

## 유의한 변수

(포함하는 설명변수가 많으므로 adjusted  $R^2$ 로 비교하자)

M1: 모든 변수 포함

→ 다중공선성 높은 변수, 회귀계수 p-value가 너무 큰 변수 제외하는 방식으로 모델 구성

M2: 공연장소 제외

M3: 공연유형 제외

→ M2, M3 비교한 결과 공연유형을 제외해도  $R^2$  변함 없기에 제외 (공연유형 다중 공선성 높았던 것도 고려하여)

M4: 공연유형, WEEKDAY\_END, WEEKDAY\_BEGIN 제거

→ 이들의 다중공선성이 높았기에 (제외해도  $R^2$  변함없기에 제외)

M5: 공연유형, WEEKDAY\_END, WEEKDAY\_BEGIN 제거/ 년도, 월 dummy화 (다른 범주형 변수들은 이전에 category로 데이터 타입 바꾸었기에 상관 x)

→ Adjusted  $r^2$ 는 변함 x

→ 월을 dummy화 했을 때 계절의 VIF까지 높이기에 년도만 dummy화 하자

M6: 공연유형, WEEKDAY\_END, WEEKDAY\_BEGIN, 내외부 여부 제외/ 년도 dummy화

최종적으로 VIF를 낮추면서 설명력이 좋은 모델은

```
'EXCCLC_EVENT_NMPR_CO~  
FCLTY_NM+REQST_EVENT_NMPR_CO+USE_DAY_CO+HOLIDAY+END_AVG_WEEKDAY+C(년도)+  
월+계절'
```