왜 NOINT를 쓰면 값이 달라지는가?
OLS, 범주형 회귀, 시계열(차분 데이터 포함)

1 핵심 직관: 설계행렬의 열공간(모형공간)이 달라진다

일반 선형회귀에서 관측벡터 $\boldsymbol{y} \in \mathbb{R}^n$, 설계행렬 $X \in \mathbb{R}^{n\times p}$, 오차 $\boldsymbol{\varepsilon}$에 대해

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}} = \underbrace{X(X^\top X)^{-1}X^\top}_{=:H}\boldsymbol{y}$$

이며 H는 *모자행렬(hat matrix)*로서 $\mathcal{C}(X)$(열공간) 위로의 직교투영이다.

절편을 포함하면 $X = [\mathbf{1}\ \ Z]$ (여기서 $\mathbf{1}$은 전부 1인 열벡터), NOINT를 쓰면 $X_0 = Z$가 된다. 따라서 적합값은

$$\hat{\boldsymbol{y}} = P_{\mathcal{C}([\mathbf{1}\ Z])}\boldsymbol{y}, \qquad \hat{\boldsymbol{y}}_0 = P_{\mathcal{C}(Z)}\boldsymbol{y}.$$

두 적합값이 같으려면 $\mathcal{C}([\mathbf{1}\ Z]) = \mathcal{C}(Z)$여야 한다. 이는 $\mathbf{1} \in \mathcal{C}(Z)$일 때에 한해 성립한다.

$\hat{\boldsymbol{y}} = \hat{\boldsymbol{y}}_0$ (절편 포함/제외의 적합값이 일치) *오직 그 때에만* $\mathbf{1} \in \mathcal{C}(Z)$.

Proof. (만약) $\mathbf{1} \in \mathcal{C}(Z)$이면 $\mathcal{C}([\mathbf{1}\ Z]) = \mathcal{C}(Z)$이므로 두 투영이 동일하다. (역으로) $\hat{\boldsymbol{y}} = \hat{\boldsymbol{y}}_0$가 모든 $\boldsymbol{y}$에 대해 성립하면 $P_{\mathcal{C}([\mathbf{1}\ Z])} = P_{\mathcal{C}(Z)}$이므로 두 열공간이 같다. 특히 $\mathbf{1} \in \mathcal{C}([\mathbf{1}\ Z])$이므로 $\mathbf{1} \in \mathcal{C}(Z)$. □

결론. NOINT는 상수방향(평균방향)으로의 투영 자유도를 제거한다. 즉, 일반적으로는 모형공간이 달라져서 *계수와 적합값이 달라진다*. 다만, *특별히* Z의 선형결합으로 $\mathbf{1}$을 재현할 수 있으면(예: 특정 범주형 더미구성) 적합값은 동일해질 수 있다.

2 단순회귀(OLS)에서의 비교: $y = \beta_0 + \beta_1 x + \varepsilon$

관측쌍 $\{(x_i, y_i)\}_{i=1}^n$에 대해 절편 포함 OLS 추정량은

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

반면 NOINT에서의 기울기 추정량은

$$\tilde{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}.$$

만약 *진짜 모형*이 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$라면(절편 존재),

$$\tilde{\beta}_1 = \frac{\sum_i x_i(\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum_i x_i^2} = \beta_1 + \beta_0 \frac{\sum_i x_i}{\sum_i x_i^2} + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}.$$

오차가 x와 독립이고 평균 0이면, X에 조건부 기대로

$$\mathbb{E}[\tilde{\beta}_1 \mid X] = \beta_1 + \beta_0 \frac{\sum_i x_i}{\sum_i x_i^2}.$$

즉, $\sum_i x_i = 0$ (중심화)인 *특수한 경우*를 제외하면 `NOINT`의 기울기는 편향된다. 이는 정리 **??**의 관점에서 $Z = \boldsymbol{x}$의 열공간에 $\mathbf{1}$이 포함되지 않아(대개 그렇다) 모형공간이 달라졌기 때문이다.

또한 `NOINT` 직선은 원점을 반드시 지난다. 데이터가 원점 근처에 없으면 잔차제곱합(SSE)과 예측오차가 커질 수 있다.

3 범주형 회귀(더미코딩)에서의 비교

수준이 k개인 요인 $G \in \{1, \dots, k\}$를 생각하자. 각 수준에 대한 지시벡터를 $n \times k$ 행렬 $D = [\boldsymbol{d}_1, \dots, \boldsymbol{d}_k]$로 두면 언제나

$$\boldsymbol{d}_1 + \cdots + \boldsymbol{d}_k = \mathbf{1}.$$

(1) 단일 요인, 교호작용 없음

절편 포함(treatment coding). 기준수준을 하나(예: level k) 고르고 $D_{-k} = [\boldsymbol{d}_1, \dots, \boldsymbol{d}_{k-1}]$를 쓰면

$$X = [\mathbf{1} \;\; D_{-k}], \qquad \text{열공간 } \mathcal{C}(X) = \mathcal{C}(D).$$

`NOINT`. 모든 더미를 쓰면 $X_0 = D$이고 역시 $\mathcal{C}(X_0) = \mathcal{C}(D)$. 따라서

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{y}}_0 \quad (\text{적합값 동일}).$$

즉, *단일 범주형 주효과만 있을 때*는 절편 포함/제외 *모두 동일한 모형공간*을 적합하므로 적합값이 같다. 차이는 *계수의 표기법*뿐이다:

$$(\text{절편 포함}) \quad \hat{\beta}_0 = \mu_k, \quad \hat{\beta}_j = \mu_j - \mu_k \ (j = 1, \dots, k-1); \qquad (\texttt{NOINT}) \quad \tilde{\gamma}_j = \mu_j \ (j = 1, \dots, k),$$

여기서 μ_j는 각 수준의 표본평균이다.

(2) 복수 요인, 교호작용 없음 (가산모형)

두 요인 A(k_A수준), B(k_B수준)에 대해 $y \sim A + B$를 보자.

- 절편 포함: $X = [\mathbf{1} \;\; D_A^{(-)} \;\; D_B^{(-)}]$로 열 수는 $1 + (k_A - 1) + (k_B - 1)$.

- `NOINT`: $X_0 = [D_A \;\; D_B]$로 열 수는 $k_A + k_B$이나,

$$\underbrace{\mathbf{1}}_{=\sum(\text{A-더미})} = \underbrace{\sum(\text{B-더미})}_{\text{또한 }\mathbf{1}}$$

때문에 *정확히 한 개*의 선형의존이 존재하여 $\operatorname{rank}(X_0) = k_A + k_B - 1$.

두 경우 모두 열공간은 동일하게 $k_A + k_B - 1$차원이므로 적합값이 동일하다. 다시 말해,

$$P_{\mathcal{C}([\mathbf{1}\;\; D_A^{(-)}\;\; D_B^{(-)}])} = P_{\mathcal{C}([D_A\;\; D_B])}.$$

다만 계수 해석은 달라진다(기준수준 대비 효과 vs 수준 자체 효과 + *합=0* 제약 등).

(3) 교호작용 포함 (포화모형/셀평균)

요인 A, B의 교호작용을 포함하면 $A \times B$의 각 셀에 대한 지시열이 생긴다.

- 절편 포함: 주효과와 교호작용을 합치면 열공간이 전체 셀평균공간과 일치한다.
- `NOINT`: 각 셀 더미만으로도 동일한 공간을 생성한다.

따라서 포화모형에서는 적합값이 동일하며, 특히 `NOINT`일 때 각 셀의 회귀계수는 그 셀의 평균과 일치한다.

4 왜 시계열에서는 보통 달라지는가?

시계열 회귀에서 Z가 시차항(예: $y_{t-1}, y_{t-2}, \ldots$)이나 추세 t 등으로만 구성되면 일반적으로 $\mathbf{1} \notin \mathcal{C}(Z)$이다. 따라서 정리 **??**에 의해 절편 포함/제외의 적합값이 달라진다. 예를 들어 AR(1) 모형

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

에서 `NOINT`는 $c = 0$을 강제하여 과정의 평균을 0으로 제한한다:

$$\mathbb{E}[y_t] = \frac{c}{1-\phi} \quad \longrightarrow \quad 0.$$

이는 데이터의 수준(level)을 설명할 자유도를 제거하는 효과다.

차분된(differenced) 자료의 특수성

원시계열 $\{Y_t\}$를 1차 차분하면

$$\Delta Y_t = Y_t - Y_{t-1},$$

그 기댓값은

$$\mathbb{E}[\Delta Y_t] = \mathbb{E}[Y_t] - \mathbb{E}[Y_{t-1}] = 0.$$

즉, 차분 자료는 이론적으로 평균이 0이다. 이 경우 절편을 포함해도 추정치가 거의 0이 되므로, `NOINT`와 절편 포함 모형의 적합값은 동일하거나 수치 오차 수준의 미세한 차이만 난다.

실무에서 차분 후 회귀(예: ARIMA의 AR 부분)를 적합할 때 평균이 정확히 0임을 알고 있다면 `NOINT`를 지정해 자유도를 절약할 수 있다. 그러나 표본평균이 0에서 크게 벗어나면 작은 절편을 허용하는 편이 안전하다.

5 실무적 요약

- **일반 OLS(연속형 공변량):** 대개 $\mathbf{1} \notin \mathcal{C}(Z)$이므로 NOINT는 *적합값과 계수*를 바꾼다. x를 중심화하면 일부 효과(기울기 편향) 완화 가능.
- **범주형 주효과(가산):** 절편 유무와 무관하게 *적합값은 동일*. 차이는 *계수의 파라미터화*와 해석.
- **교호작용(포화):** 적합값 동일. NOINT에서는 계수가 곧 셀평균.
- **시계열(수준 포함):** 절편은 평균수준을 설명한다. NOINT는 평균을 0으로 고정.
- **차분 자료:** 평균이 이론적으로 0이므로 절편 유무가 결과에 거의 영향을 주지 않는다.