

Estadística para ciencia de datos – Poisson Regression

Edison Velazco-Grupo3

2024-03-03

Introducción

Regresión Poisson

Los modelos de regresión de Poisson se utilizan mejor para modelar eventos en los que se cuentan los resultados. O, más específicamente, contar datos: datos discretos con valores enteros no negativos que cuentan algo, como la cantidad de veces que ocurre un evento durante un período de tiempo determinado o la cantidad de personas en la fila en la tienda de comestibles.

Los datos de recuento también se pueden expresar como datos de tasa, ya que el número de veces que ocurre un evento dentro de un período de tiempo se puede expresar como una cuenta sin procesar (es decir, “En un día, comemos tres comidas”) o como una tasa (“Comemos a una tasa de 0,125 comidas por hora”).

La *regresión de Poisson* nos ayuda a analizar tanto los datos de recuento como los datos de tasa al permitirnos determinar qué variables explicativas (valores X) tienen un efecto en una variable de respuesta dada (valor Y, el recuento o una tasa). Por ejemplo, una tienda de comestibles podría aplicar la regresión de Poisson para comprender y predecir mejor la cantidad de personas en una línea.

Distribución Poisson vs Distribución Normal

La distribución de Poisson se usa más comúnmente para encontrar la probabilidad de que ocurran eventos dentro de un intervalo de tiempo dado. Dado que estamos hablando de un recuento, con la distribución de Poisson, el resultado debe ser 0 o superior; no es posible que un evento ocurra un número negativo de veces. Por otro lado, la distribución normal es una distribución continua para una variable continua y podría resultar en un valor positivo o negativo.

Distribución Poisson	Se utiliza para contar datos o tasa(razón) de datos.
Distribución Normal	Usada para variables continuas

Distribución Poisson	Se utiliza para contar datos o tasa(razón) de datos. Sesgada según los valores de lambda Varianza igual que la media
Distribución Normal	Usada para variables continuas Curva en forma de campana que es simétrica alrededor de la media La varianza y la media son parámetros diferentes; media, mediana y moda son iguales

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
cat("$P(X=k)=\\frac{\\lambda^k e^{-\\lambda}}{k!}$")
```

```
## $P(X=k)=\\frac{\\lambda^k e^{-\\lambda}}{k!}$
```