

자가불소 결과

분석 데이터: HR 인재유형 데이터 vs 역량검사 데이터

작성일: 2025년 9월 17일

분석 대상: 61명 (HR 158명, 역량검사 152명 중 매칭)

1. 핵심 요약

- 데이터 매칭: HR 158명, 역량검사 152명 중 최종 매칭 58명(32.1%).
- 핵심 역량: 전략성·정체성이 인재 선발에 가장 중요한 요인으로 일관되게 도출.
- 선발 모델: 로지스틱 회귀를 사용한 예측이 가장 적합함
- ⚠ 이슈: 게으른 가연성과 성실한 가연성 구분 라벨 검토 필요.

2. 데이터 로딩 및 전처리

2.1 전체 데이터 로딩 현황

【STEP 1】 데이터 로딩

- HR 데이터: 158명 로드 완료
- 역량검사 데이터: 152명 (상위항목 & 종합점수 모두 확보)

【STEP 1-1】 신뢰가능성 필터링

- 역량검사 데이터 신뢰불가: 5명 제외
- 최종 신뢰가능 데이터: 147명

【STEP 1-2】 개발자 여부 필터링

- 개발자 필터링 여부를 코딩으로 구현함.
- 개발자: 29명 (19.7%)
- 비개발자: 118명 (80.3%)
- 분석 선택 옵션: 모든 데이터 사용 → **147명 유지**

최종 매칭 결과

- HR 인재유형 데이터: 158명
- 역량검사 데이터: 152명
- 최종 매칭 인원: 63명 → 신뢰성 및 필터링 반영 후 58명
- 매칭률: 32.1%

2.2 인재유형 분포

인재유형	인원(명)	비율(%)
게으른 가연성	6	8.5
무능한 불연성	6	8.5
성실한 가연성	31	43.7
소화성	6	8.5
유능한 불연성	13	18.3
유익한 불연성	4	5.6
자연성	5	7.0

특징:

- 위장형 소화성(1명) 분석에서 제외

- 성실한 가연성이 거의 절반을 차지하는 편중된 분포

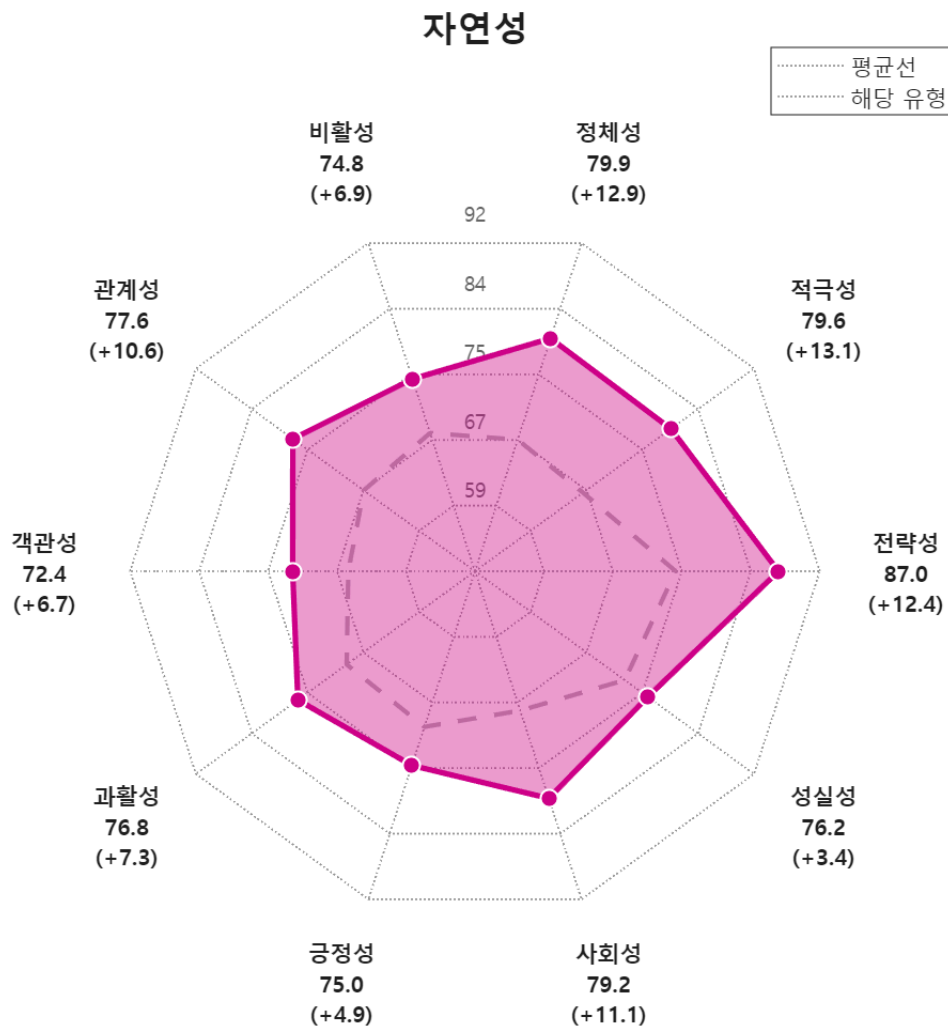
3. 인재유형별 역량 프로파일 분석

3.1 자가불소 유형 프로파일

가중평균(Weighted Mean) 적용: 클래스 불균형 반영

유형 (CODE)	비활성	정체성	적극성	전략성	성실성	사회성	관계성
자연성 (8)	74.8 (+6.9)	79.9 (+12.9)	79.6 (+13.1)	87.0 (+12.4)	76.2 (+3.4)	79.2 (+11.1)	77.6 (+10.6)
성실한 가연성 (7)	68.7 (+0.8)	67.5 (+0.4)	66.2 (-0.3)	75.4 (+0.8)	73.2 (+0.5)	68.8 (+0.8)	67.1 (+0.1)
유익한 불연성 (6)	75.4 (+7.5)	74.3 (+7.2)	81.0 (+14.5)	80.5 (+5.9)	77.8 (+5.0)	74.3 (+6.2)	73.5 (+6.0)
유능한 불연성 (5)	60.4 (-7.5)	65.2 (-1.9)	59.8 (-6.7)	75.4 (+0.8)	72.5 (-0.2)	63.4 (-4.7)	63.6 (-3.0)
게으른 가연성 (4)	66.4 (-1.5)	66.6 (-0.4)	65.2 (-1.3)	69.4 (-5.2)	73.6 (+0.9)	64.8 (-3.2)	65.2 (-1.0)
무능한 불연성 (3)	63.2 (-4.7)	55.4 (-11.7)	62.0 (-4.5)	70.1 (-4.5)	59.5 (-13.3)	59.2 (-8.9)	64.7 (-2.0)
소화성 (1)	72.6 (+4.7)	65.1 (-2.0)	65.9 (-0.6)	66.5 (-8.1)	75.2 (+2.5)	71.3 (+3.2)	64.2 (-2.0)

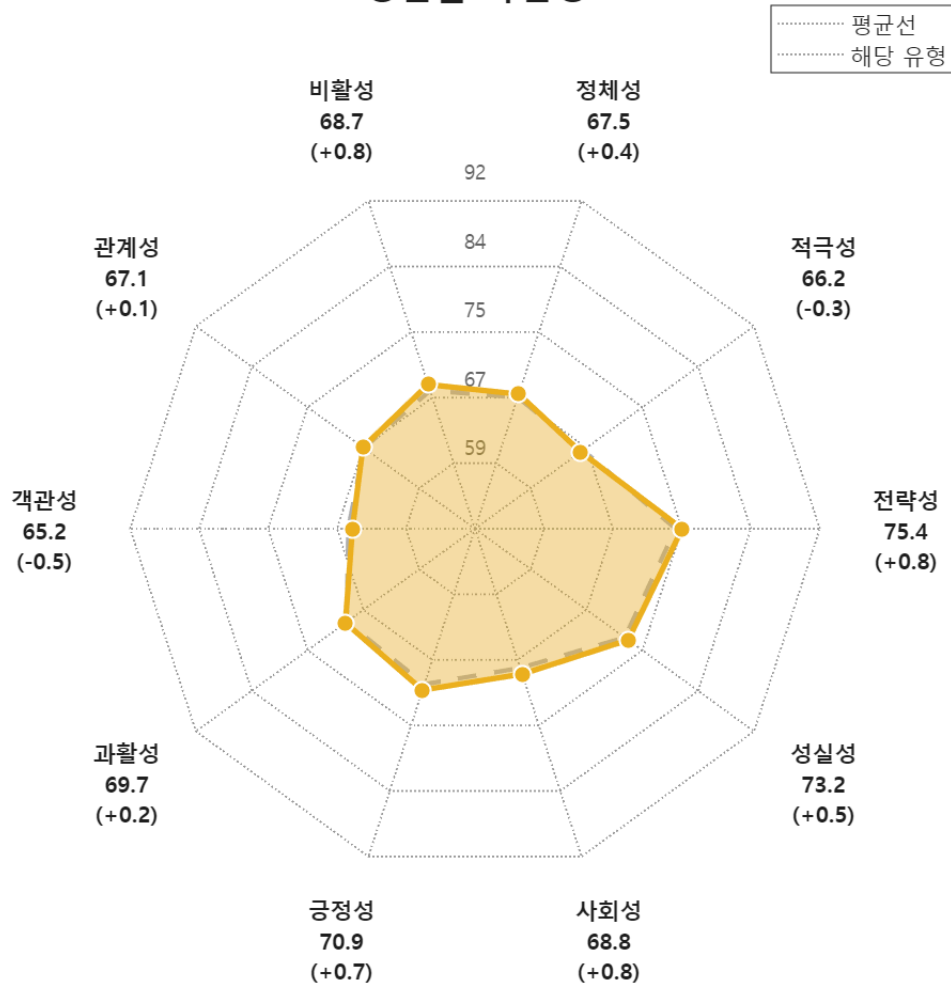
3.1. 자연성 (CODE: 8)



CODE: 8

3.2. 성실한 가연성 (CODE: 7)

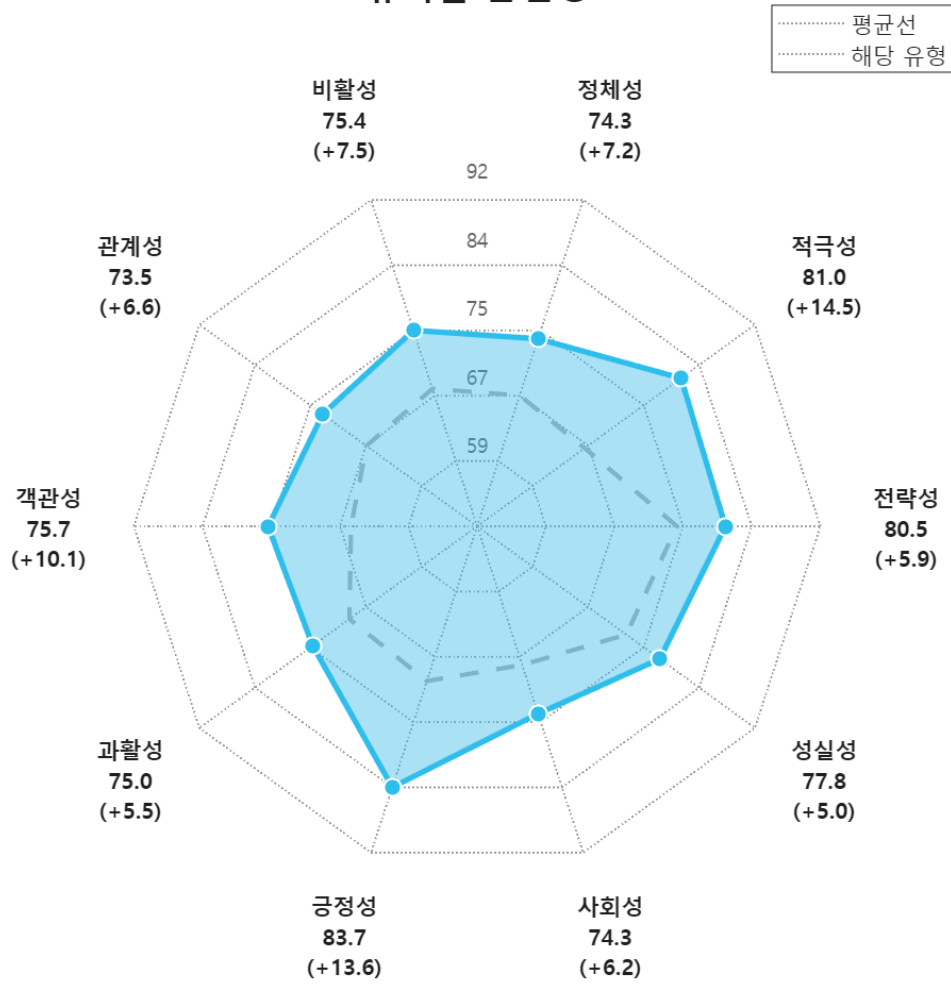
성실한 가연성



CODE: 7

3.3. 유익한 불연성 (CODE: 6)

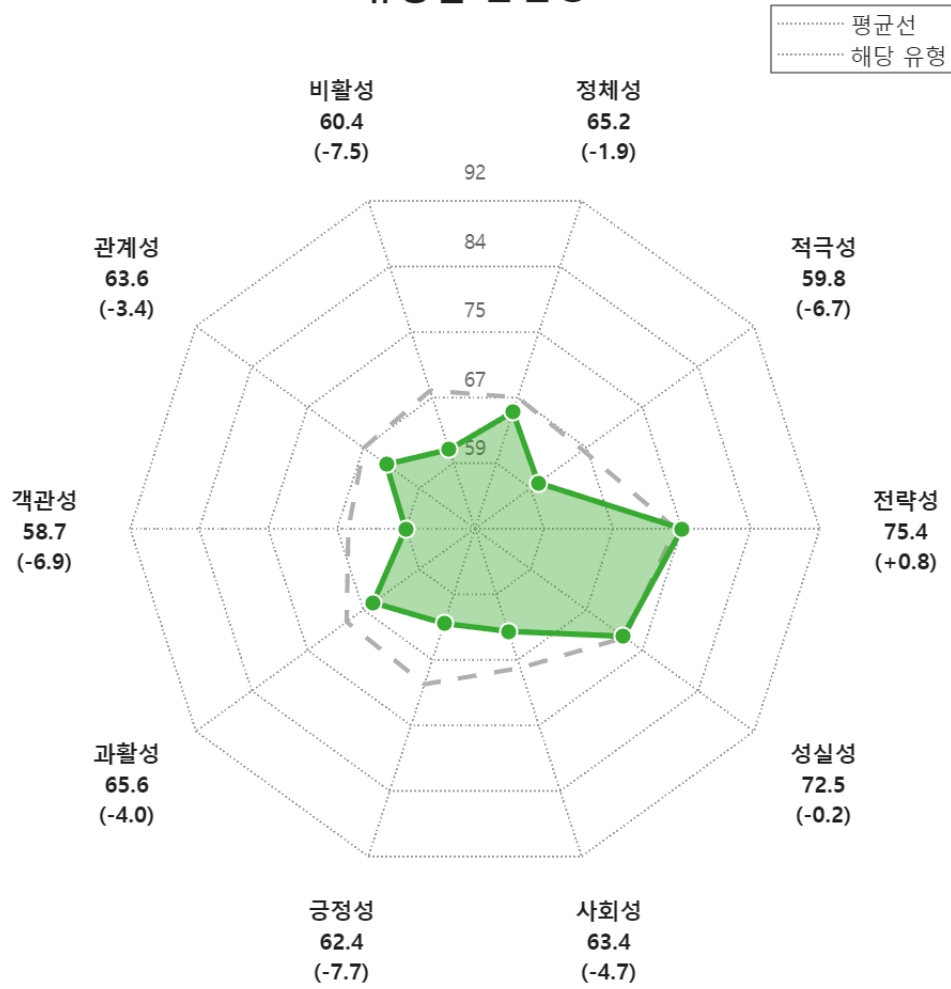
유익한 불연성



CODE: 6

3.4. 유능한 불연성 (CODE: 5)

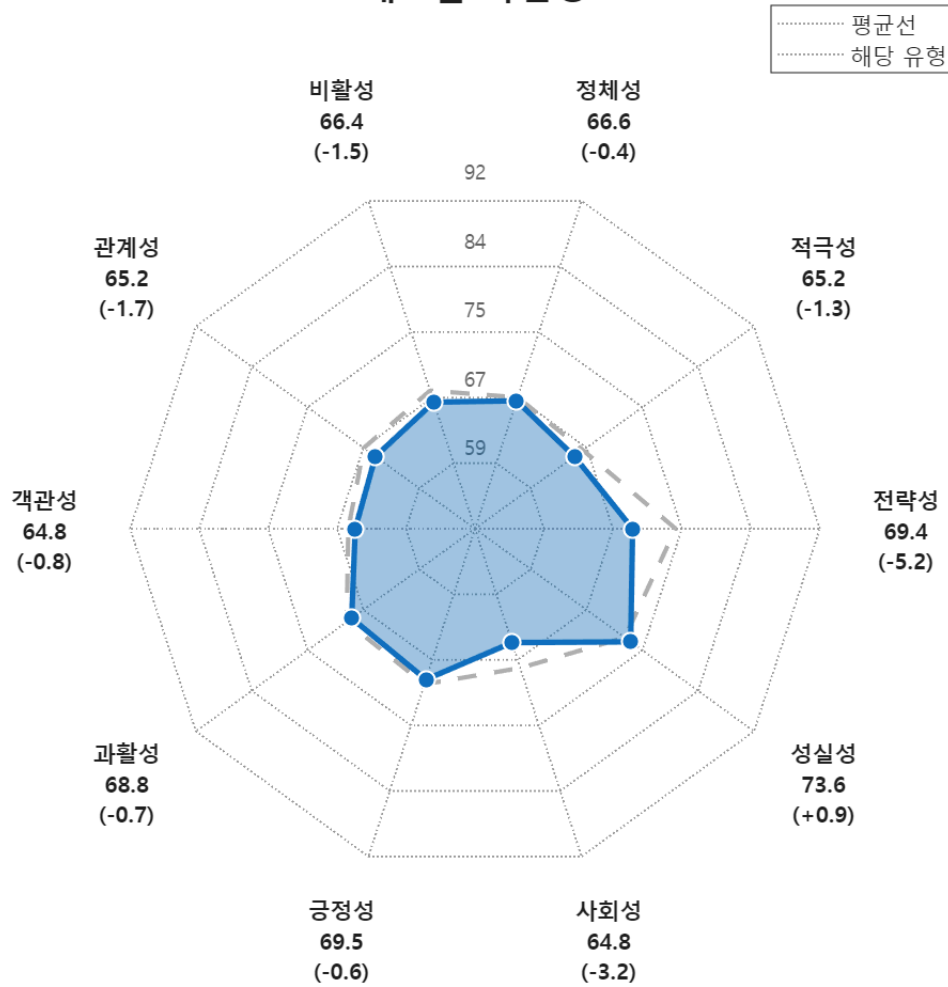
유능한 불연성



CODE: 5

3.5. 게으른 가연성 (CODE: 4)

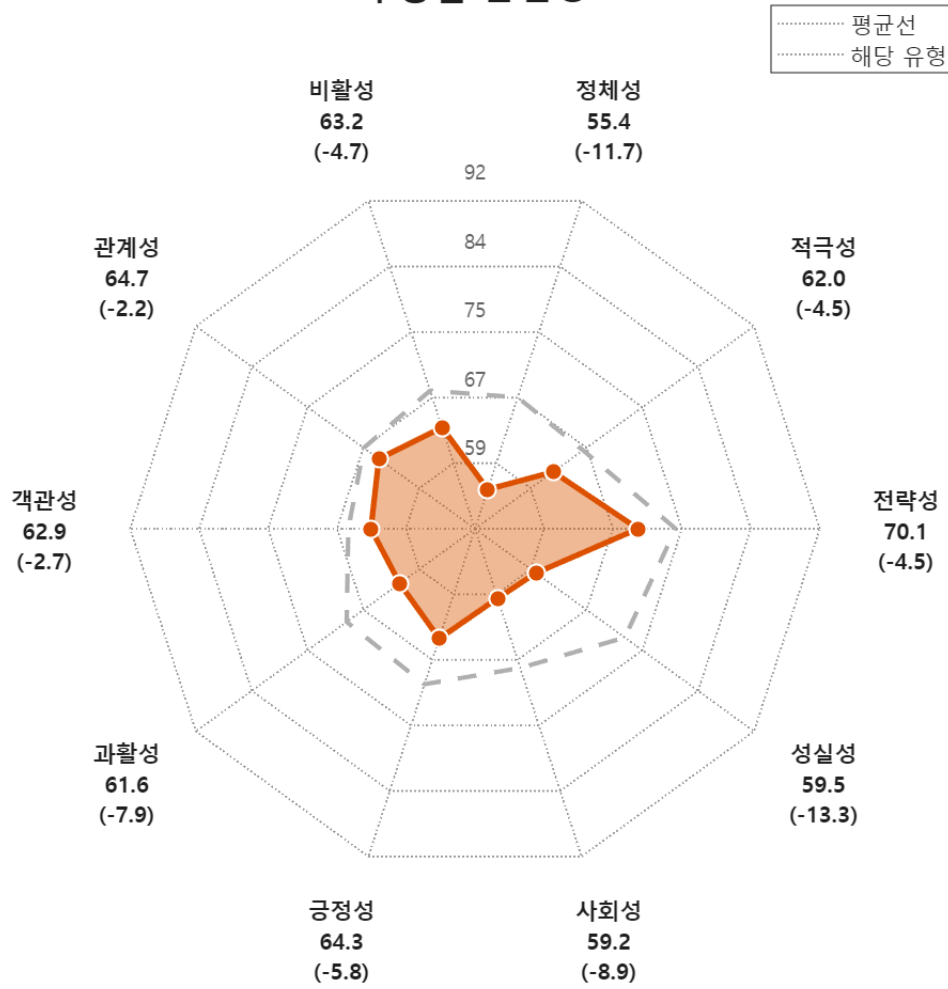
게으른 가연성



CODE: 4

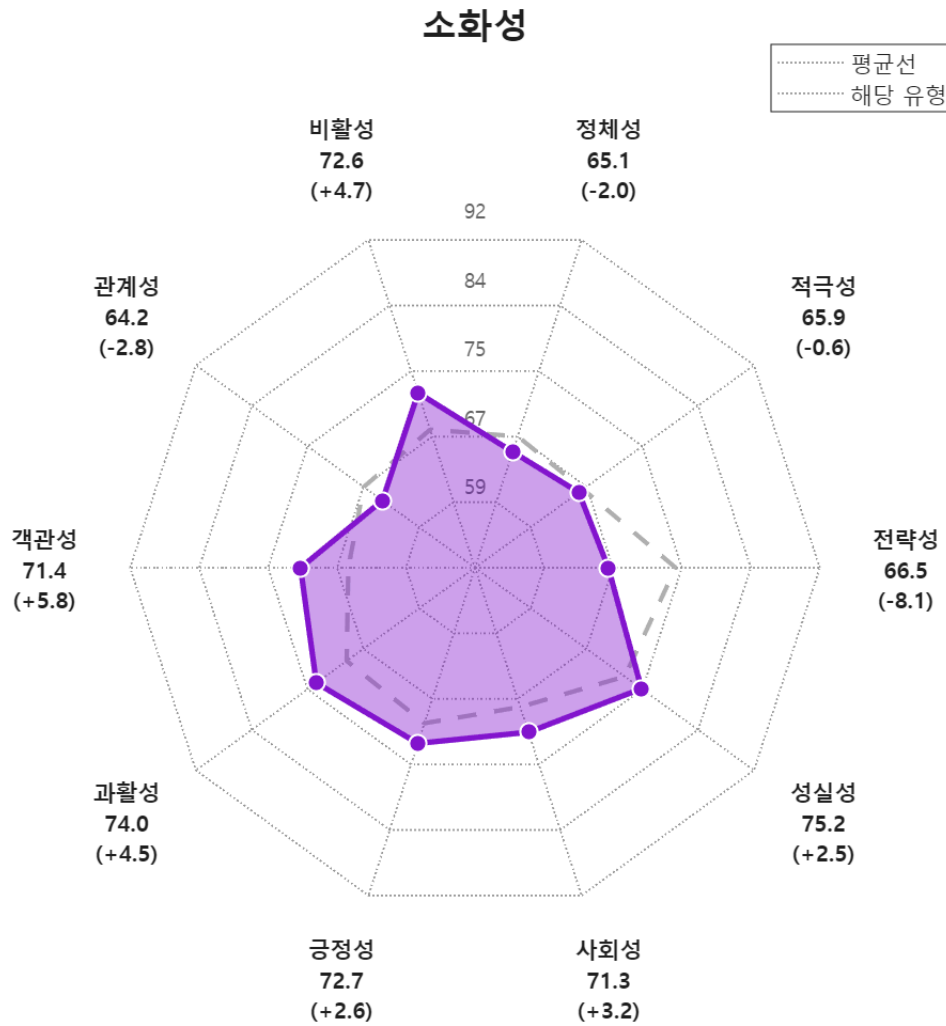
3.6. 무능한 불연성 (CODE:2)

무능한 불연성



CODE: 3

3.7. 소화성 (CODE: 1)



CODE: 1

4. 역량-성과 상관분석 결과

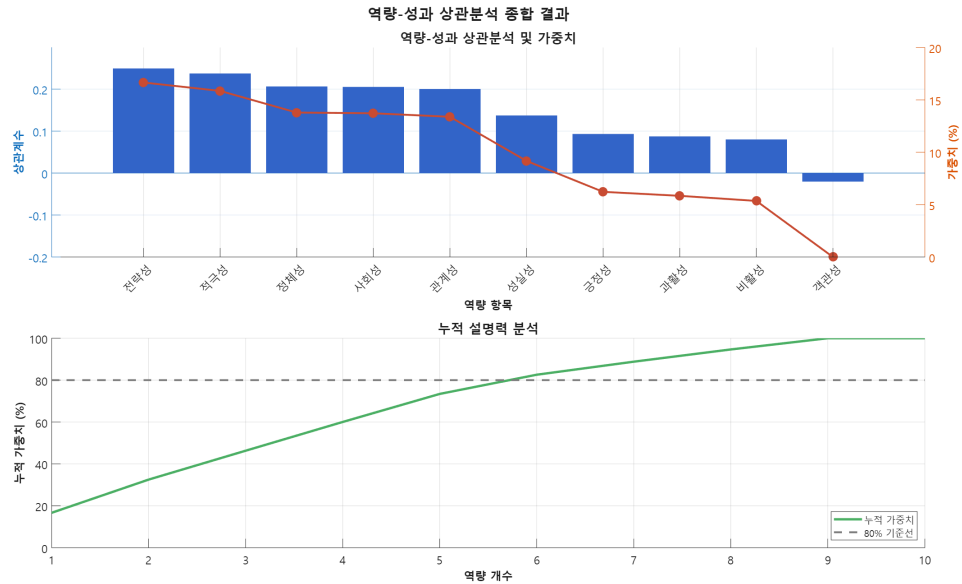
4.1. 분석 개요

- 분석 방법: Spearman 순위 상관 분석
- 목적: 역량이 성과에 미치는 영향 파악
- 상관계수 기반 가중치 산출
- ▲ 주의: 클래스 간 불균형으로 인해 결과가 왜곡될 가능성이 높음

4.1 상위 10개 성과 예측 역량

	역량	상관계수	p-값	효과크기	가중치(%)
1	전략성	0.222	0.157	0.28	17.47
2	정체성	0.217	0.167	0.27	17.08
3	적극성	0.203	0.197	-0.01	15.96
4	관계성	0.190	0.229	0.11	14.92
5	사회성	0.180	0.254	0.23	14.16

	역량	상관계수	p-값	효과크기	가중치(%)
6	성실성	0.153	0.334	0.28	12.02
7	과활성	0.073	0.647	0.15	5.72
8	비활성	0.019	0.903	0.13	1.52
9	긍정성	0.015	0.927	0.10	1.15
10	객관성	-0.038	-	-0.04	0.00



5. 인재유형 간 마할라노비스 거리 분석

5.1 분석 개요

- 대상: 7개 인재유형 (총 63명)
- 방법: 유형별 공분산 행렬 계산 후 마할라노비스 거리 측정
- 목적: 유형 간 유사성·차이 정도 정량화 → 그룹핑 및 클러스터 분석

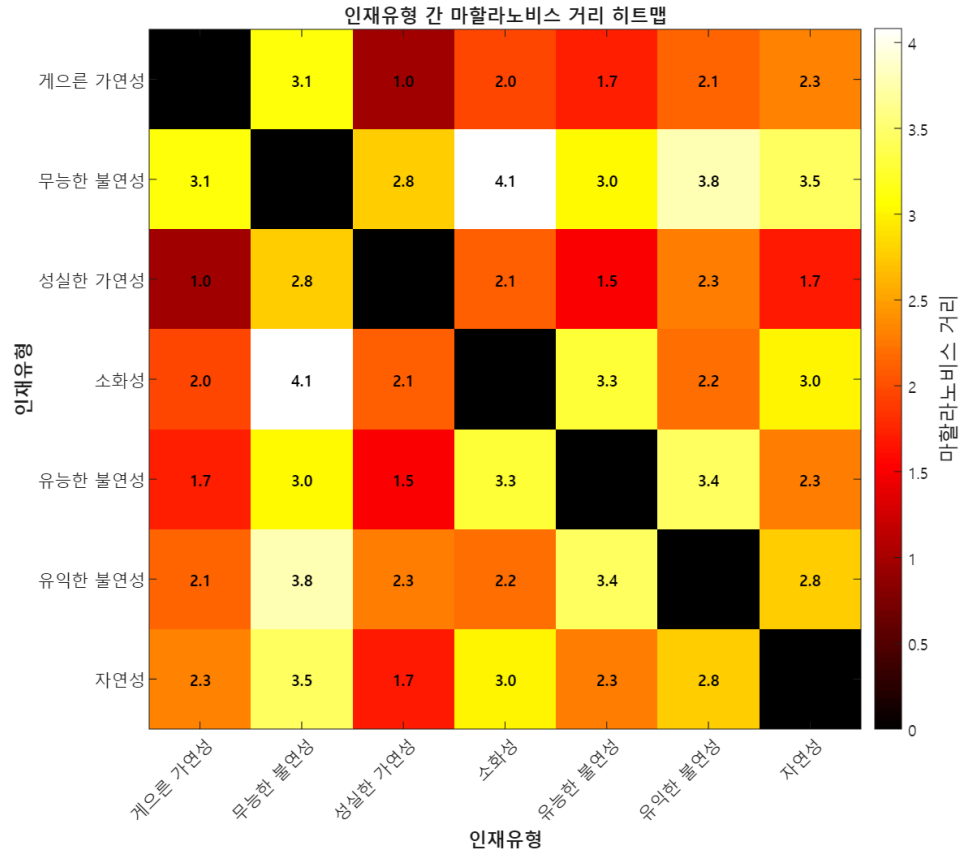
5.2 인재유형별 샘플 수

유형	샘플 수
게으른 가연성	6
무능한 불연성	5
성실한 가연성	28
소화성	6
유능한 불연성	10
유익한 불연성	4
자연성	4

5.3 마할라노비스 거리 행렬 (요약)

From \ To	게으른 가연성	무능한 불연성	성실한 가연성	소화성	유능한 불연성	유익한 불연성	자연성
게으른 가연성	-	3.09	0.98	1.96	1.71	2.13	2.30
무능한 불연성	3.09	-	2.77	4.08	3.04	3.79	3.45
성실한 가연성	0.98	2.77	-	2.09	1.51	2.29	1.69
소화성	1.96	4.08	2.09	-	3.28	2.21	3.02

From \ To	게으른 가연성	무능한 불연성	성실한 가연성	소화성	유능한 불연성	유익한 불연성	자연성
유능한 불연성	1.71	3.04	1.51	3.28	-	3.44	2.29
유익한 불연성	2.13	3.79	2.29	2.21	3.44	-	2.79
자연성	2.30	3.45	1.69	3.02	2.29	2.79	-



가장 가까운 유형 쌍: 게으른 가연성 ↔ 성실한 가연성 (거리 0.98)

→ 게으른 가연성과 성실한 가연성은 비슷한 유형

가장 먼 유형 쌍: 무능한 불연성 ↔ 소화성 (거리 4.08)

→ 무능한 불연성과 소화성은 상이한 유형

*소화성 인재에 대한 역검 점수가 높은 편이어서 이렇게 나온다고 추측됨.

5.4 성과 순위와의 관계

• 성과 순위 차이 vs 마할라노비스 거리 상관계수: 0.089

→ 매우 약한 양의 관계, 즉 유형 간 거리가 성과 차이를 일부 반영하지만 설명력은 낮음

*소화성 인재에 대한 역검 점수가 높은 편이어서 이렇게 나온다고 추측됨.

6. Logistic regrssion(Cost-Sensitive Learning)



- 로지스틱 회귀의 목적: HR에서 구분한 인재 유형을 바탕으로 뽑고 싶은 인재를 잘 예측하게 해주는 핵심 역량 도출
- 일반 회귀가 수치에 대한 예측을 한다면, 로지스틱 회귀는 yes/no에 대한 예측
- 이 사람이 우리가 뽑고 싶은 인재일까? → OO% 확률로 yes



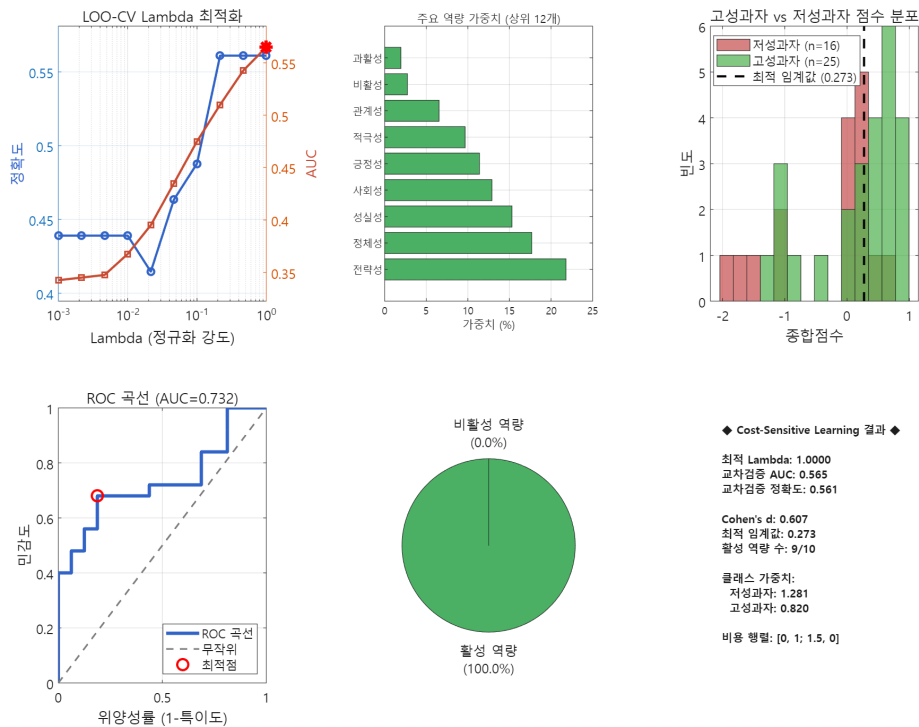
앞으로의 분석에서 뽑고 싶은 인재 유형들을 고성과자로 명명함
 앞으로의 분석에서 거르고 싶은 인재 유형을 저성과자로 명명함
 필요 시, 용어 변경 가능



왜 로지스틱 회귀를 사용했는가?

- 데이터 불균형 문제: 고성과자/저성과자 클래스 간 샘플 수 차이가 커서 복잡한 모델은 과적합 가능성이 높음
- 학습 데이터 규모: 모델 학습에 사용되는 총 샘플 수가 적어 복잡한 모델이나 비선형 모델을 적용하기 어려움
- 해석 가능성: 로지스틱 회귀는 각 변수에 대한 기여도를 직접 확인할 수 있음 → 직관적!
- 데이터 추가 확보시 모델 변경 가능

Cost-Sensitive Learning 기반 고성과자 예측 시스템 분석 결과



6.1 분석 개요

- 분석 대상: 최종 41명
 - 고성과자: **25명** (성실한 가연성, 자연성, 유익한 불연성)
 - 저성과자: **16명** (무능한 불연성, 소화성, 게으른 가연성)
 - 제외: 유능한 불연성, 위장형 소화성 (샘플 수 부족 & 분석 최적화를 위함)
- 불균형 보정
 - 클래스 가중치: 저성과자 1.281 / 고성과자 0.820
 - 비용 행렬: 저성과자 → 고성과자 오분류 비용 **1.5배** 적용
- 활용 변수: 10개 역량

6.2 모델 및 하이퍼파라미터

- 모델: Cost-Sensitive 로지스틱 회귀

- 규제 방식: Ridge (L2) 정규화
- 하이퍼파라미터 탐색: Leave-One-Out 교차검증 (LOO-CV)
 - $\lambda=0.001\sim1.000$ 구간 탐색
 - λ 증가 시 AUC-정확도 모두 상승 → 과적합 위험 없이 성능 안정화
- 최적 λ : 1.000
 - AUC = 0.565
 - 정확도 = 0.561

6.3 모델 성능

- ROC AUC: 0.732 → 양호한 분류 성능
- 최적 임계값 (Youden's J): 0.273 (중간 정도 수준)
 - 민감도 = 0.680
 - 특이도 = 0.812
- 집단 평균 비교

그룹	평균(\pm 표준편차)	표본 수
고성과자	0.184 \pm 0.725	25
저성과자	-0.288 \pm 0.855	16
차이	0.472	

- t-test: $t=1.896$, $p=0.065$ → 경계적 유의 (가중치를 이용한 종합 점수 간 차이)
- Cohen's d = 0.607 → 중~큰 효과크기

6.4 주요 역량 기여도 (가중치 %)

순위	역량	가중치(%)
1	전략성	21.80
2	정체성	17.69
3	성실성	15.30
4	사회성	12.89
5	긍정성	11.40
6	적극성	9.68
7	관계성	6.54
8	비활성	2.74
9	과활성	1.96

해석:

전략성과 정체성이 전체 기여도의 **~40%**를 차지 → 성과 예측에서 가장 핵심적인 변수

나머지 역량들도 대부분 양(+)의 계수 → 역량 수준이 전반적으로 높을수록 고성과자 확률 ↑

6.5. Bootstrap 기반 가중치 안정성 검증

Bootstrap 5000회 시행

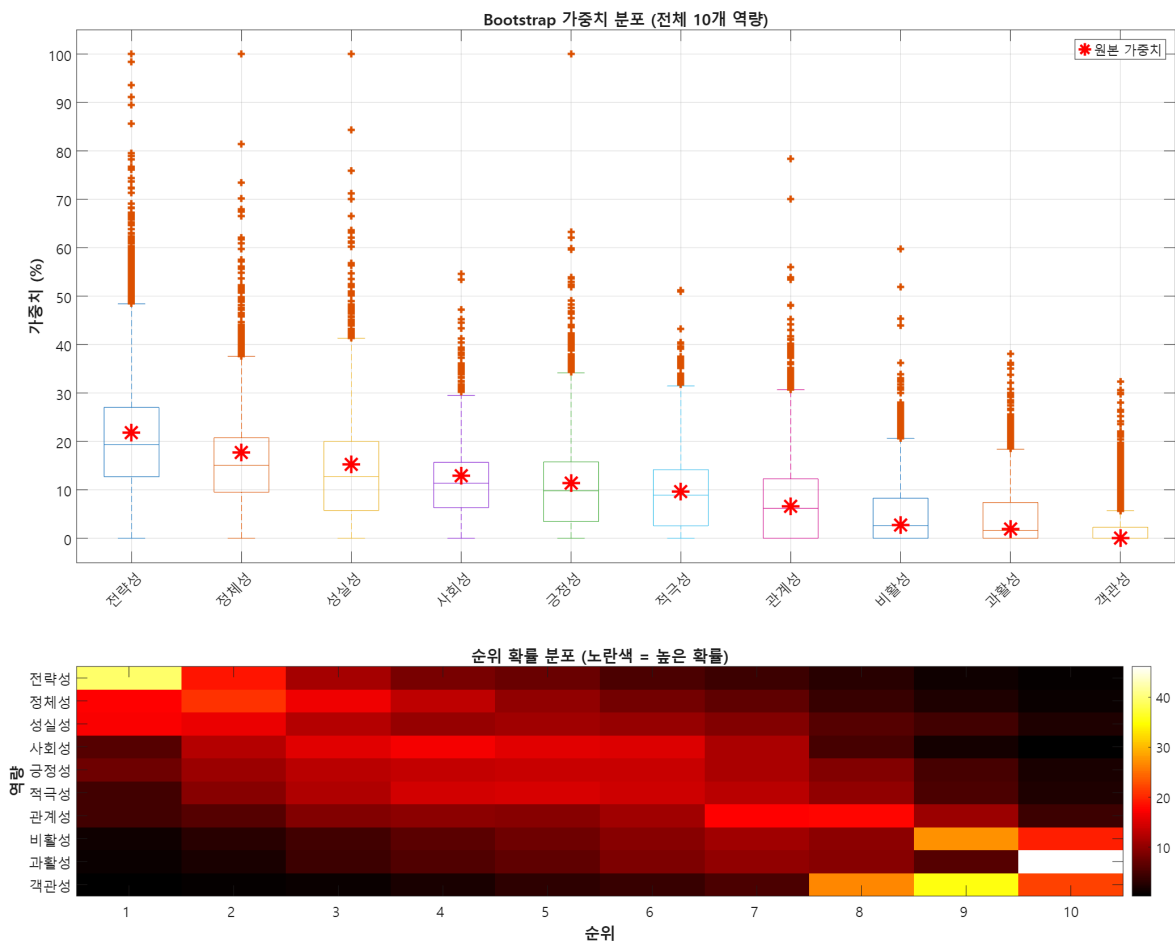


Bootstrap: 데이터 시뮬레이션 방식으로 데이터를 무작위로 뽑고 넣고를 반복(복원추출)해서 가중치가 얼마나 일관적으로 나오는지 확인하는 방법

역량	원본(%)	평균(%)	95% CI	CV	Top3 확률(%)	Top5 확률(%)
전략성	21.8	20.67	[0.00-51.73]	0.62	68.8	84.6
정체성	17.69	15.39	[0.00-35.77]	0.60	55.0	78.2
성실성	15.30	13.56	[0.00-38.33]	0.77	44.5	65.7
사회성	12.89	11.37	[0.00-26.31]	0.63	34.9	66.9
긍정성	11.40	10.60	[0.00-29.99]	0.83	31.5	59.1
적극성	9.68	9.47	[0.00-25.61]	0.79	27.2	54.8
관계성	6.54	7.65	[0.00-25.51]	1.03	19.7	39.3
비활성	2.74	4.99	[0.00-18.59]	1.21	9.1	23.1
과활성	1.96	4.34	[0.00-17.54]	1.25	7.1	21.2
객관성	0.00	1.94	[0.00-12.66]	1.92	2.1	7.2

*CV:Coefficient of Variation

*CI:Confidence Interval



히트맵 해석 (하단 그림)

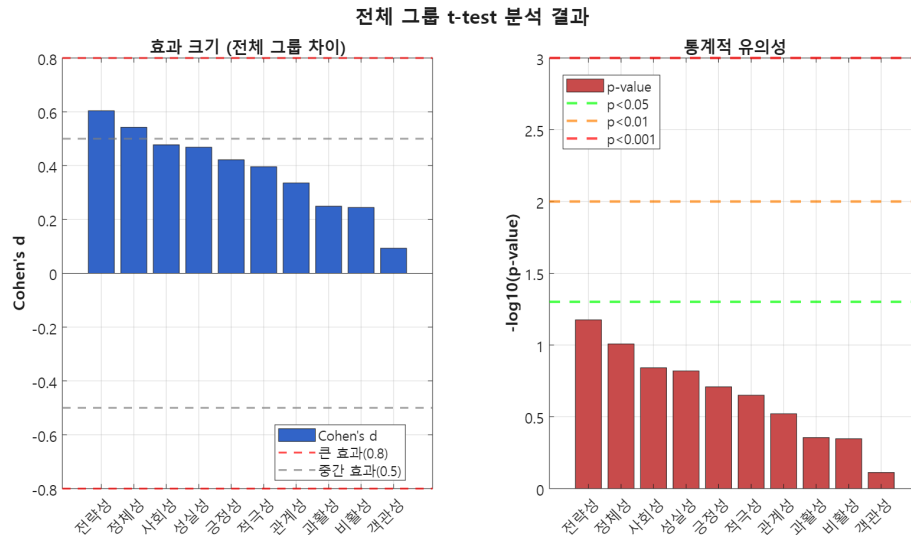
- 각 역량이 몇 위에 위치하는지 (가중치 기준) 재샘플링마다 순위를 기록한 결과입니다.
- 색이 밝을수록 → 해당 순위에 자주 등장 (순위 안정적)
- e.g. 전략성은 1~2위 근처에 많이 몰려 있어 항상 상위권에 위치함을 보여줌.
- 반대로 과활성, 객관성은 8~10위 근처에 넓게 분포 → 순위 변동이 심하고 영향력이 상대적으로 작음

결론

- 전략성·정체성·성실성은 평균 가중치가 높고 Top5 확률 >65% → 상대적으로 일관성 있는 핵심 역량
- 관계성·비활성·과활성·객관성은 CV(변동계수)가 1 이상, Top5 확률 낮음 → 모델에 따라 기여도 편차 큼 → 해석 시 주의 필요

- 전략성은 평균 가중치가 가장 높으며, 5,000번 반복 중 84.6% 확률로 Top5에 등장 → 가장 안정적 중요 변수
- 데이터 확보 후 추가적인 검증 필요

7. t-test 방식 (고성과자)



7.1 분석 설계

- 비교 집단
 - 고성과자: 자연성·성실한 가연성·유익한 불연성 (n=25)
 - 저성과자: 무능한 불연성·소화성·게으른 가연성 (n=16)
- 목적: 10개 역량 평균이 집단 간 유의하게 다른지 검증(독립표본 t-test)

7.2 결과 요약 (상위 차이 항목 중심)

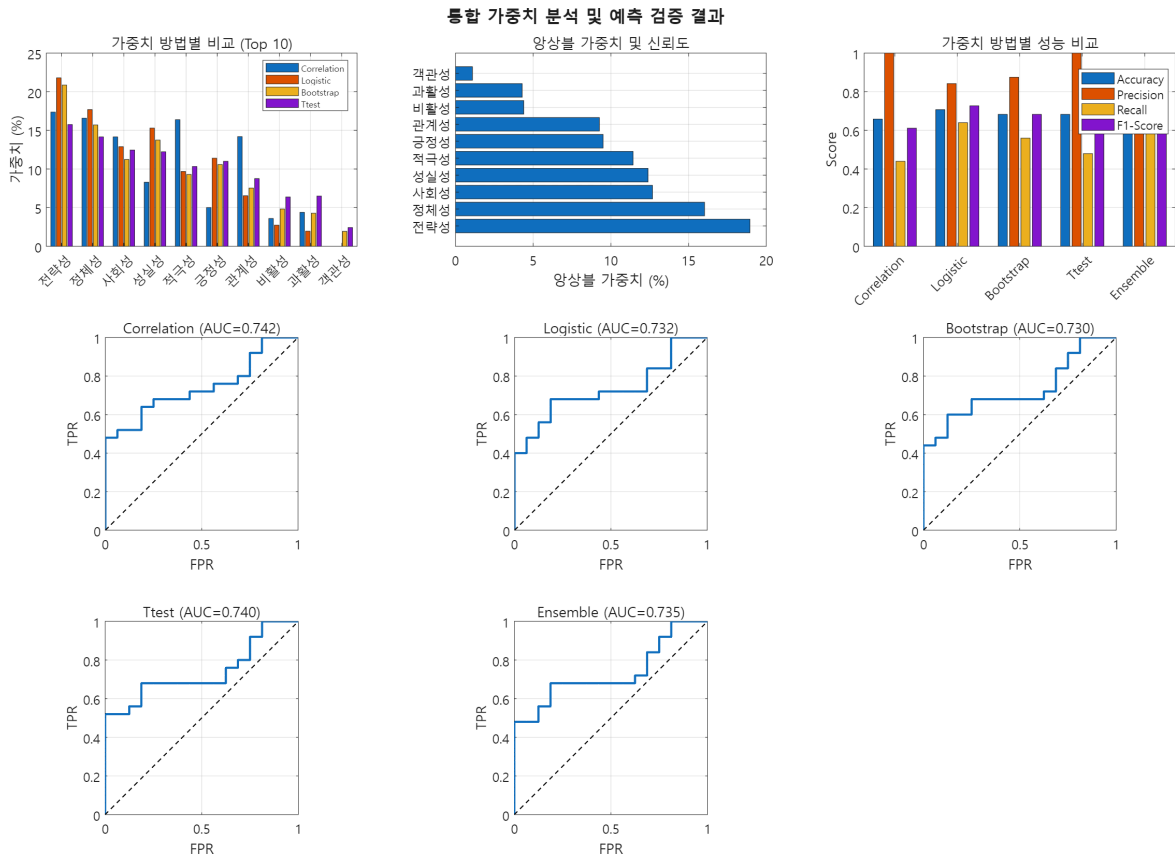
- 전략성: 고 77.1±14.6 vs 저 68.7±12.9 → **+8.4점**, $t=1.89$, $p=0.067^+$, $d=0.604$ (중간)
- 정체성: 고 69.6±12.7 vs 저 62.6±13.0 → **+6.9점**, $t=1.69$, $p=0.098^+$, $d=0.542$ (중간)
- 사회성: **+5.5점**, $p=0.144$, $d=0.477$ (작음)
- 성실성: **+4.5점**, $p=0.152$, $d=0.468$ (작음)
- 나머지(긍정성·적극성·관계성·과활성·비활성·객관성)는 **차이 작음/유의 아님**

†: 경계적 유의($0.05 \leq p < 0.10$) — 표본 확대 시 유의 가능성

7.3 해석 및 시사점

- 전략성·정체성은 **효과크기(중간 수준)**가 가장 커 **실질적 차이**가 존재할 가능성이 높음.
- 통계적 유의성은 부족하지만 **방향성과 크기는 Cost-Sensitive 모델 가중치·상관분석과 일치**.
- 데이터가 더 확보되면(특히 저성과자/유익한 불연성 표본) 전략성·정체성에서 **유의 검정 통과** 가능성이 큼.

8. 통합 가중치 분석 & 예측 검증 요약



8.1 분석개요

- 역량 중요도를 **4가지 방법**(상관, 로지스틱, Bootstrap 평균, t-test 효과크기)으로 산출
- 각각의 가중치로 만든 **예측 점수**로 성능(Accuracy/Precision/Recall/F1/AUC)을 비교
- 추가로 4가지 평균 가중치를 만들어 검증

8.2 공통 핵심 역량(가중치 일관성)

- **High 신뢰:** 전략성, 경제성, 사회성, 성실성, 적극성
 - **Medium:** 긍정성, 관계성, 비활성, 과활성
 - **Low:** 객관성
- 방법이 달라도 상위권은 **전략성·경제성**으로 수렴(그림 좌상/중앙 차트와 표 일치).

8.3 방법별 분류 성능(테스트 집합)

- **Correlation 가중치:** Acc 0.659 · Prec **1.000** · Rec 0.440 · F1 0.611 · **AUC 0.742**
 - 장점: 오탐 거의 없음(보수적). 단점: 놓침(FN) 큼.
- **Logistic(비용가중):** Acc **0.707** · Prec **0.842** · Rec **0.640** · F1 **0.727** · **AUC 0.732**
 - 균형적 성능. 실제 운영 적합.
- **Bootstrap 평균:** Acc 0.683 · Prec 0.875 · Rec 0.560 · F1 0.683 · AUC 0.730
 - 로지스틱 대비 다소 보수적.
- **t-test 가중치:** Acc 0.683 · Prec **1.000** · Rec 0.480 · F1 0.649 · **AUC 0.740**
 - 상관 방식과 유사한 특성(보수적).
- **Ensemble(평균):** Acc **0.707** · Prec **0.842** · Rec **0.640** · F1 **0.727** · **AUC 0.735**
 - 로지스틱과 동급 성능, 변동성 완충.

결론: ****로지스틱(=앙상블)**이 최적.**

8.4 임계값과 오분류 케이스

- 오분류 12건(29.3%) = FP 3, FN 9
 - **False Positive(3)**: 모두 *게으른 가연성*, 점수 0.34~0.58
의미: 선발하지 말아야 할 인재를 고성과자로 분류
 - **False Negative(9)**: 다수 *성실한 가연성*에서 발생(점수 -1.17~+0.01)
의미: 선발해야 할 인재를 놓침



게으른 가연성과 성실한 가연성에 대한 label 설정 확인 필요

8.5 F1 score 유의성 검증

- 검증 방법: 퍼뮤테이션 테스트 (5000회)
- 목적: 관찰된 F1 스코어(0.7273)가 우연히 발생할 가능성이 낮은지 확인

지표	결과
관찰된 F1 스코어	0.7273
귀무분포 평균	0.5266 ± 0.0718
95% 신뢰구간	[0.4091, 0.6818]
백분위 순위	99.5% (상위 0.5%)
Z-score	2.796
p-value	0.0054 (p < 0.01 → 통계적으로 유의)
효과 크기 (Cohen's d)	2.80 (큰 효과)

해석

- 로지스틱 모델의 F1 스코어는 무작위 분류 대비 매우 우수
- p-value < 0.01 → 통계적으로 유의
- 효과 크기(Cohen's d)가 2.8 → 실무 적용 시에도 충분히 강한 효과

8.6 결론

- 방법이 달라도 **전략성·정체성**이 일관된 1~2순위
- 로지스틱/양상불이 가장 균형적(F1 0.727) → 현 시점 가장 적합
- 데이터 추가 확보 시
 - 임계값 조정 → **재현율·정밀도 최적화**
 - 가중치 미세 조정 → **특정 역량(예: 성실성) 영향력 보완**
→ 예측 정확도 추가 개선 가능

참고:

- * *p < 0.01 (매우 유의)
- *p < 0.05 (유의)
- † p < 0.1 (경계선 유의)
- ns p ≥ 0.1 (유의하지 않음)