

Table 1: 전체 데이터셋(Sample)

공통 변수						개별 변수								
sex	age	incm	educ	ainc	D_1_1	DM1_dg	DM4_dg	DM8_dg	AC1_yr	MH1_yr	EC1_1	BE3_71	HE_BMI	BE3_31
1	61	1	5	241.667	3	0	0	0	2	2	2	2	25.987	4
1	53	1	5	150.000	2	0	0	0	2	2	2	2	19.782	4
2	71	2	4	70.000	4	0	0	0	2	1	2	2	29.312	4
2	77	3	3	124.000	3	0	1	0	2	2	2	2	23.871	4
1	60	3	6	700.000	3	0	0	0	2	2	1	2	21.657	4
1	65	4	5	650.000	2	0	0	1	2	2	1	2	27.013	1

개별 변수									문항 변수					예측 변수
BE5_1	BE3_85	DF2_pr	marri_2	LQ4_00	BP16_1	BP5	BP1	BD7_5	LQ_1EQL	LQ_2EQL	LQ_3EQL	LQ_4EQL	LQ_5EQL	EQ5D
1	2	0	1	2	4	1	3	2	1	1	1	1	1	1.000
2	2	0	1	2	8	2	3	2	1	1	1	1	1	1.000
3	2	0	3	2	8	2	3	2	1	1	1	2	1	0.913
1	2	0	3	2	7	2	4	2	1	1	1	1	1	1.000
1	1	0	1	2	4	2	3	2	1	1	1	1	1	1.000
1	1	0	1	2	8	2	3	2	1	1	1	1	1	1.000

Table 2: 모델별 각 문항 변수의 정확도

variable	xgBoost	catBoost	radomForest	lightGBM
LQ_1EQL	0.8766	0.8648	0.8716	0.8601
LQ_2EQL	0.9689	0.9660	0.9657	0.9603
LQ_3EQL	0.9278	0.9321	0.9233	0.9018
LQ_4EQL	0.7962	0.7877	0.7898	0.7749
LQ_5EQL	0.9193	0.9172	0.9099	0.9072

Table 3: 문항 변수의 class 분포와 2번 항목의 비율

class	LQ_1EQL	LQ_2EQL	LQ_3EQL	LQ_4EQL	LQ_5EQL
1	1217	1367	1317	1080	1297
2	189	44	89	311	114
3	7	2	7	22	2
ratio of 2	0.1338	0.0311	0.0630	0.2201	0.0807

Table 4: 삶의 질 예측 값 중 class 1, 2의 sensitivity 비교

variable	sensitivity of class1	sensitivity of class2
LQ_1EQL	0.8982	0.5607
LQ_2EQL	0.9696	0.7500
LQ_3EQL	0.9400	0.3793
LQ_4ELQ	0.8245	0.5732
LQ_5EQL	0.9296	0.5143

```

/* 1.운동능력 수준 */
IF LQ_1EQL in (1, 2, 3) THEN M2 = (LQ_1EQL=2) ;
IF LQ_1EQL in (1, 2, 3) THEN M3 = (LQ_1EQL=3) ;

/* 2.자기관리 수준 */
IF LQ_2EQL in (1, 2, 3) THEN SC2 = (LQ_2EQL=2) ;
IF LQ_2EQL in (1, 2, 3) THEN SC3 = (LQ_2EQL=3) ;

/* 3.일상활동 수준 */
IF LQ_3EQL in (1, 2, 3) THEN UA2 = (LQ_3EQL=2) ;
IF LQ_3EQL in (1, 2, 3) THEN UA3 = (LQ_3EQL=3) ;

/* 4.통증, 불편감 수준 */
IF LQ_4EQL in (1, 2, 3) THEN PD2 = (LQ_4EQL=2) ;
IF LQ_4EQL in (1, 2, 3) THEN PD3 = (LQ_4EQL=3) ;

/* 5.불안, 우울 수준 */
IF LQ_5EQL in (1, 2, 3) THEN AD2 = (LQ_5EQL=2) ;
IF LQ_5EQL in (1, 2, 3) THEN AD3 = (LQ_5EQL=3) ;

/* Interaction 모형 */
IF LQ_1EQL in (1, 2, 3) & LQ_2EQL in (1, 2, 3) & LQ_3EQL in (1, 2, 3) &
   LQ_4EQL in (1, 2, 3) & LQ_5EQL in (1, 2, 3)
THEN N3 = (LQ_1EQL=3 or LQ_2EQL=3 or LQ_3EQL=3 or LQ_4EQL=3 or LQ_5EQL=3);

/*>>> EQ-5D index : 삶의 질 조사도구(EQ-5D)의 질 가중치 추정 연구 보고서, 질병관리본부, 2007)*/

EQ5D = 1 - (0.05 + 0.096*M2 + 0.418*M3 + 0.046*SC2 + 0.136*SC3 + 0.051*UA2 +
             0.208*UA3 + 0.037*PD2 + 0.151*PD3 + 0.043*AD2 + 0.158*AD3 + 0.05*N3) ;

IF LQ_1EQL=1 & LQ_2EQL=1 & LQ_3EQL=1 & LQ_4EQL=1 & LQ_5EQL=1 THEN EQ5D = 1 ;

```

Figure 1: EQ-5D index 변수 생성

		LQ_1EQL 운동 능력	LQ_2EQL 자기 관리	LQ_3EQL 일상 활동	LQ_4EQL 통증/ 불편	LQ_5EQL 불안/ 우울
공 통 변 수	개 인 변 수	age: 나이(Ratio Variable) incm: 개인 소득 4분위수 educ: 세부 교육 수준 ainc: 월평균 가구소득(Ratio Variable) EC1_1: 경제활동 상태 sex: 성별				
	건 강	D_1_1: 주관적 건강인지				
개 별 변 수	개 인		marri_2			
	건 강 변 수	DM1_dg: 관절염 DM4_dg: 골다골증 DM8_dg: 통풍 AC1_yr: 1년간 손상 발생 여부 BE5_1: 1주일간 근력 운동 일수 MH1_yr: 1년간 입원 여부 BE3_71: 고강도 신체 활동 여부 HE_BMI: 체질량 지수	AC1_yr: 1년간 손상 발생 여부 BE3_31: 1주일간 걷기 일수 BE5_1: 1주일간 근력 운동 일수 BE3_71: 고강도 신체 활동 여부 DF2_pr: 우울증 현재 유병 여부 BE3_85: 중강도 신체 활동 여부: 여가	AC1_yr: 1년간 손상 발생 여부 BE3_31: 1주일간 걷기 일수 BE5_1: 1주일간 근력 운동 일수 BE3_71: 고강도 신체 활동 여부 DF2_pr: 우울증 현재 유병 여부 LQ4_00: 활동제한 여 부 BP16_1: 주중(또는 일 하는 날) 하루 평균 수면 시간 (만 12세 이상)	AC1_yr: 1년간 손상 발생 여부 BE3_31: 1주일간 걷기 일수 BE5_1: 1주일간 근력 운동 일수 BE3_71: 고강도 신체 활동 여부 DF2_pr: 우울증 현재 유병 여부 LQ4_00: 활동제한 여 부 BP16_1: 주중(또는 일 하는 날) 하루 평균 수면 시간 (만 12세 이상)	BE3_31: 1주일간 걷기 일수 BE5_1: 1주일간 근력 운동 일수 BP5: 우울증선별도구 9항목(문항3-①~⑨) 점수 합 DF2_pr: 우울증 현재 유병 여부 LQ4_00: 활동제한 여 부 BP1: 평소 스트레스 인지 정도 BD7_5: 1(성인) 1년간 음주문제 상담 여부

(age, ainc을 제외한 나머지 variable은 categorical data이다)

Figure 2: 문항 변수별 세부 변수표

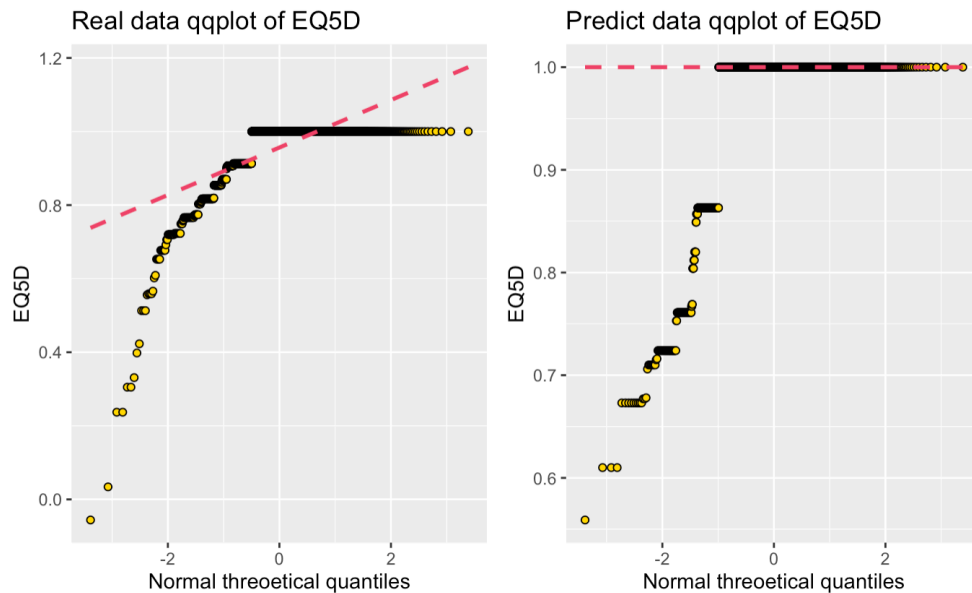


Figure 3: 실제 삶의 질 점수와 예측한 삶의 질 점수 분포

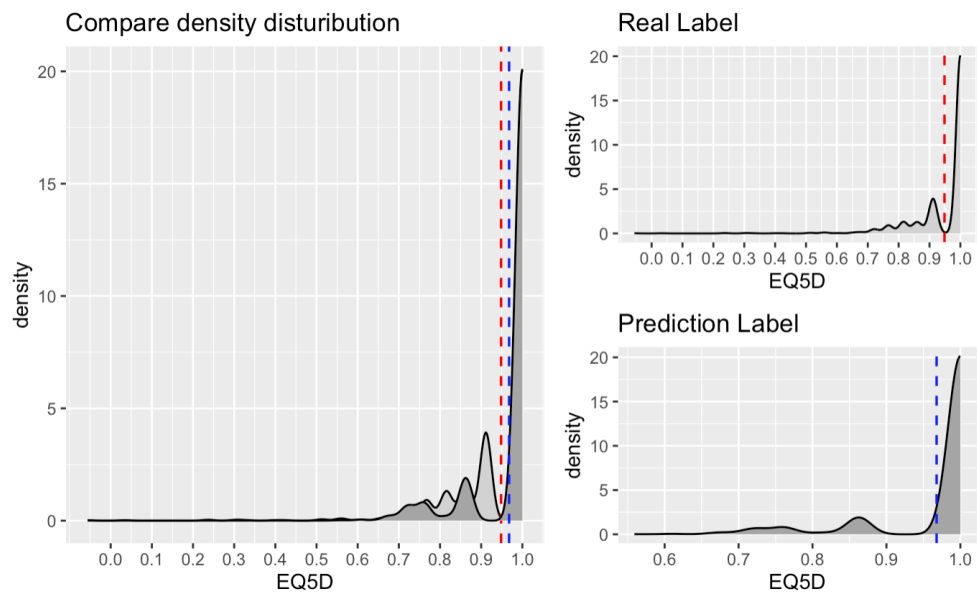


Figure 4: 실제와 예측값의 density distribution 비교