

6.5 푸아송 회귀

푸아송 회귀란

- 확률분포에 **푸아송 분포**를 사용하고
- 링크함수에 **로그함수**를 사용한 일반선형모델
- 독립변수가 여러개 있어도 상관없고
- 연속형과 카테고리형이 혼재되어 있어도 된다.

6.5.1 이 절의 예제

기온에 의해 판매 개수가 바뀐다 의 선형예측자는 B_0+B_1x 기온(℃)

6.5.2 푸아송 회귀의 구조

링크함수에 로그함수 사용

$\log[\text{맥주 판매 개수}] = B_0 + B_1x$ 기온(℃)

양변에 exp를 취한다(마이너스x). 이 식을 사용해서 판매 개수의 평균치를 예측

$\text{맥주 판매 개수} = \exp[B_0 + B_1x]$ 기온(℃)

실제 맥주 데이터를 얻었다. 이 데이터에서 *기온이 맥주 판매 개수에 영향을 주는지* 조사. 맥주 판매 개수 Y 는 강도 λ , 즉 평균값의 푸아송 분포를 따를 것.

$$\text{맥주 판매 개수} : Y \sim \text{Pois}(y | \exp[\beta_0 + \beta_1 \times \text{기온}])$$

푸아송 분포의 확률질량함수는 다음과 같다.

$$\text{Pois}(y | \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}$$

6.5.3 분석 준비