Statistics 24/25 Seminar II

Dmytro Tupkalenko

Format of the seminar:

As for the first seminar, I made the seminar in the format of a blog. The blog can be found here: Medium blog. All the sources are referenced in the blog, and here is the Github repository with all the working files: Github repository.

My seminar includes the following sections (with a short description):

Introduction

An overview of the topics covered in this article.

An Introduction to Generalized Linear Models (GLM)

I wanted to present Logistic Regression in more depth as it is discussed in the Agresti (2015) book. That is why I start by explaining what GLM is.

Binomial Distribution as a Member of Exponential Dispersion Family

I then move to the Binomial Distribution to show that it is a member of the Exponential Dispersion Family and identify its natural parameter, which will be used later in the explanation.

Preliminary Setup for Logistic Regression

Here, I discuss the Latent Variable Model and introduce the link function for Logistic Regression.

Logistic Regression: Properties and Interpretation of Parameters

Here, I describe the model itself, provide an explanation of how the model is fitted, explain how to interpret the coefficients, and discuss the use of Logistic Regression in Case-Control Studies.

Practical Exercises to Build Intuition

Here, I solve three exercises from Agresti (2015).

Example of the Use of Logistic Regression in an Applied Paper

Here, I present a paper in which Logistic Regression was used to assess the likelihood of overweight. Here is the link to the article: article.

Conclusion

Here, I conclude with the blog post.

This is the LaTeX code I used:

$$f(x;p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(x;p) = \binom{n}{x} p^x (1-p)^{n-x} \overset{\log(\cdot)}{\iff}$$

$$\log f(x;p) = \log \binom{n}{x} + x \log p + (n-x) \log(1-p) =$$

$$\log \binom{n}{x} + x \log(\frac{p}{1-p}) + n \log(1-p) \overset{e^{(\cdot)}}{\iff}$$

$$f(x;p) = e^{x \log(\frac{p}{1-p}) + n \log(1-p) + \log \binom{n}{x}}$$

Taking $\theta = \log(\frac{p}{1-p})$, we can write $\log(1-p)$ as $-\log(1+e^\theta)$

$$\theta = \log(\frac{p}{1-p}) \iff e^\theta (1-p) = p \iff p = \frac{e^\theta}{1+e^\theta} \iff 1-p = 1 - \frac{e^\theta}{1+e^\theta} = \frac{1}{1+e^\theta}$$

Then, taking $b(\theta) = n \log(1+e^\theta)$, $a(\phi) = 1$, $c(x,\phi) = \log \binom{n}{x}$ yields us a distribution in the exponential dispersion family:

$$f(x;\theta,\phi) = e^{\frac{x\theta - b(\theta)}{a(\phi)} + c(x,\phi)}$$

$$y_i^* = \sum_j \beta_j x_{ij} + \epsilon_i$$

$$P(y_i = 1) = P(y_i^* > \tau) = 1 - P\left(\sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i \le \tau\right) = 1 - F\left(\tau - \sum_{j=1}^{p} \beta_j x_{ij}\right)$$

$$P(y_i = 1) = F\left(\sum_{j=1}^{p} \beta_j x_{ij}\right), \quad \text{and} \quad F^{-1}[P(y_i = 1)] = \sum_{j=1}^{p} \beta_j x_{ij}. \tag{†}$$

$$F(z) = \frac{e^z}{1 + e^z}$$

$$\text{logit}[P(y = 1 \mid x = 1)] - \text{logit}[P(y = 1 \mid x = 0)] = [\beta_0 + \beta_1(1)] - [\beta_0 + \beta_1(0)] = \beta_1.$$

$$e^{\beta_j} = \frac{P(y = 1 \mid x = 1)}{1 - P(y = 1 \mid x = 1)} \div \frac{P(y = 1 \mid x = 0)}{1 - P(y = 1 \mid x = 0)}$$

$$\pi_i = \frac{\exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)} \quad \text{or} \quad \mathrm{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{j=1}^{p} \beta_j x_{ij}$$

$$\frac{\partial \pi_i}{\partial x_{ij}} = \beta_j \frac{\exp\left(\sum_j \beta_j x_{ij}\right)}{\left[1 + \exp\left(\sum_j \beta_j x_{ij}\right)\right]^2} = \beta_j \pi_i (1 - \pi_i)$$

$$e^{\beta} = \frac{P(y=1 \mid x=1)}{P(y=0 \mid x=1)} \div \frac{P(y=1 \mid x=0)}{P(y=0 \mid x=0)} = \frac{P(x=1 \mid y=1)}{P(x=0 \mid y=1)} \div \frac{P(x=1 \mid y=0)}{P(x=0 \mid y=0)}$$