Andrew Gard - equitable.equations@gmail.com

Poisson regression: modeling bikeshare data

Poisson regression is used to model *counts*, for instance the number of bike rentals per hour in Washington, DC. The general form for such a model is:

$$\begin{cases} \ln(\lambda) & = & \beta_0 + \sum \beta_i x_i & \quad \text{systematic component} \\ y & \sim & \text{pois}(\lambda) & \quad \text{random component} \end{cases}$$

where λ is the expected number of occurrences for specified values of the explanatory variables. The coefficients β_i are unknown and must be estimated from sample data.

$$
\begin{cases}
\ln(\lambda) & = & \beta_0 + \sum \beta_i x_i & \text{systematic component} \\
y & \sim & \text{pois}(\lambda) & \text{random component}
\end{cases}
$$

The bikeshare model has one quantitative variable, `temp` (call this x_1), and two categorical ones, `workingday` and `weathersit`, which we can encode using binary dummy variables.

$$\begin{cases} \ln(\lambda) & = & \beta_0 + \sum \beta_i x_i & \quad \text{systematic component} \\ y & \sim & \text{pois}(\lambda) & \quad \text{random component} \end{cases}$$

The bikeshare model has one quantitative variable, `temp` (call this x_1), and two categorical ones, `workingday` and `weathersit`, which we can encode using binary dummy variables. `workingday` is particularly simple:

$$x_2 = \begin{cases} 1 & \text{if} \quad \texttt{workingday} = \texttt{TRUE} \\ 0 & \text{if} \quad \texttt{workingday} = \texttt{FALSE} \end{cases}$$

Becauase `weathersit` has 4 possible values (`clear, cloudy/misty, light rain/snow, and heavy rain/snow), we need 3 binary variables to encode it.`

$$x_3 = \begin{cases} 1 & \text{if} \quad \texttt{weathersit} = \texttt{cloudy/misty} \\ 0 & \text{if} \quad \texttt{weathersit} \neq \texttt{cloudy/misty} \end{cases}$$

$$x_4 = \begin{cases} 1 & \text{if} \quad \texttt{weathersit} = \texttt{lightrain/snow} \\ 0 & \text{if} \quad \texttt{weathersit} \neq \texttt{lightrain/snow} \end{cases}$$

$$x_5 = \begin{cases} 1 & \text{if} \quad \texttt{weathersit} = \texttt{heavyrain/snow} \\ 0 & \text{if} \quad \texttt{weathersit} \neq \texttt{heavyrain/snow} \end{cases}$$

If the weather is clear, all three of these will be 0. Otherwise, exactly one of them will be 1.

`weathersit`	x_3	x_4	x_5
`clear`	0	0	0
`cloudy/misty`	1	0	0
`light rain/snow`	0	1	0
`heavy rain/snow`	0	0	1

Overall, our bikeshare model will need 6 coefficients (`intercept`, `temp`, `workingday`, and 3 `weathersit`).

$$\begin{aligned}\ln(\lambda) &= \beta_0 + \beta_1\texttt{temp} + \beta_2\texttt{workingday} + \beta_3(\texttt{cloudy/misty}) + \\ &\quad \beta_4(\texttt{light rain/snow}) + \beta_5(\texttt{heavy rain/snow})\end{aligned}$$

Overall, our bikeshare model will need 6 coefficients (`intercept`, `temp`, `workingday`, and 3 `weathersit`).

$$
\begin{aligned}
\ln(\lambda) &= \beta_0 + \beta_1 \texttt{temp} + \beta_2 \texttt{workingday} + \beta_3 (\texttt{cloudy/misty}) + \\
&\qquad \beta_4 (\texttt{light rain/snow}) + \beta_5 (\texttt{heavy rain/snow}) \\
\ln(\lambda) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5
\end{aligned}
$$

Overall, our bikeshare model will need 6 coefficients (`intercept`, `temp`, `workingday`, and 3 `weathersit`).

$$\begin{aligned}
\ln(\lambda) &= \beta_0 + \beta_1 \texttt{temp} + \beta_2 \texttt{workingday} + \beta_3 (\texttt{cloudy/misty}) + \\
&\qquad \beta_4 (\texttt{light rain/snow}) + \beta_5 (\texttt{heavy rain/snow}) \\
\ln(\lambda) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \\
\lambda &= e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5}
\end{aligned}$$

```r
model %>%
  broom::tidy() %>%
  gt::gt()
```

term	estimate	std.error	statistic	p.value
(Intercept)	3.885009924	0.003230805	1202.489746	0.000000e+00
temp	2.129054163	0.004788834	444.587172	0.000000e+00
workingday1	-0.008888259	0.001943032	-4.574427	4.775243e-06
weathersitcloudy/misty	-0.042218993	0.002131712	-19.805205	2.684687e-87
weathersitlight rain/snow	-0.432089576	0.004022716	-107.412409	0.000000e+00
weathersitheavy rain/snow	-0.760994643	0.166681086	-4.565573	4.981322e-06

That is,

$$\lambda = e^{3.89+2.13x_1-.01x_2-.04x_3-.43x_4-.76x_5}$$

$$\lambda = \left(e^{3.89}\right)\left(e^{2.13x_1}\right)\left(e^{-.01x_2}\right)\left(e^{-.04x_3}\right)\left(e^{-.43x_4}\right)\left(e^{-.76x_5}\right)$$

$$\lambda = \left(e^{3.89}\right)\left(e^{2.13x_1}\right)\left(e^{-.01x_2}\right)\left(e^{-.04x_3}\right)\left(e^{-.43x_4}\right)\left(e^{-.76x_5}\right)$$

- When all of the variables are zero (minimum recorded temperature, clear non-workday), the expected number of bikers is $e^{3.89} \approx 48.9$.

$$\lambda = \left(e^{3.89}\right)\left(e^{2.13x_1}\right)\left(e^{-.01x_2}\right)\left(e^{-.04x_3}\right)\left(e^{-.43x_4}\right)\left(e^{-.76x_5}\right)$$

- When all of the variables are zero (minimum recorded temperature, clear non-workday), the expected number of bikers is $e^{3.89} \approx 48.9$.
- For each additional unit of `temp`, the expected number of bikers increases by a factor of $e^{2.13} \approx 8.4$.

$$\lambda = \left(e^{3.89}\right)\left(e^{2.13x_1}\right)\left(e^{-.01x_2}\right)\left(e^{-.04x_3}\right)\left(e^{-.43x_4}\right)\left(e^{-.76x_5}\right)$$

- When all of the variables are zero (minimum recorded temperature, clear non-workday), the expected number of bikers is $e^{3.89} \approx 48.9$.
- For each additional unit of `temp`, the expected number of bikers increases by a factor of $e^{2.13} \approx 8.4$.
- On a working day, the number of expected bikers decreases by a factor of $e^{-.01} \approx 0.99$.

$$\lambda = \left(e^{3.89}\right)\left(e^{2.13x_1}\right)\left(e^{-.01x_2}\right)\left(e^{-.04x_3}\right)\left(e^{-.43x_4}\right)\left(e^{-.76x_5}\right)$$

- When all of the variables are zero (minimum recorded temperature, clear non-workday), the expected number of bikers is $e^{3.89} \approx 48.9$.
- For each additional unit of `temp`, the expected number of bikers increases by a factor of $e^{2.13} \approx 8.4$.
- On a working day, the number of expected bikers decreases by a factor of $e^{-.01} \approx 0.99$.
- If the weather is cloudy/misty, the number of expected bikers decreases by a factor of $e^{-.01} \approx 0.96$.

$$\lambda = \left(e^{3.89}\right)\left(e^{2.13x_1}\right)\left(e^{-.01x_2}\right)\left(e^{-.04x_3}\right)\left(e^{-.43x_4}\right)\left(e^{-.76x_5}\right)$$

- When all of the variables are zero (minimum recorded temperature, clear non-workday), the expected number of bikers is $e^{3.89} \approx 48.9$.
- For each additional unit of `temp`, the expected number of bikers increases by a factor of $e^{2.13} \approx 8.4$.
- On a working day, the number of expected bikers decreases by a factor of $e^{-.01} \approx 0.99$.
- If the weather is cloudy/misty, the number of expected bikers decreases by a factor of $e^{-.01} \approx 0.96$.
- If the weather is light rain/snow, the number of expected bikers decreases by a factor of $e^{-.43} \approx 0.65$.

$$\lambda = \left(e^{3.89}\right)\left(e^{2.13x_1}\right)\left(e^{-.01x_2}\right)\left(e^{-.04x_3}\right)\left(e^{-.43x_4}\right)\left(e^{-.76x_5}\right)$$

- When all of the variables are zero (minimum recorded temperature, clear non-workday), the expected number of bikers is $e^{3.89} \approx 48.9$.
- For each additional unit of `temp`, the expected number of bikers increases by a factor of $e^{2.13} \approx 8.4$.
- On a working day, the number of expected bikers decreases by a factor of $e^{-.01} \approx 0.99$.
- If the weather is cloudy/misty, the number of expected bikers decreases by a factor of $e^{-.01} \approx 0.96$.
- If the weather is light rain/snow, the number of expected bikers decreases by a factor of $e^{-.43} \approx 0.65$.
- If the weather is heavy rain/snow, the number of expected bikers decreases by a factor of $e^{-.76} \approx 0.48$.