Foundations of Data Science: Regression and inference - Generalised linear models
We want to investigate the relationship between the number of bikes hired in an hour and the mean temperature during that hour.

Is there a problem with using ordinary least squares linear regression to do this?

Data sources:
- Edinburgh Just Eat Bikes data 2020
- Edinburgh temperature observations, Met Office via MIDAS
Overview

Monday

1. The maximum likelihood principle
2. Application of max likelihood to a simple example
3. Application of max likelihood to linear regression

Today

0. Recap + prediction uncertainty
1. Max likelihood with non-normal distributions
2. Poisson regression
3. Logistic regression and generalised linear models
Foundations of Data Science: Regression and inference - Recap of max likelihood applied to linear regression
Application of max likelihood to linear regression

\[ Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \]

\[ \varepsilon_i \sim N(0, \sigma^2) \]

\[ (x_i, y_i) \]

\[ \text{residual} \]

OR

\[ Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) \]

\[ \ln p(\gamma = y_1, \ldots, y_n; x_1, \ldots, x_n | \beta_0, \beta_1, \sigma^2) \]

\[ = \sum_{i=1}^{n} \left( -\frac{1}{2} \ln \pi_0 \sigma^2 - \frac{1}{2 \sigma^2} (y_i - \beta_0 - \beta_1 x_i)^2 \right) \]
Log likelihood of coefficients

Data from Wauters and Dhondt 1989

Peter Trimming, Wikimedia Commons, CC BY 2.0
Bootstrap inference of coefficients

Data from Wauters and Dhondt 1989
(Log) Likelihood function

Bootstrap samples

![Graphs showing log likelihood function and bootstrap samples.](image-url)
Uncertainty in predictions (with Bootstrap)

\[ y = \beta_0^{(1)} + \beta_1^{(1)} \cdot 210 \]
\[ y = \beta_0^{(2)} + \beta_1^{(2)} \cdot 210 \]

Bootstrap sample #1
Bootstrap sample #2
We want to investigate the relationship between the number of bikes hired in an hour and the mean temperature during that hour.

Is there a problem with using ordinary least squares linear regression to do this?

Are there any techniques described in the course so far that could fit the data?
\[ \ln y = \beta_0 + \beta_1 x \]

\[ y = e^{\beta_0 + \beta_1 x} \]
Foundations of Data Science: Regression and inference - Max likelihood of univariate non-normal distributions
Max likelihood for models other than the normal

We don't have to assume the data is normally distributed.

E.g. Poisson distribution

\[
P(\gamma = k) = \frac{\lambda^k e^{-\lambda}}{k!},
\]

\[k = 0, 1, 2, \ldots\]

\[
E(y) = \lambda = \mu
\]

\[
V(y) = \lambda = \sigma^2
\]
E.g. Number of goals in World Cup football matches

\[ P(Y = k) = \frac{2.5^k e^{-2.5}}{k!} \]

\[ P(Y = 0) = 2.5^0 \frac{e^{-2.5}}{0!} = e^{-2.5} = 0.082 \]

\[ P(Y = 1) = 0.205 \]

\[ P(Y = 2) = 0.257 \]
Number of deaths by horse kicks in the Prussian army

\[ \mathbf{y} = (0, 2, 2, 0, \ldots, y_1, y_2, \ldots, y_{280}) \]

\[ n_k = \sum_{i=1}^{n} I(y_i = k) \]

Bortkewitsch 1898

<table>
<thead>
<tr>
<th>k</th>
<th>n_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>144</td>
</tr>
<tr>
<td>1</td>
<td>91</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Log likelihood calculation of Poisson distribution

\[ \log \text{likelihood} \ l = \ln P(\mathbb{Y} = \mathbf{y}_1, \ldots, \mathbf{y}_n | \lambda) \]

\[ = \sum_{i=1}^{n} \ln P(\mathbf{Y} = \mathbf{y}_i) \]

\[ = \sum_{i=1}^{n} \ln \lambda^{y_i} e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} \]

\[ = \sum_{i=1}^{n} \left( y_i \ln \lambda + (-\lambda) - \ln y_i! \right) \]

\[ l(\lambda) = \ln \lambda \sum_{i=1}^{n} y_i - n\lambda - \sum_{i=1}^{n} \ln y_i! \]
\[ l = \ln P(Y = y_1, \ldots, y_n) = \ln \lambda \sum_{i=1}^{n} y_i - n\lambda - \frac{n}{2} \ln \lambda \]

\[ \frac{d l}{d \lambda} = 0 \]

\[ \Rightarrow \lambda_{MLE} = \frac{1}{n} \sum_{i=1}^{n} y_i \]
Foundations of Data Science: Regression and inference - Poisson regression
\[ Y_i \sim \text{Poisson}\left(e^{\beta_0 + \beta_1 x_i}\right) \]

\[ \ln \lambda = \beta_0 + \beta_1 x \]
Results with statsmodels GLM

Generalized Linear Model Regression Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model:</td>
<td>GLM</td>
<td>Df Residuals:</td>
<td>8299</td>
</tr>
<tr>
<td>Model Family:</td>
<td>Poisson</td>
<td>Df Model:</td>
<td>1</td>
</tr>
<tr>
<td>Link Function:</td>
<td>Log</td>
<td>Scale:</td>
<td>1.0000</td>
</tr>
<tr>
<td>Date:</td>
<td>Wed, 01 Mar 2023</td>
<td>Deviance:</td>
<td>1.3111e+05</td>
</tr>
<tr>
<td>Time:</td>
<td>06:46:41</td>
<td>Pearson chi2:</td>
<td>1.40e+05</td>
</tr>
<tr>
<td>No. Iterations:</td>
<td>5</td>
<td>Pseudo R-squ. (CS):</td>
<td>1.000</td>
</tr>
<tr>
<td>Covariance Type:</td>
<td>nonrobust</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| coef  | std err | z     | P>|z| | [0.025 0.975] |
|-------|---------|-------|-------|----------------|
| const | 1.7861  | 0.006 | 304.092 | 0.000 | 1.775 1.798 |
| air_temperature | 0.1373 | 0.000 | 323.057 | 0.000 | 0.136 0.138 |

\[ \ln \lambda = \beta_0 + \beta_1 x \]

\[ \lambda = e^{\beta_0 + \beta_1 x} \]

\[ e^{0.1373} = 1.14 \]
Poisson regression

\[ l = \ln \ P(Y = y_1, \ldots, y_n) \]

\[ l(\beta_0, \beta_1) = \sum_{i=1}^{n} (\beta_0 + \beta_1 x_i) y_i - \sum_{i=1}^{n} e^{\beta_0 + \beta_1 x_i} - \sum_{i=1}^{n} \ln y_i! \]

\[\Rightarrow\] Optimize \( \beta_0 \) and \( \beta_1 \).
To my Valentine, Poison Regression

Roses are red
Violets are blue
Some things aren't normal
and nor are you
Foundations of Data Science: Regression and inference - Logistic regression and generalised linear models
What distribution would we use to model the data here?

**Bernoulli** - Parameter $p$

How would the parameter of that distribution depend on $x$ (Age)?

**Logistic function**

$p = \text{Logistic} (\beta_0 + (\beta_1x))$
### Generalised linear models (GLMs)

<table>
<thead>
<tr>
<th>Type of Regression</th>
<th>Distribution</th>
<th>Link Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>Normal</td>
<td>$\mu = \beta_0 + \beta_1 x$, $\sigma^2$</td>
</tr>
<tr>
<td>Poisson regression</td>
<td>Poisson</td>
<td>$\ln \lambda = \beta_0 + \beta_1 x$</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>Bernoulli</td>
<td>$\ln \left( \frac{P}{1-P} \right) = \beta_0 + \beta_1 x$</td>
</tr>
</tbody>
</table>
Link functions

Expected value $\mu = E(Y|X)$ of a Bernoulli dist is $p$

$\mu = E(Y|X)$ " " Poisson dist is $\lambda$

In general the link function is denoted $g(\mu)$ where $\mu = E(Y|X)$ for that distribution:

$$g(\mu) = \beta_0 + \beta_1 x$$

To make predictions, we invert the link function:

$$\mu = g^{-1}(\beta_0 + \beta_1 x)$$
Foundations of Data Science: Regression and inference - And finally...
Max likelihood $\rightarrow$ Bayesian Inference

Bayes Theorem:

$$P(\Theta | Y = y) = \frac{P(Y = y | \Theta) P(\Theta)}{P(Y = y)}$$

Posterior

Likelihood

Prior

Evidence

Horsekick posterior

$$P(Y = y) = \int P(Y = y | \Theta) P(\Theta) \, d\Theta$$
Summary

Motivated the probabilistic basis of inference using max likelihood.

Important: think of what distribution should describe the data.

Links to future courses:
- MLG (derivation of standard ML methods)
- MLPR (Bayesian approach; application to new problems)
- MCI (Causal inference)