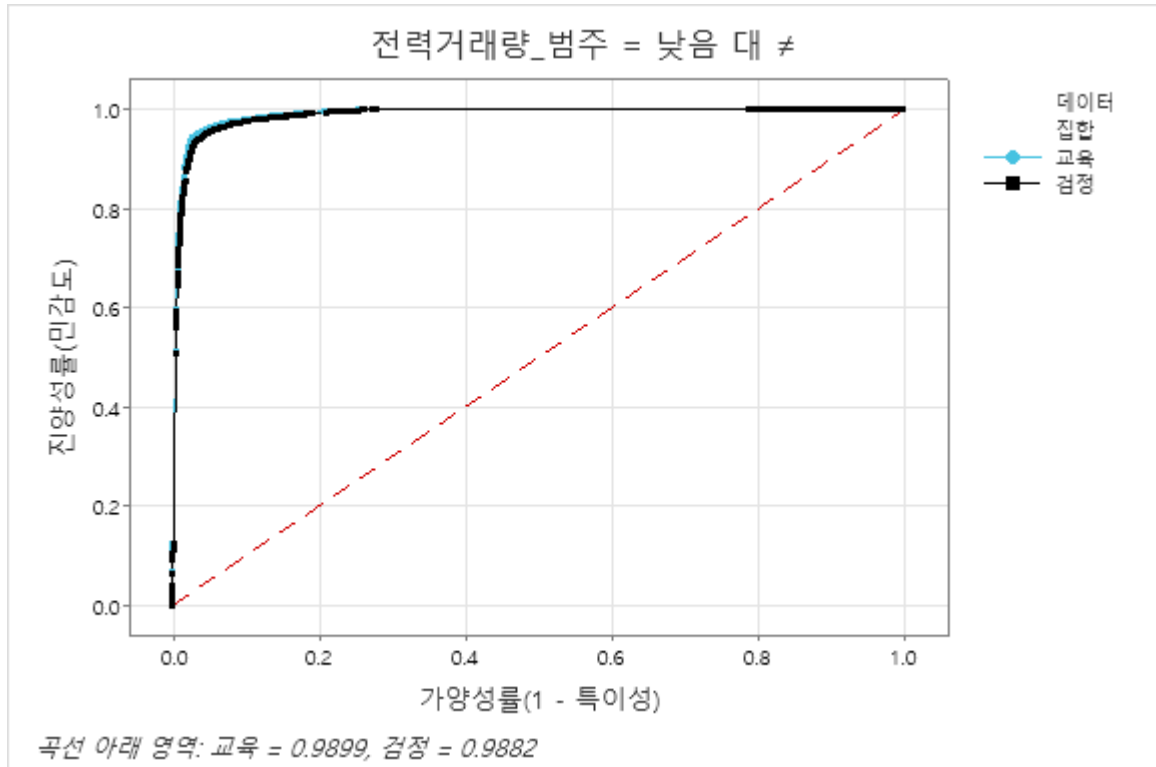
실제 등급	카운트	낮음	높음	매우 낮음	매우 높음	보통	정답률(%)
낮음	78251	72723	59	878	0	4591	92.9
높음	23594	0	21629	0	886	1079	91.7
매우 낮음	115672	2913	2	112756	0	1	97.5
매우 높음	2772	0	354	0	2418	0	87.2
보통	71736	3213	3675	2	29	64817	90.4
모두	292025	78849	25719	113636	3333	70488	93.9

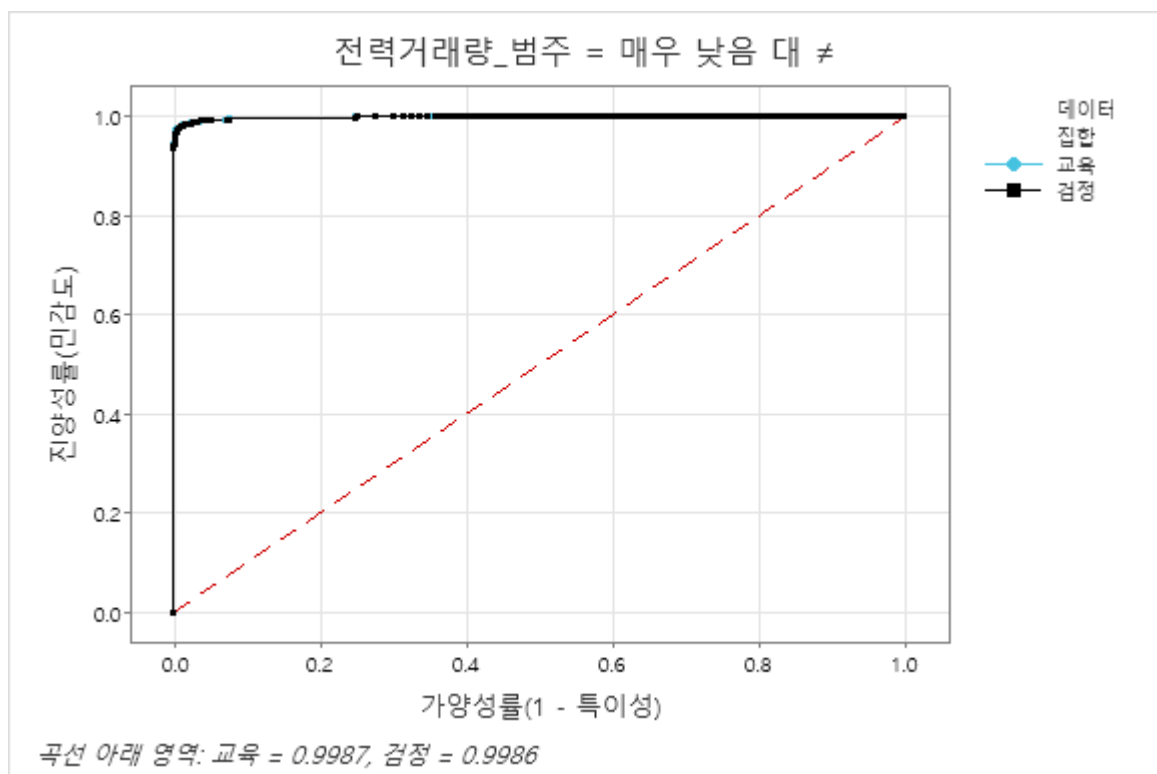
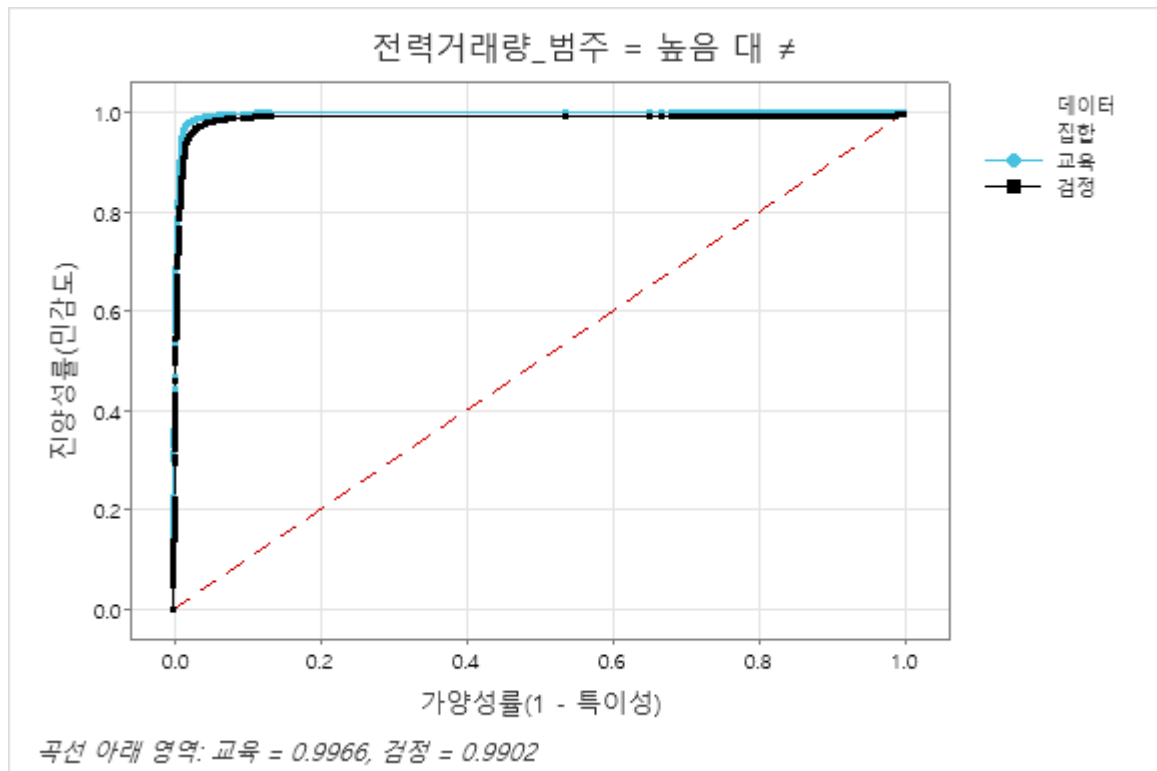
오분류

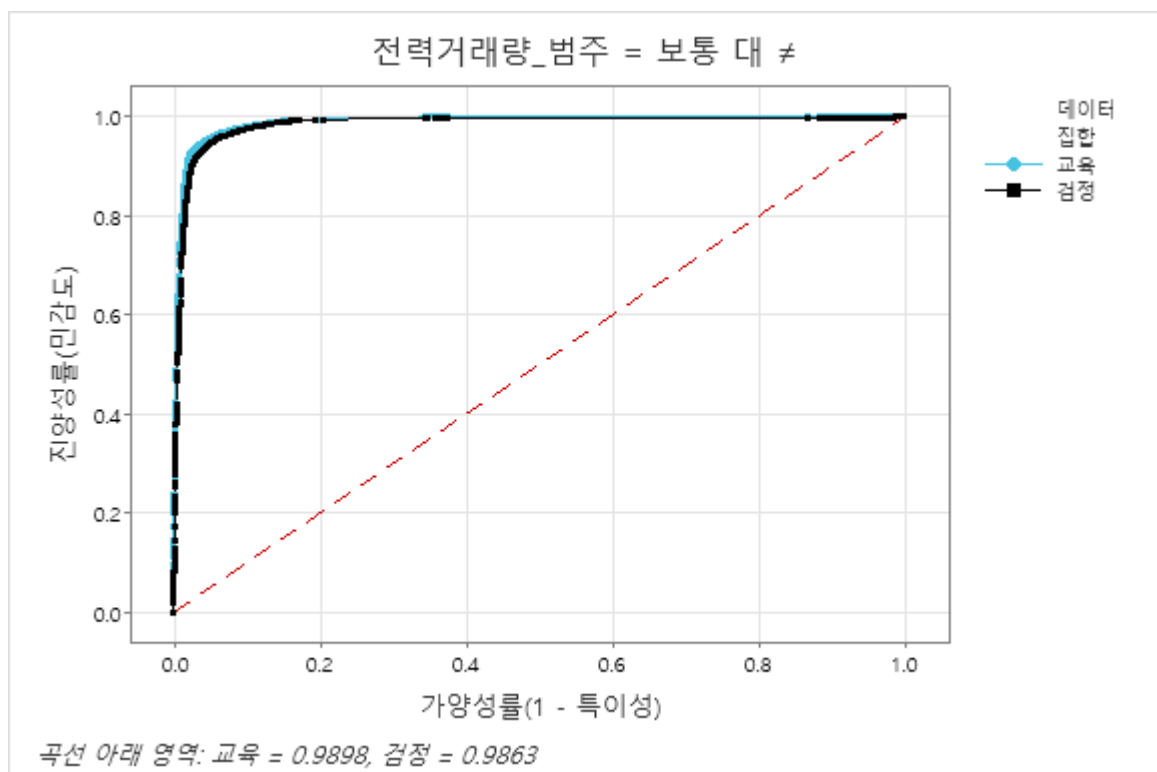
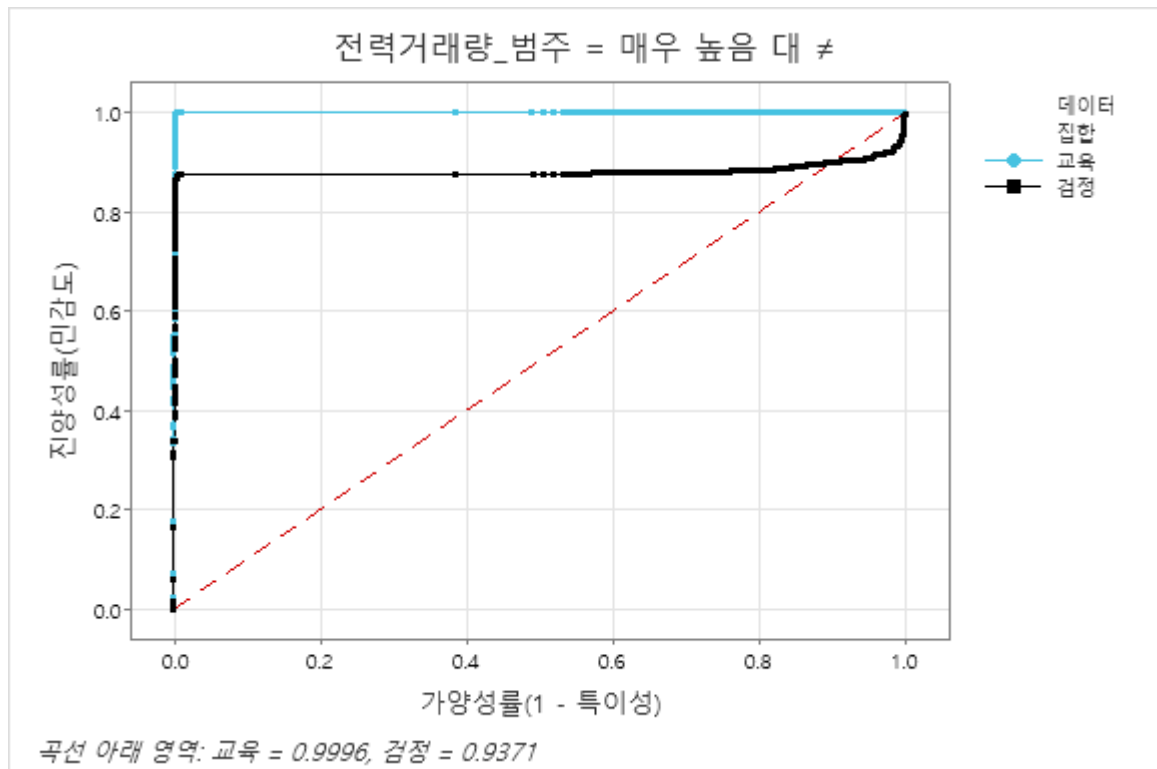
입력 오분류	비용	예측된 등급				
실제 등급		낮음	높음	매우 낮음	매우 높음	보통
낮음			1.00	1.00	1.00	1.00
높음		1.00		1.00	1.00	1.00
매우 낮음		1.00	1.00		1.00	1.00
매우 높음		1.00	1.00	1.00		1.00
보통		1.00	1.00	1.00	1.00	

실제	등급	교육			비용	검정			비용
		카운트	오분류됨	오차율(%)		카운트	오분류됨	오차율(%)	
낮음		182296	10760	5.9	0.0590	78251	5528	7.1	0.0706
높음		55001	2465	4.5	0.0448	23594	1965	8.3	0.0833
매우 낮음	낮음	269750	6396	2.4	0.0237	115672	2916	2.5	0.0252
매우 높음	높음	6470	6	0.1	0.0009	2772	354	12.8	0.1277
보통		167916	13675	8.1	0.0814	71736	6919	9.6	0.0965
모두		681433	33302	4.9	0.0420	292025	17682	6.1	0.0807

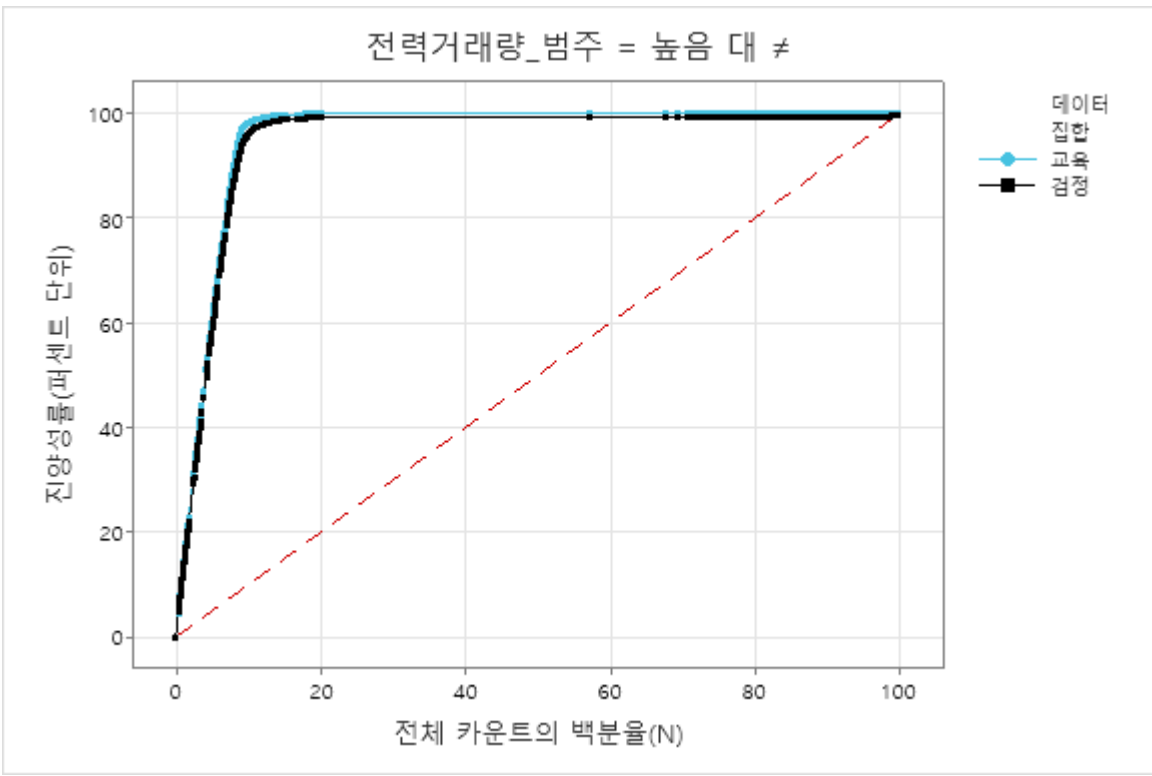
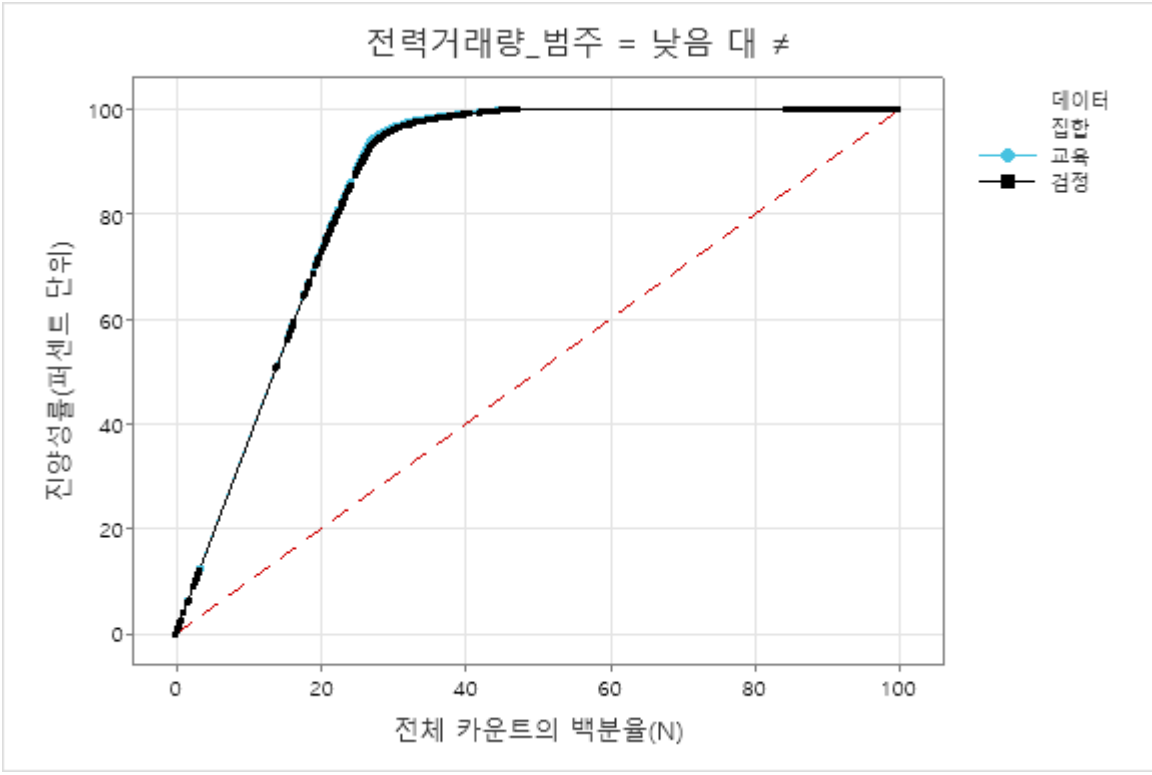
수신자 검사 특성(ROC) 곡선

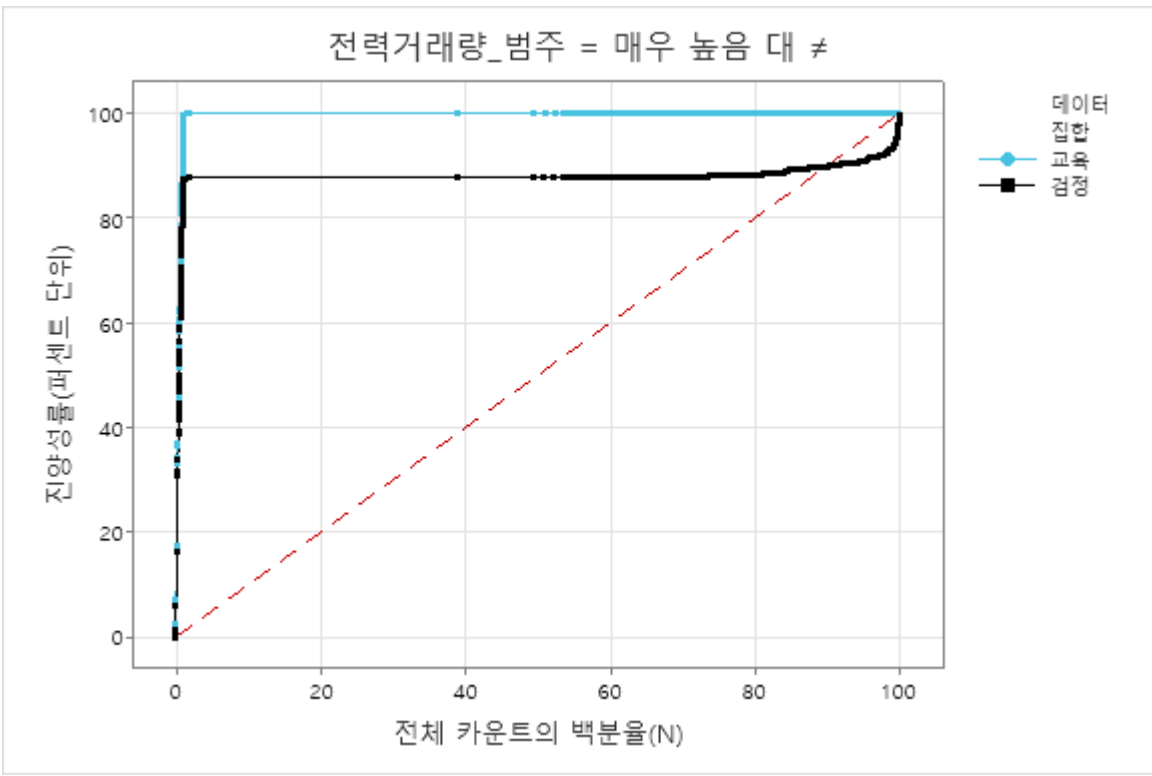
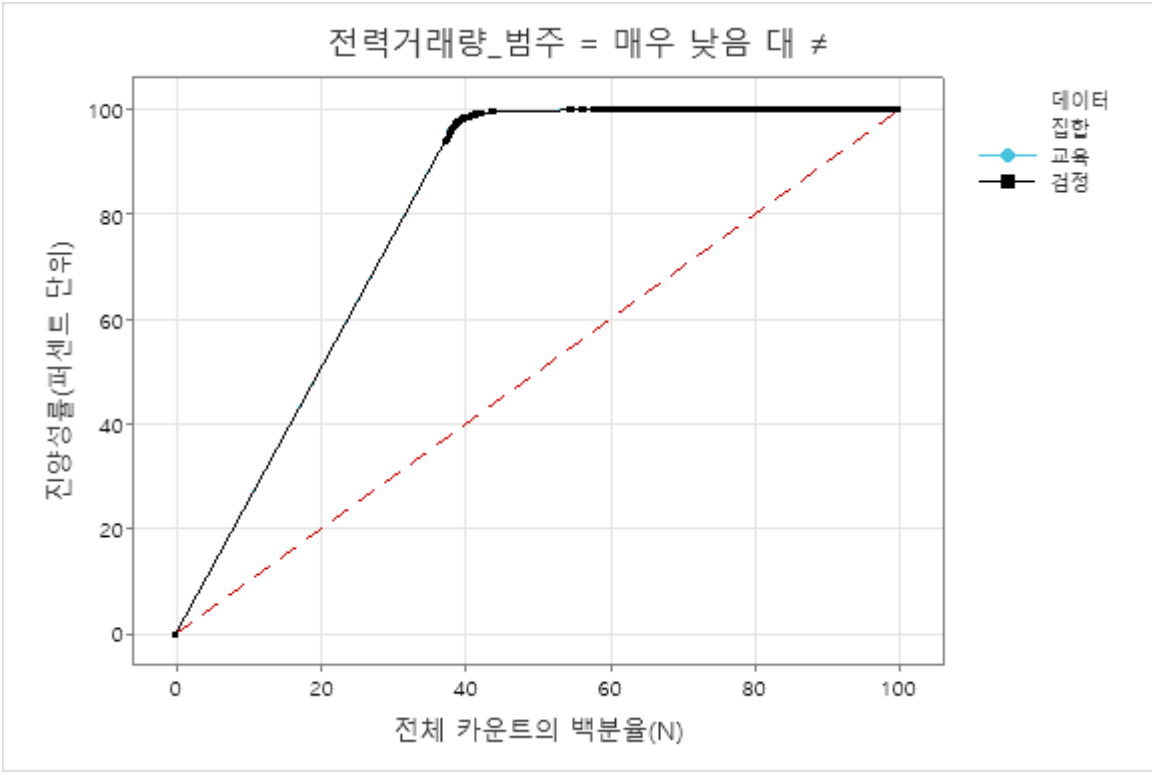


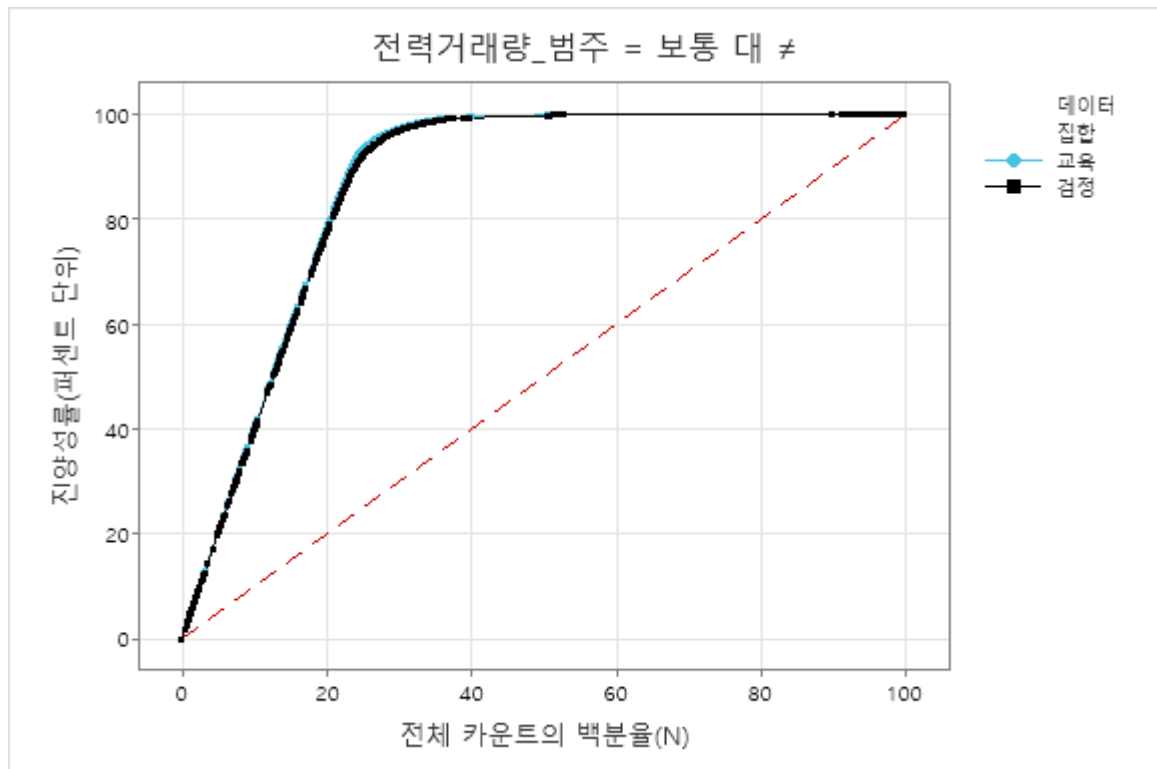




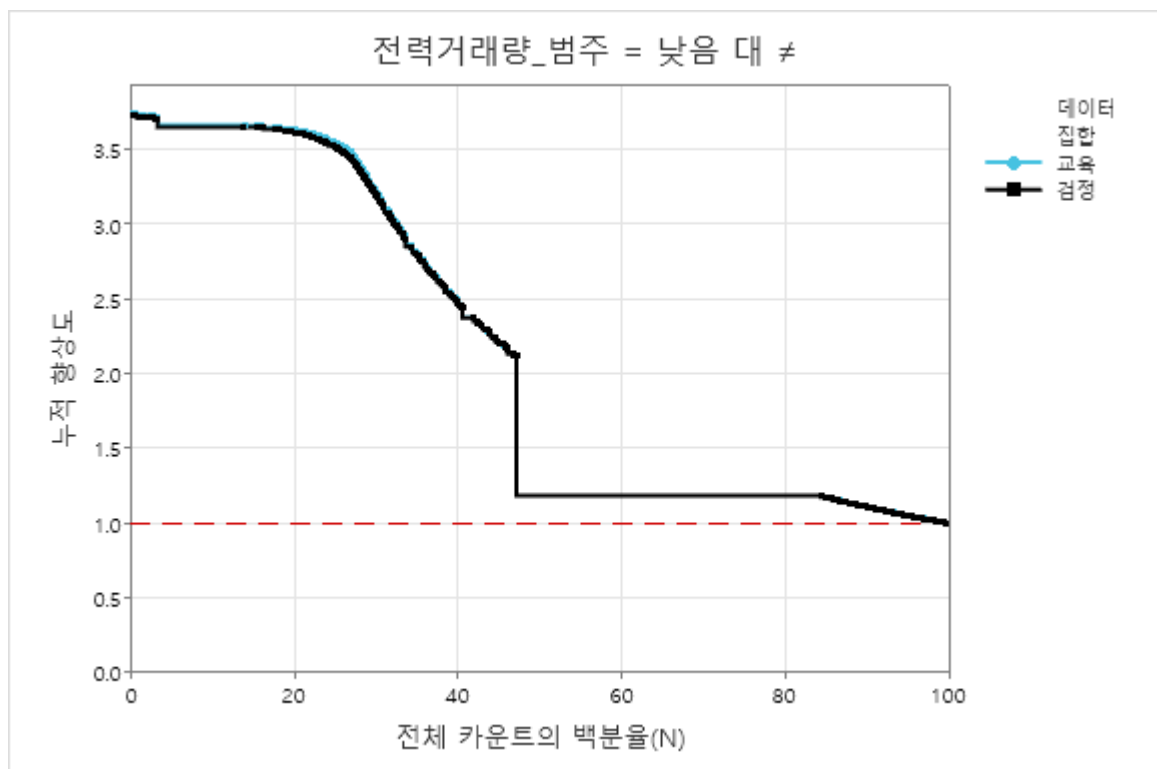
이익 차트

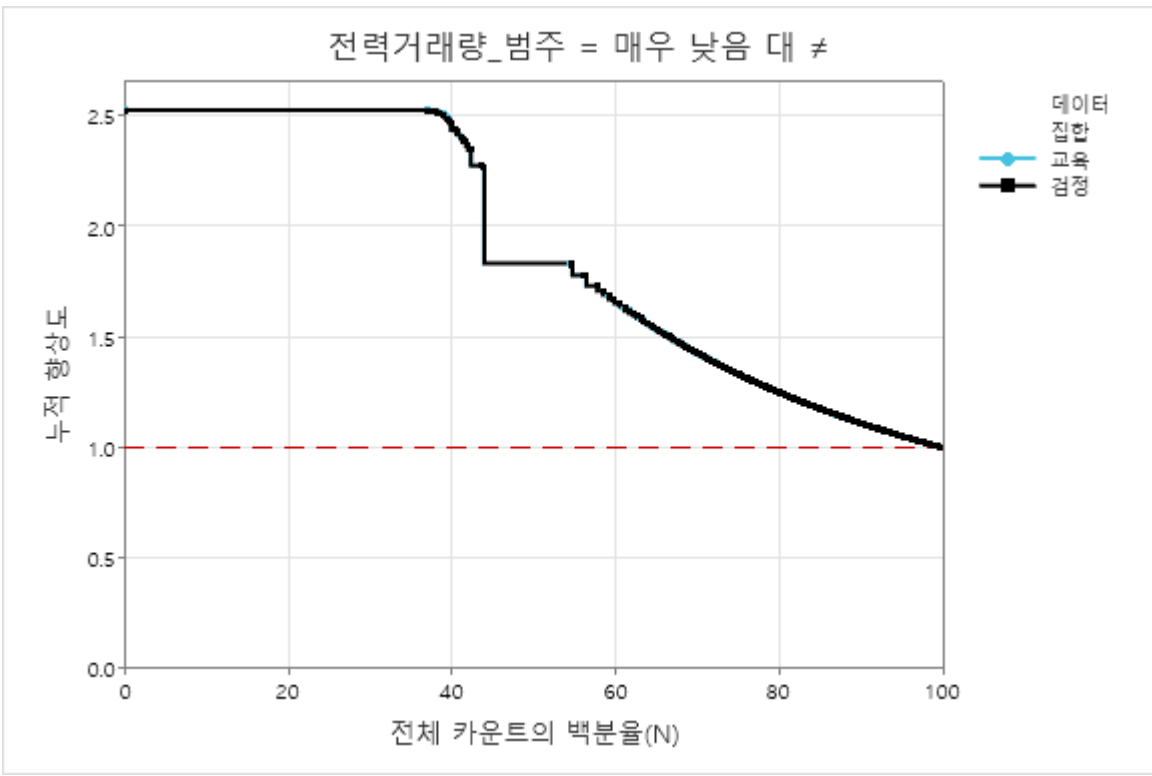
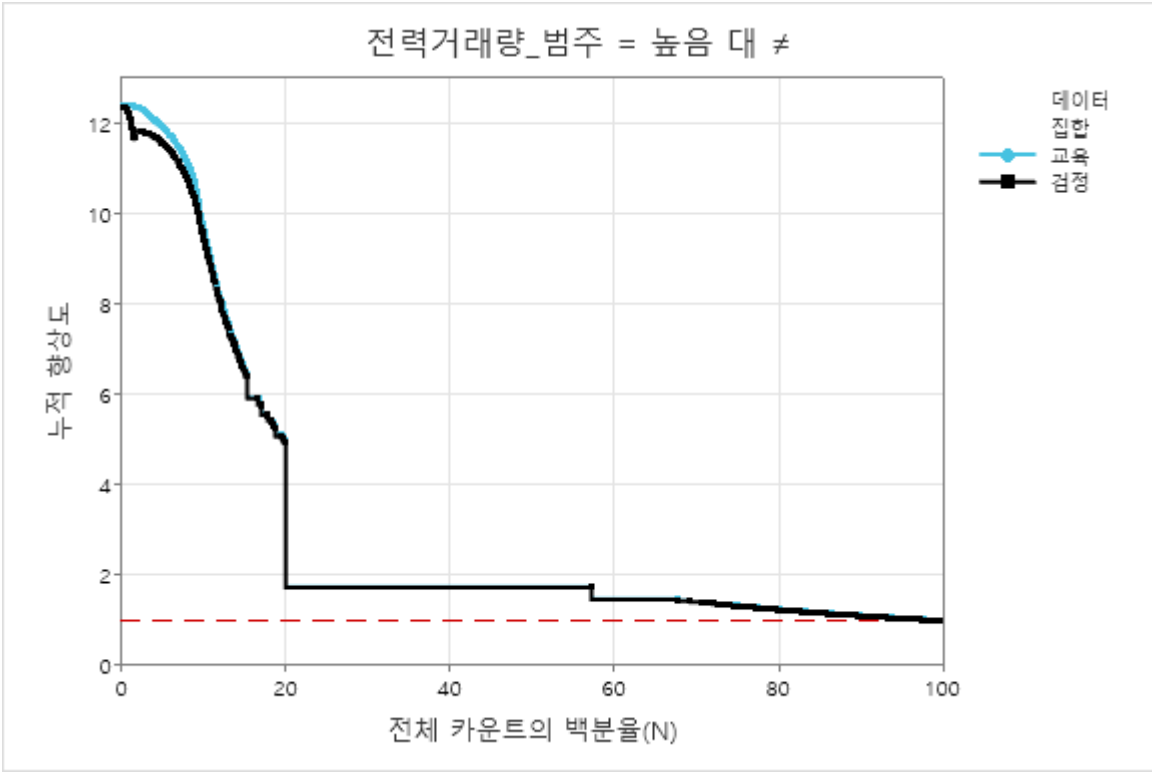


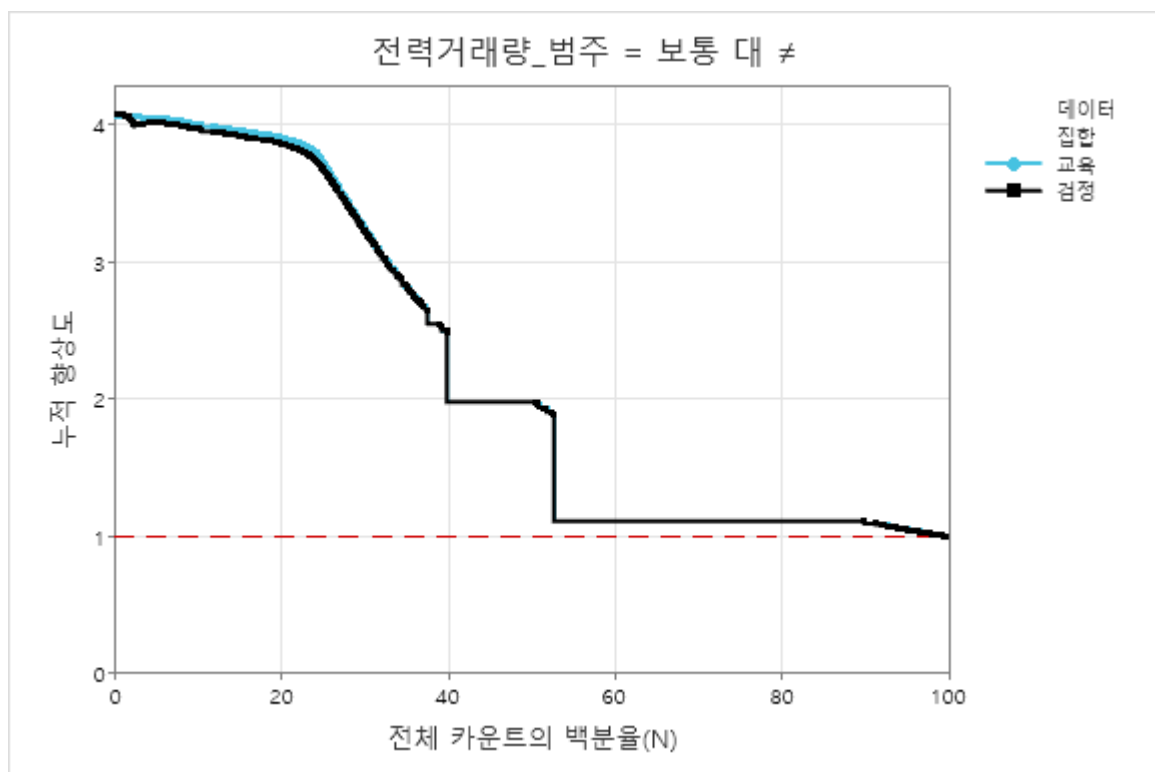
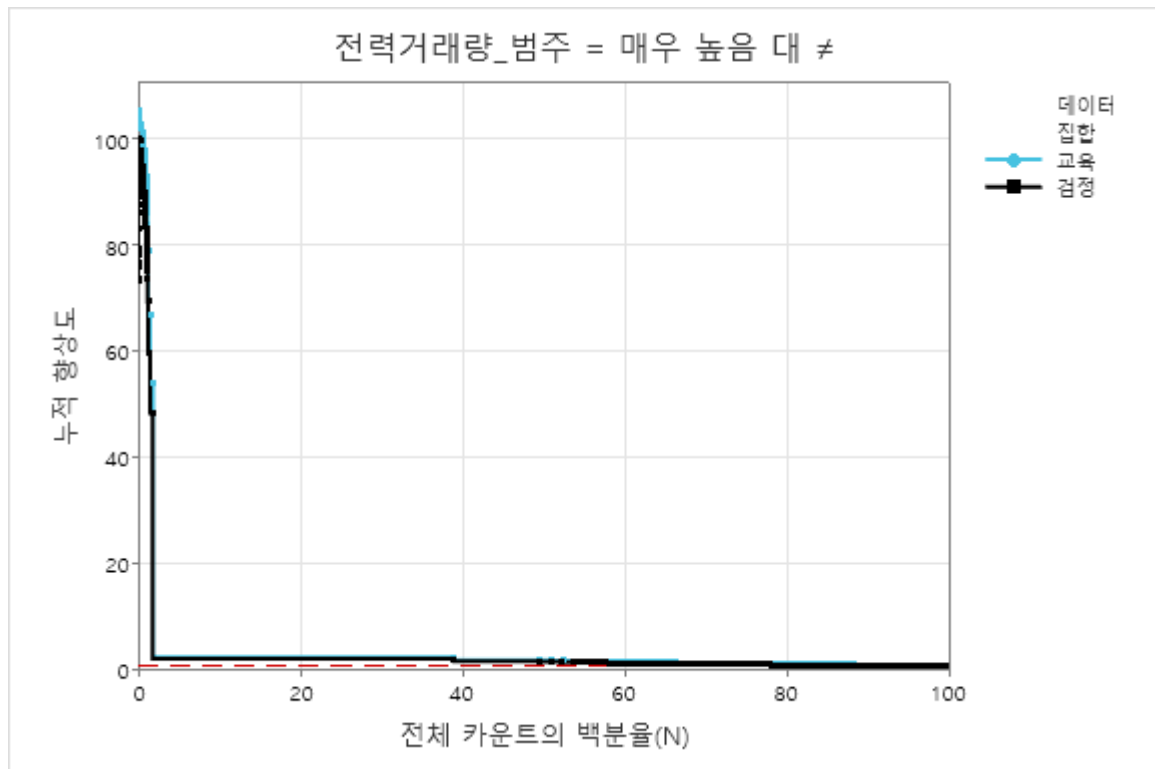




향상도 차트







ALL_DATA(범주형_시간추가_결측치열모두제거) (1).CSV

MARS® 회귀 분석: 전력거래량(MWh) 대 기온(°C), 풍속(m/s), 풍향(16방위), 습도(%), 증기압(hPa), 이슬점온도(°C),

현지기압(hPa), 해면기압(hPa), 전운량(10분위), 중하층운량(10분위), 시정(10m), 지면온도(°C), 지역, 년도, 월, 일, 요일, 시간

방법

최적 트리 선택 기준	최대 R-제곱
모형 검증	70/30% 교육/검정 집합
최대 기본 함수 수	30
매듭 사이의 최소 관측치 수	자동
사용된 행	974890
사용되지 않은 행	3926

반응 정보

데이터 집합	N	N의 비율(%)	평균	표준 편차	최소값	Q1	중위수	Q3	최대값
교육	682420	70.00	3799.16	3953.52	0	333.25	2798	6220	18838
검정	292470	30.00	3798.31	3951.49	0	334.00	2805	6224	18880

회귀 방정식

BF1 = 지역이(가) 경기도, 경상남도, 경상북도, 인천광역시, 전라남도, 충청남도인 경우

BF3 = 지역이(가) 강원도, 경기도, 경상북도, 부산광역시, 울산광역시, 충청남도인 경우

BF5 = 지역이(가) 충청남도인 경우

BF7 = 지역이(가) 경상북도, 부산광역시, 인천광역시, 전라남도, 전라북도인 경우

BF9 = 기온(°C)이(가) 누락되지 않은 경우

BF11 = 최대(0, 기온(°C) - 17.8667) * BF9

BF12 = 최대(0, 17.8667 - 기온(°C)) * BF9

BF13 = 요일이(가) 5, 6인 경우

BF15 = 시간이(가) 1, 2, 3, 4, 5, 6, 7, 8인 경우

BF17 = 지역이(가) 경기도, 대구광역시, 부산광역시, 서울특별시, 인천광역시, 전라남도, 전라북도, 제주도, 충청남도인 경우

BF19 = 월이(가) 3, 4, 5, 6, 9, 10, 11인 경우

BF21 = 시간이(가) 1, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21인 경우

BF23 = 증기압(hPa)이(가) 누락되지 않은 경우

BF25 = 최대(0, 증기압(hPa) - 30.5) * BF23

BF26 = 최대(0, 30.5 - 증기압(hPa)) * BF23

BF27 = 현지기압(hPa)이(가) 누락되지 않은 경우

BF29 = 최대(0, 현지기압(hPa) - 1011.78) * BF27

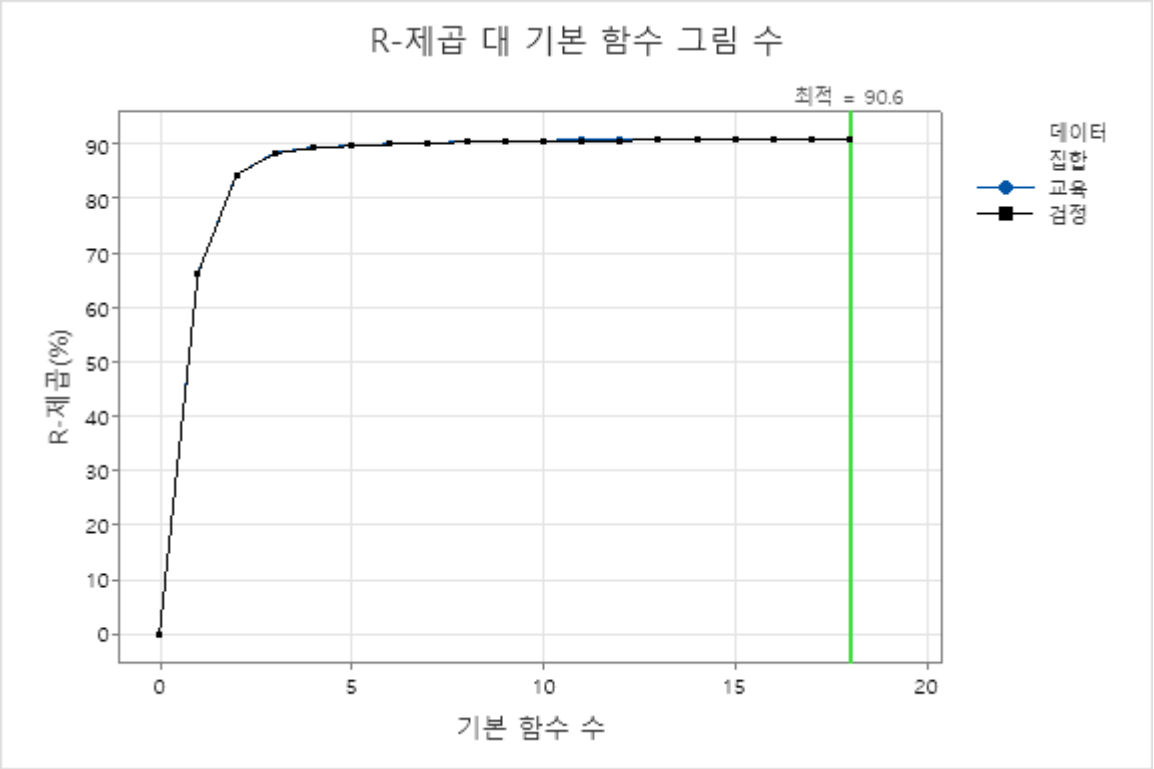
BF30 = 최대(0, 1011.78 - 현지기압(hPa)) * BF27

전력거래량(MWh) = -0.941251 + 5068.19 * BF1 + 3060.26 * BF3 + 3924.54 * BF5 + 818.541 * BF7 +

373.353 * BF9 + 21.0036 * BF11 + 40.3074 * BF12 - 501.845 * BF13 - 462.316 * BF15 + 333.15 * BF17 -

200.523 * BF19 + 226.377 * BF21 + 167.448 * BF23 - 66.3699 * BF25 - 21.0296 * BF26 -

259.889 * BF27 + 23.3931 * BF29 + 8.73601 * BF30

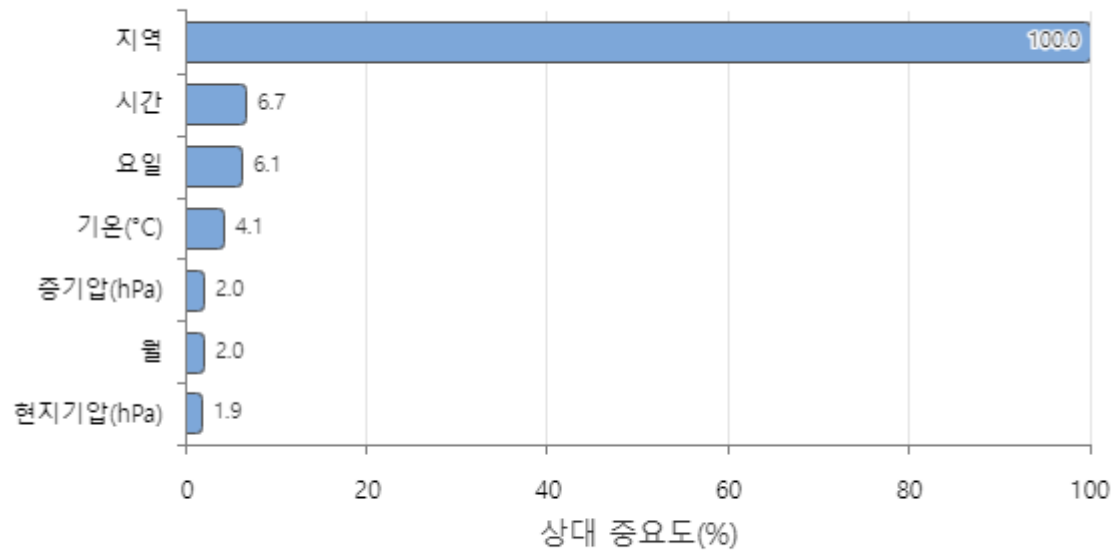


모형 요약

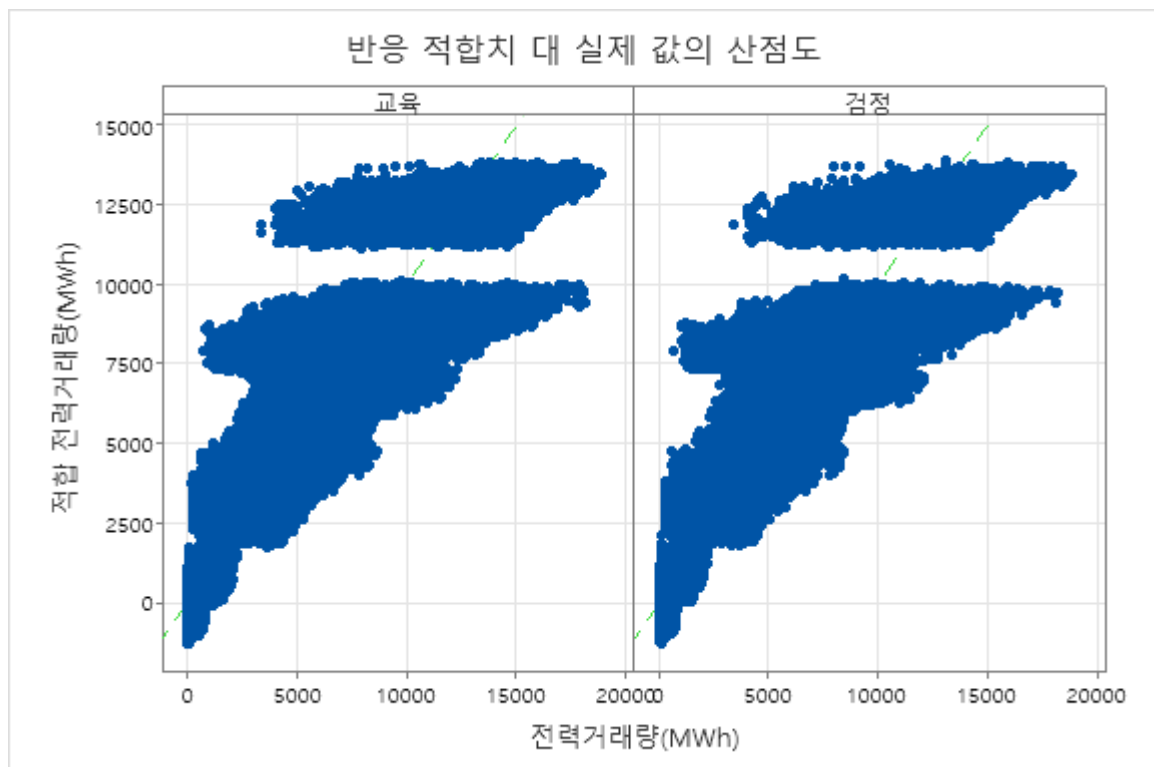
전체 예측 변수 18
 중요 예측 변수 7
 최대 기본 함수 수 30
 최적의 기본 함수 수 18

통계량	교육	검정
R-제공	90.65%	90.60%
루트 평균 제곱 오차(RMSE)	1208.5871	1211.3565
평균 제곱 오차(MSE)	1460682.8009	1467384.4721
평균 절대 편차(MAD)	815.4029	816.1765

상대 변수 중요도

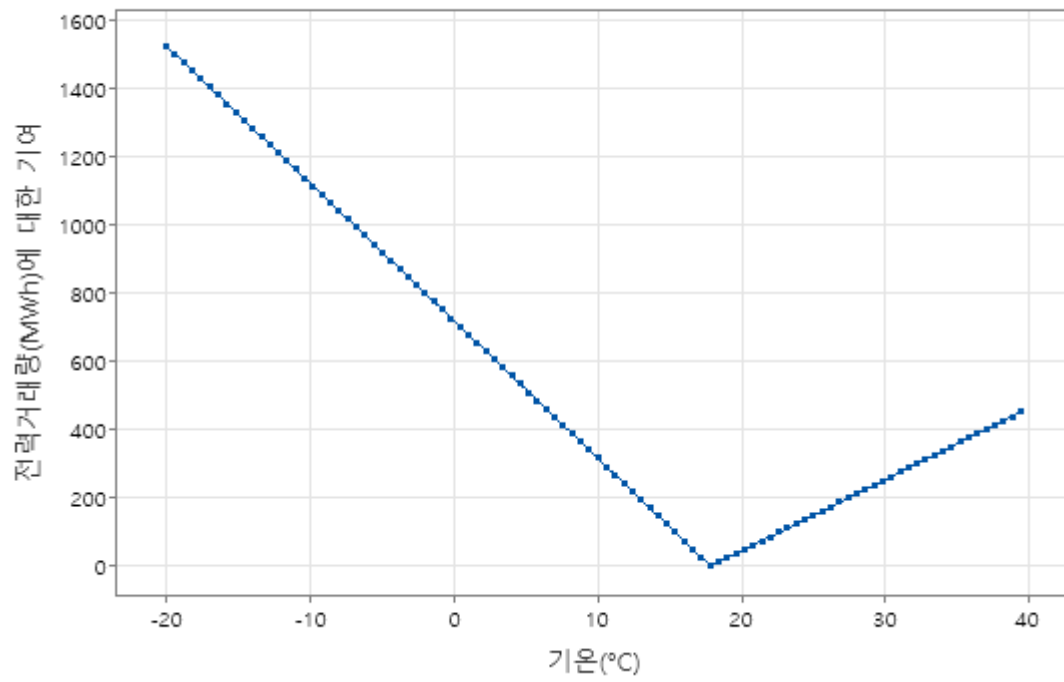


변수 중요도는 예측 변수의 기본 함수에서 모형 예측 능력을 측정합니다. 상대 중요도는 상위 예측 변수를 기준으로 한 % 개선으로 정의됩니다.

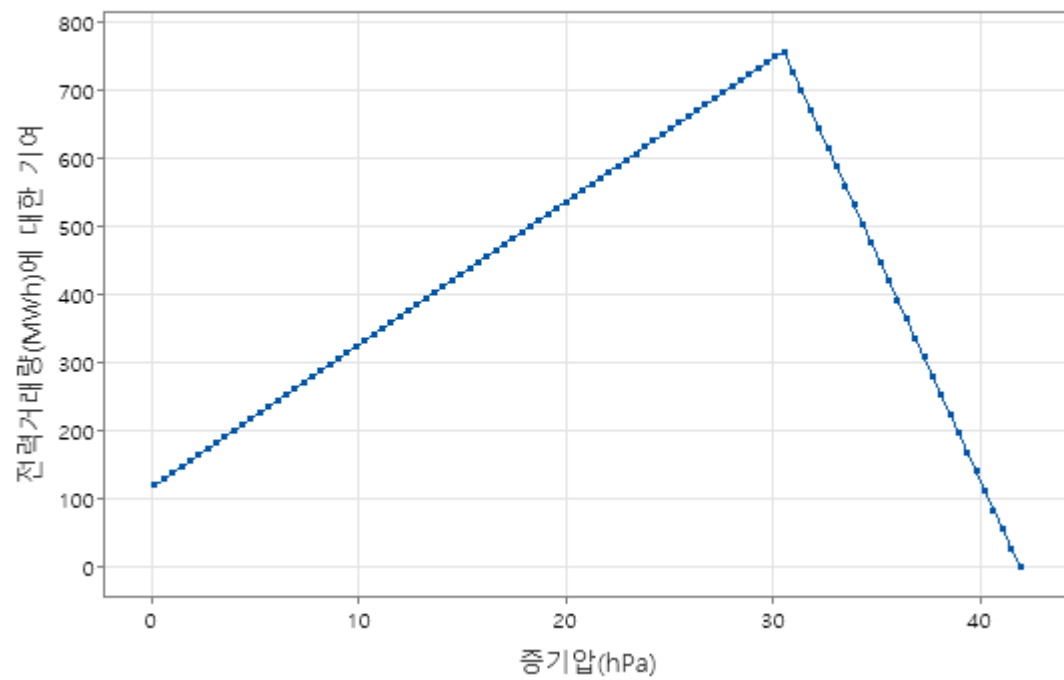


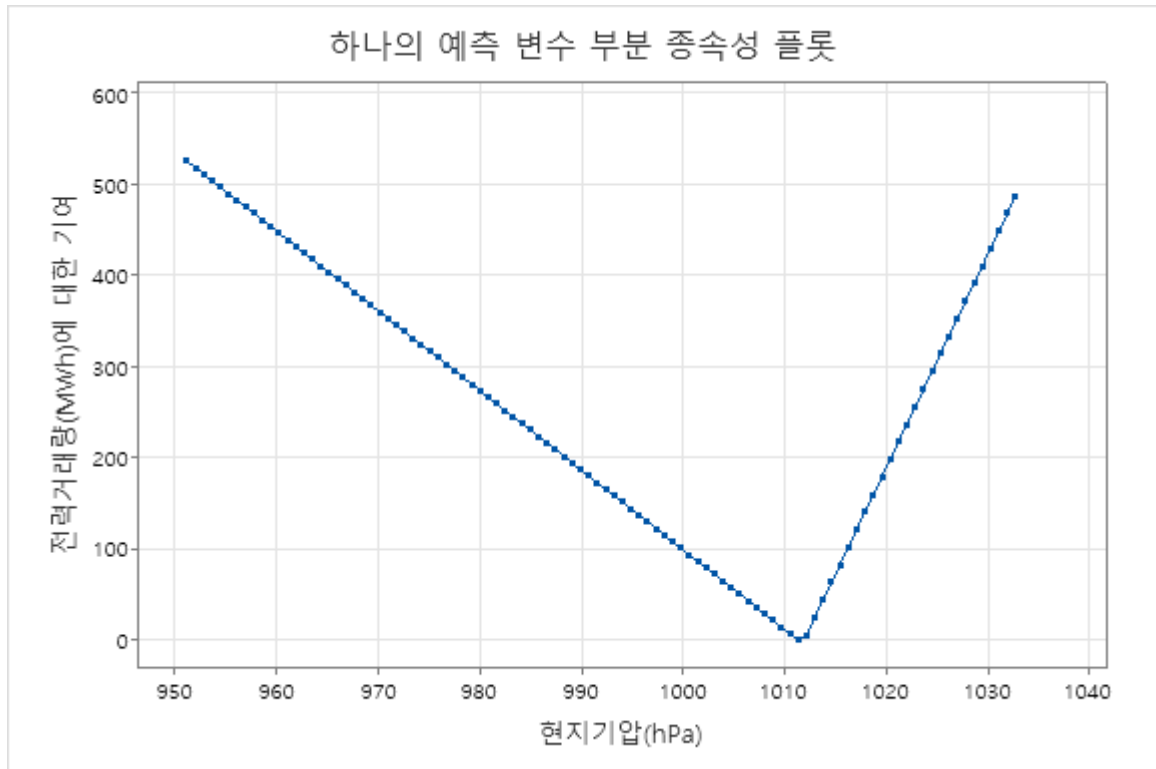
하나의 예측 변수 부분 종속성 플롯

하나의 예측 변수 부분 종속성 플롯



하나의 예측 변수 부분 종속성 플롯





ALL_DATA(범주형_시간추가_결측치열모두제거) (1).CSV

TreeNet® 분류: 전력거래량_범주 대 기온(°C), 풍속(m/s),
 풍향(16방위), 습도(%), 증기압(hPa), 이슬점온도(°C),
 현지기압(hPa), 해면기압(hPa), 전운량(10분위), 중하층운량(10분위),
 시정(10m), 지면온도(°C), 지역, 년도, 월, 일, 요일, 시간

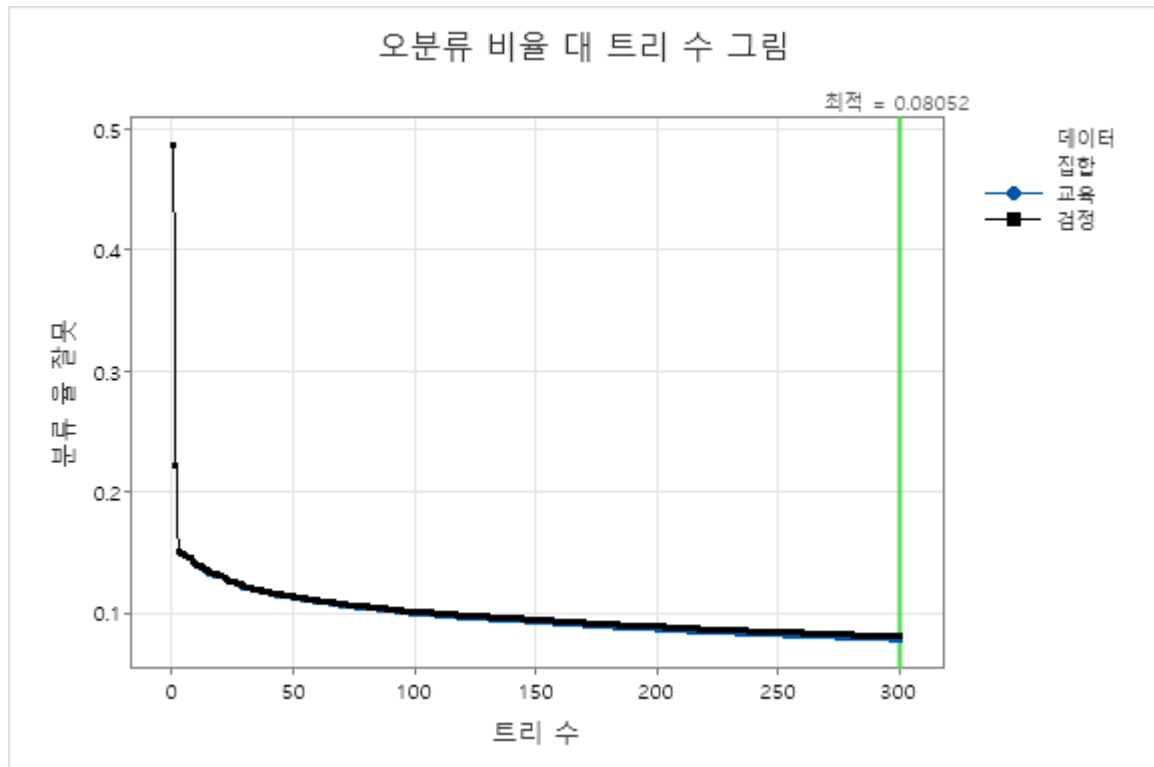
방법

최적 트리 수 선택 기준	최소 오분류율
모형 검증	70/30% 교차/검정 집합
학습률	0.1
하위 표본 선택 방법	완전 랜덤
하위 표본 부분	0.5
트리당 최대 터미널 노드 수	6
최소 단말 노드 크기	3
노드 분할을 위해 선택된 예측 변수 수	총 예측 변수 수 = 18
사용된 행	973458
사용되지 않은 행	5358

다항 반응 정보

변수	등급	교육		검정	
		카운트	%	카운트	%
전력거래량_범주	낮음	182296	26.75	78251	26.80

높음	55001	8.07	23594	8.08
매우 낮음	269750	39.59	115672	39.61
매우 높음	6470	0.95	2772	0.95
보통	167916	24.64	71736	24.57
모두	681433	100.00	292025	100.00

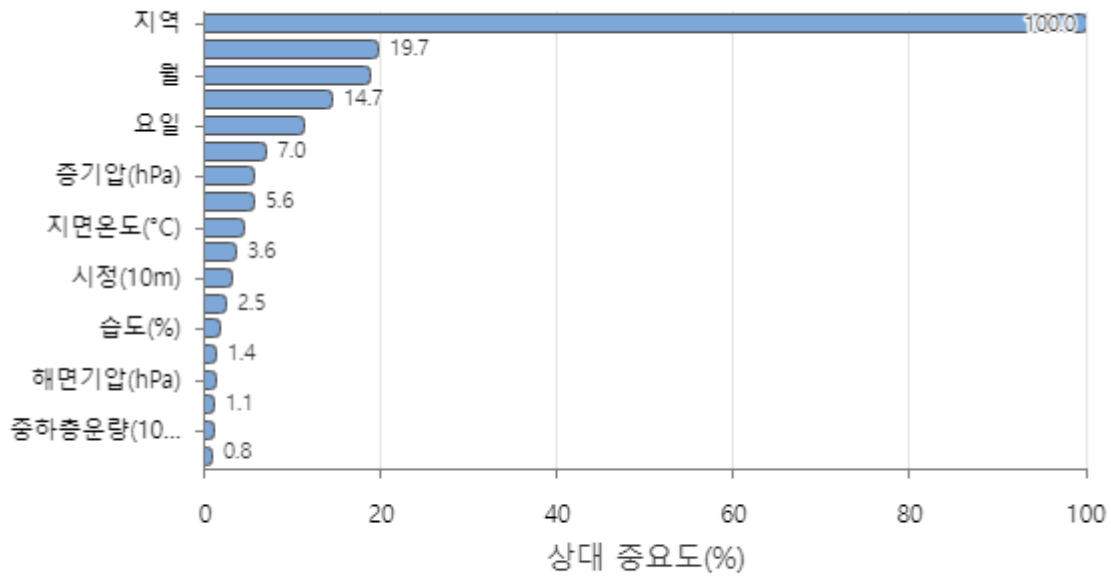


모형 요약

전체 예측 변수 18
중요 예측 변수 18
성장한 트리 수 300
최적의 트리 수 300

통계량	교육	검정
평균 음수 로그 우도	0.2008	0.2053
오분류 비율	0.0782	0.0805

상대 변수 중요도



변수 중요도는 예측 변수에 분할을 적용할 때 모델 개선을 측정합니다. 상대 중요도는 상위 예측 변수를 기준으로 한 % 개선으로 정의됩니다.

오차 행렬

예측된 등급(교육)

실제 등급	카운트	낮음	높음	매우 낮음	매우 높음	보통	정답률(%)
낮음	182296	163716	108	2529	0	15943	89.81
높음	55001	0	45231	0	641	9129	82.24
매우 낮음	269750	6083	0	263663	0	4	97.74
매우 높음	6470	0	2212	0	4257	1	65.80
보통	167916	11129	5513	0	5	151269	90.09
모두	681433	180928	53064	266192	4903	176346	92.18

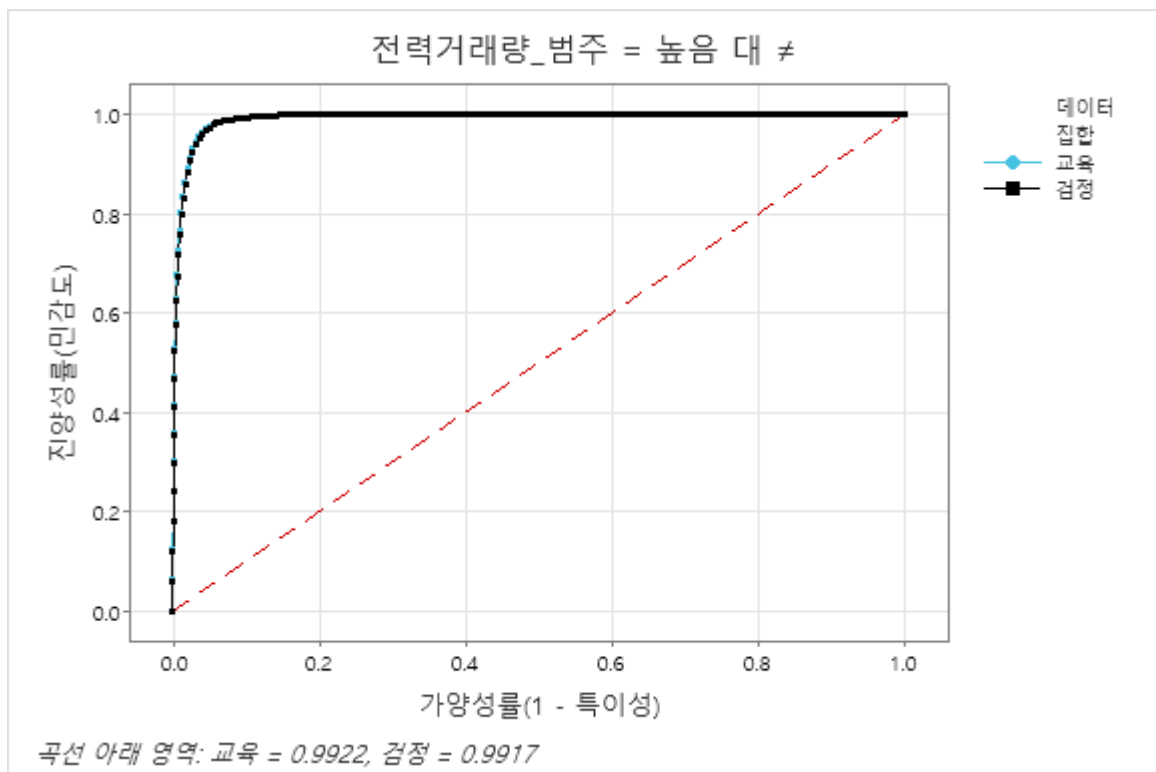
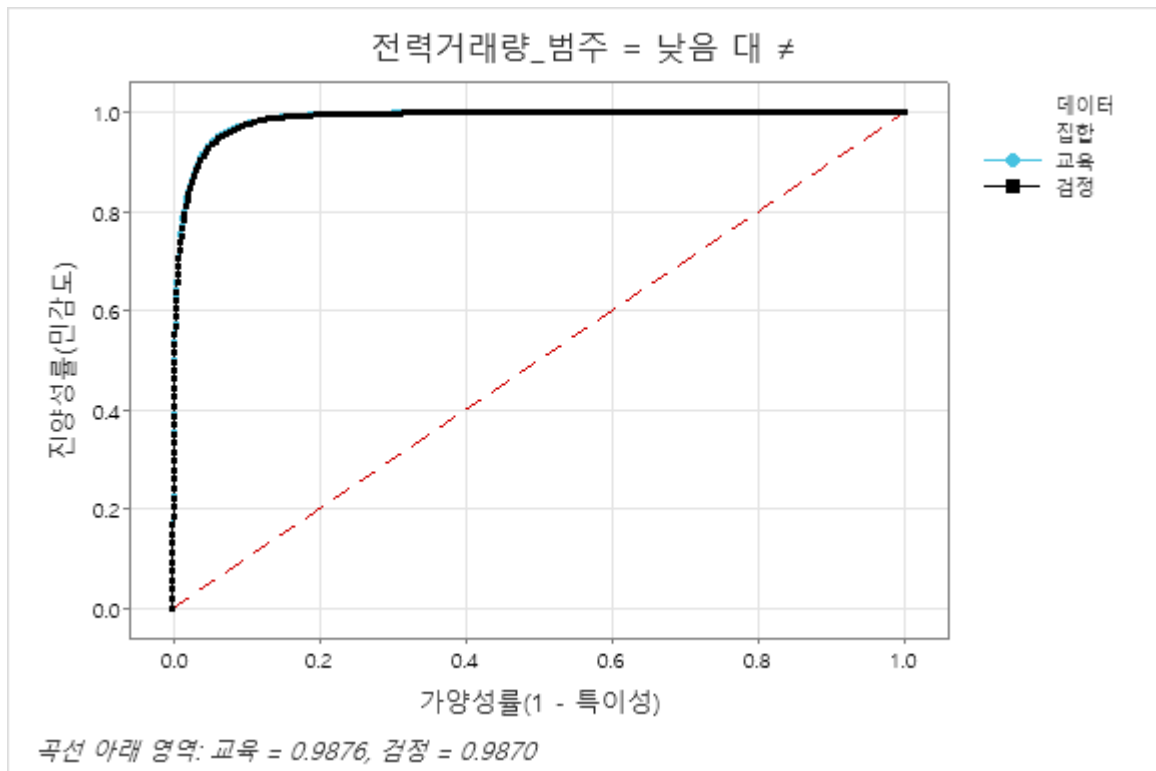
예측된 등급(검정)

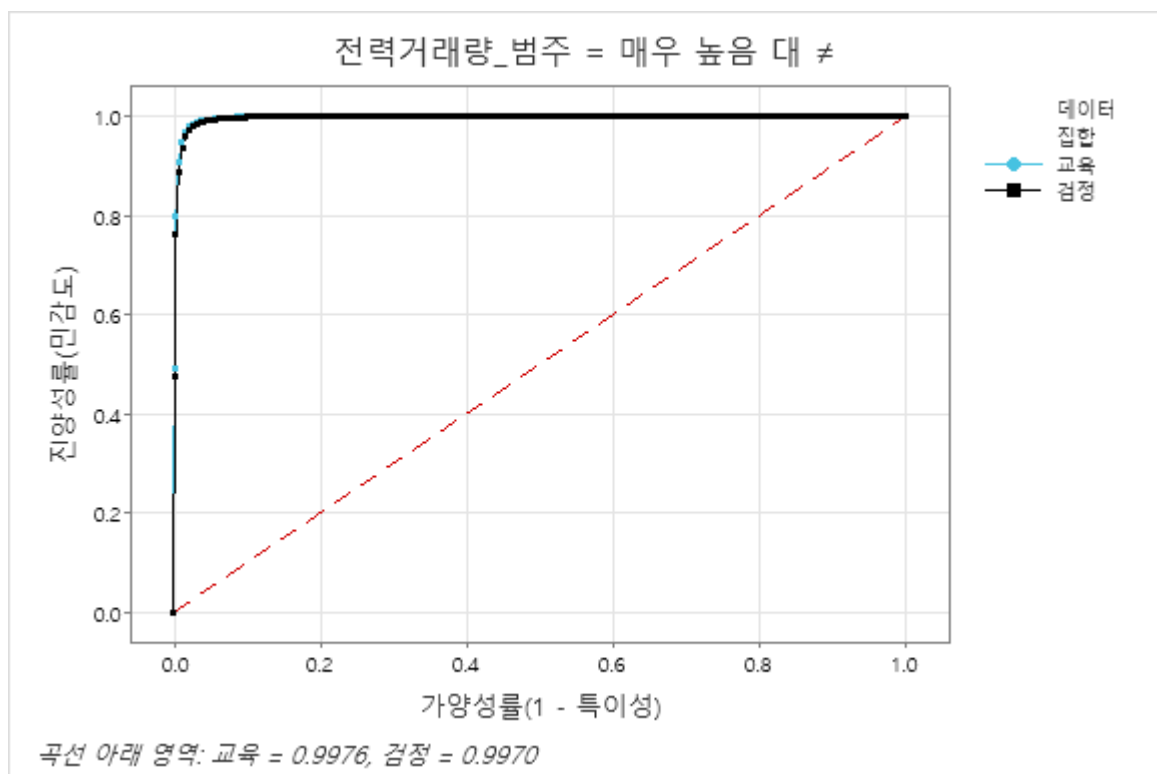
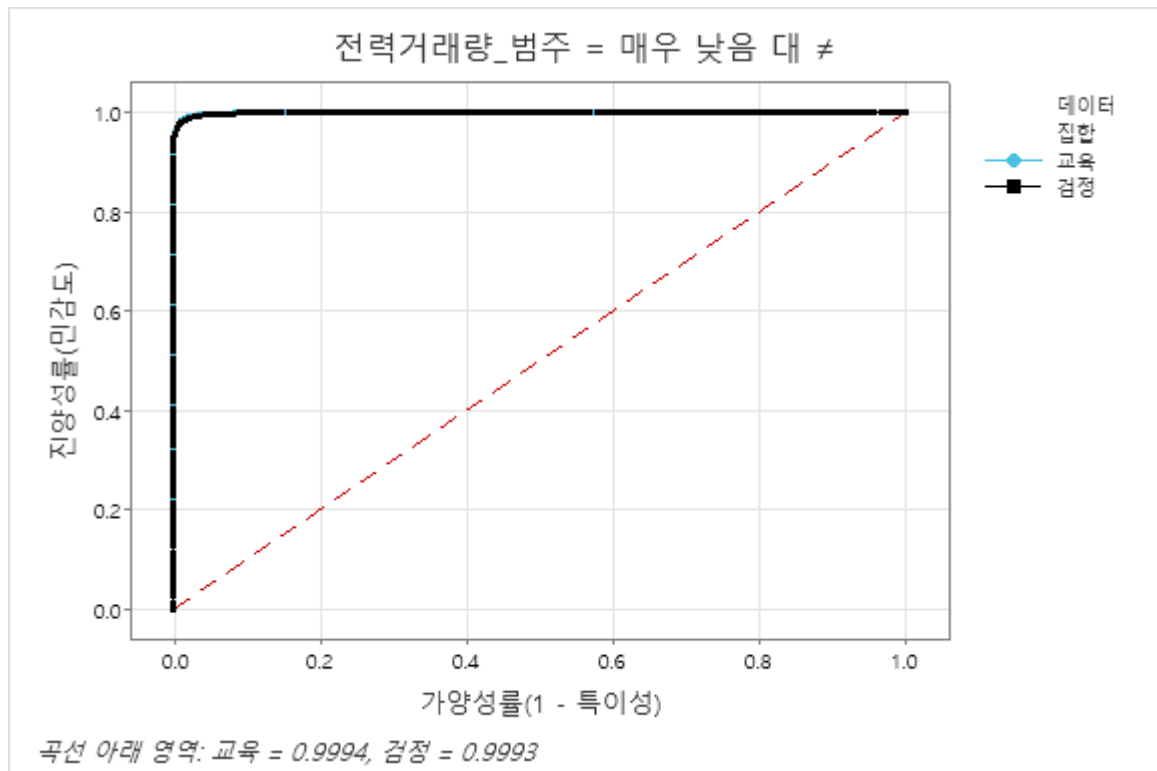
실제 등급	카운트	낮음	높음	매우 낮음	매우 높음	보통	정답률(%)
낮음	78251	70002	33	1163	0	7053	89.46
높음	23594	0	19334	0	305	3955	81.94
매우 낮음	115672	2703	1	112967	0	1	97.66
매우 높음	2772	0	1061	0	1709	2	61.65
보통	71736	4790	2440	0	7	64499	89.91
모두	292025	77495	22869	114130	2021	75510	91.95

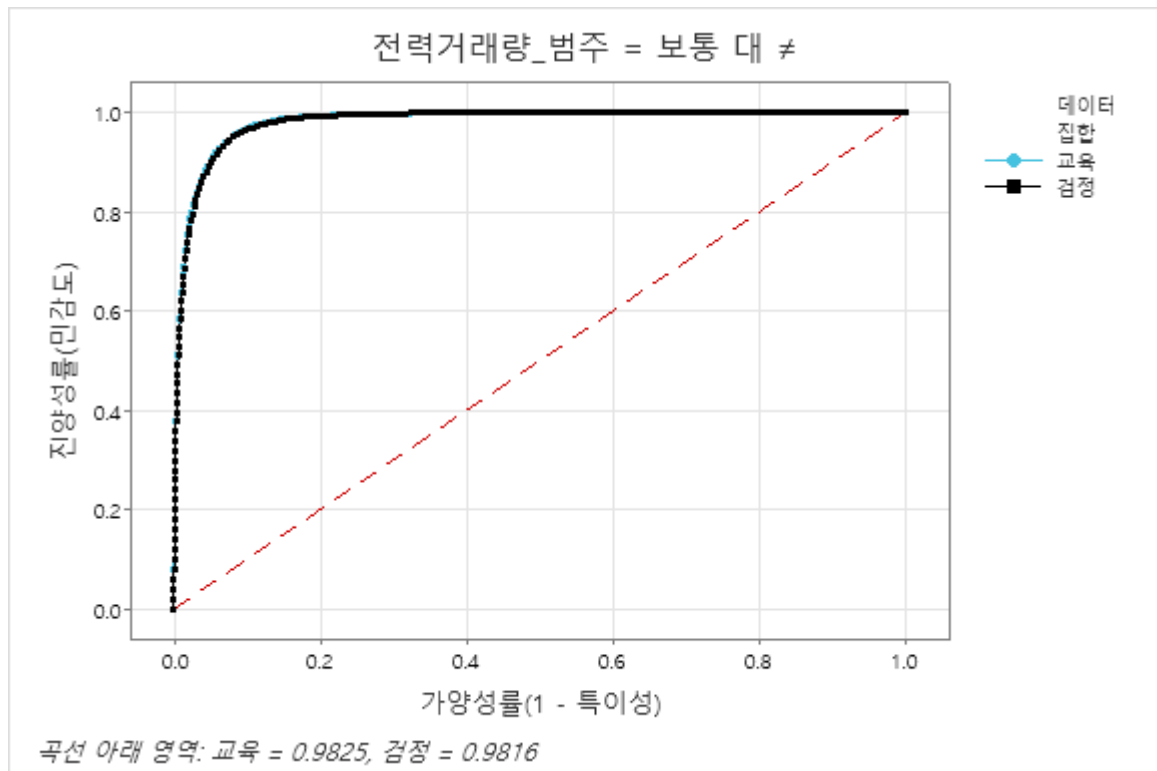
오분류

교육				검정			
실제 등급	카운트	오분류됨	오차율(%)	카운트	오분류됨	오차율(%)	
낮음	182296	18580	10.19	78251	8249	10.54	
높음	55001	9770	17.76	23594	4260	18.06	
매우 낮음	269750	6087	2.26	115672	2705	2.34	
매우 높음	6470	2213	34.20	2772	1063	38.35	
보통	167916	16647	9.91	71736	7237	10.09	

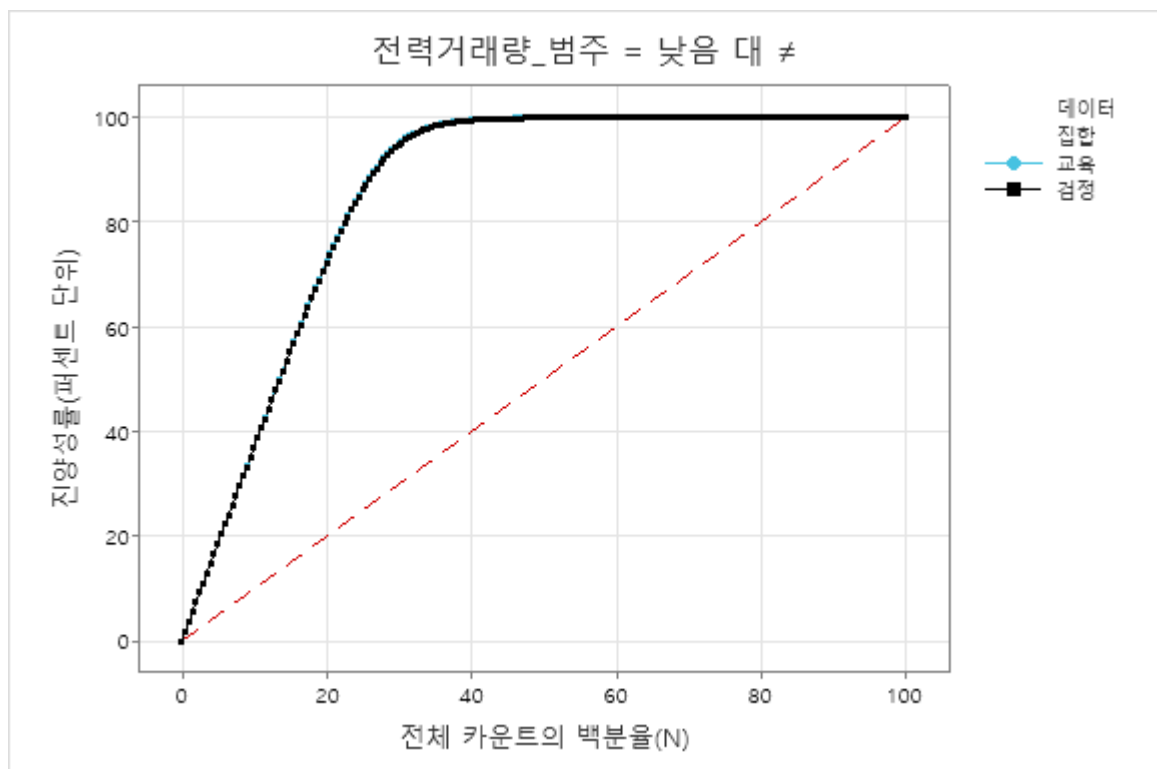
모두 681433 53297 7.82 292025 23514 8.05
수신자 검사 특성(ROC) 곡선

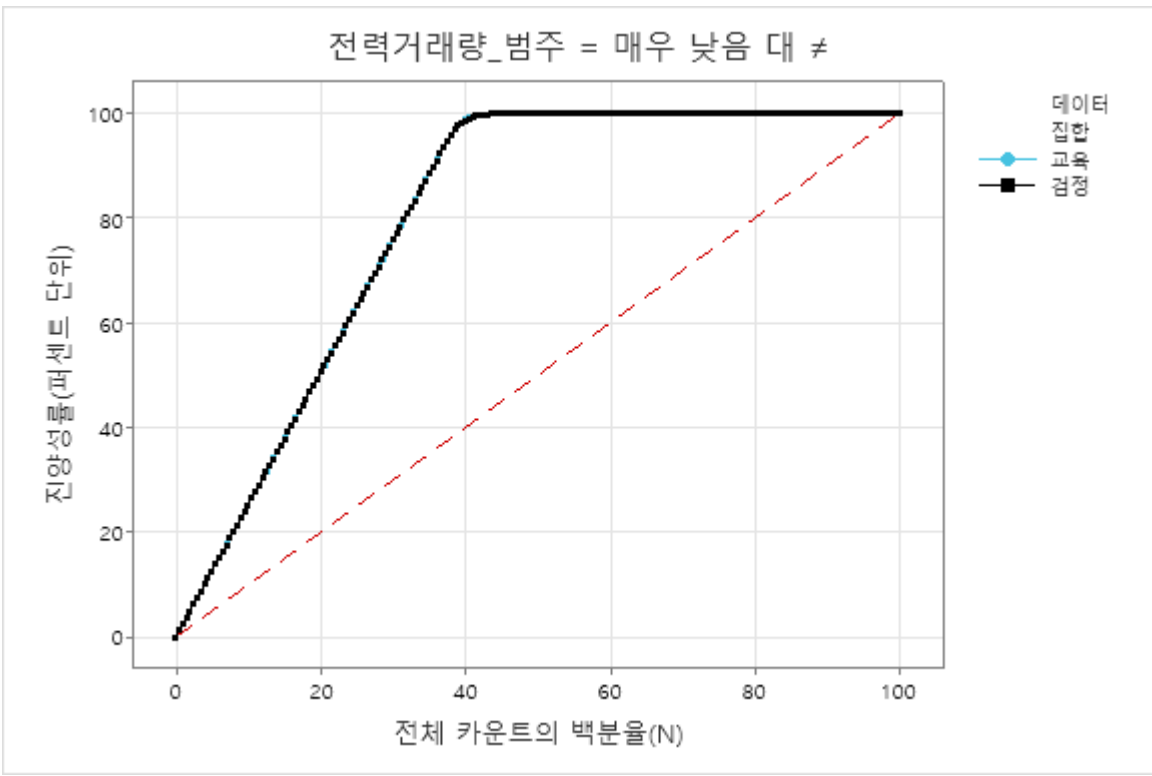
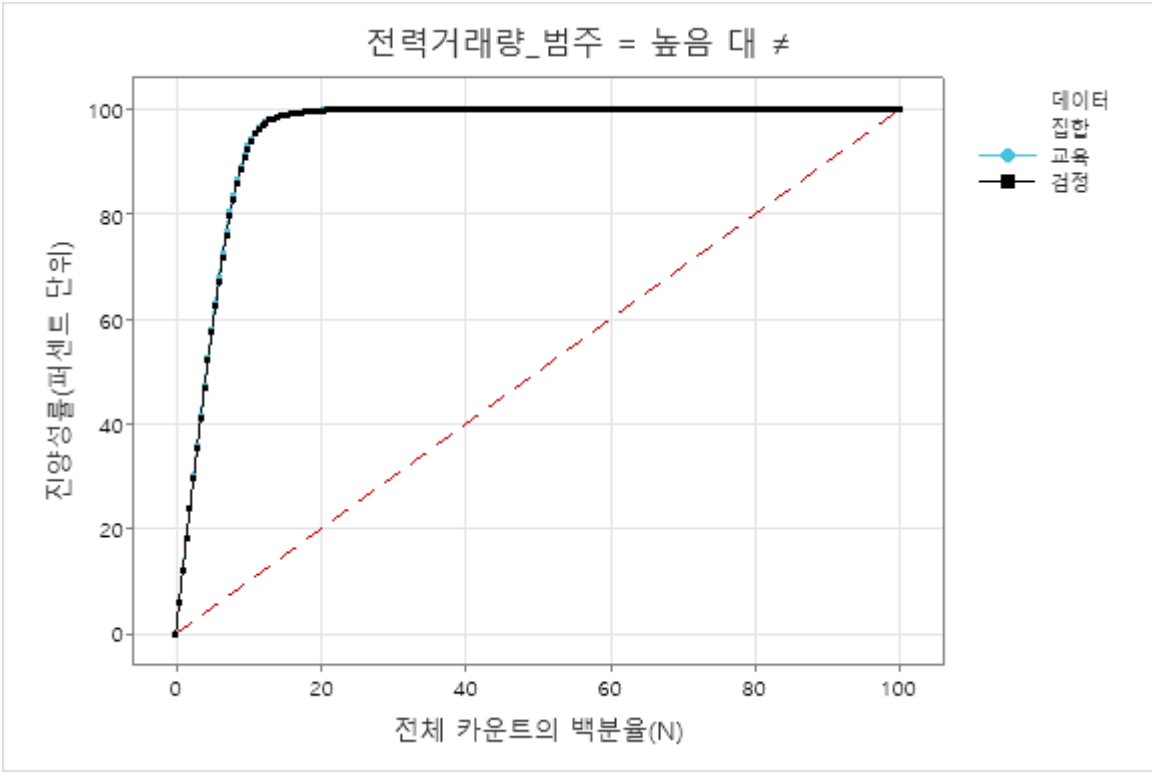


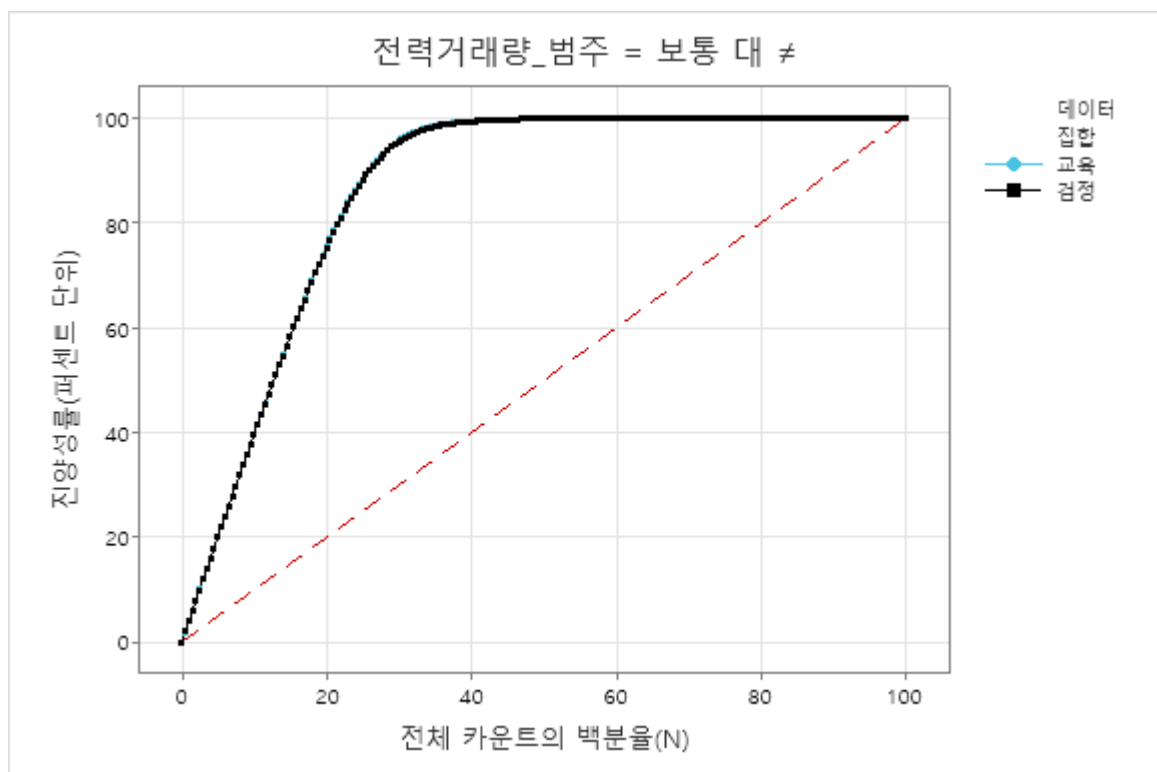
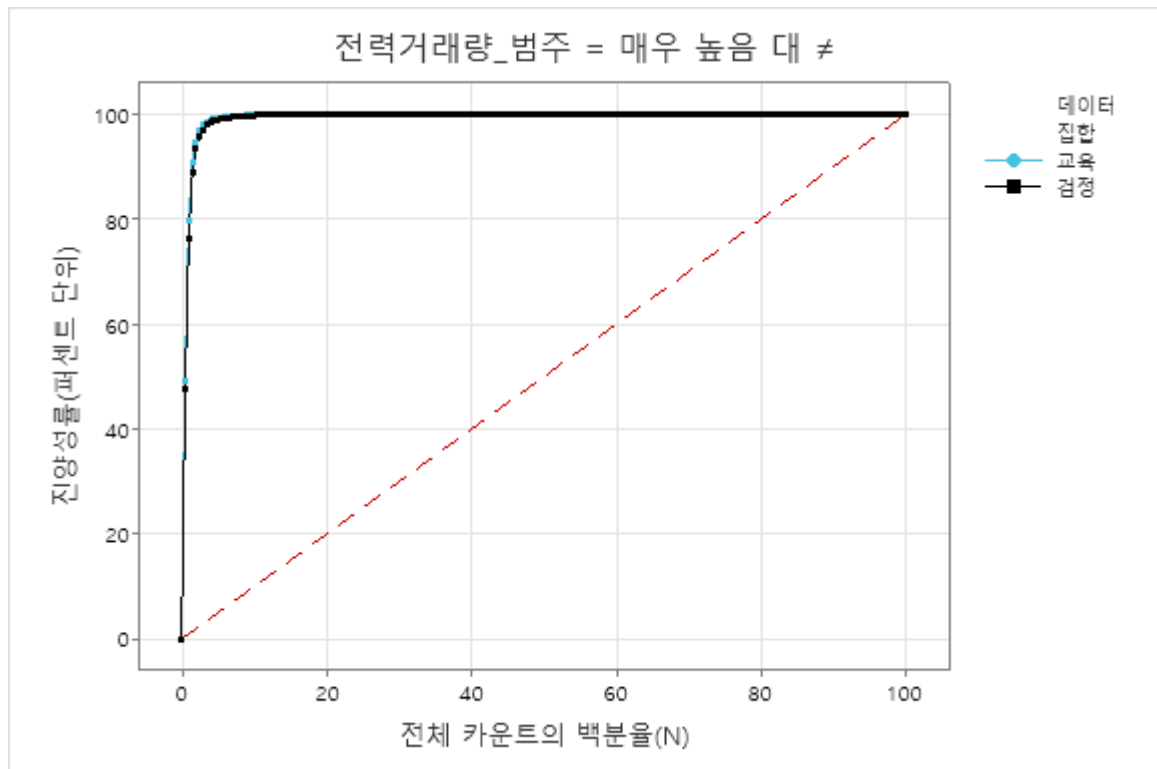




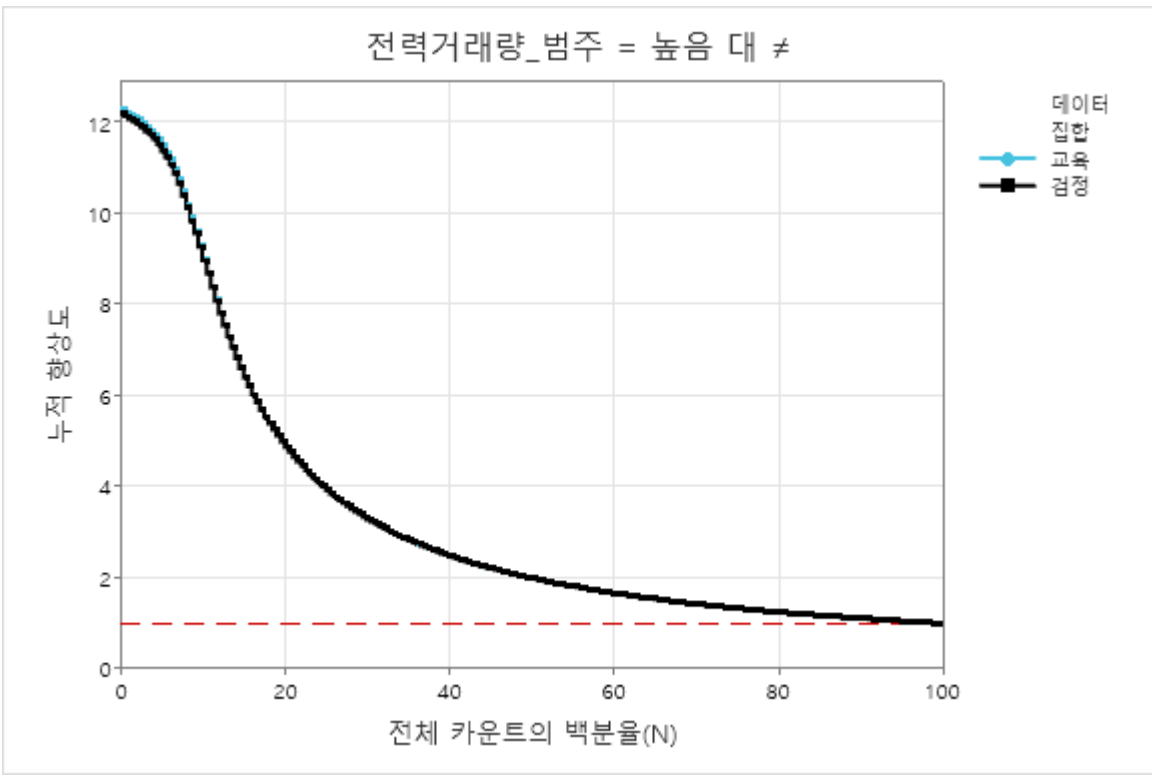
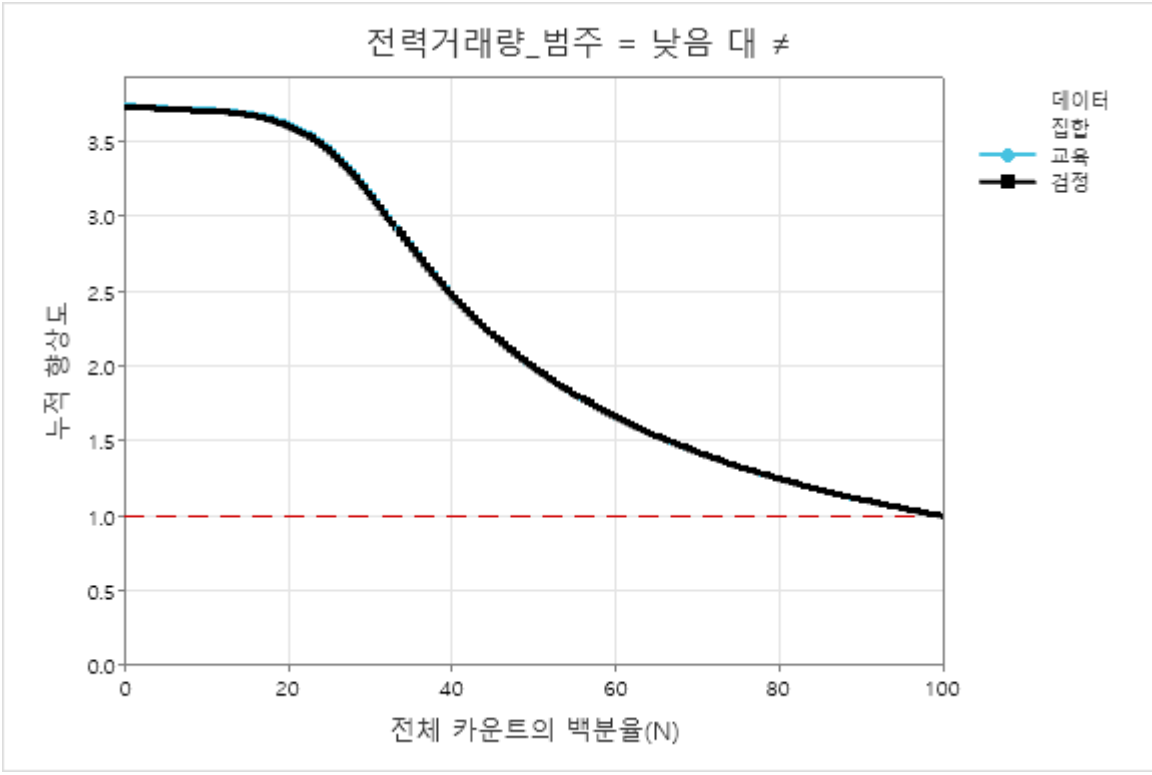
이익 차트

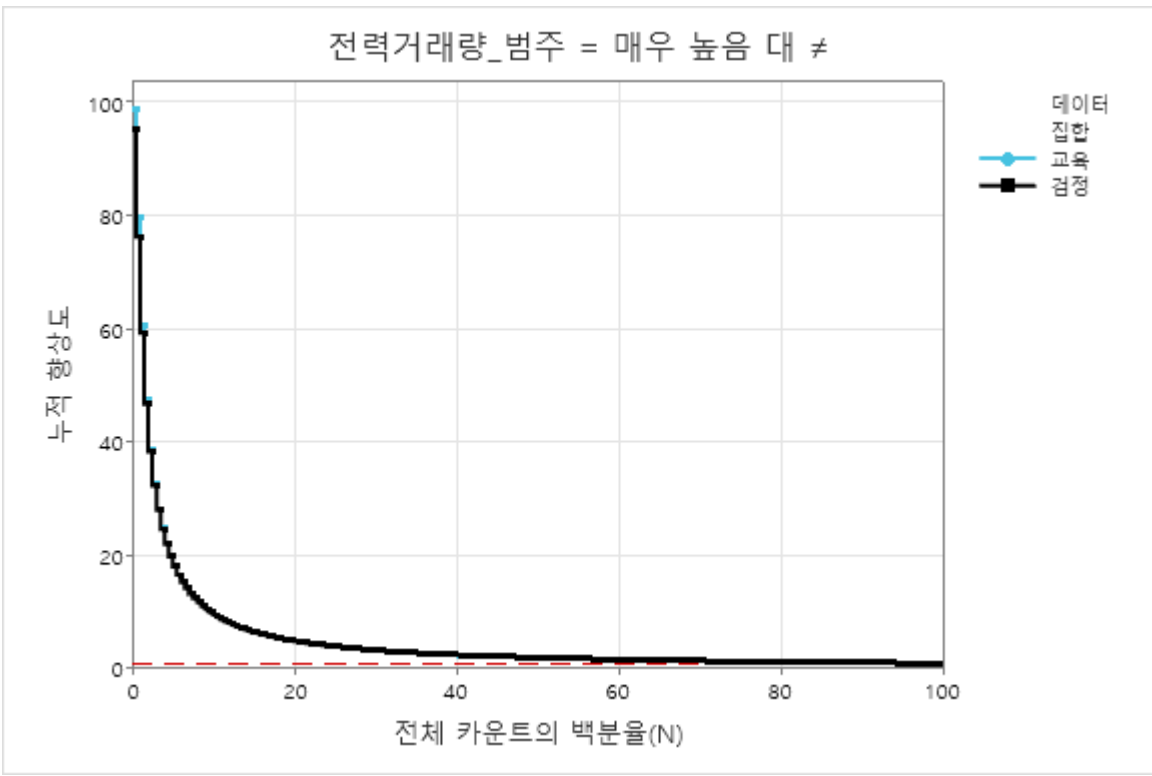
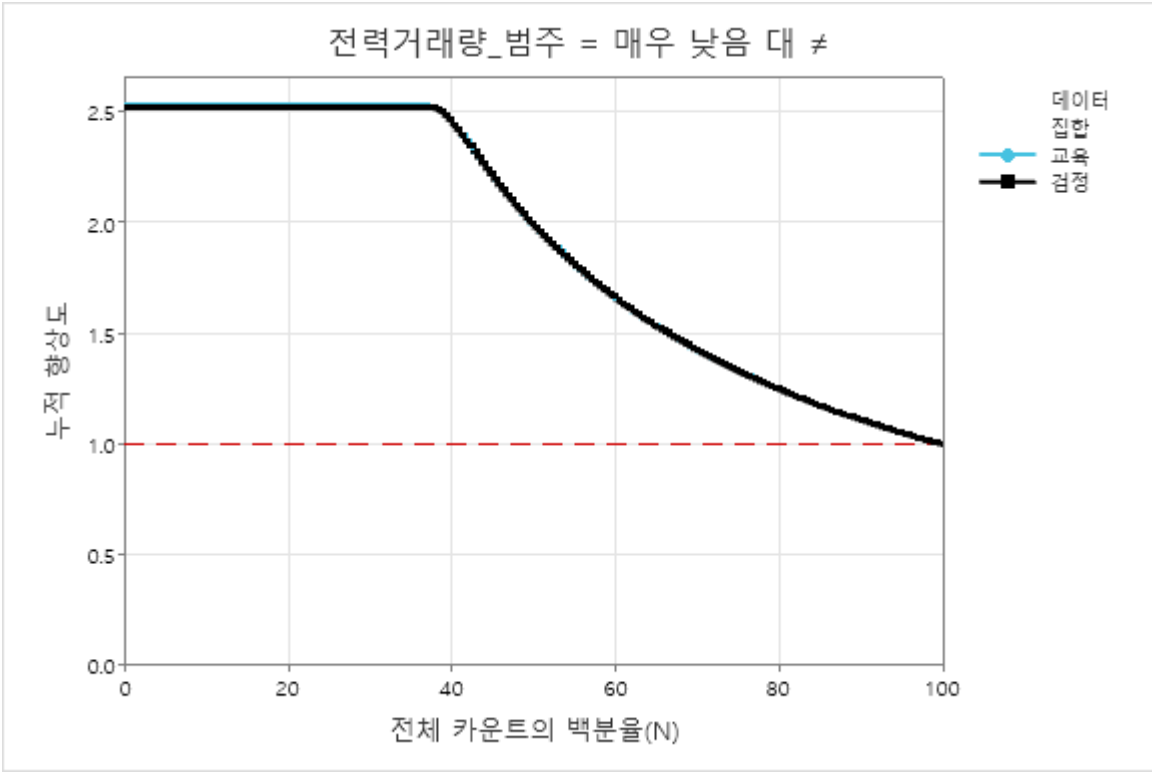


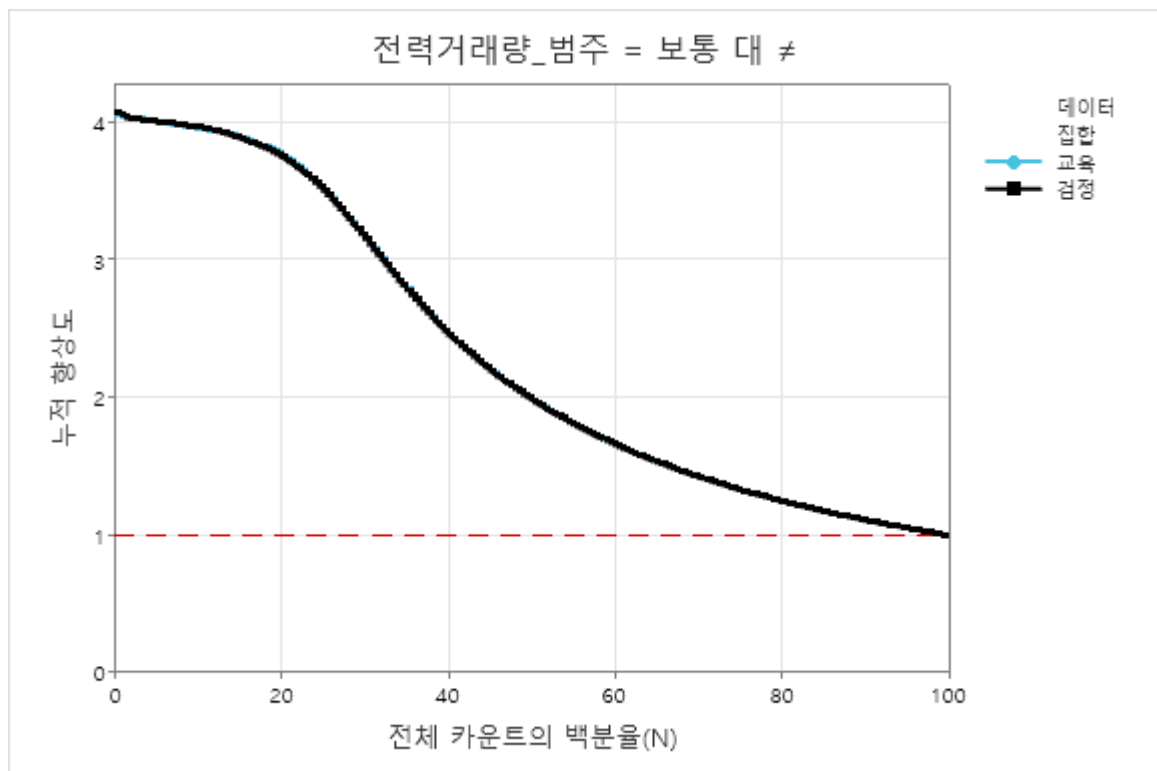




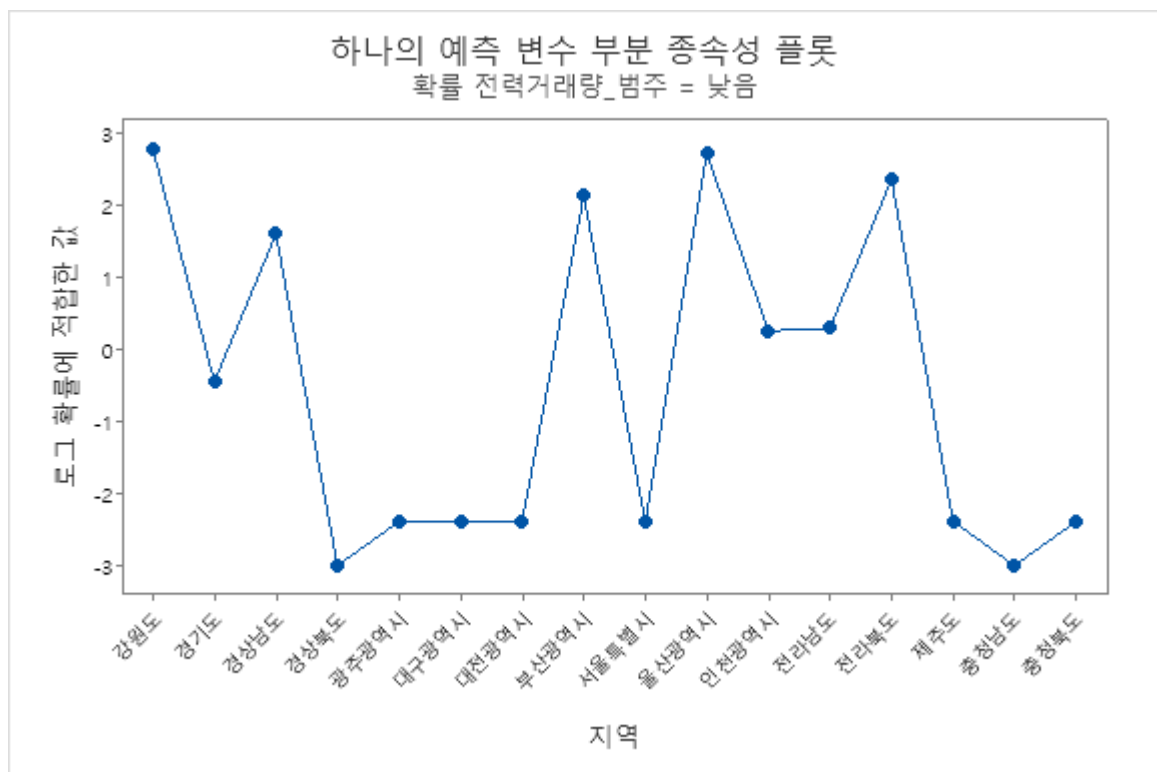
항상도 차트



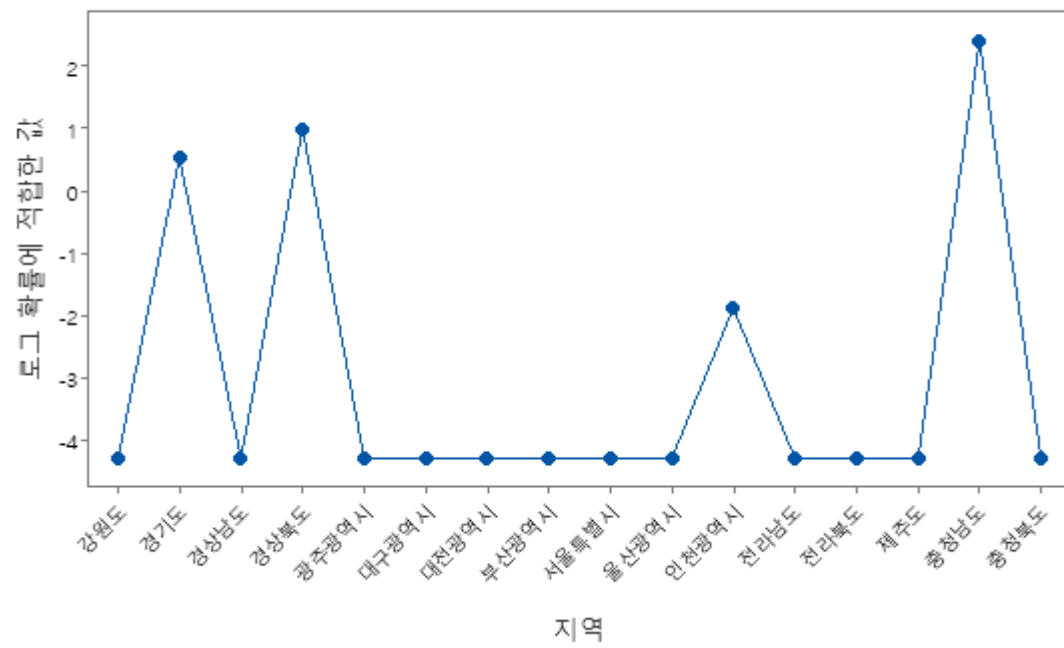




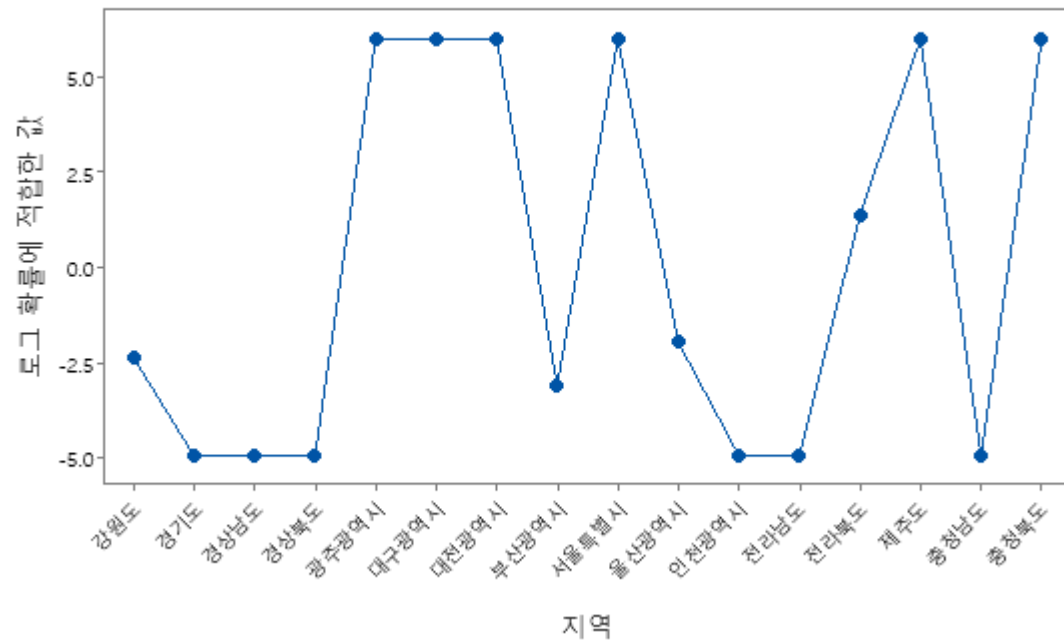
하나의 예측 변수 부분 종속성 플롯



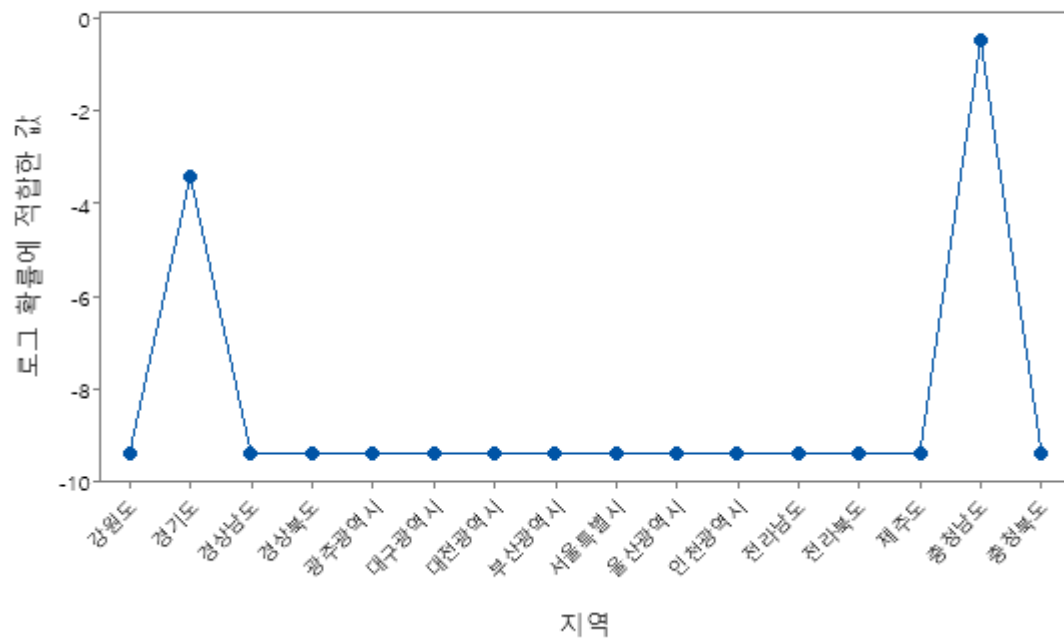
하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 높음



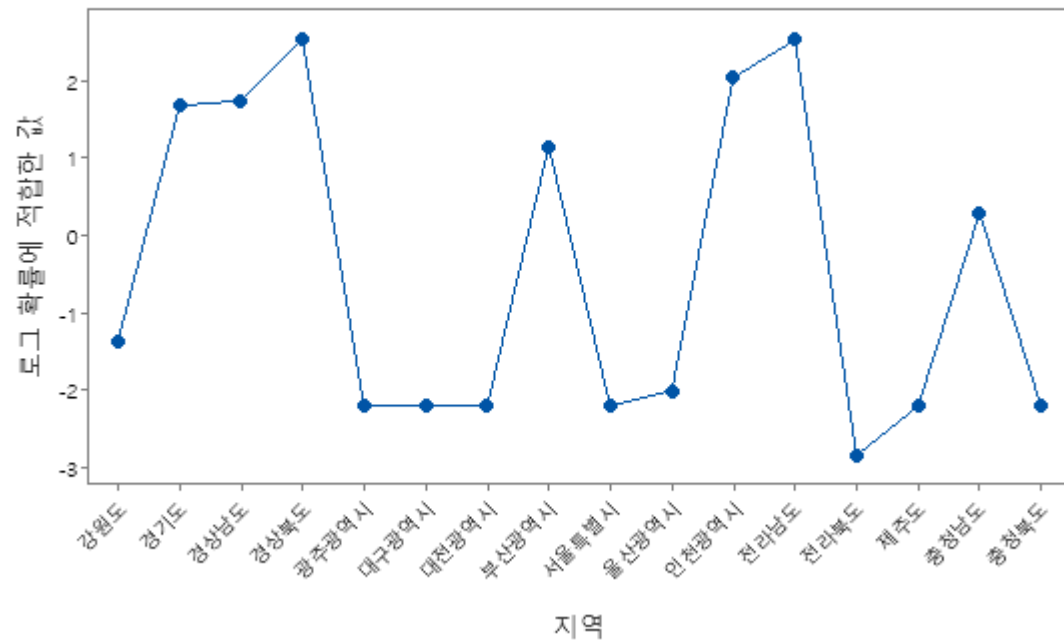
하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 매우 낮음



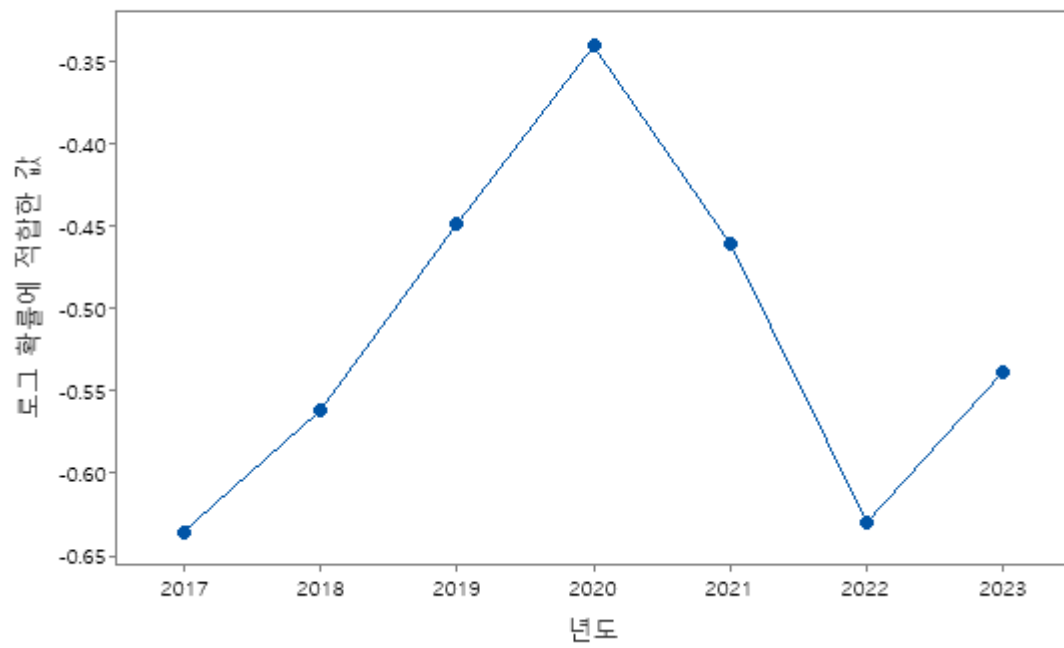
하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 매우 높음



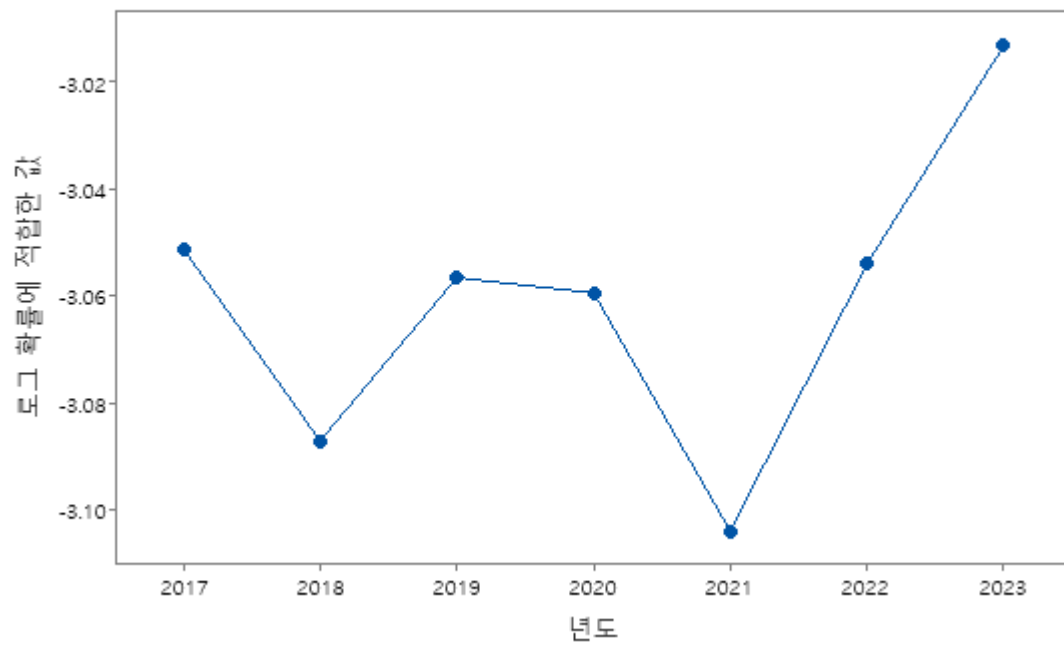
하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 보통



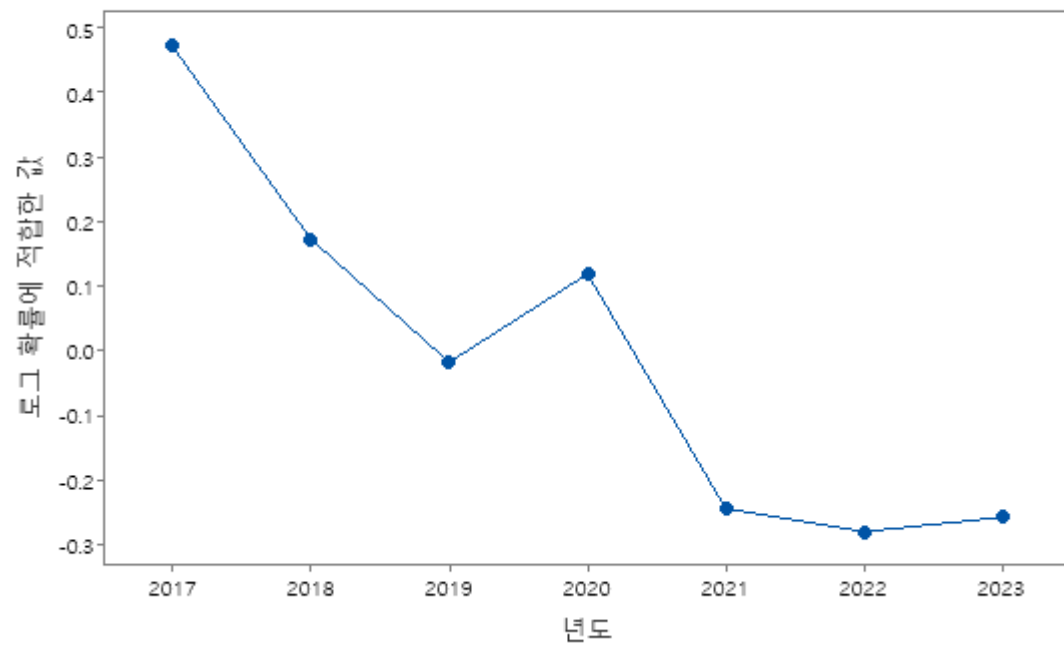
하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 낮음



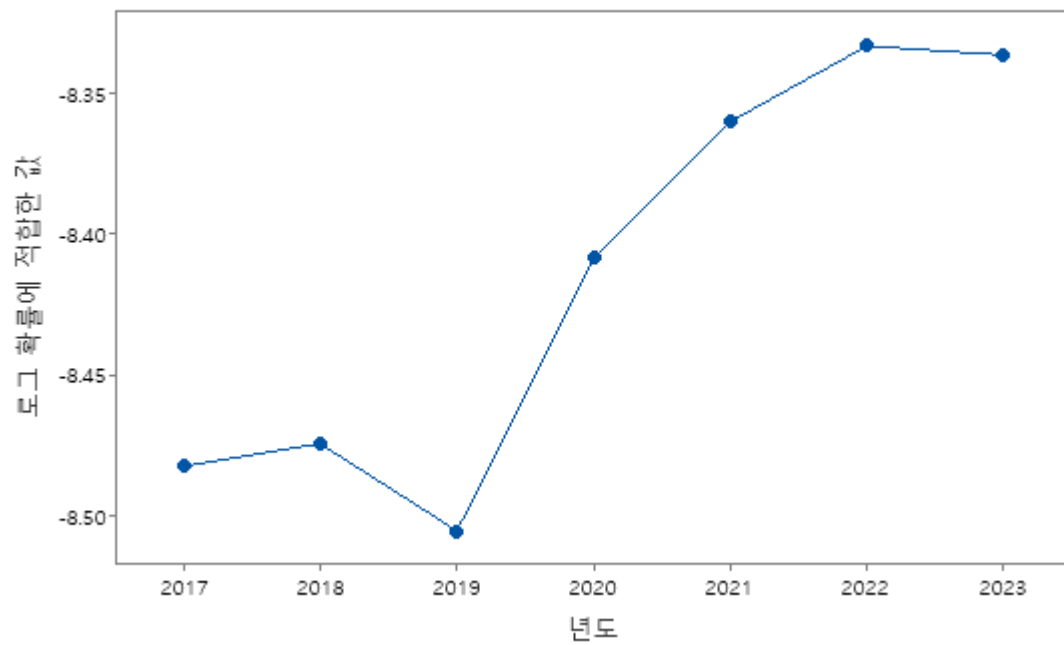
하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 높음



하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 매우 낮음



하나의 예측 변수 부분 종속성 플롯
 확률 전력거래량_범주 = 매우 높음



하나의 예측 변수 부분 종속성 플롯
확률 전력거래량_범주 = 보통

