**역량검사 타당성 검증 및 CSR 효과 분석 계획서**

**분석 목적**

1. **직무별 역량검사 가중치 고도화**: 직무 특성에 따른 차별적 예측 모델 구축
2. **역량검사 타당도 검증**: 역검과 성과/성장의 인과관계 및 예측력 평가
3. **CSR과 성과 관계 규명**

**1단계: 기초 분석 (Descriptive & Correlation Analysis)**

**1-1. 데이터 품질 검증**

• 결측치 확인

• 이상치 탐지 (IQR, Z-score 기준)

• 데이터 분포 확인 (정규성, 왜도, 첨도): Histogram 시각화

• 평가자별 관대성/엄격성 편향 탐지: 평가자 A의 평가 패턴 확인 및 처리 방안 고려

**1-2. 기초 통계 분석**

**전체 상관관계 매트릭스**

* **역검 총점 ↔ 조직성과등급/개인기여도**
* 역검 하위지표 ↔ 역량진단 하위요인 간 매칭 분석
* 시차별 상관관계 (역검→6개월후/1년후 성과)

**집단별 차이 검증**

* 채용전형별 예측력 비교 (역검 100% 선발자 vs 경력직 채용자)
* 부서별 성과 분포 차이: 직무 환경 별 평가에 대한 분포 차이 확인

**2단계: 에측 모델링**

**2-1. 혼합효과모형 (Multilevel Modeling)**

# 레벨 구조

Level 1 (개인): 역검점수, 근속년수, 연령, 채용전형

Level 2 (팀/부서): 부서코드, 직군

Level 3 (시점): 측정년도

# 모형 예시

성과점수 ~ 역검점수 + 통제변수 + (1|평가자) + (1|부서) + (1|연도)

**주요 분석 내용:**

* **급내상관계수(ICC)**: 평가자/부서/시점별 분산 기여도  
  분산기여도 확인 후 level 재조정
* **고정효과**: 역검 지표별 순수 예측력 (β계수)
* **랜덤효과**: 평가자별 편향, 부서별 성과 기준 차이

**3단계: 예측 모델링 (Machine Learning)**

**3-1. 성과 예측 모델 구축**

**모델 후보군:**

* **Random Forest**: 비선형 관계, 상호작용 효과 탐지  
  (Bag 방식 차용)
* **Gradient Boosting**: 순차 학습을 통한 예측 정확도 향상  
  (Ada boost, Logit boost)
* **Support Vector Machine**: 고차원 데이터에서의 분류 성능
* **Deep Neural Network**: 복잡한 패턴 학습

Baseline 모델(Random Forest, GBM, SVM) 유의성 검정 후→ 심화 모델(DNN) 구조로 단계별 적용

**타겟 변수 설정:**

* **이진분류**: 고성과자 vs 일반성과자 (상위 20% 기준)
* **다중분류**: S/A/B/C 등급 예측

**3-2. 모델 평가 및 해석**

**성능 지표:**

* **분류**: Accuracy, AUC-ROC
* **회귀**: RMSE, MAE, R²

**검증 방식:**

* **교차검증**: Stratified K-fold (k=5),
* Nested cross-validation
  + 외부 루프(Outer Loop): 모델의 일반화 성능 추정
  + 내부 루프(Inner Loop): 최적 하이퍼파라미터 탐색  
    → 과적합 방지 및 모델 성능 추정의 편향 최소화

**해석 가능성:**

* **SHAP (SHapley Additive exPlanations)**: 개별 예측에 대한 기여도 분해
* **LIME**: 국소적 모델 해석
* **Feature Importance**: 전역적 변수 중요도 순위
* **Permutation Importance:** Feature 값 무작위 셔플링 후 성능 저하 정도로 중요도 측정

**📈 4단계: CSR 데이터 분석 방안**

**4-1. CSR 지표 구성**

**신규 추가 문항 (25년):**

* 조직시너지 1문항
* 자부심(로열티) 1문항
* C(소통) 2문항, S(전략) 2문항, R 1문항

**4-2. CSR-성과 상관**

* 소통점수 ↔ 개인성과
* 전략점수 ↔ 개인성과
* 성찰점수 ↔ 개인성과
* 조직시너지 ↔ 팀성과/개인성과
* 자부심 ↔ 개인성과

**4-3. CSR-성과 회귀**

**종속변수 (Y):**

* 개인성과 (예: 성과평가 점수, KPI 점수 등)

**독립변수 (X): CSR 요인**

* X1: 소통점수
* X2: 전략점수
* X3: 성찰점수
* X4: 자부심

**공변량 (Covariates):**

* 성별 (더미코딩: 남=0, 여=1)
* 연령 (연속변수)
* 근속년수 (연속변수)

MEM 설계에 대한 고려 필요

**5단계: 직무별 가중치 고도화**

**5-1. 직무별 차별 모델링**

직무군별 독립 모델:

- 개발직: 기술역량 + 문제해결 중심

- 영업직: 소통역량 + 성취동기 중심

- 기획직: 전략적사고 + 분석적사고 중심

- 지원직: 협업능력 + 정확성 중심

**5-2. 가중치 최적화**

* **Grid Search**: 각 역검 지표별 가중치 조합 탐색
* **베이지안 최적화**: 효율적 하이퍼파라미터 탐색
* **교차검증 기반**: 과적합 방지한 최적 가중치 도출

**6단계: 통제변수 및 공변량 설정**

**핵심 통제변수**

**개인 특성:**

* 인구통계: 성별, 연령, 학력
* 경력정보: 근속년수, 채용전형, 이전 경력
* 직무정보: 직군, 직급, 담당업무

**조직 맥락:**

* 팀 크기, 팀 성과
* 조직 변화 (인사이동, 팀 변경)
* 평가 시점 (분기별 성과 변화)

**공변량 처리 전략**

**Model 1**: 기본 통제변수만 포함 (성별, 연령, 근속년수)

**Model 2**: 직무 관련 변수 추가 (직군, 직급, 채용경로)

**Model 3**: 조직 맥락 변수 추가 (팀특성, 리더특성)

**Model 4**: 시점 변수 추가 (경영환경, 정책변화)

→ 현실성 고려 및 모델 비교를 통해 최적 통제 수준 결정

**분석 제약사항 및 한계점**

**데이터 한계**

* **시점 문제**: 23년 상반기 역량진단 누락
* **표본 편향**: 재직자만 분석
* **측정 도구**: 시기별 역검 버전 변화 영향

**해석상 주의점**

* **인과관계**: 상관관계를 인과관계로 해석 금지
* **타당도 문제**: 우리가 측정하는 성과는 정말 성과를 잘 반영하는가?, 우리가 측정하는 CSR은 정말 CSR을 잘 반영하는가?
* **일반화**: 타 조직 및 타 회사 적용 시 맥락적 차이 고려
* **동적 변화**: 조직문화/전략 변화에 따른 지속적 모니터링 필요