2024-2학기 machine learning in finance 텀프로젝트 : 변수명 정리				
fold 0	**train**	dataset	train_data_f0	fold 0의 train 데이터(더미변수화, 스케일링 등 전처리 완료, charges는 로그 변환함)
		label	y_train_f0	train_data_f0, train_f0_1, train_f0_2의 타겟 변수(charges, 로그변환됨)
		단변량분석	X_age_f0	fold 0의 age만 X로 설정한 변수
			X_bmi_f0	fold 0의 bmi만 X로 설정한 변수
		다변량분석	f0X0	전체 feature 사용한 X(y label만 분리)
			f0X1	age, bmi, children, smoker_yes만 포함된 X.(모든 피쳐를 사용한 다중선형회귀 시 유의하지 않은 변수 제거)
			f0X2	f0X1에 (smoker_yes * bmi) interaction feature 추가한 X.
			f0X3	f0X1에 (sex_male * smoker_yes) interaction feature 추가한 X.
			f0X4	f0X1에 (children * smoker_yes) interaction feature 추가한 X.
			train_f0_1	train_data_f0에서 지역컬럼 통합(남부), region_northwest 제거(통계적으로 유의하지 않음), smoker_bmi 추가, age_squared 추가
		벤치마크	f0X5	train_f0_1의 X(label인 charges 제거한 것)
		PCA	f0X6	f0X5에 PCA 적용된 X변수
	test	dataset	test_data_f0	fold 0의 test 데이터(더미변수화, 스케일링 등 전처리 완료, charges는 로그 변환함)
		label	y_test_f0	test_data_f0의 타겟 변수(charges, 로그변환됨)
		동일한 전처리 후	test_f0	train과 동일한 전처리를 거친 테스트 데이터셋(train_f0_1과 형식 같음)
			X_test_f0	test_f0의 X

fold 1	**train**	dataset	train_data_f1	fold 1의 train 데이터(더미변수화, 스케일링 등 전처리 완료, charges는 로그 변환함)
		label	y_train_f1	train_data_f1, train_f1_1, train_f1_2의 타겟 변수(charges, 로그변환됨)
		단변량분석	X_age_f1	fold 1의 age만 X로 설정한 변수
			X_bmi_f1	fold 1의 bmi만 X로 설정한 변수
		다변량분석	f1X0	전체 feature 사용한 X(y label만 분리)
			f1X1	age, bmi, children, smoker_yes만 포함된 X.(모든 피쳐를 사용한 다중선형회귀 시 유의하지 않은 변수 제거)
			f1X2	f1X1에 (smoker_yes * bmi) interaction feature 추가한 X.
			f1X3	f1X1에 (sex_male * smoker_yes) interaction feature 추가한 X.
			f1X4	f1X1에 (children * smoker_yes) interaction feature 추가한 X.
			train_f1_1	train_data_f1에서 지역컬럼 통합(남부), region_northwest 제거(통계적으로 유의하지 않음), smoker_bmi 추가, age_squared 추가
		벤치마크	f1X5	train_f1_1의 X(label인 charges 제거한 것)
	test	dataset	test_data_f1	fold 1의 test 데이터(더미변수화, 스케일링 등 전처리 완료, charges는 로그 변환함)
		label	y_test_f1	test_data_f1의 타겟 변수(charges, 로그변환됨)
		동일한 전처리 후	test_f1	train과 동일한 전처리를 거친 테스트 데이터셋(train_f1_1과 형식 같음)
			X_test_f1	test_f1의 X

fold2, fold3도 위와 동일(f 뒤에 붙은 숫자만 다름)